

Genomic descriptors of biodiversity – A review

Genomische Deskriptoren der Biodiversität – Eine Übersicht

Gábor Mészáros

University of Natural Resources and Life Sciences Vienna (BOKU), Division of Livestock Sciences, Gregor-Mendel-Straße 33, 1180 Vienna, Austria,
gabor.meszaros@boku.ac.at

Received: 9 November 2017, received in revised form: 11 June 2018, accepted: 1 July 2018

Summary

The characterization of livestock genetic diversity has experienced extensive changes with the availability of dense nucleotide markers. Among the various forms of markers, the single nucleotide polymorphisms (SNP) have arguably the largest influence. A wide range of indicators for the assessment of genetic diversity was developed, or the existing methods were improved, enabling us to make informed decisions on the management of livestock populations. This review discusses the selected aspects of diversity assessment, with special attention to the SNP based methods.

One of the core concepts in genomics of diversity is the linkage disequilibrium (LD), as it was shaped by demographic events during the development of breeds and species. These events, either natural or artificial, left detectable signals within the livestock genomes. Further changes were induced by human activity when mating related animals, leading to fixing or improving the desired traits in the breed, but reducing their genetic variability. The assessment of relatedness is also pivotal to construct meaningful mating plans and to avoid the negative consequences of inbreeding depression that might be detrimental especially in small, endangered populations. Both LD and relatedness are of interest on their own, as well as in their follow-up applications deriving overall measures of effective population size.

Keywords: Single nucleotide polymorphism, linkage disequilibrium, population structure, inbreeding, livestock

Zusammenfassung

Die Charakterisierung der genetischen Diversität der Nutztiere wurde durch die Einführung von Nukleotidmarkern stark verändert. Es gibt eine Anzahl von verschiedenen Genommarkern, wobei die Einzelnukleotid-Polymorphismus-Marker (SNP) am bedeutsamsten sind. Eine hohe Anzahl an Indikatoren zur Beurteilung der genetischen Diversität wurde entwickelt bzw. die bereits bestehenden Methoden wurden optimiert. Diese Methoden tragen dazu bei, dass Selektionsentscheidungen in Populationen leichter getroffen werden können. Diese Übersichtsarbeit befasst sich mit der Beurteilung von genetischer Diversität unter besonderer Berücksichtigung von SNP-Methoden.

Das wichtigste genomische Diversitätskonzept ist das Kopplungsungleichgewicht (LD). Durch den demographischen Wandel haben sich verschiedene Rassen und Spezies entwickelt. Das Konzept des Kopplungsungleichgewichts hilft diesen natürlichen oder künstlichen Wandeln anhand sogenannter detektierbarer Signale im Genom der Nutztiere zu verstehen. Zusätzlich zu natürlichen Ereignissen hat der Mensch durch gezielte Selektion und gezielte Paarung in einer Population bestimmte gewünschte Merkmale verbessert bzw. fixiert und dadurch die genetische Varianz reduziert. Die Bestimmung des Verwandtschaftsgrades in einer Population unterstützt die Erstellung von Anpaarungsplänen und verhindert dadurch die negativen Konsequenzen durch Inzuchtdepression, speziell in kleinen und gefährdeten Populationen. Kopplungsungleichgewicht als auch der Grad der Verwandtschaft sind wichtige Diversitätsparameter einer Population, können aber auch zur weiteren Ableitung von der effektiven Populationsgröße, genutzt werden.

Schlagworte: Einzelnukleotid-Polymorphismus, Linkage-Disequilibrium, Populationsstruktur, Inzucht, Nutztiere

1. Introduction

Biological diversity can be described as the extent of variation contained within nature. The overall concept of diversity has been around for some time and is regarded as an important component for ensuring and assessing adaptation mechanisms. Given the constant and consistent need of food, domesticated animals were kept, selected and bred to fulfill these demands. During the process, specialized breeds of livestock were formed, with the desirable, but compared to their wild relatives, narrower genetic diversity. Currently, the needs and challenges within the livestock sector are similar to those of our predecessors. Ideally, we would like to have populations that are well adapted, have high fitness and production. The overall fitness could be broken down to several other aspects such as better survival of the wild and prolonged productive life of the livestock populations, good health and reproduction. These characteristics should be satisfactory in the animals' current environment, but they should maintain this adaptation when any of the external factors change. Also, the variance within the selected population is essential in achieving the selection response and genetic gain; thus, diversity is a major parameter for genetic improvement.

In order to cope with these challenges, it is essential to monitor and maintain genetic diversity within populations, that is, to ensure allelic variation within breeds. The across breed diversity is equally important, to ensure that there is a large enough genetic distance between breeds to maximize the production potential within different environments.

Until recently, the assessment of genetic diversity relied on the existence and accuracy of pedigree records. The pedigrees were a central tool in the assessment of relatedness between individuals, and consequently, the avoidance of mating overly related individuals that could lead to occurrence of recessive disorders, decline in fertility and overall fitness. Based on the rate of inbreeding (and other factors), the effective population size could be assessed, as a comparison metric of real and idealized populations. This metric then provided a reference on the overall genetic diversity of populations. However, the problem with the pedigrees is that they neither exist nor are accurate for many of the livestock breeds, which negates or seriously hampers any effort to assess genetic diversity.

With the availability of molecular markers such as microsatellites, and more recently single nucleotide polymorphisms (SNP) and whole genome sequence data, it is possible to analyze diversity on a genomic level. The dense

marker genotypes negate the necessity to have historical records, and thus, all problems associated with pedigrees. It allows a more precise assessment of relatedness accounting for Mendelian sampling, and consequently, more accurate inbreeding levels. The dense marker data also enables new types of analyses, such as more precise heterozygosity estimates and ability to pinpoint exact regions of the genomes undergoing selection.

