E

联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Продовольственная и сельскохозяйственная организация Объединенных Наций | Organización de las Naciones Unidas para la Alimentación y la Agricultura | منظمة الأغذية والزراعة للأمم المتحدة

# COMMISSION ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE

**Item 4 of the Provisional Agenda**

**INTERGOVERNMENTAL TECHNICAL WORKING GROUP ON ANIMAL GENETIC RESOURCES FOR FOOD AND AGRICULTURE**

**Twelfth Session**

**Rome, 18 - 20 January 2023**

## METHODS FOR ESTIMATION OF WITHIN-POPULATION GENETIC VARIATION

**Note by the Secretariat**

1. Increased genetic diversity in an organism or group of organisms is generally considered beneficial, as it promotes resiliency and allows greater options for adaptation. Genetic diversity can be measured and evaluated on landscape, species, population, individual and genic scales. The *Proposed headline and component indicators for the Post-2020 Global Biodiversity Framework*[1] of the Convention on Biological Diversity address the maintenance of genetic variation in its Goal A.

2. Milestone A.3 is that "Genetic diversity of wild and domesticated species is safeguarded, with an increase in the proportion of species that have at least 90 percent of their genetic diversity maintained." The Headline indicator A.0.4 proposed for this milestone is "The proportion of populations within species with a genetically effective population size > 500."

3. Genetic diversity, including effective population size, can be estimated by using demographic and/or pedigree data and these approaches have been used for many years in the characterization and management of animal genetic resources for food and agriculture. Application of these methods is restricted to populations for which such data are available, which limits their utilization for many livestock breed populations, especially in developing regions. Genomic data may also be used to estimate genetic diversity and recent developments in biotechnologies have greatly decreased the costs of obtaining molecular information, thus increasing the potential availability of such data. In conjunction with these developments in genomic analysis, new methods for estimation of genetic diversity have been developed.

[1] CBD/WG2020/3/INF/2.

4. To support its Member Countries in the use of genomics, FAO and partners developed the *Draft practical guide on genomic characterization of animal genetic resources*,[2] which was presented to the Commission on Genetic Resources for Food and Agriculture (Commission) at its Eighteenth Regular Session. The Commission took note of the practical guide and requested FAO to […] disseminate them and to encourage countries to make full use of them, according to their specific needs.[3] The practical guide included discussion of genomic methods for estimation of genetic diversity. The Commission requested FAO to undertake, subject to the availability of financial resources, a feasibility study on the availability of, access to, and optimal use of genomic and/or breed demographic and pedigree data to estimate parameters that may be suitable to complement breed population size data as indicators for monitoring the genetic diversity within livestock breeds.[4]

5. In response to the request by the Commission, FAO utilized Regular Programme Funds to contract with the University of Natural Resources and Applied Life Sciences, Vienna, Austria (BOKU) to convene an expert group to undertake the study. This document summarizes the study. The document has been prepared by Pamela Burger (Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Vienna, Austria), Licia Colli (Università Cattolica del Sacro Cuore, Piacenza, Italy), Ino Curik (Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia), Coralie Danchin-Burge (Institut de l'Elevage, Paris, France), Christian Looft (Department of Animal Breeding and Husbandry, University of Applied Science Neubrandenburg, Germany), Gábor Mészáros, Johann Soelkner and Chang Xu (BOKU, Vienna, Austria), Dominique Ouedraogo (Joseph KI-ZERBO University, Burkina Faso), Ben Rosen (Animal Genomics and Improvement Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, Maryland, United States of America), Yuri Tani Utsunomiya (Department of Production and Animal Health, School of Veterinary Medicine, São Paulo State University, Brazil), Jack Windig (Wageningen Livestock Research, Animal Breeding and Genomics, Wageningen University and Research, Netherlands) and FAO staff members. The content of the review is entirely the responsibility of the authors, and does not necessarily represent the views of FAO or its Members.

[2] CGRFA-18/21/10.2/Inf.2.
[3] CGRFA-18/21/Report, paragraph 74.
[4] CGRFA-18/21/Report, paragraph 76.

**Review of demographic, pedigree and genomic measures of genetic variation within livestock populations**

Pamela Burger, Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Vienna, Austria

Licia Colli, Università Cattolica del Sacro Cuore, Piacenza, Italy

Ino Curik, Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

Coralie Danchin-Burge, Institut de l'Elevage, Paris, France

Christian Looft Department of Animal Breeding and Husbandry, University of Applied Science Neubrandenburg, Germany

Gabor Meszaros, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Dominique Ouedraogo, University Center of Ziniaré / Joseph KI-ZERBO University, Burkina Faso

Ben Rosen, Animal Genomics and Improvement Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, Maryland, United States of America

Johann Soelkner, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Yuri Tani Utsunomiya, Department of Production and Animal Health, School of Veterinary Medicine, São Paulo State University, Brazil

Jack Windig, Wageningen Livestock Research, Animal Breeding and Genomics, Wageningen University and Research, Netherlands

Chang Xu, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

## I. EXECUTIVE SUMMARY

Genetic diversity within livestock populations is crucial for keeping them genetically healthy, and providing opportunities for improving their productivity and their capacity to adapt to changes in their production environments. Increased genetic diversity of populations is considered beneficial for the fitness and welfare of individuals, while a decrease of genetic variation will lead to an increase in inbreeding and thus the occurrence of inbreeding depression and genetic defects. It is therefore important to provide indicators of diversity when monitoring livestock populations, as is being done by using the data in the Domestic Animal Diversity Information System (DAD-IS) and/or local databanks. Population size data in terms of total number of animals is not sufficient for that purpose.

This document provides an overview on the use of breed demographic, pedigree and genomic data to estimate parameters that may be suitable to complement breed population size information indicators for monitoring the genetic diversity within livestock populations. It also gives recommendations on strategies of reporting within population diversity indicators in DAD-IS.

Effective population size (Ne) is the key indicator proposed and it can be estimated from demographic, pedigree and genomic information sources. From demographic data, Ne is typically derived from the expected change of inbreeding per generation (ΔF) based on numbers of male and female breeding animals. With pedigree data, inbreeding coefficients are calculated and Ne size is again derived from ΔF over generations. With genomic data, two most prominent methods to calculate Ne are (i) the use of ΔF, and (ii) estimation based on levels of linkage disequilibrium (LD).

The conclusions and recommendations of the authors of this review, regarding the monitoring of all livestock populations are as follows:

- Genetic variation within livestock populations should be measured, using at least one indicator of Ne.
- The monitoring should be performed on a regular basis at intervals that consider the generation interval of the population/species.
- An estimate of Ne should be included in the data fields of DAD-IS and the method employed should be indicated. Effective population size is the proposed indicator for genetic variation in the Draft post-2020 Global Biodiversity Framework of the Convention on Biological Diversity (CBD).[5]
- Molecular estimates of Ne are considerably more accurate compared to demographic indicators and pedigree-based measures, especially if pedigree information is of low quality.
- Whenever possible, samples of DNA for genetic variation analysis of a population, should be collected from at least 100 animals, to be genotyped with an array of 50 000 (50K) single nucleotide polymorphism (SNP) markers (or equivalent density). The animals in the genotyped sample should be selected including both sexes, as well as old and young individuals, from multiple generations (i.e. 50 young and 50 old).
- For the assignment of breeds to risk-status categories based on demographic information, the thresholds presented in the FAO guidelines on *In vivo conservation of animal genetic resources*[6] should be used.

[5] CBD/WG2020/3/INF/2.

[6] **FAO.** 2013. *In vivo conservation of animal genetic resources.* FAO Animal Production and Health Guidelines. No. 14. Rome. https://www.fao.org/3/i3327e/i3327e.pdf

## II. BACKGROUND

**Broad considerations on measures of genetic variation**

High genetic diversity in individual organisms and populations is generally considered beneficial for their health and fitness. Genetic diversity is both the basis for genetic selection and for adaptation to changing environmental conditions, such as those caused by climate change. Genetic changes related to these two processes are not possible without genetic variation. At the same time, a decrease of genetic variation may lead to an increase in inbreeding and the occurrence of genetic defects. There are many ways in which genetic variation can be measured and monitored and with the development of molecular techniques, even more possibilities have arisen.

**Processes influencing genetic variation**

*Mutation*, a (generally) random change of the DNA sequence, is the ultimate source of all genetic variation. Theoretically, the effect of beneficial mutations is predicted to be generally small (Orr, 2004), and genome wide association studies (GWAS) confirm that quantitative trait loci (QTL) with large effects are rare. At the same time, the frequency of detrimental mutations with a large or even lethal effect tends to remain small or be quickly reduced in the population by selection (purging). As a result, traits tend to be influenced by many loci, each with generally a low minor allele frequency and a small effect. Because the frequency of mutation occurrence is low (i.e. in the order of $10^{-6}$ per locus per generation), allele frequencies change very slowly due to this source of variation. Therefore, mutations have tended to be mostly ignored in animal breeding, at least until the most recent years (Mulder *et al.*, 2019).

