



Food and Agriculture
Organization of the
United Nations

Monitoring within-breed genetic variation at global level

G. Leroy¹, R. Baumung¹, G. Mészáros², I. Curik³, J. Windig⁴, B. Rosen⁵, Y.T. Utsunomiya⁶, P. Burger⁷, L. Coli⁸, P. Boettcher¹, J. Soelkner², C. Looft⁹, C. Danchin¹⁰, A. Stella¹¹

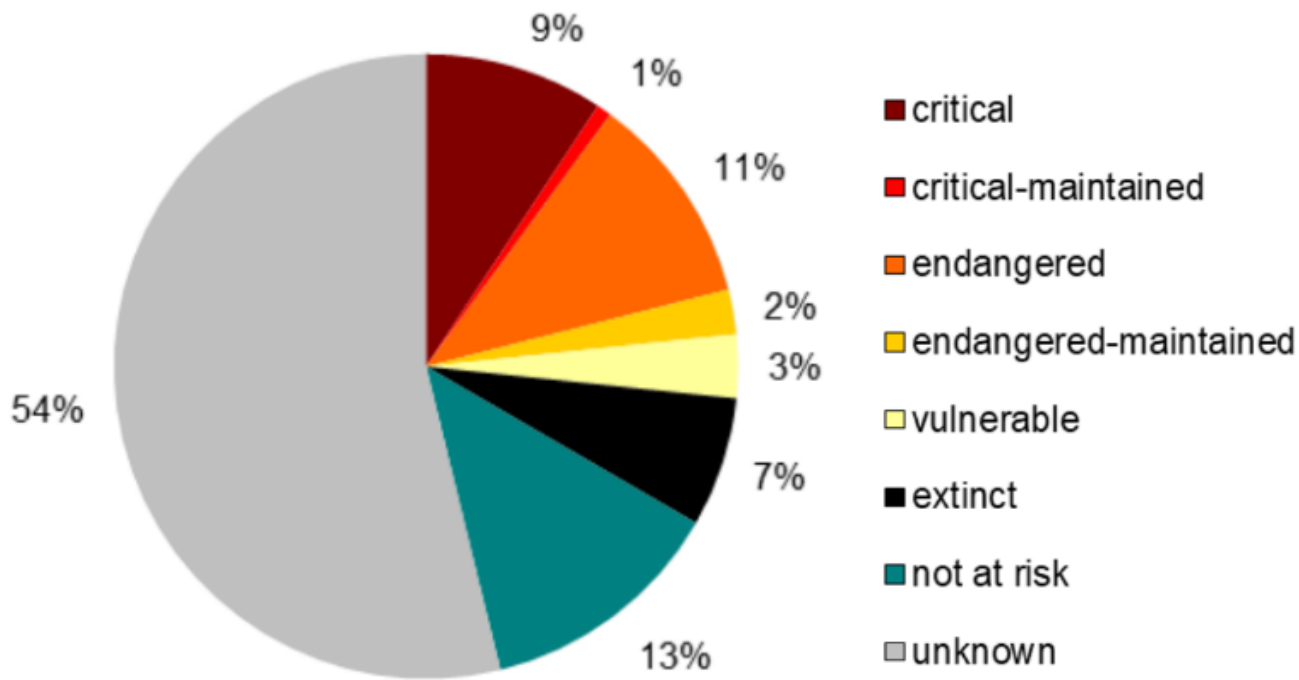
¹FAO, Italy; ²BOKU, Austria; ³University of Zagreb, Croatia; ⁴WUR, the Netherlands;

⁵USDA, USA; ⁶UNESP, Brazil; ⁷University of Veterinary Medicine, Austria;

⁸Università Cattolica del Sacro Cuore, Italy; ⁹University of Applied Science, Germany;

¹⁰IDELE, France; ¹¹CNR, Italy.

