# The significance of Effective Population Size (Ne) for monitoring genetic variability in breeding programs

Coralie Danchin, Johannes Geibel, Christian Reimer, Steffen Weigend, and the EURC EAB consortium, opinion paper

April 2024

There is a consensus that the genetic variability within breeds must be preserved. Indeed, high levels of inbreeding lead to two types of detrimental effects, the increased manifestation of genetic defects and a loss of beneficial variation resulting in inbreeding depression.

There is a particular concern for local breeds since most of them have limited numbers of breeding animals, therefore a restricted gene pool, and an often-decreasing trend in number of animals per breed. On the other side of the spectrum, the high selection intensity applied in the breeds that are the most used for farming, such as the Holstein breed in dairy cattle or Large White in pigs, does also negatively affect genetic variability.

Breeding organizations are therefore advised to monitor vulnerability of breeds towards loss of genetic variation using simple and accurate indicators. Regular monitoring of genetic variability should be fully integrated with the implementation of approved breeding programs by officially recognized breed societies[1].

Effective population size, Ne[2], is the most favored indicator today:

- it is fairly easy to understand - the bigger the better.
- it can be calculated either by using demographic, pedigree, or genomic data - although the definition of these calculators contains genetic assumptions that require special considerations, such as panmixia.
- it measures accurately the situation, with the limitations that its best use is when Ne can be compared over time.
- It is directly related to the increase of the inbreeding rate over generations.

[1] EU Animal Breeding legislation (EU) 2016/1012

[2] Equivalent size of a breed that would be structured with an equal number of males and females, which would contribute the same to the next generation by having the same number of offspring, with no other forces involved such as mutation or migration. This population would generate the same amount of inbreeding per generation than the breed we compared it with.

There are various methods to calculate $N_e$ for each type of data. Based on the FAO report[3], the following recommendations are given when a breeding organization wants to determine which data/method to use:

- Choice of data and temporality

The organization needs to assess which data are easiest to collect for a representative sample of the breed, across time, for a temporality equal at least to three generations interval of the breed (for species with generation interval that excess a year) or at least three years (short interval species).

Following the genetic variability of breed overtime is worthwhile if it can be done periodically. If the infrastructure of the breed organization allows for it, our recommendation is to use two successive generation intervals, for instance, in ruminants, two successive periods of 4 years.

- About genomic data

Genomic data are often presented as a way to overcome the impossibility of gathering reliable pedigree data; however, according to FAO recommendations, this means that a sample of at least 100 animals, with a balanced sex ratio, representative of the breeds' genetic variability and age balanced should be sampled, genotyped and analyzed periodically. Therefore, the breed organization needs to plan for specific fundings and staff devoted to this task on a regular basis. To increase the number of data, randomly sampled genotypes can come as a « by-product » of the increasing implementation of genomic selection in numerous breeds, including local ones, and routine parentage control based on SNP instead of microsatellites.

- Pedigree data

The use of pedigree data for calculating Ne instead of genomic data depends on the availability of accurate pedigrees and associated costs, the representativity of the animals recorded versus the total number of existing animals of the breed, and the pedigree depth. As far as the representativity of the breed is concerned, it is not relevant to suggest a minimum percentage of the breed with pedigrees since it will depend mostly on how the breed is structured. As an example, if only 10% of the breed is pedigree recorded (selection nucleus), it might be sufficient to monitor its genetic variability if most of the males used in the breed are coming from this selection nucleus. On the other hand, if a breed is structured in different nuclei that poorly interact with each other and only one nucleus has data, the use of molecular data could be more interesting.

[3] CGRFA/WG-AnGR-12/23/4/Inf.3

As far as pedigree depth is concerned, our recommendations, based on ten-years activity of the French Genetic Variability Observatory's, are to favor these data when the pedigree completeness level for the recorded animals, known as equivalent complete generations[4], is at least equal to 2.5. This value is equivalent to a pedigree where the parents, grand-parents and half the great-grand-parents are known. This value is sufficient to detect poor management practices and is even more valuable when implementing regular monitoring of the breed.

- About demographic data

Demographic data are the only choice possible when neither pedigree data nor periodic molecular data are available. In order to increase the level of accuracy on the genetic health of the breed, our recommendations would be to include other metrics such as the evolution of the breed population as well as the evolution of the number of farmers.

- Choice of methods for calculating Ne

Once the data set (demographic, molecular or pedigree) is chosen, there are a wide array of methods for calculating Ne which can be confusing. Here are some recommendations on the methods that can be used.

Ne based on demographic data are the least reliable; when they are the only data available, our recommendation is to have rough estimates of the variance of the family size in the breed, at least in cattle breeds which use AI and where the female contribution is very limited compared with males.

With pedigree dataset, a possibility is to use a software which calculates 6 types of Ne indicators[5]. Freely available software for this purpose is PopRep (https://popreport.fli.de/) In general, our recommendation is to choose a method based on kinship rates to avoid any bias linked to population substructure, as well as a method that corrects for pedigree depth.

As mentioned in the FAO report and the EURC EAB survey (to be published), it does not seem that any breeding organization uses molecular data on a regular basis to calculate Ne estimates. For this reason, there is no consensus on which methods to use depending on the situations, as well as how to parameter the software depending on the breed's situation. Since genomic relationships are already calculated for other uses such as optimum genomic contributions selection, a pragmatical solution would be to use these results for calculating a genomic Ne.

[4] Proportions of known ancestors summed over each generation.

5 NegutF: Gutiérrez based on inbreeding rate, NegutP: Fernandez method based on kinship rate, NeN, demographic method based on breeders, NeH, Hill demographic method, NeFa Falconer method based on the reference population and its parents, using inbreed rate, NeFalP, same method but using kinship rates.