*Genetic drift* is the random change in allele frequencies that occurs when a new generation is born. In diploid organisms an offspring inherits only one of the two copies of a locus from each parent. Consequently, purely by chance one of the copies may get lost if none of the offspring of an individual inherits that copy. In a population this process leads to random changes in allele frequencies. However, when an allele reaches a frequency of 100 percent or 0 percent, its frequency cannot change anymore, i.e. the allele is either fixed or eliminated. After many generations, without other processes influencing frequencies, all alleles will become either fixed or eliminated by genetic drift. Thus, in the long run, genetic drift always decreases genetic variation. However, the speed at which this happens depends on the number of animals reproducing in a population. The smaller the number of reproducing animals the greater the genetic drift and the faster the variation is lost. Genetic drift acts over the whole genome and influences all variable loci in a similar way - the direction of allele frequency change being random.

*Selection*, unlike genetic drift, is non-random. It acts on specific loci that are associated with traits under selection. Moreover, the direction of allele frequency change is predictable, i.e. the frequencies of beneficial alleles will increase, and those of detrimental alleles will decrease. Selection can be either artificial or natural. With artificial selection, humans choose animals that better fit the breeding goal; while natural selection favors those animals with a greater fitness/survival. Like genetic drift, selection tends to decrease genetic variation in the long run, fixing beneficial alleles and eliminating (purging) detrimental alleles. Although selection acts on specific loci, allelic frequencies of neighboring loci may also change, in a process known as “hitchhiking”. When a specific allele coincides more often with the allele under selection, i.e. is in linkage disequilibrium (LD) (see Box 1.), its frequency will change as well.

BOX 1
**Glossary on measures and concepts**

**Ascertainment bias:** systematic distortion in estimates of molecular genetic parameters (such as allelic frequencies) due to irregularities in the process used to identify the markers used for the genotypic assay. For instance, many single nucleotide polymorphism markers (SNPs) in large panels were selected according to their high minor allele frequency in international transboundary breeds and can thus underestimate the relative diversity in other breeds.

**Coancestry (coefficient):** (abbreviated *f* and also known as the "kinship" or "kinship coefficient") the probability that a randomly selected allele from two individuals (at the same locus) is identical by descent (IBD) from a common ancestor.

**Effective population size (Ne):** the size of a hypothetical idealized population (population with equal numbers of males and females, contributing uniform numbers of progeny, and not subject to other forces that change genetic diversity, such as mutation, migration and selection) that would generate the same amount of genetic drift or change of inbreeding as the population under study.

**Genetic drift:** (or simply "drift") the change in the frequency of an allele in a population due to random sampling of alleles during gamete formation.

**Genetic erosion:** the decrease of genetic variation in a population due to genetic drift and inbreeding.

**Homozygosity/Heterozygosity:** the condition in which the two alleles at a given locus are the same/different.

**Identity by descent (IBD):** homozygosity due to a common ancestor.

**Identity by descent (IBS):** homozygosity due to chance, as a function of the frequency of the homozygous allele in the general population.

**Inbreeding:** the mating of relatives. **Inbreeding coefficient** (F) is the probability that the alleles at any given locus are identical because they were each inherited from a common ancestor of the two parents.

**Inbreeding depression:** the reduction in fitness or performance due to the negative effects of inbreeding.

**Kinship:** see coancestry.

**Linkage disequilibrium (LD):** a non-random association between the alleles carried at different loci by an individual. This usually occurs because two loci are located closely together on the same chromosome.

**Microsatellite: (also known as STR *simple tandem repeat* or SSR *simple sequence repeat*) t**andem DNA repeat of a 2 to 5 base pair unit. In most cases, the repeat unit is the dinucleotide CA. The number of repeats of a given microsatellite locus is often polymorphic within populations, in which case the microsatellite may serve as a genetic marker. .

**Principal component analysis (PCA):** a method for analysis of a set of variables, such as allele frequencies, by calculation of a new set of statistically independent coordinates that each corresponds to a weighted combination of the original variables in such a way that each coordinate captures as much variation in the original variables as possible. In many datasets, a small number of coordinates may explain a large proportion of the initial variability, thus increasing efficiency. Plotting the distribution of individuals or breeds in a graph of the first two or three coordinates allows for simple visualization of the pattern of diversity.

**Runs of homozygosity (ROH):** contiguous regions of the genome that are homozygous across all sites.

**Single nucleotide polymorphism (SNP):** variation resulting from a point mutation and most often

corresponding to a biallelic (having two different alleles) marker.

*Migration* usually introduces genetic variation into populations. If the migrants carry alleles that are not present, or are very rare in the population in which they enter, the genetic variation in the introgressed population will increase. It is important to note that if a population is split into several isolated subpopulations, the allele frequencies will change randomly within each subpopulation due to genetic drift. Consequently, subpopulations will differentiate genetically as soon as there is some form of isolation. In other words, subdivision can maintain genetic variation between populations while genetic variation within populations may disappear. Migration through crossing of populations, in turn, will generally increase variation within populations, but will decrease variation between populations. Given the potential impact that this process may have on genetic variation within and between populations, it is important to get an overview of the genetic history and structure of the populations under study.

These four processes (mutation, genetic drift, selection and migration) do not act independently. For example, under strong selection, the number of breeding animals used will be small relative to the overall population, and because of that genetic drift will be increased. Another example is that migration of high production sires into a population will initially increase genetic variation, but subsequent selection for these sires and their offspring will reduce genetic variation in the long term, often at the expense of the alleles of the original population. In a stable population, a balance may be reached when the rate at which alleles are generated by mutation equals the rate at which alleles are eliminated by drift and selection. Where this equilibrium lies depends on the (effective) size of the populations and the type of alleles considered, e.g. neutral, deleterious or beneficial, recessive or dominant, and with different strengths of their effect. Typically, in a large population at equilibrium, recessive detrimental alleles can be present at low frequencies.

**Effects of loss of genetic variation**

Various drivers such as population fragmentation or rapid change in population size can influence the processes described above, leading to reduction of genetic variation (Leroy *et al.*, 2018). In practice, such a reduction is expected to have two main outcomes: (i) a decrease in fitness, and (ii) a loss of adaptive potential. Reduction of response to selection or adaptive potential has been well studied in animal models (Hoffmann *et al*., 2017), however there is limited literature on actual impacts in livestock populations, outside of the fixation of specific discrete traits. Meta-analyses have suggested that genetic drift and selection have little effect on the adaptive potential in nature (Wood *et al.*, 2017). Therefore, this section of the document will focus on the effects of genetic erosion on fitness and production traits.

The expression of detrimental alleles is the main mechanism behind the reduction of fitness in populations that experience a loss of genetic variation. In all populations, detrimental alleles can be present at low frequencies under the mutation/selection/drift balance. This applies especially to recessive alleles that are almost never expressed in a population when at low frequencies (i.e. $p^2$, the frequency of homozygotes that express the disease, can be so small that no homozygous individuals occur). Consequently, phenotypic selection cannot remove detrimental recessive alleles efficiently, because most or all of them occur in the heterozygote form. Due to genetic drift or the selection hitchhiking effect, an increase in frequency of detrimental recessive alleles can occur unexpectedly, leading to a random emergence of deleterious phenotypes or maladaptive traits, despite selection working against them. There are currently a large number of detrimental traits with associated genes and variants cataloged, for both livestock and companion animal species (OMIA, 2022). An example is found in Holstein cattle, where two genetic defects called BLAD (bovine leukocyte adhesion deficiency) and CVM (Complex vertebral malformation) re-emerged around the year 2000. These defects could be linked to a single sire called “Bell”, who had high genetic merit for productivity and whose lineage was therefore spread all around the world.

The emergence of deleterious phenotypes or maladaptive traits in relation to a limited number of variants with large detrimental effects is generally differentiated from the phenomenon of inbreeding depression, although the underlying mechanisms are basically the same (i.e. expression of detrimental alleles).