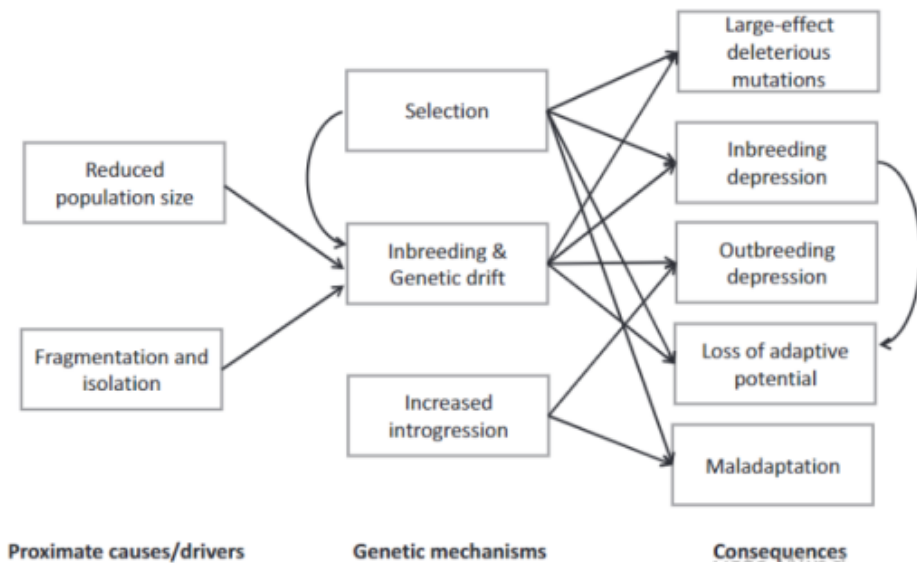
Monitoring of animal genetic resources at global level



Proportions of the world's breeds by risk status category (FAO, 2023)

- Monitoring of trend as part of strategic priorities of the Global Plan of Action for Animal Genetic Resources
- National data and information systems often underdeveloped -> demographic risk status unknown
- Genetic variation indicators to be considered

Why monitor genetic variation?



Drivers, mechanisms and consequences of genetic erosion (Leroy et al. 2017)

- Genetic variation relates to a diversity of processes (mutation, genetic drift, selection and migration)
- Loss of genetic variation (genetic erosion) has two main consequences:
 - Decrease in fitness (inbreeding depression/increased expression of detrimental traits)
 - Loss of adaptive potential (limited effect)
- In practice
 - Impact on animal welfare and productivity
 - Increased genetic/demographic stochasticity: -> increased risk of extinction

Data sources for monitoring of genetic variation

- Classical approaches through **demographic** or **pedigree** proxies
- **Genomics**: more precise insight into the drivers, mechanisms, and consequences (including neutral and non-neutral variation)
- Inclusion of related indicators in monitoring systems neglected
- Multiple challenges impede the inclusion of related indicators into monitoring systems
 - *Lack of technological and organisational capacities*

Context



- In 2021, the Commission on Genetic Resources requested FAO to investigate the integration in the global monitoring system DAD-IS (www.fao.org/dad-is) of fields related to indicators of genetic variability <https://www.fao.org/3/cc3758en/cc3758en.pdf>
- In 2023, publication of *Genomic characterization of animal genetic resources – Practical guide*, which updates 2011 guidelines on *Molecular genetic characterization of animal genetic resources* <https://www.fao.org/3/cc3079en/cc3079en.pdf>

Questions

- What are the constraints for the collection of data related to indicators of genetic variation?
- What are the best indicators for the monitoring of genetic variation?
- Which information related to indicators should be integrated into DAD-IS?

Properties and challenges: **demographic** information

Main properties	Challenges	
	Data collection	Applicability of results
Insight on demographic stochasticity and underlying causes behind the changes in genetic variation	Collection of data through breed censuses , surveys or animal identification systems	<ul style="list-style-type: none">• Estimates of genetic variation are basic interpolations of genetic variation.• Often underestimate loss of genetic variation