The population structure and genetic distances established groundbreaking works of Nei (1977) and Weir and Cockerham (1984) reliant on overall heterozygosity levels that can also be more precisely computed for entire populations or in some cases for individual SNPs. In case of crossbred animals, the precise admixture levels can be computed for individual animals or even chromosomes, instead of reliance on average values.

Given its obvious advantages, the dense molecular marker data should be the method of choice to analyze livestock populations. This paper reviews how SNP based genomic markers could be used to investigate selected aspects of genetic diversity. Chapter 2 introduces the Linkage disequilibrium (LD) as a core concept to all follow up applications. In Chapter 3, the estimation of inbreeding from the classical (pedigree) perspective is introduced, as well as using dense marker data. In Chapter 4, the effective population size is discussed as one of the most often used metric in livestock diversity, with computation options based also on LD and inbreeding levels. In Chapter 5, an overview about changes of allele frequencies in a population are given and Chapter 6 concludes the paper.

2. Linkage disequilibrium

Linkage disequilibrium (LD) stands for the nonrandom association between alleles of different loci. The term was first used by Lewontin and Kojima (1960). Contrary to its name, however, detecting LD does not ensure linkage nor lack of equilibrium. It could be loosely defined as the measure of connectedness between alleles, a probability to be inherited together or a correlation of occurrence between allele pairs across a given population.

The LD could be expressed using various coefficients, reviewed in Slatkin (2008). The historically first measurement unit was the "disequilibrium coefficient" defined as:

$$D_{AB} = p_{AB} - p_A \times p_B$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). The D' was developed by Lewontin (1964) to make up for some of the difficulties in interpretation of D . The D' is the ratio of D to its maximum possible absolute value, given the allele frequencies, with values ranging between -1 and 1. The extreme values mean that at least one of the possible allele combinations is missing. However, the results of the D' could be inflated if the analyzed population is small or one of the alleles is rare.

The mean squared correlation of allele frequencies at different loci (r^2) is another metric of LD that was shown to be a more robust measure of LD. It is defined as:

$$r^2 = \frac{(\hat{p}_{AB} - \hat{p}_A \times \hat{p}_B)^2}{\hat{p}_A \times (1 - \hat{p}_A) \times \hat{p}_B \times (1 - \hat{p}_B)}$$

The r^2 is the most commonly used descriptor of LD in livestock genomics. Its values range between 0 and 1, with high values indicating that the two alleles are good surrogates for each other. It was found that the mean value of r^2 was determined almost entirely by the population size (N), recombination fraction (c) and time, measured proportional to N (Hill and Robertson, 1968).

The linkage disequilibrium is a function of demographic events within the species, mainly bottlenecks as a mechanism underlying selection and genetic drift, and recombination rates. The directly visible outcome of this effect is the on average higher LD between alleles close to each other and its gradual decline with increasing distance on the genome. Regarding the species-specific LD levels, the overall conclusion is that less intensively selected species show lower LD across the genome. This is confirmed by low LD levels in humans, with LD extending to only a few kilo-bases (Hinds et al., 2005). In sheep (Kijas et al., 2012; Al-Mamun et al., 2015) and goats (Brito et al., 2015; Md-ladla et al., 2016), the LD levels are generally lower than other livestock due to comparably lower selection intensity. In contrast, the LD is higher in cattle (Bohmanova et al., 2010; Espigolan et al., 2013; Porto-Neto et al., 2014), pigs (Du et al., 2007) and dogs (Lindblad-Toh et al., 2005).

LD is generally estimated within the chromosome. However, LD between chromosomes and breeds can vary due to differences in recombination rates, heterozygosity, genetic drift and effects of selection between chromosomes and breeds (Qanbari et al., 2010). As the inheritance of genomic region is based on haplotypes rather than single

nucleotides, one would expect higher LD within established haplotype blocks. Similarly, in an inbreeding scenario, within the runs of homozygosity (ROH) segments, the LD is increased, as any one of these continuous, homozygous segments originated from a single ancestor, shaped by recombination events. Thus, one could expect a connectedness between the higher chromosome-wise LD levels and the haplotype block structure and the ROH distribution, as shown in Al-Mamun et al. (2015).

LD is also a function of inter marker distance, as the LD between adjacent markers is generally high, but decreases quickly with increasing distance. The extent of LD decay could differ between breeds and species. When comparing taurine and indicine cattle breeds, the LD is initially higher in taurine, which is attributed to a smaller effective population size and a stronger bottleneck during breed formation. The LD decays at a faster rate in the taurine breeds, however, with generally higher background LD rates in indicine cattle (Pérez O'Brien et al., 2014; Porto-Neto et al., 2014). A relatively low LD was detected for the Frizarta sheep, with r^2 values of 0.18, with average inter marker distance of 31kb. As expected, the LD was decaying with increasing distance. The rate of decay among the chromosomes was variable, with more rapid decay in shorter and slower decay in longer chromosomes (Kominakis et al., 2017).

In population genetics, LD provides information for population history and both natural and artificial selection. The LD in various distances is the basis for calculation of one of the possible measurements of effective population size (Hill, 1981; Hayes et al., 2003; Tenesa et al., 2007; Mészáros et al., 2015). Also, the elevated LD values in specific genomic regions are considered as signatures of selection. The logic underlying this strategy is straightforward. When a mutation arises, it does so on an existing background haplotype characterized by complete LD between the new mutation and the linked polymorphisms (Bamshad and Wooding, 2003). Selection against recurrent deleterious mutations also reduces variation at linked loci (Charlesworth et al., 1993). This mechanism, known as “background selection”, causes the continuous removal of linked sequences along with deleterious mutations, resulting in a reduced effective population size (Kim and Stephan, 2002).