Inbreeding depression is defined as a decrease in fitness or performance of individuals in relation to mating between relatives. It is generally measured in terms of its impact on a quantitative trait (such as milk production, litter size or fitness in general) and it is caused by detrimental effects at many homozygous loci, each with a very small impact (Goddard 2022), which are thus difficult to detect (Charlesworth & Willis 2009). Inbreeding depression is known in all livestock species, and generally affects all investigated traits (Leroy, 2014; Doekes *et al*., 2021). Moreover, the causal loci seem to be spread across the whole genome (Pryce *et al.* 2014). The combined effects of all detrimental alleles present in a population form what is known as the "genetic load". The fates of detrimental alleles and populations affected by them depends on the strength of their effects. Those with a strong or even lethal effect can usually be quickly eliminated. However, when drift increases their frequency faster than selection can eliminate them, the population may go extinct. Alleles with small effects will be less quickly eliminated; they may even become fixed in the population without the population going extinct. However, if the loss of genetic variation continues, other detrimental alleles may accumulate. After prolonged inbreeding fertility may permanently decrease (e.g. observed in Kakapo, Northern Rhinoceros, Saarloos Wolfdog), leading to smaller population sizes, more inbreeding, and eventually to a population "meltdown". This process is known as the "extinction vortex". This is one reason for the importance of monitoring genetic variation, so that preventive measures can be taken before populations enter the extinction vortex.

**Measuring genetic variation of populations**

As outlined above, efficient monitoring of genetic variation is a prerequisite for the sustainable management of animal genetic resources. The monitoring of trends and risks, indeed, allows the raising of awareness so breeders can take action before decrease of genetic variation impacts the fitness of a given breed and puts its very existence at stake. Three sources of information are available to monitor genetic variation. Demographic parameters provide insight into factors driving the loss of genetic variation (e.g. low population size, overuse of popular sires). Pedigree approaches use the knowledge on ancestry to infer the transmission of variants at a locus, allowing the estimation of parameters related to genetic variation such as inbreeding coefficients. Finally, molecular approaches provide direct insight on genomic variation, with information on allele frequencies and homozygosity. These approaches have different merits and limitations, and provide a diversity of measures with different properties, from which a choice has to be made when investigating and monitoring populations.

To be useful, measures of genetic variation need to have certain properties. In the first place, they need to be reliable, easy to understand, and measure the real situation properly. Furthermore, measures should be comparable across species, data sources, populations and time. It is also important that they are easy to determine, easy to interpret, and informative as an early warning indicator in case genetic variation decreases to dangerous levels. As an illustration, measures such as homozygosity or inbreeding may intuitively seem among the best measures of genetic variation. On one hand, both concepts are of utmost importance for the study of genetic variation; homozygosity refers to the occurrence of alleles that are identical (identity by state) at a certain locus for a given individual, while inbreeding is defined here as a measure of shared ancestry in the paternal and maternal lineages (Templeton and Read, 1994). The coefficient of inbreeding (usually abbreviated as F) is defined as the probability of homozygosity due to a common ancestor (the probability of identity by descent). On the other hand, measures such as homozygosity or coancestry can be hard to compare across populations, for example due to different pedigree depths or different marker sets. Genetic load may also be quite different between populations, depending on their history of inbreeding and selection, and consequently a high homozygosity or inbreeding level may entail a different risk depending on the population.

Taking the advantages and limitations of these basic measures of genetic variation into context, the effective population size (Ne) has been considered as one of the best metrics of genetic variation and erosion. The Ne is defined as the size of an "idealized" population (population of constant size in which any member has the same chance to reproduce with any other member) that would result in the same amount of genetic drift or change of inbreeding as the population under study. Besides being quite intuitive in terms of interpretation, a big advantage of Ne is that it can be estimated on the basis of

various data sources (i.e. demography, pedigree, molecular marker sets) or signals (e.g. inbreeding, allelic variance). The Ne is also related to the inbreeding within a population through the formula $Ne = 1/2\Delta F$, with $\Delta F$ being the change in the rate of inbreeding per generation in the population. This is very useful, since the increase in inbreeding approaches the rate at which alleles are eliminated or fixed in a population (genetic drift), while also being directly linked to the increase in risk that genetic defects will be expressed. In other words, Ne and $\Delta F$ both measure the loss of genetic diversity occurring in a population. However, Ne as a measure of genetic variation is not without limits. First, as in most populations, genetic variability change is not constant over time; the estimate of Ne can be highly dependent on the period chosen for the analysis. For instance, in many pedigreed dog breeds inbreeding rates have decreased over the past twenty years, as a result of measures taken by breeding organizations to reduce mating of relatives. Consequently, for a given breed, estimates of inbreeding rates considering the most recent generations are low or even negative compared to the estimates considering a longer time period (Windig & Hulsegge, 2021). Also, as the different methods used to estimate Ne consider different assumptions and time periods, mixing different approaches may lead to misleading interpretations of the results.

## III. DEMOGRAPHIC APPROACHES

Demographic measures do not provide genetic information, but give indications on the underlying causes behind the changes in genetic variation present in a population. This information may allow understanding how genetic erosion and inbreeding rates are shaped, and can even provide rough approximates of Ne. However, demographic parameters cannot replace direct information on genetic variation such as from pedigree or molecular information. This section explains how demographic parameters can be used to infer information on genetic variation of the populations.

**Population census size**

Population size is a direct indicator of the endangerment status related to demographic stochasticity, that is, extinction due to environmental and random catastrophic events (Frankham, 1995a,b; 1996). Census size is also positively correlated to the genetic diversity of a population and its evolutionary potential. In large populations, more genetic variants are expected to be present than in small populations (Gandini *et al*., 2004). As not all animals reproduce and therefore contribute to future genetic diversity, the number of active breeding females and males is therefore a better indicator than the total number of animals of any age. Demographic trends over time provide additional clues to the fate of a population in terms of genetic variation. High variation of family size, which can be caused by artificial insemination (AI) systems in livestock production, also contributes strongly to the decrease of genetic diversity (Bruford *et al*., 2015). The reproductive capacity of a species affects the ability of a population to recover. If the habitat of an endangered breed is geographically limited, outbreaks of infectious diseases may put the species at high risk of extinction. The FAO guidelines on *In vivo conservation of animal genetic resources* (FAO, 2013) describe in detail the impact of the major factors influencing risk of extinction and categorizes demographic endangerment status. Categories of risk status of living populations are Critical, Endangered, Vulnerable, Not at risk, Unknown. In its most basic form, assignment to risk status depends on thresholds for census size and reproductive capacity. They can also account for the number of breeding males and females, population trend, and level of pure breeding (see Tables 3, 4, 5 of the document, pp. 47-48).

DAD-IS is a tool for monitoring the diversity of animal genetic resources and currently has information on about 8 900 breeds of 37 livestock species in 182 countries, including their risk status. Demographic factors that can be entered are population size (minimum-maximum), trend (increasing, stable, decreasing), breeding males, breeding females, breeding females registered in herdbooks, purebred breeding females, herd size (average), the use of AI, breeding males in AI, males in natural service, and reliability of demographic data (very reliable, reliable, not reliable). The input of minimal and maximal population data is mandatory, all other factors are optional. For a large part of the national breed populations in the system only these two indicators are given, which restrict the information to infer

actual genetic diversity of those national breed populations. For instance, data on breeding females and males (used for artificial insemination and natural mating) are completed for only approximately 30 percent of the 15 000 livestock populations in DAD-IS, and are rarely in an updated state (accessed 15 September 2022).

**Effective population size**

As previously discussed, the Wright-Fisher idealized population considered for the estimation of Ne is under random mating and no influence of mutation, migration, and selection (Wright, 1931). Some demographic parameters can provide information on the influence of those factors and therefore be used to approximate Ne.

*Census size*. Because population census size is correlated with genetic variation, it can also be used to estimate Ne in a very rough manner. The equation Ne = 0.10*Nc, where Nc is census size has been proposed for the CBD post-2020 Global Biodiversity Framework, as it is around the mean across results from a wide range of species (Hoban *et al*., 2020). However, this equation may not be appropriate for livestock populations; Hall (2016) reported a median value of 0.03 for the ratio of Nc/Ne across a collection of studies involving more than 500 breeds.

*Sex ratio, in connection to population size*. Data on the actual number of reproducing males and females has been used as a first proxy indicator of Ne (Allendorf *et al*., 2022). The most commonly used estimator for Ne is based on the unequal sex ratio using the formula:

$$Ne = 4 (Nm * Nf) / (Nm + Nf),$$

where, Nm is the number of reproducing males and Nf is the number of reproducing females.
In livestock species, the number of breeding males is usually much lower than the number of females used for reproduction, so Ne is largely determined by the number of males used.

*Variance of family size, prolificacy of females*. Unequal contribution of potential parents to the genetic make-up of the next generation is another important factor that defines Ne. The corresponding formula is the following:

$$N_e = (8Nm - 4) / (Vkm + Vkf + 4),$$

where, Nm is the number of breeding animals and Vkm and Vkf are the variances of the number of offspring of potential male and female parents, respectively, including non-reproducing individuals. Very high Vkm values occur when few sires have a large number of offspring, as is usually the case with AI. Since Vkm is in the denominator of the formula, Ne will be small in such cases. When Vkm = Vkf, the formula is naturally simplified.