Properties and challenges: **pedigree** information

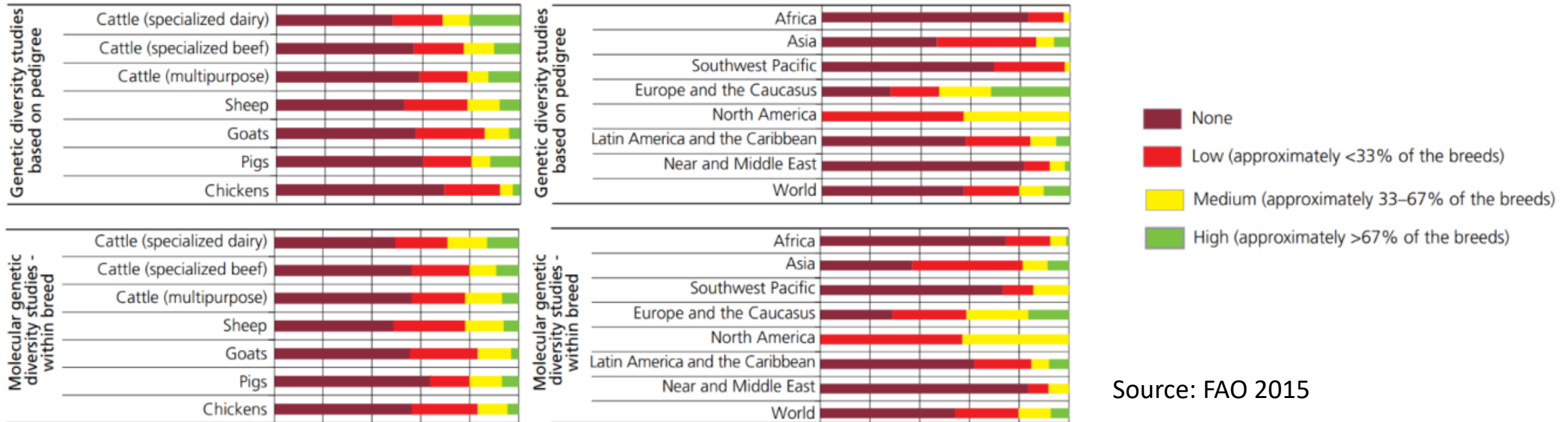
Main properties	Challenges	
	Data collection	Applicability of results
Inference on the genetic variation of neutral loci, based on knowledge of parent-offspring relationships	Requires registration of pedigree information to be as complete as possible	<ul style="list-style-type: none">• Do not consider mutation, mendelian sampling and selection• Prone to bias related to incomplete or incorrect pedigree

Properties and challenges: **genomic** information

Main properties	Challenges	
	Data collection	Applicability of results
Direct information on genomic variation , but provide no direct information on demographic stochasticity	Requires accurate sampling in terms of individuals and markers	<ul style="list-style-type: none">• Choice of appropriate parameters used for analyses (requires skill)• May yield inaccurate results if parameters are incorrect

Data sources for genetic variation monitoring

Frequency of responses for characterization and monitoring activities



Source: FAO 2015

- For less than 44% of breeds regular monitoring of population size
- In majority of countries, pedigree or molecular diversity studies have never been implemented
- Hall (2016) aggregated pedigree N_e estimates from 321 breeds and 31 countries, and genomic N_e estimates from 203 breeds and 30 countries

Challenges for data collection

Logistic and organizational capacities:

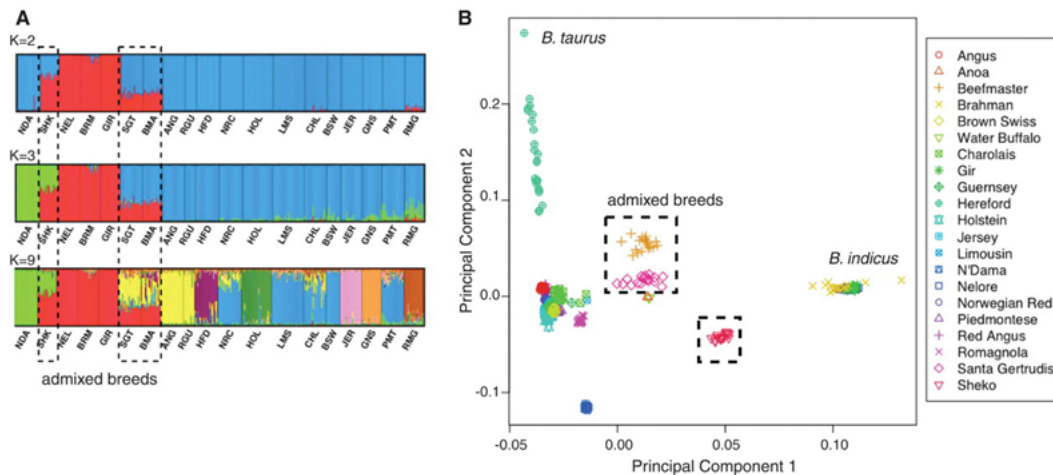
- Absence of animal identification system, regular breed census, and/or breeding organization
- Lack of technological and human capacities for sampling, extraction, storage and data analysis

Costs

- Funding of continuing costs e.g. for maintenance of monitoring systems and related human resources

Specificities of genomic information

Genetic structure of worldwide cattle breeds, according to:
(A) model-based clustering; and (B) principal component analysis



Source: adapted from The Bovine HapMap Consortium, et al. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926): 528–532. <https://doi.org/10.1126/science.1167936>.