The analysis of dense marker data is based on the assumption that SNPs are in sufficient linkage disequilibrium to regions of interests. In genome wide association studies (GWAS), the significant SNPs are assumed to be in the vi-

cinity of the quantitative trait locus (QTL) or the gene influencing the phenotype under study. This interpretation of the power to correctly define an association on the genome lead to the concept of “useful LD” (Kruglyak, 1999), which should be $r^2 > 0.33$ according to the suggestion of Ardlie et al. (2002) to limit the required sample size. The D' threshold for the useful LD should be higher, due to the tendency of D' to overestimate the magnitude of LD.

3. Inbreeding

Inbreeding is a result of mating between two individuals that share at least one common ancestor. Therefore, the concepts of relatedness and inbreeding are connected to each other. Historically, both measures were evaluated based on pedigrees, based on the path coefficient methodology proposed by Wright (1922).

The offspring of two related individuals would be inbred. The inbreeding coefficient is calculated as:

$$F_Z = \frac{a_{X,Y}}{2} = \sum_{i=1}^N \left(\frac{1}{2}\right)^{(n_i+n'_i+1)} (1 + F_A)$$

where F is the inbreeding coefficient of an animal Z , $a_{X,Y}$ is the relationship coefficient between the parents of animal Z , n and n' are the number of generations from sire and dam of animals Z to a common ancestor A . If the common ancestor A is inbred itself, the inbreeding coefficient F_A should be worked out from the pedigree (Wright, 1922).

A different measure of inbreeding is the so-called F -statistics, mostly F_{IS} , introduced by Wright (1951). F_{IS} ranges between -1 and 1 , with the positive indicating that the mate pair is more related than the average relatedness in the population, thus we would expect lower proportion of heterozygotes as expected from the Hardy-Weinberg proportions. Negative F_{IS} values indicate less related individuals and excess of heterozygotes. However, there are several unexpected properties of F_{IS} showing inbreeding around zero in closely related populations, as the average population relatedness is to be considered. Moreover, the coefficient could be negative in very small populations due to allele frequency differences between males and females (Balloux and Williams, 2004). Due to these features, Kardos et al. (2016) discourages the use of F_{IS} as a measure of individual inbreeding in a population.

With the availability of genotype marker data, the inbreeding levels could also be considered from the genomic perspective. The genomic measure of inbreeding is most commonly described by continuous homozygous segments of the genome, called “runs of homozygosity” (ROH). Such homozygous runs are assumed to be a consequence of inbreeding; thus, the ROH are chromosome segments identical-by-descent. The inbreeding coefficient (F_{ROH}) is then calculated as the proportion of the autosomal genome covered with ROH:

$$F_{ROH} = \sum L_{ROH} / L_{auto}$$

where $\sum L_{ROH}$ is the total length of all of an individual's ROHs above a specified minimum length and L_{auto} is the length of the autosomal genome covered by SNPs. If the chromosomes of the analyzed organism contain centromeres, these should be excluded from L_{auto} because they do not contain SNPs and their inclusion might inflate estimates of autozygosity if both flanking SNPs are homozygous (McQuillan et al., 2008).

The length of ROH segments is strongly influenced by the number of generations separating the inbred individual from the common ancestor(s). Inbreeding due to recent ancestors usually generates quite long identical by descent (IBD) chromosome segments, whereas IBD segments deriving from more distant ancestors tend to be shorter on average because of a higher number of meioses, and recombination events, separating the inbred individual from the ancestor (Kardos et al., 2016). Under the assumption of genetic and physical distance of approximately $1 \text{ cM} = 1 \text{ Mb}$, the minimum ROH lengths denote the age of inbreeding. ROH segment lengths of 4, 8 and 16 Mb refer to a common ancestor 12, 6 and 3 generations ago, respectively (Horigan et al., 2011; Curik et al., 2014). The possibility to look for ROH segments of different length also means different sets of quality control criteria. However, there is an ongoing discussion and lack of standards on what the criteria should be. The most widely used criteria are reviewed in Peripolli et al. (2016), suggesting cautious and critical interpretation when comparing studies, analyzing the density of the SNP chip used, the minimum length of ROH, the number of genotyping errors allowed, and the minimum number of SNPs allowed in a single ROH, as they are likely to greatly affect ROH-based estimates of autozygosity.

There are several advantages of genomic measures of inbreeding. First and foremost, it is possible to calculate them solely from the genotype data, even in the absence of pedigree re-

cords. Contrary to non-genomic approaches, the inbreeding could be calculated for animals, whose pedigree is dubious, incomplete, or entirely missing (Mészáros et al., 2015). The marker-based inbreeding was also shown to be the best measure using computer simulations (Kardos et al., 2015).

Studies involving ROH are increasingly common and provide valuable information about how the genome's architecture can disclose a population's genetic background (Peripolli et al., 2016). In addition to determination of inbreeding levels, ROH could be used for a number of other purposes. Perhaps, the most related to inbreeding is the study of genetic regions with adverse effects. As deleterious variants tend to be recessive, and thus required to occur in homozygous state to manifest, the examination of ROH regions in affected and control animals could reveal the approximate location of the deleterious allele, as shown in Drögemüller et al. (2011). Even if the selected variant is not harmful, the selection increases homozygosity around the target locus, fixing the genomic region within the population, and creating the so-called ROH islands.