*Variable population size across generations (effect of bottlenecks)*. Fluctuations in population size over time (generations) also affect Ne. This is especially the case when the number of reproducing individuals becomes very small and later increases again (bottleneck). To account for this case, the following formula can be used:

$$Ne = t / (1/Ne_1 + 1/Ne_2 + 1/Ne_3 + \ldots + 1/Ne_t) ,$$

Where, Ne is the effective population size over a period of t generations, and $Ne_1$, $Ne_2$, $Ne_3$, and $Ne_t$ are the effective population sizes of all generations considered in the period of interest. This formula should be used when a genetic bottleneck is suspected. This is the case when the Ne varies greatly (decreases and increases) in successive generations. Note that $Ne_1$, $Ne_2$, $Ne_3$, and $Ne_t$ can be estimated using either of the two formulas presented above, taking into account sex ratio or variation of family size. The same formula should be applied for all generations.

It has to be noted that demographic estimates of Ne, and especially the sex-ratio based methods, are merely approximations, and generally yield overestimated values of Ne (Leroy *et al.*, 2013). On the other hand, data required for the sex-ratio Ne estimate (e.g. numbers of males and females reproducing) are usually much easier to obtain than for the family size variation method. For instance, DAD-IS does not include a field for variance of family size. For details and extensions of the Ne formulas presented here, see Falconer and Mackay (1996), Caballero (2020), Allendorf *et al.*, (2022).

**Generation interval as a demographic parameter**

Generation interval (L) is defined as the average age of parents at the birth of their offspring. It is an important parameter affecting Ne and inbreeding rate during the observation period. In contrast to genetic improvement, where a decrease in L leads to a higher response to selection per year, an increase in L is desirable in management of genetic variation. In most cases an increase in L leads to an increase in Ne or a decrease in the rate of inbreeding over time. While most demographic indicators for Ne do not take into account the generation interval, indicators based on pedigree data usually do if the birth years of the animals are available.

**Use of demographic measures for genetic variability monitoring**

Despite providing only crude estimates of genetic variability, demographic information is of paramount importance for the management of livestock breeds in general. First, because a large number of factors (e.g. epidemics, low number of farms keeping the breed, localized distribution of breeds, low number of animals, high average age of livestock owners) that can contribute to population extinction are more linked to demographic parameters than to genetic variability. Second, because demographic data can provide valuable information on the causes behind the loss of genetic variation. Despite being probably simpler to collect than pedigree or molecular information, the collection of demographic information is still challenging in many countries. When an animal identification system or regular breed census is not available, cost-effective survey approaches should be implemented to estimate population census size and demographic related measures on a regular basis (FAO, 2004).[7] The thresholds in *In vivo conservation of animal genetic resources* (FAO, 2013) constitute a solid reference to classify demographic risk status and decide on management or support measures based on this classification.

## IV. PEDIGREE BASED APPROACHES

Over the last two centuries, centralized genealogical registries have become one of the pillars of breeding programmes, and for the unification of breeders around a specific livestock breed population, allowing the assessment of kinship relationships among registered individuals. This makes pedigree data a powerful source of information for the analysis of genetic variation and structure of a given registered population. This section will focus on the properties of pedigree approaches and related indicators for the monitoring of genetic variation.

**Properties and limits of pedigree approaches**

In terms of genetic variability, the knowledge of parent-offspring relationships allows the tracing of gene transmission from generation to generation for an entire population, based on the principle that each parent transfers 50 percent of its DNA to an offspring. It is therefore possible to infer the polymorphism of a given locus, neutral and without mutation, using the probability of either gene identity (a measure of the proportion of genes that are identical within an individual or a population), which allows the estimation of inbreeding and kinship coefficients, or gene origins (a measure of the proportion of genes that have been inherited from a specific individual or subpopulation).

Pedigree approaches have limitations, however. The first limitation is that analyses are restricted to a specific subset of individuals, i.e. those that are currently registered in a herdbook or have been registered in the past (Leroy, 2011). Therefore, an exhaustive breed analysis requires that all individuals

[7] FAO, 2022. Collection and estimation of population size data for risk classification in DAD-IS = A sampling methodology. Other documents.

are registered in the herdbook, which is not the case with most breeds. Also, investigations are limited to the period of time during which genealogies have been followed. Depending on the breed, this period may range from a few years to more than a century. As pedigree knowledge may differ over different ancestral lineages due to incomplete registrations, or to the fact that a herdbook may be open for animals whose parents were not registered, heterogenous pedigree knowledge may introduce a bias in pedigree analysis, as relationships between some individuals may not be fully taken into account. Different metrics are used to assess pedigree depth and gaps. Unknown pedigrees may be inferred to help overcome this bias, while some metrics have been designed to take into account these issues. *Equivalent complete generations* (EqG), i.e. the proportion of ancestors that are known, summed over each generation, constitutes probably the most common metric to assess pedigree completeness. Other metrics such as number of complete generations traced or the maximum number of generations traced can also be used to assess contrasts between pedigree depth and pedigree gaps.

Existence of pedigree errors may constitute another source of bias in the calculation of indicators of genetic variation. However, among livestock breeds, the extent of pedigree errors is typically less than 10 percent (Leroy *et al.*, 2012), whereas Oliehoek and Bijma (2009) observed that errors will have an impact of relevance to conservation decisions only if this percentage exceeds 15 percent.

Finally, it is important to underline that pedigree approaches are based on expectations, thus ignoring stochastic variation, and generally assume neutrality, and therefore cannot be used to reliably infer genetic variation for loci under selection.

**Measures of genetic variation**

When considering gene identity, two metrics are of utmost importance, namely the inbreeding (*F*) and kinship, a.k.a. coancestry, ($\phi$) coefficients. Both indicators measure the probability of two alleles of a given locus to be IBD, either for a single individual, or between two individuals, respectively. By definition, the coancestry coefficient between two individuals is equal to the expected inbreeding coefficient of a potential offspring of the two individuals. Averaged over a population, both metrics are expected to differ proportionally to an eventual deviation from Hardy-Weinberg equilibrium. As a consequence, the scientific literature recommends to focus on the minimization of average coancestry for monitoring and conservation of genetic variability at the breed level (Baumung and Solkner, 2003; Caballero and Toro, 2000). As IBD is among other things a function of genetic drift that accumulates over generations, pedigree inbreeding and coancestry coefficients tend to increase with pedigree knowledge, which implies that the use of those metrics for comparison of populations needs to consider differences in pedigree completeness among the populations being considered.

Approaches using probabilities of gene origin generally consider the contributions of founders, ancestors or genomes (using their respective effective numbers), which may be of interest for the identification of genetic bottlenecks or to investigate the impact of genetic drift within a given breed (Boichard *et al.*, 1997), those parameters being differentially affected by the extent of pedigree knowledge within the studied population.

Different approaches may be used to investigate population structure with pedigree information. As previously underlined, the simple comparison between average inbreeding and kinship may allow one to identify deviations from Hardy Weinberg equilibrium, which are often linked to the existence of subpopulations or intentional inbreeding (Leroy *et al.*, 2013). Deeper investigation may be considered, using principal component analysis (PCA) (see later) or phylogenetic approaches applied to the coancestry matrix (Pirault *et al.*, 2013).

Pedigree based Ne metrics logically use the increase in the rate of IBD over time and generations. They differ on the fact that either inbreeding or kinship are considered and how the period of time or heterogeneity in pedigree completeness are considered. Most common approaches consider the simple increase or linear regression of IBD over a predetermined period (Leroy *et al.*, 2013; Windig & Hulsegge, 2021), or correct individual IBD coefficients by their EqG (Cervantes *et al.*, 2011; Leroy *et*

*al.*, 2020) (Table 1). A variety of software has been developed to analyze pedigree data and compute pedigree metrics, including Ne (See Annex 1).

**Table 1. Examples of approaches estimating effective population size with pedigree data** :

| Parameter used | Effective population size formula | IBD rate formula | Time scale considered |
|---|---|---|---|
| Inbreeding rate across generations | $N_{eF} = 1/2\overline{\Delta F}$ | $\Delta F_t = \frac{F_{t+1} - F_t}{1 - F_t}$ | Adjustable over a given period by linear regression |
| Individual inbreeding rate | | $\Delta F_i = 1 - \sqrt[EqG_i - 1]{1 - F_i}$ | Depends on pedigree knowledge |
| Individual kinship rate | $N_{eC} = 1/2\overline{\Delta C}$ | $\Delta C_{ij} = 1 - \sqrt[(EqG_i + EqG_j)/2]{1 - C_{ij}}$ | Depends on pedigree knowledge |
| Restricted individual kinship rate | | $\Delta C_{Rij} = 1 - \sqrt[(EqG_{Ri} + EqG_{Rj})/2]{1 - C_{Rij}}$ | Adjustable on a chosen number of generations |

t: Generation number; $F$, $\Delta F$: inbreeding and inbreeding rate; $C_{(R)}$, $\Delta C_{(R)}$: kinship ($_R$: restricted) and kinship rate; $EqG_{(R)}$: Equivalent Complete Generations ($_R$: restricted).