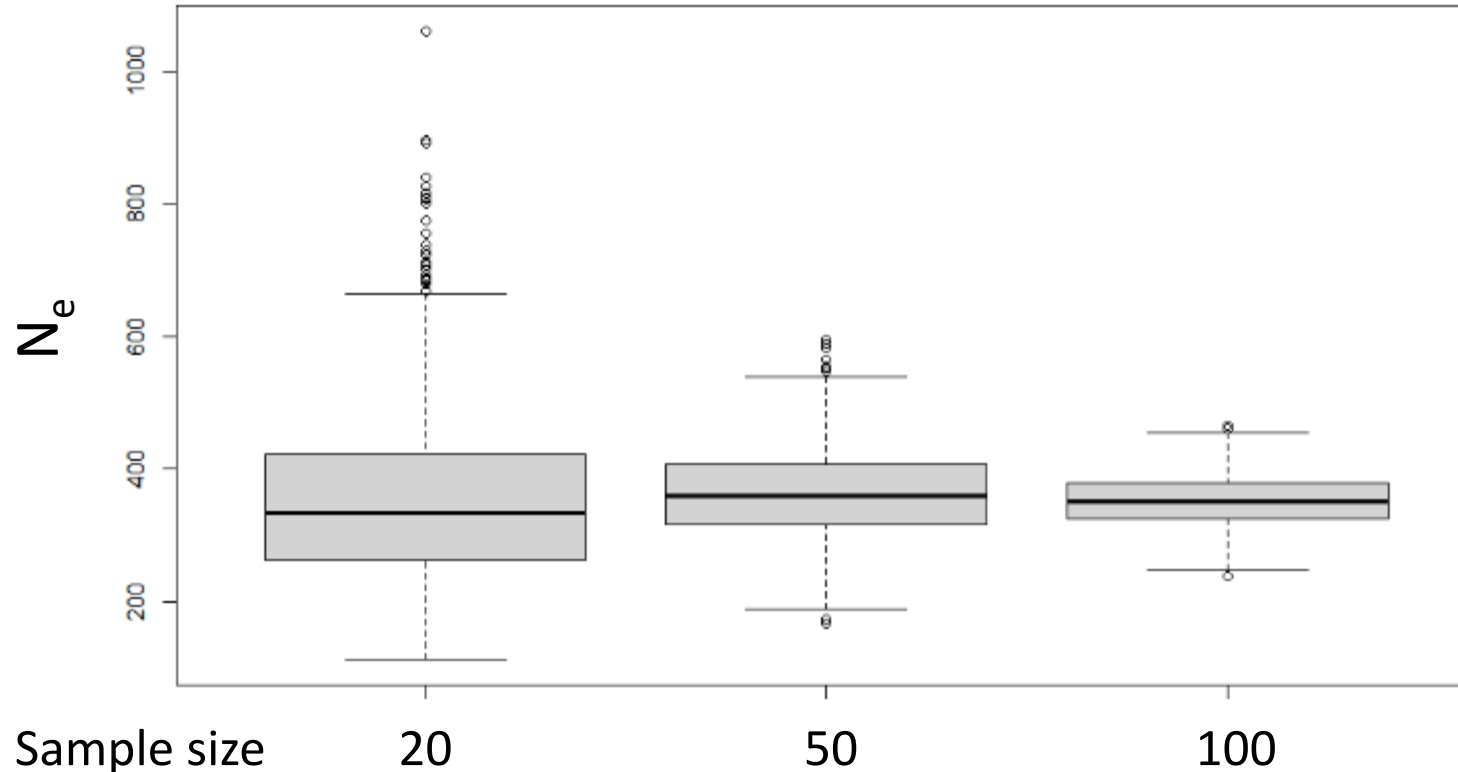
Processing

- Diversity of genomic tools (SNPs, WGS, ...)
- Specific steps to be considered from the field to data analysis

Further applications beyond monitoring of genomic variability

- Assessment of population structure and between-breed genomic variation
- Reconstruction of history and demographic modelling
- Analysis of non-neutral variability (selection signatures, GWAS...)

Which sample size is sufficient for molecular monitoring?



A sample of 100 individuals (50k SNP) provide reliable estimates

Estimation of effective population size of a goat breed based on sampling of various sizes (source: A. Stella)

What data source to prioritize?

