Digital sequence information on animal genetic resources for food and agriculture

Submission of information on the use of "digital sequence information on genetic resources for food and agriculture" and potential implications for the conservation and sustainable use of genetic resources for food and agriculture, including exchange, access and the fair and equitable sharing of the benefits arising from their use

by the **ABS Task Force of the European Regional Focal Point on Animal Genetic Resources**
Elżbieta Martyniuk[1], Beate Berger[2], Danijela Bojkovski[3], Sipke Joost Hiemstra[4], Alwin Kopše[5], Vera Matlova[6], Nina Sæther[7], Jan Tomka[8]

[1] Warsaw University of Life Sciences / National Research Institute of Animal Production
[2] AREC Raumberg-Gumpenstein, Institute of Organic Farming and Biodiversity of Farm Animals
[3] University of Ljubljana, Biotechnical Faculty
[4] Centre for Genetic Resources, the Netherlands (CGN), Wageningen University and Research
[5] Unit for International Affairs, Sustainable Development, Food Systems, Federal Office for Agriculture, Switzerland
[6] National Centre for Animal Genetic Resources, Institute of Animal Science Praha Uhrineves, Czech Republic
[7] Norwegian Genetic Resource Centre, Norwegian Institute of Bioeconomy Research
[8] Institute for Animal Husbandry Systems, Breeding and Product Quality / Research Institute for Animal Production Nitra, Slovakia

1. INTRODUCTION

The issue of Digital Sequence Information (DSI) was raised during the negotiations of ABS regime. In preparation for the 3rd meeting of the Ad Hoc Open-Ended Working Group on Access and Benefit-Sharing (Bangkok, 14-18 February 2005), the EU submitted a study that addressed the use of DSI among other matters (CBD, 2005). This study provided a review of the state and trends in the development of genomics, proteomics and biotechnology and an assessment of their potential implications on a new international regime. The paper addressed the challenges and potential opportunities for the development of an international regime resulting from the growth of bioinformatics and international electronic transfers of genetic data. The author suggested that the genomes and proteomes of biological organisms constitute a significant gap within the existing international policy framework established under the United Nations system.

A process to reassume discussion on this topic both by the Parties to the Convention on Biological Diversity (decision CBD/COP/DEC/XIII/16) and the Parties of the Nagoya Protocol (decision CBD/NP/MOP/DEC/2/14), as well as by the members of the Commission on Genetic Resources for Food and Agriculture was initiated in late 2016 early 2017 period.
Future debate on this matter will need to face a number of technical challenges, such as lack of agreement on the definition of DSI, as well as different types and technical scope of the DSI. The debate should lead to better understanding of existing terminology related to DSI on genetic resources, availability and extent of use of DSI data and their implications on implementation of the objectives of the Convention and the Nagoya Protocol.

The submission prepared by the ABS Task Force provides some technical information about existing sources, management and availability of DSI on animal genetic resources, as well as information on the applications of DSI in animal breeding and conservation of animal genetic resources. Taking into account the all-embracing applications of the DSI on animal genetic resources, the submission is not exhaustive, but is meant to provide general information on this issue.

2. DSI DATABASES

Genomics can be briefly defined as "the study of genes and their function" and is concerned with the mapping and analysis of the entire genetic make-up of an organism constituting its genome. Genomics provides the foundation for the science of proteomics which is concerned with the mapping and analysis of the protein make-up within an organism (the proteome) (CBD, 2005).

2.1 Key DSI databases

A quantitative trait locus (*QTL*) is a section of DNA which correlates with variation in a phenotype. Once a region of DNA is identified as contributing to a phenotype, it can be sequenced so that the nucleotide order of a given DNA fragment can be determined. The DNA sequence of any gene in this region can then be compared to a database of DNA for genes whose function is already known.

QTL mapping identifies which molecular marker (Single Nucleotide Polymorphism (SNP) or Amplified fragment length polymorphism (AFLP)) correlate with an observed trait, and usually represents an early step in identifying and sequencing respective genes responsible for the trait variation.
There are three major sequence repositories: the National Center for Biotechnology Information, the European Bioinformatics Institute and the DNA Data Bank of Japan, which share the same sequence information.

- **National Center for Biotechnology Information (NCBI, USA)**
  https://www.ncbi.nlm.nih.gov

It includes more than 30 databases, related to genes, genomes and maps, proteins and chemicals, as well as bibliographic records from MEDLINE and other sources. The major NCBI Entrez Database provides integrated access to nucleotide and protein sequences, complete genomes and schematics of entire chromosomes, as well as associated mapping information, for example:

- ✓ **GenBank** https://www.ncbi.nlm.nih.gov/genbank/
  GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is a part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

- ✓ **The High Throughput Genomic (HTG) Sequences db**
  https://www.ncbi.nlm.nih.gov/genbank/htgs/

The High Throughput Genomic (HTG) Sequences db was created to accommodate a growing need to make unfinished DNA sequences generated by the high-throughput sequencing, rapidly available to the scientific community.