The ROH regions can also reveal some rare events on the genome, such as uniparental disomy (UPD). In short, UPD is a consequence of karyotypic anomalies during the meiosis and/or mitosis occurring independently or in combination, ultimately affecting chromosome distribution (Eggermann et al., 2015). Another event is the hemizygous deletion, occurring when a part of a chromosome is deleted, and this deletion is inherited by the offspring (McCarroll et al., 2006). As the SNP genotyping algorithm is not capable to detect the partially missing genotype, the results would show as a homozygous region (Huie et al., 2002), with unusual signal intensity statistics.

The genomic perspectives of inbreeding depression in wild populations are further reviewed in Kardos et al. (2016). Current applications of ROH in livestock are reviewed in Peripolli et al. (2016). The inbreeding depression in livestock is reviewed by Leroy (2014).

4. Effective population size

The effective population size (N_e) is a theoretical concept describing any given population in terms of genetic size, rather than actual number of individuals, that is, the census population (N). Formally, the N_e could be defined as the number of individuals from a Wright-Fisher population (finite and constant population size, random mating, no mutation, no selection and non-overlapping generations)

that would manifest the same extent of genetic drift and inbreeding as the population in question. The theory was initially developed and later extended by Sewall Wright (Wright, 1931). The concept of effective population size is used in the population diversity studies as a benchmark to predict the rate of inbreeding and loss of genetic variation. The ratio of effective and census population could also be used to describe the relationship between the estimated and counted population size (N_e/N). Such a ratio could be as low as 0.1, an estimate from wild populations (Frankham, 1995). The N_e/N ratio is about 0.03 in livestock (Hall, 2016), a much lower value given the high census population size in livestock originating from relatively few ancestors. In addition to the extent of genetic variation within the population, the N_e also predicts the effectiveness of spreading both beneficial and harmful alleles, with quicker spread in populations with lower effective size (Charlesworth, 2009). The N_e is affected by various factors, and it could be computed in different ways. The sex ratio in the population (Wright, 1931) is one of the most frequently used definitions of N_e , computed as:

$$N_e = \frac{4N_f N_m}{N_f + N_m}$$

Here, the N_f is the number of females, N_m the number of males. The less frequently occurring sex has the biggest influence on N_e . Other demographic estimates of N_e involve characteristics such as reproductive success (number of offspring), effective number of breeders or fluctuations in population size (Ardren and Kapuscinski 2003).

The inbreeding, here interpreted as the correlation between maternal and paternal alleles of an individual, could also be used to compute N_e . The change in inbreeding levels per generation (ΔF) computed either from pedigree or genomic data is used to estimate N_e as:

$$N_e = \frac{1}{2\Delta F}$$

Another possibility of calculating N_e of a genotyped population based on LD in autosomes (Hill, 1981; McEvoy et al., 2011) is:

$$N_{eLD} \approx \frac{1}{4c} * \left(\frac{1}{r_{LD}^2} - \alpha \right)$$

where r_{LD}^2 is the squared correlation of allele frequencies at a pair of loci, α is 2 when the impact of mutation is considered and 1 otherwise. Variable c is the genetic distance between loci in Morgans. Assuming that the population has been constant in size, the approximation of N_{eLD} is true for t generations ago, where $t = 1/(2c)$ (Hayes et al., 2003). In case of unlinked loci, the recombination frequency between them is 50%, that is, $c = 0.5$, pointing towards contemporary effective population size (Sved et al., 2013; Waples et al., 2016). Additional factors, such as the mode of inheritance (autosomal or sex linked), population age structure, changes in population size, migration and selection processes also influence N_e (Charlesworth, 2009).

The N_e calculation based on the numbers of males and females is referenced by the Food and Agriculture Organization of the United Nations (FAO). The recommendation, as a rule of thumb, is to keep the N_e above 50 to ensure that inbreeding levels are below $F = 0.01$ per year. This inbreeding threshold, however, should be viewed as a short-term criterion, rather than the ultimate goal, as continuous inbreeding leads to a gradual attrition of genetic variation. A population with consistent N_e around 50 will lose about one-fourth of its genetic variation after 20 to 30 generations, and along with it, much of its capacity to adapt to the changing conditions (Barthelmes, 1983).

The 50/500 concept was proposed as a benchmark when considering risk of extinction. Populations with N_e below 50 are at an extreme risk of extinction, while those with N_e less than 500 are in a long-term risk of extinction. In a recent study, however, Frankham et al. (2014) argue that a threshold with $N_e > 50$ is too low to prevent inbreeding depression, and $N_e > 100$ should be considered instead. Furthermore, the authors argue that $N_e > 1000$ would be required to retain evolutionary potential. While Frankham et al. (2014) refer to wild populations with considerably higher N_e , the thresholds are stunningly high for livestock. The contradiction between the suggested and observed N_e values demonstrates well the elusive concept of the N_e . While the N_e is much lower than 1000 in all commercial livestock populations, they achieve considerable genetic gains each year, which could be compared to evolutionary potential of wild populations. The question arises where the biological limits of any given species are. With the selection based on genetic merit using conventional breeding values, genomic selection or the recent promise of new frontiers uncovered by genome editing reminds us time and time again about our poor understanding of what is possible to achieve.

Apart from the presented methods, the effective population size could be assessed by the changes of allele frequencies over time (Pollak, 1983; Waples, 1989; Jorde and Ryman, 1995) or heterozygote excess in small populations (Balloux and Williams, 2004). Further perspectives of N_e estimation are summarized for pedigree-based estimators in Caballero (1994), wildlife specific aspects in Frankham (1995), the effects of spatial-temporal stratifications in Waples (2010) and the molecular genetic approaches in Charlesworth et al. (1993), Charlesworth (2009) and Lanfear et al. (2014).