Depends on the context:

- For a population with accurate pedigree information over several generations, demographic and pedigree data may allow effective monitoring, that can be complemented with genomic data
- For a population with no information, basic characterization of demographic parameters complemented with genomic characterization are recommended

What are the properties of a good indicator?

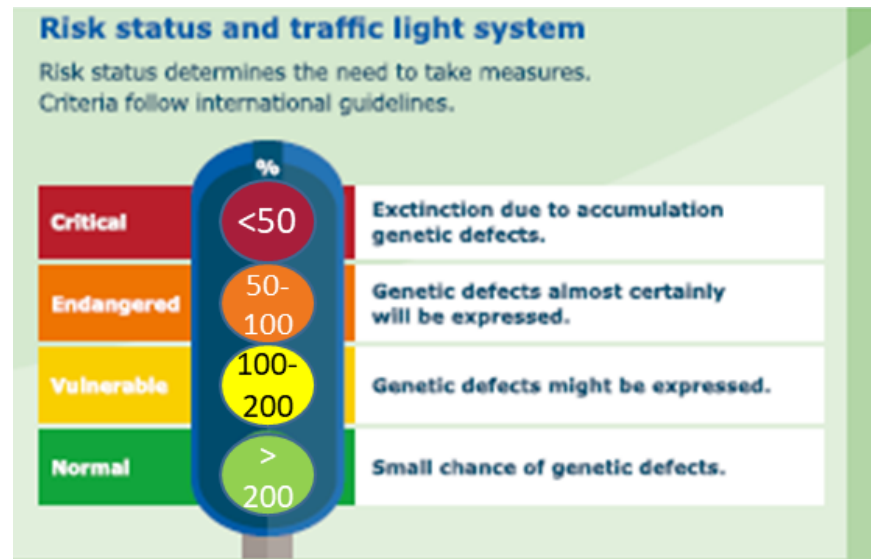
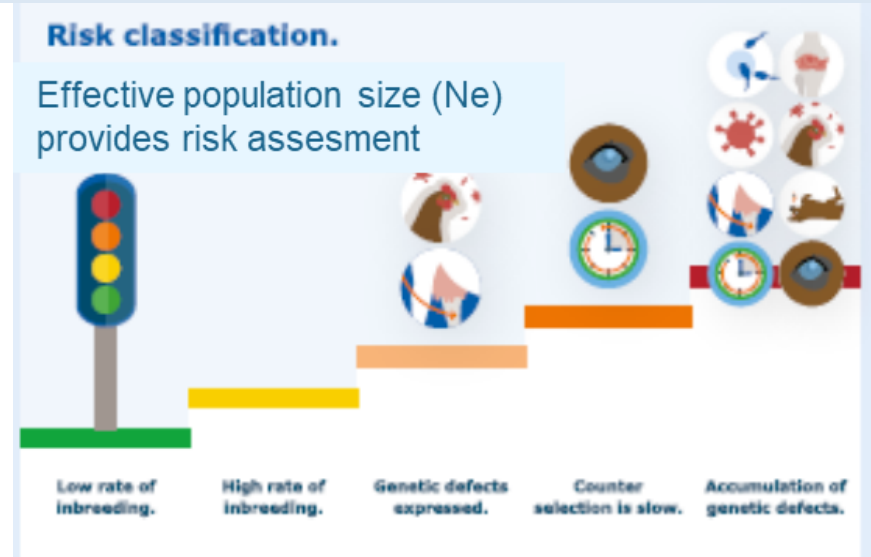
Property	Definition	What does it mean for genetic variation
Meaningful/relevant	Represent important information and reflect intervention's outcomes	Should reflect the amount and change in genetic variation
Adequate	Measure change over time and progress	Change in drivers, mechanisms and consequences of genetic variation -> change in indicator
Reliable	Consistently measured across time, approaches and data collectors	Allow comparison for different populations, data sources, or approaches
Understandable/synthetic	Easy to comprehend and interpret	Stakeholders understand the indicator and can react accordingly
Practical/feasible	Reasonable in terms of the data collection cost, frequency, and timeliness	Possibility to use proxies

Main indicators of genetic variation

Name	Definition	Advantages	Limits
Gene diversity	Proportion of heterozygous genotypes – reflects current state of genetic variability	Easy to measure if genotyping is feasible	Comparison across different marker sets difficult
Inbreeding/ coancestry coefficient	Probability that two alleles at any given locus are IBD	Can be approached with pedigree data and genomic data	Comparison difficult across different populations
Effective population size	Size of an “idealized” population that would result in the same amount of genetic drift/change of inbreeding	Can be approached with demographic , pedigree and genomic data	Comparison difficult across different approaches or data sources