**Considerations on time scale**

As previously indicated, inbreeding and coancestry coefficients, as well as Ne measures to a lesser extent, are affected by the pedigree knowledge and time period considered, which may bring bias when monitoring evolution of indicators or comparing different breeds. For instance, in populations with deep pedigree knowledge, changes in breeding practices may lead to relative decreases in inbreeding and kinship, and thus impossibility to estimate Ne. For the sake of comparison, a possibility can be for instance to consider only a restricted number of generations (Leroy *et al.*, 2020). Besides this, given the existence of random fluctuations of inbreeding (and to a lesser extent kinship) levels, that are likely to occur in populations of limited size, it is advised to consider periods larger than two generations and use regression approaches (Pérez-Enciso 1995) rather than estimating Ne on the basis of two single time points.

**Use of pedigree metrics for genetic variability monitoring**

According to the *Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture* (FAO, 2015) the proportions of breeds covered at least partially by pedigree recording was 40 and 51 percent for exotic and locally adapted breeds, respectively, considering the "Big Five" species (cattle, goats, sheep, pigs and chickens). Also, for these five species, between 68 and 91 percent of countries indicated that genetic diversity studies based on pedigree have either not been implemented at all, or only to a small extent (less than 33 percent of breeds). In DAD-IS, herdbooks are reported for fewer than 20 percent of national breed populations. Therefore, genealogical estimates are expected to be potentially available for only a minority of breeds.

For breeds that have generalized genealogical registries, pedigree approaches may however provide useful indicators for breed managers. Pedigree is a common source of data for scientific literature focusing on characterization and monitoring of genetic variability in livestock breeds. For instance, the review of Hall (2016) aggregated from about 90 studies pedigree estimates of Ne from 321 breeds, 31 countries (including 7 countries outside Europe and North America), and 5 species (cattle, sheep, horses, pigs and goats). Especially for cattle and horses, an increasing number of breeding organizations have integrated pedigree metrics as decision support tools for breeders, allowing for instance assessment of the potential inbreeding of the offspring from a planned mating (e.g. ICBF, 2022). Monitoring systems providing indicators of genetic variability at the breed level in a regular manner are quite rare, however, although some examples exist (see Box 2).

BOX 2
**VARUME: a genetic variability observatory for the ruminants and equid species in France.**

The aim of the project VARUME (Genetic Variability of RUMinants and Equine species) is to set up an observatory of the genetic variability of the French ruminant and equine species, based on pedigree data. It publishes indicators that assess breed genetic variability on a regular basis, by using a common method. In ruminants, the indicators are published on a yearly basis for all cattle and goat breeds with sufficient pedigree information (e.g. mean of at least 2.5 generations of depth) and every 3 years for meat sheep breeds. For instance, in 2021, over 80 individual breed PDF reports were generated. The reports are sent to each breeding organization and are made available publicly on the French Livestock Institute website (IDELE). The main users of the reports are the breeding organizations that are managing breeds with limited genetic variability, typically either endangered breeds or highly selected breeds such as the dairy cattle breeds. The reports allow them to evaluate the efficiency of their management practices to maintain the breeds' genetic variability. Journalists, researchers and teachers are also using the reports for topics linked to genetic variability. Last but not least, the reports are needed by the French Ministry of Agriculture, as some VARUME indicators can be chosen by the breeding organizations as an indicator of the efficiency of their selection programme. For the equid species, the indicators are calculated on demand by the breeding organizations by using the same methodology. In the equid sector, the breeding organizations generally prefer tools to assess genetic variability at the individual level, such as the calculation of the inbreeding level of a horse and its main contributing ancestor.

While combinations of different estimates may provide interesting insights on historical bottlenecks or substructure within populations, indicators considering the current levels or trends in coancestry probably constitute the most useful metrics for the monitoring of the genetic variability at the breed level. Even with the existence of metrics taking into account heterogeneity in pedigree knowledge and pedigree completeness, the corresponding genetic/time scale should be considered for the interpretation of the results and comparison with other breeds.

## V. GENOMIC APPROACHES

In contrast to demographic and pedigree data, genomic data provide direct information on the genetic variation present in a population. The field of genomics has developed rapidly in recent years, and will continue to do so. There are a myriad of methods available, not only to determine an animal's genotype for molecular markers, but also in mathematics and software to infer genetic variation from the molecular data. In this section, common genomic approaches for assessment of genetic variation and their properties are reviewed.

**Marker sets and sampling for genetic variability monitoring**
The technologies that enable the efficient sequencing of whole genomes have been accessible and affordable for little more than a decade. As such, the widespread use of genome-wide data to characterize genetic diversity is a relatively recent development. The availability of high quality reference genomes for nearly all livestock species provides many options for assaying diversity, from SNP arrays to various whole genome sequencing (WGS) approaches. A more thorough review of these options can be found in the FAO guide on *Genomic characterization of animal genetic resources* (FAO, 2022a).

Differences in the genomes of individuals are known as DNA sequence variants. There are many types of variants, which can differ in the number of nucleotides, structure, and complexity. Prior decades saw the widespread use of microsatellite markers for the characterization of genetic diversity. They possessed advantages such as being multiallelic and easily assayed via a combination of widespread technologies, namely the polymerase chain reaction (PCR) amplification and capillary electrophoresis. However, genotyping more than a handful of microsatellites per sample can be a very laborious task, thus limiting their utility in representing the diversity and complexity contained in entire livestock

genomes. Current technologies are much less laborious, and provide a much more detailed insight into genetic variation of livestock for a much lower price. Therefore the use of microsatellites should be reconsidered, and only done when practical circumstances strongly justify it, such as when a new population is to be added to an existing dataset of microsatellite genotypes. In general microsatellites should be phased out whenever possible, in favor of newer technologies such as SNP and WGS (FAO, 2022a).

The most abundant type of genetic variants are the SNPs, comprising single base-pair differences. To review, the combination of two parental alleles at a specific locus is called a genotype. With SNPs alleles can be either A, C, G or T, corresponding to the four nucleotides of DNA, i.e. adenine, cytosine, guanine and thymine, respectively. Similar to other genetic markers or genes, animals carrying two different parental alleles (e.g. genotype = TG) are called heterozygotes, while animals inheriting identical alleles from each parent (e.g. TT or GG) are called homozygotes. Furthermore, the specific arrangement or combination of inherited alleles across different loci in a parental chromosome copy is known as a haplotype. SNPs are typically bi-allelic with alleles that can occur in various frequencies within a population. Many analyses remove SNP loci that are polymorphic in less than 1 percent of the samples.

Even when WGS data provide nearly all nucleotides in the genome of each individual, smaller sets of SNPs are often used to represent the genome. Because alleles are inherited from parents linearly on chromosomes, organized in haplotypes, DNA markers are intended to serve as tags that indirectly capture information about their neighboring unobserved variant sites. Decreasing the resolution of a genomic analysis from the whole genome to a smaller number of variants has the objective to reduce both expenses and complexity. Importantly, the decision on which type and how many variants to select depends largely on the objectives and demands of the intended analysis, in addition to the project budget.

Naturally, both SNP arrays and WGS approaches come with a range of advantages and disadvantages. While commercial SNP arrays are available for the most common livestock species, there are economically marginalized species (e.g. yak, reindeer, guinea fowl) or more isolated populations, for which either no SNP arrays exist or the populations are genetically differentiated from the breeds that were used to design the commercial SNP arrays. In the latter cases, low coverage WGS might be the best way to avoid a potential problem called “ascertainment bias”, which is defined as the systematic deviation of population genetic statistics from theoretical expectations (Lachance & Tishkoff, 2013). Many SNPs that are polymorphic in the population on which the SNP array was originally designed may be less variable or even monomorphic in more distant populations. Vice versa, rare alleles existing in the target population might not be captured by the commercial SNP array. This is because the small sample size typically used to develop SNP arrays is more likely to identify loci where both alleles are relatively common than loci with rare alleles. Another reason apply WGS would be to gain information about the genetic load, which, as explained earlier, is the presence of disadvantageous (harmful) alleles in the selected individuals (Huber *et al*., 2020). Applying two complementary methods, Genomic Evolutionary Rate Profiling (GERP) scores and simple predictions of functional effects including loss-of-function (LOG) mutations, it is possible to estimate mutational load, a proxy for genetic load, from genomic data (Kutschera *et al*., 2022).