5. Allele frequency changes in populations

On the population level, there is an ongoing competition between the genetic drift, that is, the random fluctuation of allele frequencies, and selection, that is, the tendency to fix certain alleles in the population. While the genetic drift is likely to affect small populations by the spread of non-beneficial mutations, the fixation of selected regions of the genome in large populations likely contains beneficial effects (Hallatschek et al., 2007).

These two counteracting processes, together with inbreeding, played a central role in the domestication processes. While the genetic drift led to random changes, the partially controlled processes of relaxed natural selection, natural selection in captivity and the fully controlled artificial selection led to the formation of livestock breeds as we know them today (Mignon-Grasteau et al., 2005).

The genetic drift was described as one of the major effects on changes of allele frequencies, counteracted by selection, if this is more intensive than commonly observed in natural populations (Lacy, 1987). In livestock, we usually observe artificial selection, but many small breeds are typically vulnerable to the effects of genetic drift (Willi et al., 2006). This is because of the small N_e , which is inversely correlated with the increase of evolutionary constraints, inbreeding and drift load. From the evolutionary perspective, there are two immediate consequences for small breeds with a small N_e : 1. In breeds without genetic variation, the evolution is constrained, no matter of the intensity of selection, unless it removes the newly arisen deleterious alleles; 2. even if genetic variation is present, only small selection responses could be expected (Kristensen et al., 2015). This happens because with a decrease of N_e , the impact of genetic drift increases, and loci start to behave as neutral, when selection coefficients become smaller or

equal to $1/[2N_e]$ (Wright, 1931). We have to note, however, that some of the major breeds, for example, Holstein, were estimated to have a strikingly low N_e around 100, without any detrimental consequences.

All livestock breeds constantly face challenges within their environment, such as diseases, climatic conditions and pressure to maintain/increase production. In order to cope with their environment, the individuals express particular traits in response to local conditions (phenotypic plasticity). When the changed conditions persist, the response is the evolution of the breed/species via genetic selection leading to adaptation. These detectable responses compared to a situation expected purely by chance are called “selection signatures” (Kim and Stephan, 2002; Pertoldi et al., 2016).

The selection pressure increases the frequency of favorable alleles that increase the chances of survival, but also the frequency of neutral alleles in linkage disequilibrium with the favorable alleles (Smith and Haigh, 1974; Flori et al., 2009). Theoretical and empirical research of selection signatures indicates that selection substantially affects the levels of neutral polymorphism, either acting against deleterious mutations or favoring advantageous mutations. This means that we can use patterns of polymorphism at neutral sites to detect selection acting at the molecular level (Payseur and Nachman, 2002). Population genetic theory predicts that beneficial mutations are either lost by genetic drift or increase in frequency until they eventually become fixed in a population (Schlötterer, 2003).

A selection signature tends to drastically reduce the variation within a population but will not lead to a reduction in species specific differences. Conversely, negative selection acting on multiple loci will reduce variability between species more drastically than variability within species (Nielsen, 2005). Neutral variants linked to deleterious mutations will indirectly experience selective pressure to be removed from populations. This idea, termed as background selection (Charlesworth et al., 1993), predicts that genomic regions of reduced recombination will exhibit decreased polymorphism levels. Such genetic hitchhiking (Smith and Haigh, 1974) also predicts a reduction of polymorphism in regions of low recombination.

Apart from a within population analysis, it is possible to evaluate selection signatures by comparing genome patterns in multiple populations. Because background selection is a balancing process that involves recurrent deleterious mutations, all populations are expected to respond in a roughly similar fashion. Alternatively, genetic hitchhiking may involve the fixation of beneficial mutations in one population

only, or the fixation of different beneficial mutations in different populations. Thus, population-specific deviations from neutrality at particular loci may identify candidate regions for genetic hitchhiking (Payseur and Nachman 2002). Analysis of these similarities or differences between the selection signatures leads to a better understanding of the genotype-phenotype map. The study of a large number of polymorphisms spread across the genome reveals aspects of the genetic structure of the population, including, in some cases, evidence of adaptive selection across the genome (Gibbs et al., 2003; Weir et al., 2005).

In animal breeding, directional selection results in loss of variation within breeds, but at the same time increases between-breed differences, given the diverse breeding goals. Selection will also increase the frequency of alleles of neutral markers in linkage disequilibrium with the favorable alleles (Smith and Haigh, 1974). In earlier works, microsatellite markers were used to detect the selection signatures in *Drosophila melanogaster* (Kauer et al., 2003; Schöfl and Schlötterer, 2004). When commercial SNP chips became available also for non-model animals, the research was quickly expanded within animal breeding. Selection signatures based on SNP markers were discovered in dairy cattle (Flori et al., 2009; Qanbari et al., 2010; Stella et al., 2010; Glick et al., 2012; Ramey et al., 2013; Pérez O'Brien et al., 2014; Pintus et al., 2014; Randhawa et al., 2014). Other works compared the differences in selection signatures of dairy and beef cattle (Prasad et al., 2008; Hayes et al., 2009) or *Bos Taurus* and *Bos Indicus* (Porto-Neto et al., 2014) using SNP markers or restricted genomic regions. Significant efforts were made to combine different methodologies to increase the power to detect selection signatures (Schwarzenbacher et al., 2012; Utsunomiya et al., 2013; Qanbari and Simianer, 2014; Randhawa et al., 2014).

A comprehensive overview of the selection signature methods from the last fifty years, explaining their conceptual motivations and statistical interpretations is outlined in Vitti et al. (2013). Strategies and approaches to detect positive signatures of selection are assessed in Qanbari and Simianer (2014).

6. Conclusion

The concept of evaluation of biological diversity is central to proper management of livestock breeds. An accurate assessment of such diversity characteristics enables to make qualified decisions on the application of protective measures.