Effective population size







- N_e determines genetic stability of a population
- Linked to drivers, mechanisms and consequences of changes in genetic variation
 - Genetic drift / Random change of allele frequencies
 - Change in inbreeding (ΔF)/frequency of heterozygotes
 - Probability of expression of genetic defects
 - Genetic stochasticity/risk of population
- Easy to understand and interpret:
 - e.g. *effective population size of Holstein cattle is around or less than 100*
- Can be estimated with diverse approaches and data sources
- Estimate of N_e is essential for monitoring of genetic variation
 - *Use of thresholds (e.g. $N_e < 50$)*



Examples of approaches estimating effective population size

Data source	Abbrev	Parameter used	Effective population size formula		Time scale considered	Reference
Demographic	N_{eS}	Sex ratio population size	$N_{es} = \frac{4MF}{M + F}$		Previous generation	Wright (1931)
Demographic	N_{eV}	Variance of family size	$N_e = 8Nm / (V_{km} + V_{kf} + 4)$		Previous generation	Hill (1979)
Pedigree	N_{eF}	Individual inbreeding rate N_{eC}	$N_{eF} = 1/2\overline{\Delta F}$	$\Delta F_t = \frac{F_{t+1} - F_t}{1 - F_t}$	Adjustable on a chosen number of generations	Falconer (1996)
Pedigree	N_{eCi}	Individual coancestry rate	$N_{eCi} = 1/2\overline{\Delta C}$	$\Delta C_{ij} = 1 - \frac{(EqG_i + EqG_j)^{1/2}}{\sqrt{1 - C_{ij}}}$	Depends on pedigree knowledge	Cervantes et al. (2011)
Pedigree	N_{eCir}	Restricted coancestry kinship rate	$N_{eCir} = 1/2\overline{\Delta C}$	$\Delta C_{Rij} = 1 - \frac{(EqGR_i + EqGR_j)^{1/2}}{\sqrt{1 - C_{Rij}}}$	Adjustable on a chosen number of generations	Leroy et al. (2020)
Genomic	N_{eLD}	Variance of gene frequency changes	$N_{eVa} = 1/2(\hat{V} - \frac{1}{2S_{t-1}} - \frac{1}{2S_t})$		Adjustable on the basis of populations sampled	Nei and Tajima (1981)
Genomic	N_{eLD}	Linkage disequilibrium	$N_{eLD} = \frac{1}{3(r^2 - \frac{1}{S})}$	$N_{eLD} = \frac{1}{3(r^2 - \frac{1}{S})}$	Adjustable according to marker distance	Waple (2006)

Effective population size / what about the reliability?

Breed	N_{eS}	N_{eCi}	N_{eLD}
	8582	339	303
	2439	148	145
	6458	100	98
	1215	99	92
	6654	90	142
	23504	82	81

Source: VARUME, IDELE

	Cattle	Sheep	Horses	Pigs
$N_{eLD} \times N_{eS}$	0.23	-0.02	0.27	
$N_{eLD} \times NeF$				
$NeF \times (\text{Log10}) N_{eS}$	0.53	-0.1		0.46

Correlations among published N_e values considering 321 breeds (Hall 2016)

- Large discrepancies according to approaches and data sources used
- Divergences in hypotheses, sampling, time scale, marker set and density...

What should be considered for DAD-IS?

- The majority of countries lack the capacity to collect data related to genetic variability
- N_e has the advantage of being a relevant and synthetic indicator that can be estimated with diverse approaches and data sources, including simple ones
- N_e is the proposed indicator for genetic variation in the Kunming-Montreal Global Biodiversity Framework adopted by the CBD
- Over-estimation of N_e may be a risk, especially for the comparison of breed populations and in relation to the current endangerment thresholds, which are based on census size
- Adding an option on reporting on N_e in DAD-IS has been recommended by experts, if reported with additional information on approach used, timeline, and data sources

Discussion and conclusions

General conclusions and recommendations

- Measure genetic variation with at least one indicator of N_e
- Monitor populations regularly
- Use Molecular tools to estimate N_e
 - more accurate than demographic and pedigree based measures
 - especially with low quality pedigree
- Genotype
 - at least 100 animals
 - with a 50K SNP array (or equivalent)
 - Sample both sexes
 - From multiple generations

Thank you!



*See you at the Global Conference, 25-27
September 2023*