A major question relates to how many samples are needed to ensure that the sample represents the population and that the estimates are not biased. Different studies (Bhati *et al.*, 2020, Schmidtmann *et al.*, 2021, Mastrangelo *et al.*, 2018, Mukherjee *et al.*, 2018) are based on sometimes very different sample numbers (from tens to hundreds of animals). The sample size required depends on the objectives of the study and analyses to be undertaken. Based on these literature results, and because representative sampling cannot always be guaranteed for small populations, for the purpose of evaluating genetic variation it is suggested to sample at least 100 animals (possibly least related and preferentially sex and age balanced) for SNP-genotyping with a 50K array. The cost of genotyping 100 animals with a 50K array will be around USD 2 500 for most livestock species in many countries. The cost of sampling is not included in this estimate, however, and will be variable based on the strategy chosen.

Concerning the assessment of whether SNP genotyping or WGS should be applied, it must be noted that WGS generally generates more information, but the processing of the data is significantly more difficult compared to the processing of SNP data and costs are therefore much higher. A greater level of technical capacity is required as well. There are numerous very well established software programs available for processing SNP data in diversity studies that can be applied in a user-friendly manner (FAO, 2022a). Nevertheless, if the number of samples is very small, it is recommended to move from SNP analysis to WGS.

**Population structure assessment**

An overview of the population structure is very beneficial, especially for populations with unknown genetic variation and diversity. There are well-known measures allowing the assessment of the degree of genetic structure within a population, such as the F-statistics (Wright, 1951). Population structure can uncover the patterns of relatedness, population substructure, or identify individuals that do not truly belong to the breed being studies (e.g. crossbreds, mislabelled individuals from other populations). Establishing a visual picture of the genetic distribution of animals can provide a more complete view of the population structure and is therefore recommended.

The most widely known and used form of such population structure visualizations is the principal component analysis (PCA). The principle of this approach is to reduce a complex dataset into a limited number of uncorrelated variables, called "principal components". The two or three principal components that explain the most variation of the dataset are plotted for each individual. The relative distance on this plot shows the similarity between the individuals (e.g. Gurgul *et al.*, 2021). Another approach is to use Bayesian inference to assign individuals on a predetermined number of origins (Pritchard *et al.*, 2000), which can identify animals with putative origins that differ from expectations.

**Genomic methods and tools to estimate inbreeding levels**

While pedigree-based approaches have historically been the most commonly used approach to estimate inbreeding levels, with the development of dense marker sets, the genomic measures of inbreeding have become superior to pedigree-based estimates in most situations (Alemu *et al*., 2021). Several well-established methods are available to estimate the inbreeding levels of breeds by using genomic data. These methods consider, for instance, the correlation between uniting gametes, genomic relationships, excess of homozygosity or homozygous-by-descent segments. While different tools are being used in research studies, the most recognized methods to estimate genomic inbreeding levels are based on the use of the so-called "runs of homozygosity" (ROHs). The ROHs in the genome of an individual occur because of the inheritance of haplotypes from a common ancestor its two parents. Genomic tools consider all common ancestors simultaneously in the analysis of ROH. When identical haplotypes are passed down to offspring they manifest as long stretches of the genome that are entirely homozygous, potentially for millions of bases.

The genomic inbreeding based on ROHs is calculated as:

$$F_{ROH} = \sum ROH_i / L_{autosome}$$

Where, $F_{ROH}$ is the genomic inbreeding coefficient, $\sum ROH_i$ is the sum of the lengths of all ROH segments for individual i, and L is the length of the autosome genome covered by SNPs. In other words, the genomic inbreeding is the proportion of the genome that is autozygous (i.e. homozygous by descent).

On average, shorter ROHs come from demographic events dating back to a longer time in the past, because there has been more time for recombination to break the homozygous stretches, whereas longer segments are from more recent common ancestors. For this reason, one can choose to calculate inbreeding levels based on all ancestral information, or just on the recent past, which tend to mirror recent selection decisions. Following the rule of thumb that 1 centiMorgan corresponds to 1 Mb, ROH lengths of 16, 8, 4, 2, and 1 Mb refer to a common ancestor 3, 6, 12, 25, and 50 generations in the past, respectively. Also, in case of maladaptive carrier mating, the ROH segments could contain genetic

defects (e.g. Charlier *et al*., 2008), or otherwise contribute to lower performance or health status via inbreeding depression. Considering ROHs is especially important in small and endangered breeds, as such segments can more easily be inherited from both parents and in turn decrease the fitness of the population.

Several established and easy-to-use software programs are readily available for estimation of ROHs from SNP data (see Annex 1). Regardless of the software, some considerations must be given to data quality control prior to the analysis, as well as settings of crucial parameters for the analysis itself. These quality control and other settings are of utmost importance, as they can change the outcome of the analysis. The FAO guidelines on *Genomic characterization of animal genetic resources* (FAO, 2022a) provide step-by-step guidance for this quality control procedure. Some of these settings for parameters may differ based on the SNP data density or the purpose of the analysis (recent vs. ancient inbreeding). More thorough discussions of these topics, reviewing existing literature, including exact parameter settings and demonstrations of their effects are in Meyermans *et al.*, (2020), Ceballos *et al.*, (2018), Peripolli *et al.*, (2016) and other publications.

**Genomic methods and tools to estimate effective population size**

The two most prominent methods to calculate Ne based on genomic data are the use of inbreeding (ΔF), and the estimation based on the level of LD.

The $Ne_{\Delta F}$ is an indicator derived from inbreeding levels that can be estimated based on genomic inbreeding levels calculated by using ROHs. As discussed previously, the change in the inbreeding levels of animals between generations (ΔF) is an indicator of Ne, similarly to the computation of change in F from conventional pedigree data. Accordingly, the same equation $Ne = 1/2\Delta FROH$ can be used to compute NeΔF.

Established software are available to compute FROH, but no specific software tool is available for the entire process, up to NeΔF. When the FROH is computed for the entire population, however, one can compute the average ΔFROH for each generation, and follow up with the computation of the corresponding NeΔF, by regressing ΔFROH on generation number.

Estimation of Ne based on LD is an alternative way to provide valuable information on the population when pedigree information is not available or incomplete. Linkage Disequilibrium describes a non-random association of alleles at different loci and is a function of the recombination rate between these loci on a physical map. It can result from various demographic changes, like admixture and genetic drift, (Wright, 1931) or the "hitchhiking effect" during selection (Charlesworth *et al*., 1997; Maynard & Haigh, 2007). According to Sved (1971), LD between unlinked loci can arise from genetic drift from an isolated population with random mating. Hence, Ne can be estimated by the known relationship of and variance in LD among various loci, which can be calculated by allele frequency, which may reflect the genetic drift over generations. While there are many measures of LD, the most commonly used one is pairwise values. This LD-based estimation method has been widely used for conservation programmes and the study of the evolution of local populations (Waples & Do, 2010). With the rapid development of genotyping methods, software for prediction based solely on genomic data has also become available, with various software available for estimating historical changes from genotype data (see Annex 1).

**Molecular monitoring of breeds**

According to the *Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture* (FAO, 2015), among the main species (cattle, chicken, goat, pig, sheep), between 73 to 84 percent of countries indicated that within-breed molecular genetic diversity studies - had either not been implemented at all, or only to a low extent (less than 33 percent of breeds). This result suggests that obstacles likely exist for general monitoring of molecular monitoring of breeds worldwide.

A new survey is needed, however, given that the adoption is likely to have increased substantially in recent years. For preparation of this document, a small-scale assessment of 100 randomly selected breeds

from 11 members of the Intergovernmental Technical Working Group on Animal Genetic Resources was undertaken. Based on a review of the scientific literature and consultation with National Coordinators for the Management of Animal Genetic Resources from these countries, a genomic study has been undertaken within the country for 47 of the 100 breeds. When taking into consideration studies undertaken in other countries' national populations of transboundary breeds, 72 of the 100 breeds have been assessed genomically. These results suggest increased adoption of genomic tools since 2015.

Molecular approaches have reportedly been used at a similar level to pedigree data for the computation of Ne in scientific literature, and in the review of Hall (2016), about 30 studies had estimated NeLD for 203 breeds, 30 countries (including 9 countries outside Europe and North America), and 5 species (cattle, sheep, horses, pigs and goats).

Despite the accuracy that molecular measures provide, to our knowledge there is no monitoring system in place using this source of information to follow up genetic variation of livestock in a regular manner. This continuous monitoring is recommended, however, as the genetic variation of livestock breeds is expected to change over time. Depending on the result of such monitoring, breeders and breeding organizations could take action if necessary to maintain genetic variation and minimize increases in inbreeding.