This paper reviews the genomic measures of inbreeding, effective population size and allele frequency changes within the population in the form of genetic drift and selection. These and other indicators of genetic diversity could be accurately assessed, given the availability of dense molecular marker across the genome. The SNP marker based genomic data is widely considered to be the most accurate and affordable source to estimate such indicators, as it does not rely on any other requirement than the availability of a DNA sample. As a direct consequence, the approach also enables to characterize populations that would be otherwise extremely challenging, due to the lack of conventional information sources.

References

- Al-Mamun, H.A., Clark, S.A., Kwan, P. and C. Gondro (2015): Genome-Wide Linkage Disequilibrium and Genetic Diversity in Five Populations of Australian Domestic Sheep. *Genetics Selection Evolution* 47, 90.
- Ardlie, K.G., Kruglyak, L. and M. Seielstad (2002): Patterns of Linkage Disequilibrium in the Human Genome. *Nature Reviews Genetics* 3, 299–309.
- Ardren, W.R. and A.R. Kapuscinski (2003): Demographic and Genetic Estimates of Effective Population Size (N_e) Reveals Genetic Compensation in Steelhead Trout. *Molecular Ecology* 12, 35–49.
- Balloux, F. and R. Williams (2004): Heterozygote Excess in Small Populations and the Heterozygote-Excess Effective Population Size. *Evolution* 58, 1891–1900.
- Bamshad, M. and S.P. Wooding (2003): Signatures of Natural Selection in the Human Genome. *Nature Reviews Genetics* 4, 99–111.
- Barthelmes, D. (1983): FAO Fisheries Technical Paper No 217: Conservation of the Genetic Resources of Fish: Problems and Recommendations. Report of the Expert Consultation on the Genetic Resources of Fish, Rome 9–13 June 1980. FAO, Rome, ISBN 92-5-101173-7.
- Bohmanova, J., Sargolzaei, M. and F.S. Schenkel (2010): Characteristics of Linkage Disequilibrium in North American Holsteins. *BMC Genomics* 11, 421.
- Brito, L.F., Jafarikia, M., Grossi, D.A., Kijas, J.W., Porto-Neto, L.R., Ventura, R.V., Salgorzaei, M. and F.S. Schenkel (2015): Characterization of Linkage Disequilibrium, Consistency of Gametic Phase and Admixture in Australian and Canadian Goats. *BMC Genetics* 16, 67.
- Caballero, A. (1994): Developments in the Prediction of Effective Population Size. *Heredity* 73, 657–679.
- Charlesworth, B. (2009): Effective Population Size and Patterns of Molecular Evolution and Variation. *Nature Reviews Genetics* 10, 195–205.
- Charlesworth, B., Morgan, M.T. and D. Charlesworth (1993): The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics* 134, 1289–1303.
- Curik, I., Ferenčaković, M. and J. Sölkner (2014): Inbreeding and Runs of Homozygosity: A Possible Solution to an Old Problem. *Livestock Science* 166, 26–34.
- Drögemüller, C., Reichart, U., Seuberlich, T., Oevermann, A., Baumgartner, M., Boghenbor, K.K., Stoffel, M.H. et al. (2011): An Unusual Splice Defect in the Mitofusin 2 Gene (MFN2) Is Associated with Degenerative Axonopathy in Tyrolean Grey Cattle. *PLoS ONE* 6, e18931.
- Du, F.X., Clutter, A.C. and M.M. Lohuis (2007): Characterizing Linkage Disequilibrium in Pig Populations. *International Journal of Biological Sciences* 3, 166–178.
- Eggermann, T., Soellner, L., Buiting, K. and D. Kotzot (2015): Mosaicism and Uniparental Disomy in Prenatal Diagnosis. *Trends in Molecular Medicine, Special Issue: Nurturing the Next Generation*, 21, 77–87.
- Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R., Gordo, D.G., Tonussi, R.L., Cardoso, D.F. et al. (2013): Study of Whole Genome Linkage Disequilibrium in Nelore Cattle. *BMC Genomics* 14, 305.
- Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. and M. Gautier (2009): The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *PLOS ONE* 4, e6595.
- Frankham, R. (1995): Effective Population Size/adult Population Size Ratios in Wildlife: A Review. *Genetics Research* 66, 95–107.
- Frankham, R., Bradshaw, C.J.A. and B.W. Brook (2014): Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses. *Biological Conservation* 170, 56–63.
- Glick, G., Shirak, A., Uliel, S., Zeron, Y., Ezra, E., Seiroussi, E., Ron, M. and J.I. Weller (2012): Signatures of Contemporary Selection in the Israeli Holstein Dairy Cattle. *Animal Genetics* 43, 45–55.
- Hall, S.J.G. (2016): Effective Population Sizes in Cattle, Sheep, Horses, Pigs and Goats Estimated from Census and Herdbook Data. *Animal* 10, 1778–1785.

- Hallatschek, O., Hersen, P., Ramanathan, S. and D.R. Nelson (2007): Genetic Drift at Expanding Frontiers Promotes Gene Segregation. *Proceedings of the National Academy of Sciences* 104, 19926–19930.
- Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L. and M.E. Goddard (2009): A Genome Map of Divergent Artificial Selection between *Bos Taurus* Dairy Cattle and *Bos Taurus* Beef Cattle. *Animal Genetics* 40, 176–184.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. and M.E. Goddard (2003): Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Research* 13, 635–643.
- Hill, W.G. (1981): Estimation of Effective Population Size from Data on Linkage Disequilibrium. *Genetics Research* 38, 209–216.
- Hill, W.G. and A. Robertson (1968): Linkage Disequilibrium in Finite Populations. *Theoretical and Applied Genetics* 38, 226–231.
- Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A. and D.R. Cox (2005): Whole-Genome Patterns of Common DNA Variation in Three Human Populations. *Science* 307, 1072–1079.
- Howrigan, D.P., Simonson, M.A. and M.C. Keller (2011): Detecting Autozygosity through Runs of Homozygosity: A Comparison of Three Autozygosity Detection Algorithms. *BMC Genomics* 12, 460.
- Huie, M.L., Anyane-Yeboah, K., Guzman, E. and R. Hirschhorn (2002): Homozygosity for Multiple Contiguous Single-Nucleotide Polymorphisms as an Indicator of Large Heterozygous Deletions: Identification of a Novel Heterozygous 8-Kb Intragenic Deletion (IVS7–19 to IVS15–17) in a Patient with Glycogen Storage Disease Type II. *The American Journal of Human Genetics* 70, 1054–1057.
- Jorde, P.E. and N. Ryman (1995): Temporal Allele Frequency Change and Estimation of Effective Size in Populations with Overlapping Generations. *Genetics* 139, 1077–1090.
- Kardos, M., Luikart, G. and F.W. Allendorf (2015): Measuring Individual Inbreeding in the Age of Genomics: Marker-Based Measures Are Better than Pedigrees. *Hereditas* 115, 63–72.
- Kardos, M., Taylor, H.R., Ellegren, H., Luikart, G. and F.W. Allendorf (2016): Genomics Advances the Study of Inbreeding Depression in the Wild. *Evolutionary Applications* 9, 1205–1218.
- Kauer, M.O., Dieringer, D. and C. Schlötterer (2003): A Microsatellite Variability Screen for Positive Selection Associated with The ‘out of Africa’ habitat Expansion of *Drosophila Melanogaster*. *Genetics* 165, 1137–1148.
- Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B. et al. (2012): Genome-Wide Analysis of the World’s Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology* 10, e1001258.
- Kim, Y. and W. Stephan (2002): Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics* 160, 765–777.
- Kominakis, A., Hager-Theodorides, A.L., Saridaki, A., Antonakos, G. and G. Tsiamis (2017): Genome-Wide Population Structure and Evolutionary History of the Frizarta Dairy Sheep. *Animal* 3, 1–9.
- Kristensen, T.N., Hoffmann, A.A., Pertoldi, C. and A.V. Stronen (2015): What Can Livestock Breeders Learn from Conservation Genetics and Vice Versa? *Frontiers in Genetics* 5, 00038.
- Kruglyak, L. (1999): Prospects for Whole-Genome Linkage Disequilibrium Mapping of Common Disease Genes. *Nature Genetics* 22, 139–144.
- Lacy, R.C. (1987): Loss of Genetic Diversity from Managed Populations: Interacting Effects of Drift, Mutation, Immigration, Selection, and Population Subdivision. *Conservation Biology* 1, 143–158.
- Lanfear, R., Kokko, H. and A. Eyre-Walker (2014): Population Size and the Rate of Evolution. *Trends in Ecology & Evolution* 29, 33–41.
- Leroy, G. (2014): Inbreeding Depression in Livestock Species: Review and Meta-Analysis. *Animal Genetics* 45, 618–628.
- Lewontin, R.C. (1964): The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 49, 49–67.
- Lewontin, R.C. and K. Kojima (1960): The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* 14, 458–472.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M. et al. (2005): Genome Sequence, Comparative Analysis and Haplotype Structure of the Domestic Dog. *Nature* 438, 803–819.
- McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S. et al. (2006): Common Deletion Polymorphisms in the Human Genome. *Nature Genetics* 38, 86–92.