## VI. USE OF GENETIC VARIATION MEASURES FROM DIFFERENT DATA SOURCES

On the basis of the information presented in the sections above, it is important to describe the extent to which the different sources of information display different properties that affect their utility for population monitoring (Table 2). For instance, demographic parameters provide unique and indispensable information for monitoring extinction risk related to demographic stochasticity, and also reveal the existence of factors that can drive changes in genetic variation. In the case of exhaustive genealogical registration, all demographic parameters of importance can be extracted from the pedigree file. Genomic information, if obtained with appropriate marker sets, sampling and approaches, is likely to provide the most accurate picture of genetic variation. However, if the purpose is population monitoring, genomic approaches should always be used in complementarity to demographic or pedigree analyses.

TABLE 2
**Properties and challenges related to the use of demographic, pedigree and molecular approaches for population monitoring**

| Data source | Main properties | Challenges | |
|---|---|---|---|
| | | Data collection | Applicability of results |
| Demographic information | Provides insight on demographic stochasticity and underlying causes behind the changes in genetic variation | Requires collection of data through breed censuses, surveys or animal identification systems | Estimates of genetic variation are basic interpolations which often underestimate loss of genetic variation |
| Pedigree information | Provides inferences on the genetic variation of selectively neutral loci and assuming no mutation, based on knowledge of parent-offspring relationships | Requires registration of pedigree information to be as complete as possible | Results do not consider mutation, Mendelian sampling and selection, and are prone to bias related to incomplete or incorrect pedigree |

| | | | |
|---|---|---|---|
| Genomic information | Yields information directly on genomic variation, but provide no direct information on demographic stochasticity | Requires accurate sampling in terms of individuals and markers | Choice of appropriate parameters used for analyses requires skill, and may yield inaccurate results if parameters are incorrect |

The priority assigned to a given approach depends largely on the situation of the breed under study. For instance, for a population with a very large number of registered animals with complete and correct pedigree information over several generations, demographic and genealogical measures will sufficiently describe the situation of the breed in terms of genetic variation, although complementary DNA-based analyses can provide additional details about other aspects of the population. For another population with only a small number of animals, and for which the pedigree relationships are largely unknown, basic characterization of demographic parameters complemented with genomic characterization would be necessary to obtain accurate assessment of the diversity and genetic structure of the population.

The use of measures for comparison across time periods and between different breeds and subsequent decision-making needs to be carefully considered. For instance, it has been shown that demographic predictors of Ne tend to provide upwardly biased estimates relative to genealogical or molecular approaches, due to the fact that the demographic approaches do not consider several factors of importance (Leroy *et al.*, 2013; Hall 2016). Therefore, the same rules of decision (e.g. Ne thresholds to prompt a given action) should not be applied indiscriminately for demographic versus pedigree/molecular estimates. Therefore, if measures of inbreeding or Ne from different approaches are to be used as global indicators of genetic diversity and/or collected in a breed monitoring database like DAD-IS, complementary information (i.e. estimation approach used, citation of a reference providing more information on details such as the subpopulation sampled) should be provided along with the data for the estimates.

## VII. CONCLUSIONS AND RECOMMENDATIONS

This review describes the potential of applying genomic, breed demographic and pedigree data to estimate parameters for monitoring genetic diversity within livestock breeds. The monitoring of genetic diversity has been possible for years based on conventional data sources, such as demographic data and pedigrees. Other possibilities have recently arisen with the increased availability and affordability of molecular techniques, most importantly SNP arrays. Regardless of the technique used for the monitoring, the inbreeding levels and Ne are important parameters, the knowledge of which could benefit the management of any livestock population.

The conclusions and recommendations of the author team of this review, regarding the monitoring of all livestock populations are as follows:

- Genetic variation within livestock populations should be measured, using at least one indicator of Ne.
- Livestock populations shall be monitored on a regular basis at time intervals considering the generation interval, although further study is needed to determine the optimal frequency of monitoring.
- An estimate of Ne should be included in the data fields of the Domestic Animal Diversity Information System (DAD-IS). Effective population size is the proposed indicator for genetic variation in the Draft CBD post-2020 Global Biodiversity Framework.
- Molecular tools provide the opportunity to obtain estimates of Ne that are considerably more accurate compared to demographic indicators and pedigree based measures, especially if pedigree information is of low quality.
- Whenever possible, samples of DNA for genetic diversity analysis of a population, should be collected from at least 100 animals, to be genotyped with a 50K SNP array (or equivalent

density). The animals in the genotyped sample should be selected including both sexes, as well as old and young individuals representing multiple generations.

- For the assignment of breeds to risk-status categories based on demographic information, the thresholds presented in the FAO guidelines on *In vivo conservation of animal genetic resources*[8] should be used.

## VIII. REFERENCES

**Allendorf, F. W., Funk, W.C., Aitken, S.N., Byrne, M., & Luikart, G.** 2022. *Conservation and the genomics of populations.* Oxford University Press.

**Alemu, S.W., Kadri, N. K., Harland, C., Faux, P., Charlier, C., Caballero, A., & Druet, T.** 2021. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. *Heredity*, 126: 410–423 https://doi.org/10.1038/s41437-020-00383-9

**Barker, J.S.F.** 1999. Conservation of livestock breed diversity. *Animal Genetic Resources*, 25: 33-43.

**Baumung, R. & Solkner, J.** 2003. Pedigree and marker information requirements to monitor genetic variability. *Genetics, Selection, Evolution*, 35: 369–383.

**Bhati, M., Kadri, N. K., Crysnanto, D., & Pausch, H.** 2020. Assessing genomic diversity and signatures of selection in Original Braunvieh cattle using whole-genome sequencing data. *BMC genomics*, 21(1): 27. https://doi.org/10.1186/s12864-020-6446-y

**Boichard, D., Maignel, L., & Verrier, E.** (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution*, 29(1): 5-23.

**Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F. J., Anaral, A.J. *et al*.** 2015. Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Frontiers in Genetics*, 6: 314.

**Caballero, A., & Toro, M.A.** 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetics research*, *75*(3): 331-343

**Ceballos, F.C., Hazelhurst, S. & Ramsay, M.** 2018. Assessing runs of homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, 19: 106 https://doi.org/10.1186/s12864-018-4489-0

**Cervantes, I., Goyache, F., Molina, A., Valera, M., & Gutiérrez, J.P.** 2011. Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics*, 1281: 56-63.

**Charlesworth, B., Nordborg, M., & Charlesworth, D.** 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70(2): 155-174.

**Charlesworth, D., & Willis, J.H.** 2009. The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11): 783-796.

**Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., ... & Georges, M.** 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics*, *40*(4): 449-454.

**FAO.** 2004. *Secondary guidelines for the development of management plans of the animal genetic resources at the national level*. FAO Publications, Rome.

**Doekes, H.P., Bijma, P., & Windig, J.J.** 2021. How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes*, 12(6): 926.

**Falconer, D.S., & Mackay, T.F.C.** 1996. *Introduction to Quantitative Genetics*, Ed 4. Longmans Green, Harlow, Essex, UK.

**FAO.** 2004. *Secondary guidelines for the development of management plans of the animal genetic resources at the national level*. FAO Publications, Rome.

**FAO.** 2013. *In vivo conservation of animal genetic resources.* FAO Animal Production and Health Guidelines. No. 14. Rome.

[8] **FAO.** 2013. *In vivo conservation of animal genetic resources.* FAO Animal Production and Health Guidelines. No. 14. Rome. https://www.fao.org/3/i3327e/i3327e.pdf

**FAO.** 2015. *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture.* edited by B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. www.fao.org/3/a-i4787e/index.html

**FAO.** 2022a. *Genomic characterization of animal genetic resources – Practical guide*. P. Ajmone-Marsan, P. Boettcher, L. Colli, C. Ginja, J. Kantanen & J.A. Lenstra, eds., FAO Animal Production and Health Guidelines, No. 32. Rome.

**FAO.** 2022b. *Innovations in cryoconservation of animal genetic resources – Practical guide.* J. Boes, P. Boettcher & M. Honkatukia, eds., FAO Animal Production and Health Guidelines, No. 30. Rome.

**Frankham, R.** 1995a. Conservation genetics. *Annual Review Of Genetics*, *29*(1): 305-327.

**Frankham, R.** 1995b. Inbreeding and extinction: a threshold effect. *Conservation Biology*, 9(4): 792-799.

**Frankham, R.** 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10(6): 1500-1508.

**Gandini, G.C., Ollivier, L., Danell, B., Distl, O., Georgoudis, A., Groeneveld, E., Martyniuk, E., van Aerndonk, J.A.M. & Woolliams, J. A.** 2004. Criteria to assess the degree of endangerment of livestock breeds in Europe. *Livestock Production Science*, 91(1-2): 173-182.