- McEvoy, B.P., Powell, J.E., Goddard, M.E. and P.M. Visscher (2011): Human Population Dispersal 'Out of Africa' Estimated from Linkage Disequilibrium and Allele Frequencies of SNPs. *Genome Research* 21, 821–829.
- McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N. et al. (2008): Runs of Homozygosity in European Populations. *The American Journal of Human Genetics* 83, 359–372.
- Mdladla, K., Dzomba, E.F., Huson, H.J. and F.C. Muchadeyi (2016): Population Genomic Structure and Linkage Disequilibrium Analysis of South African Goat Breeds Using Genome-Wide SNP Data. *Animal Genetics* 47, 471–482.
- Mészáros, G., Boison, S.A., Pérez O'Brien, A.M., Ferencaković, M., Curik, I., Da Silva, M.V.B., Utsunomiya, Y.T., Garcia, J.F. and J. Sölkner (2015): Genomic Analysis for Managing Small and Endangered Populations: A Case Study in Tyrol Grey Cattle. *Frontiers in Genetics* 6, 173.
- Mignon-Grasteau, S., Boissy, A., Bouix, J., Faure, J.M., Fisher, A.D., Hinch, G.N., Jensen, P. et al. (2005): Genetics of Adaptation and Domestication in Livestock. *Livestock Production Science* 93, 3–14.
- Nei, M. (1977): F-Statistics and Analysis of Gene Diversity in Subdivided Populations. *Annals of Human Genetics* 41, 225–233.
- Nielsen, R. (2005): Molecular Signatures of Natural Selection. *Annual Review of Genetics* 39, 197–218.
- Payseur, B.A. and M.W. Nachman (2002): Natural Selection at Linked Sites in Humans. *Gene* 300, 31–42.
- Pérez O'Brien, A.M., Utsunomiya, Y.T., Mészáros, G., Bickhart, D.M., Liu, G.E., Van Tassell, C.P., Sonstegard, T.S., Da Silva, M.V.B., Garcia, J.F. and J. Sölkner (2014): Assessing Signatures of Selection through Variation in Linkage Disequilibrium between Taurine and Indicine Cattle. *Genetics, Selection, Evolution* 46, 19.
- Peripolli, E., Munari, D.P., Da Silva, M.V.B., Lima, A.L.F., Irgang, R. and F. Baldi (2016): Runs of Homozygosity: Current Knowledge and Applications in Livestock. *Animal Genetics* 48, 255–271.
- Pertoldi, C., Randi, E., Ruiz-Gonzalez, A., Vergeer, P. and J. Ouborg (2016): How Can Genomic Tools Contribute to the Conservation of Endangered Organisms. *International Journal of Genomics* 12, e4712487.
- Pintus, E., Sorbolini, S., Albera, A., Gaspa, G., Dimauro, C., Steri, R., Marras, G. and N.P. Macciotta (2014): Use of Locally Weighted Scatterplot Smoothing (LOWESS) Regression to Study Selection Signatures in Piedmontese and Italian Brown Cattle Breeds. *Animal Genetics* 45, 1–11.
- Pollak, E. (1983): A New Method for Estimating the Effective Population Size from Allele Frequency Changes. *Genetics* 104, 531–548.
- Porto-Neto, L.R., Kijas, J.W. and A. Reverter (2014): The Extent of Linkage Disequilibrium in Beef Cattle Breeds Using High-Density SNP Genotypes. *Genetics, Selection, Evolution* 46, 22.
- Prasad, A., Schnabel, R.D., McKay, S.D., Murdoch, B., Stothard, P., Kolbehndari, D., Wang, Z., Taylor, J.F. and S.S. Moore (2008): Linkage Disequilibrium and Signatures of Selection on Chromosomes 19 and 29 in Beef and Dairy Cattle. *Animal Genetics* 39, 597–605.
- Qanbari, S. and H. Simianer (2014): Mapping Signatures of Positive Selection in the Genome of Livestock. *Livestock Science, Genomics Applied to Livestock Production* 166, 133–143.
- Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. and H. Simianer (2010): The Pattern of Linkage Disequilibrium in German Holstein Cattle. *Animal Genetics* 41, 346–56.
- Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. and J.F. Taylor (2013): Detection of Selective Sweeps in Cattle Using Genome-Wide SNP Data. *BMC Genomics* 14, 382.
- Randhawa, I.A.S., Khatkar, M.S., Thomson, P.C. and H.W. Raadsma (2014): Composite Selection Signals Can Localize the Trait Specific Genomic Regions in Multi-Breed Populations of Cattle and Sheep. *BMC Genetics* 15, 34.
- Schlötterer, C. (2003): Hitchhiking Mapping – Functional Genomics from the Population Genetics Perspective. *Trends in Genetics* 19, 32–38.
- Schöfl, G. and C. Schlötterer (2004): Patterns of Microsatellite Variability Among X Chromosomes and Autosomes Indicate a High Frequency of Beneficial Mutations in Non-African D. Simulans. *Molecular Biology and Evolution* 21, 1384–1390.
- Schwarzenbacher, H., Dolezal, M., Flisikowski, K., Seefried, F., Wurmser, C., Schlötterer, C. and R. Fries (2012): Combining Evidence of Selection with Association Analysis Increases Power to Detect Regions Influencing Complex Traits in Dairy Cattle. *BMC Genomics* 13, 48.
- Slatkin, M. (2008): Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. *Nature Reviews Genetics* 9, 477–485.

- Smith, J.M. and J. Haigh (1974): The Hitch-Hiking Effect of a Favourable Gene. *Genetical Research* 23, 23–35.
- Stella, A., Ajmone-Marsan, P., Lazzari, B. and P. Boettcher (2010): Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. *Genetics* 185, 1451–1461.
- Sved, J. A., Cameron, E.C. and A.S. Gilchrist (2013): Estimating Effective Population Size from Linkage Disequilibrium between Unlinked Loci: Theory and Application to Fruit Fly Outbreak Populations. *PLoS ONE* 8, e69078.
- Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E and P.M. Visscher (2007): Recent Human Effective Population Size Estimated from Linkage Disequilibrium. *Genome Research* 17, 520–526.
- Utsunomiya, Y.T., do Carmo, A.S., Carvalheiro, R., Neves, H.H.R., Matos, M.C., Zavarez, L.B., Pérez O'Brien, A.M. et al. (2013): Genome-Wide Association Study for Birth Weight in Nellore Cattle Points to Previously Described Orthologous Genes Affecting Human and Bovine Height. *BMC Genetics* 14, 52.
- Vitti, J.J., Grossman, S.R. and P.C. Sabeti (2013): Detecting Natural Selection in Genomic Data. *Annual Review of Genetics* 47, 97–120.
- Waples, R.S. (1989): A Generalized Approach for Estimating Effective Population Size from Temporal Changes in Allele Frequency. *Genetics* 121, 379–391.
- Waples, R.S. (2010): Spatial-Temporal Stratifications in Natural Populations and How They Affect Understanding and Estimation of Effective Population Size. *Molecular Ecology Resources* 10, 785–796.
- Waples, R.K., Larson, W.A. and R.S. Waples (2016): Estimating Contemporary Effective Population Size in Non-Model Species Using Linkage Disequilibrium across Thousands of Loci. *Heredity* 117, 233–240.
- Weir, B.S. and C.C. Cockerham (1984): Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358–1370.
- Willi, Y., Van Buskirk, J. and A.A. Hoffmann (2006): Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics* 37, 433–458.
- Wright, S. (1922): Coefficients of Inbreeding and Relationship. *The American Naturalist* 56, 330–338.
- Wright, S. (1931): Evolution in Mendelian Populations. *Genetics* 16, 97–159.