**Goddard, M.E.** 2022. Heterosis can be explained by thousands of partially recessive, deleterious mutations with very small effects at sites that are conserved across species. In: *12th World Congress on Genetics Applied to Livestock Production*, Rotterdam, The Netherlands. www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/67_013.pdf

**Gurgul, A., Jasielczuka, I. Miksza-Cybulska, A., Kawęcka, A. Szmatoła, T. & Krupiński, J.** 2021. Evaluation of genetic differentiation and genome-wide selection signatures in Polish local sheep breeds. *Livestock Science*, 251: 104635. https://doi.org/10.1016/j.livsci.2021.104635

**Hall, S.J.G.** 2016. Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. *Animal*, 10(11): 1778-1785.

**Hedrick, P.W., & Kalinowski, S.T.** 2000. Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics*, 31(1): 139–162. https://doi.org/10.1146/annurev.ecolsys.31.1.139

**Hoban, S., Bruford, M., Jackson, J.D.U., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, PA., Paz-Vinas, I. *et al*.** 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248: 108654.

**Hayes, B.J., Visscher, P.M., McPartlan, H.C., & Goddard, M.E.** 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome research*, 13(4): 635-643.

**Hoffmann, A.A., Sgrò, C.M., & Kristensen, T.N.** 2017. Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution*, 32(7): 506-517.

**Huber, C.D., Kim, B.Y., & Lohmueller, K.E**. 2020. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genetics*, 16(5): e1008827. https://doi.org/10.1371/journal.pgen.1008827

**ICBF.** 2022. Making use of the ICBF inbreeding checker. In: I*rish Cattle Breeding Federation*. Cork. Cited 3 October 2022. https://www.icbf.com/?p=18224

**Kutschera, V. E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D., Khoonsari, P. E., Nystedt, B., Dalén, L., & Díez-Del-Molino, D.** 2022. GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. *BMC Bioinformatics*, 23(1): 228. https://doi.org/10.1186/s12859-022-04757-0

**Lachance, J., & Tishkoff, S.A.** 2013. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. BioEssays : news and reviews in molecular, cellular and developmental biology, 35(9): 780–786. https://doi.org/10.1002/bies.201300014

**Leroy, G.** 2011. Genetic diversity, inbreeding and breeding practices in dogs: results from pedigree analyses. *The Veterinary Journal*, 189(2): 177-182.

**Leroy, G., Danchin-Burge, C., Palhière, I., Baumung, R., Fritz, S., Meriaux, J. C., & Gautier, M.** 2012. An ABC estimate of pedigree error rate: application in dog, sheep and cattle breeds. *Animal Genetics*, 43(3): 309-314.

**Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E., & Danchin-Burge, C.** 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genetics Selection Evolution*, 45(1): 1-10.

**Leroy, G., Carroll, E.L., Bruford, M.W., DeWoody, J.A., Strand, A., Waits, L., & Wang, J.** 2018. Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, 11(7): 1066-1083.

**Leroy, G., Gicquel, E., Boettcher, P., Besbes, B., Furre, S., Fernandez, J. & Baumung, R.** 2020. Coancestry rate's estimate of effective population size for genetic variability monitoring. *Conservation Genetics Resources*, 12(2): 275-283.

**Mastrangelo, S., Ciani, E., Ajmone Marsan, P., Bagnato, A., Battaglini, L., Bozzi, R., Carta, A. *et al.*** 2018. Conservation status and historical relatedness of Italian cattle breeds. *Genetics Selection Evolution*, 50(1): 1-16.

**Maynard, J., & Haigh, J.** 2007. The hitch-hiking effect of a favourable gene. *Genetics Research*, 89(5-6): 391-403.

**Meyermans, R., Gorssen, W., Buys, N., & Janssens, S.** 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21(1): 1-14.

**Mukherjee, A., Mukherjee, S., Dhakal, R., Mech, M., Longkumer, I., Haque, N., Vupru, K., *et al.*** 2018. High-density Genotyping reveals genomic characterization, population structure and genetic diversity of Indian Mithun (*Bos frontalis*). *Scientific Reports*, 8(1): 10316. https://doi.org/10.1038/s41598-018-28718-x

**Mulder, H.A., Lee, S.H., Clark, S., Hayes, B.J., & Van Der Werf, J.H.** 2019. The impact of genomic and traditional selection on the contribution of mutational variance to long-term selection response and genetic variance. *Genetics*, 213(2): 361-378.

**Oliehoek, P. A., & Bijma, P.** 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genetics Selection Evolution*, 41(1): 1-11.

**OMIA.** 2022. *Online Mendelian Inheritance in Animals*. University of Sydney. Cited 20 October 2022. www.omia.org

**Pérez-Enciso, M.** 1995. Use of the uncertain relationship matrix to compute effective population size. *Journal of Animal Breeding and Genetics*, 112(1-6): 327-332.

**Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., & Baldi, F**. 2017. Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, 48(3): 255-271.

**Pirault, P., Danvy, S., Verrier, E., & Leroy, G.** 2013. Genetic structure and gene flows within horses: A genealogical study at the French population scale. *PLoS One*, 8(4): e61544.

**Pritchard, J.K., Stephens, M., & Donnelly, P.** 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2): 945-959.

**Pryce, J. E., Haile-Mariam, M., Goddard, M. E., & Hayes, B. J.** 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution*, 46(1): 1-14.

**Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M., Hinrichs, D., & Thaller, G.** 2021. Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. *Genetics Selection Evolution*, 53(1): 23. https://doi.org/10.1186/s12711-021-00613-6

**Sved, J.A.** 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, 2(2): 125-141.

**Templeton, A. R., & Read, B.** 1994. Inbreeding: one word, several meanings, much confusion. *EXS*, 68: 91-105.

**Waples, R. S., & Do, C.H.I**. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3(3): 244-262.

**Windig, J. J., & Hulsegge, I.** 2021. Retriever and pointer: Software to evaluate inbreeding and genetic management in captive populations. *Animals*, 11(5): 1332.

**Wood, J.L., Yates, M.C., & Fraser, D.J.** 2016. Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, 9(5): 640-657.

**Wright, S.** 1931. Evolution in Mendelian populations. *Genetics*, 16(2): 97.

**Wright, S.** 1951. The genetical structure of populations. *Annals of Eugenics*. 15: 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Annex 1
**Examples of pedigree and genomic analysis software**

**PEDIGREE ANALYSIS SOFTWARE**

ENDOG
Gutiérrez, J. P., & Goyache, F. (2005). A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding and genetics*, *122*(3), 172-176.
https://webs.ucm.es/info/prodanim/html/JP_Web.htm

Pedig
Boichard, D. (2002). Pedig: a fortran package for pedigree analysis suited for large populations. In *7th world congress on genetics applied to livestock production* (p. Inconnu).
https://www6.jouy.inrae.fr/gabi_eng/Support-Expertise/Software/Pedig

PMx
Lacy, R. C., Ballou, J. D., & Pollak, J. P. (2012). PMx: software package for demographic and genetic analysis and management of pedigreed populations. *Methods in Ecology and Evolution*, *3*(2), 433-437.
https://scti.tools/pmx/

PyPedal
Cole, J. (2012). PyPedal, an open source software package for pedigree analysis. *Eur. Assoc. Anim. Prod. Proc*, *18*, 239.
http://pypedal.sourceforge.net/

Retriever
Windig, J. J., & Hulsegge, I. (2021). Retriever and pointer: Software to evaluate inbreeding and genetic management in captive populations. *Animals*, *11*(5), 1332.
https://genebankdata.cgn.wur.nl/software/Retriever/Retriever.html

**GENOMIC ANALYSIS SOFTWARE**

PLINK
Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.
https://www.cog-genomics.org/plink/

cgaTOH
Zhang L, Orloff MS, Reber S, Li S, Zhao Y, *et al.*, (2013) cgaTOH: Extended Approach for Identifying Tracts of Homozygosity. PLOS ONE
http://www.cs.kent.edu/~zhao/TOH/

detectRuns
Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network).
https://orca.cardiff.ac.uk/id/eprint/108906/

RZooRoH
Druet T, Gautier M. A model-based approach to characterize individual inbreeding at both global and local genomic scales. Mol Ecol. 2017;20.
https://cran.r-project.org/web/packages/RZooRoH/vignettes/zooroh-vignette.pdf

*SneP*
Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in genetics*, *6*, 109.
https://sourceforge.net/projects/snepnetrends/

GONe
Coombs, J. A., Letcher, B. H., & Nislow, K. H. (2012). GONe: software for estimating effective population size in species with generational overlap. *Molecular Ecology Resources*, *12*(1), 160-163.
https://bcrc.bio.umass.edu/pedigreesoftware/