✓ **The GSS database** https://www.ncbi.nlm.nih.gov/nucgss

The GSS database is a collection of unannotated short single-read primarily genomic sequences from GenBank including random survey sequences, clone-end sequences and exon-trapped sequences.

✓ **The SNP database** https://www.ncbi.nlm.nih.gov/snp

The dbSNP is a database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants.

✓ **The Gene** https://www.ncbi.nlm.nih.gov/gene

The Gene database provides detailed information for known and predicted genes defined by nucleotide sequence or map position. Currently, the Gene contains over 17 million entries and includes data from all major taxonomic groups. The Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

- **The European Bioinformatics Institute (EBI)** http://www.ebi.ac.uk/

The **European Molecular Biology Laboratory (EMBL-EBI)**, publicly-funded non-profit institute is housed at six sites in Europe whose expertise covers the whole spectrum of molecular biology. The EBI, an international interdisciplinary research organisation funded by 23 member states and two associate member states, as a part of the EMBL maintains the world's most comprehensive range of freely available molecular data resources, i.a.

✓ PRIDE Archive - proteomics data repository http://www.ebi.ac.uk/pride/

The PRIDE PRoteomics IDEntifications (PRIDE) database is a centralized, standards compliant, public data repository for proteomics data, including protein and peptide identifications, post-translational modifications and supporting spectral evidence.

✓ The IPD-MHC Database https://www.ebi.ac.uk/ipd/mhc/

The Immuno Polymorphism Database (IPD) provides a centralised repository for sequences of the Major Histocompatibility Complex (MHC) from a number of different species. Through a number of international collaborations, IPD is able to provide the MHC sequences of different species. The sequences provided by each group are curated by experts in the field, and then submitted to the central database.

✓ The European Variation Archive http://www.ebi.ac.uk/eva/

This is an open-access database of all types of genetic variation data from all species. All users can download data from any study, or submit their own data to the archive. It enables also queries on all variants in the EVA by study, gene, chromosomal location or dbSNP identifier the Variant Browser.

- ✓ ENSEMBL http://www.ensembl.org/index.html
  Ensembl is a genome browser for vertebrate genomes. Based at the European Bioinformatics Institute (EMBL-EBI), it creates, integrates and distributes reference datasets and analysis tools that further enables genomics. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data.

- **DNA Data Bank of Japan** http://www.ddbj.nig.ac.jp/

This is an annotated collection of all publicly available nucleotide and protein sequences. DDBJ Center internationally contributes as a member of INSDC to collect and to provide nucleotide sequence data with ENA/EBI in Europe and NCBI in USA. DDBJ collects sequence data mainly from Japanese researchers, as well as from researchers in any other countries.

2.2 Other sources of DSI

- **National Animal Genome Research Program (NRSP -8)**
  https://www.animalgenome.org/repository/

This United States program attempts to identify DNA sequences or quantitative trait loci (QTLs) associated with disease resistance or susceptibility and production traits in livestock and poultry species. These markers are useful in selection strategies in most, if not all, livestock and poultry species. NRSP-8 data repository contains 1,081 data files at the current location, i.a.

The Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb)
https://www.animalgenome.org/cgi-bin/QTLdb/index
Strives to collect all publicly available trait mapping data, *i.e.* QTL (phenotype/expression, eQTL), candidate gene and association data (GWAS), and copy number variations (CNV) mapped to livestock animal genomes, in order to facilitate locating and comparing discoveries within and between species.

Animal Trait Correlation Database https://www.animalgenome.org/cgi-bin/CorrDB/index
This database is designed to collect all published livestock genetic/phenotypic trait correlation data. Currently, this database has an initial collection of **3,635** correlation data on **276** economically important traits of cattle, relating to meat production, milk production, growth, health and others traits.

The Bovine SNP Database, for example, contains 114,958 cattle SNPs data https://www.animalgenome.org/tools/q_bovsnp.html

- **Livestock Genomics** http://www.livestockgenomics.csiro.au/

CSIRO Animal, Food and Health Sciences (Australia) livestock genomics web site, aims to facilitate access to data generated by cattle and sheep genome mapping and sequencing projects, and provides access to interactive genome maps of cattle and sheep.

- **Bovine Genome Database** http://bovinegenome.org/

The Bovine Genome Database project hosted at the University of Missouri is to support efforts of bovine genomics researchers by providing data mining, genome navigation and annotation tools for the bovine reference genome based on the Hereford cow. It provides

tools for data mining (BovineMine), sequence database searching (BLAST), genome browsing (JBrowse) and annotation (Apollo).
The BGD project is supported by the European Union's 7th Framework Programme for research, technological development and by the USDA National Institute of Food and Agriculture.

- **UCSC Genome Browser Gateway** http://genome.ucsc.edu/

The UCSC Genome Browser was created at the University of California Santa Cruz (UCSC, USA), and is free available for academic, nonprofit, and personal use free for academic, nonprofit, and personal use. Started in 2000 at the UCSC Genomics Institute as a part of International Human Genome Project, the website has grown to include a broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data.

2. 3. **Metadata sources**

Except „single" databases there are also metadata sources, such as:

- **Nucleic Acids Research *Database*** https://academic.oup.com/nar

This is an open-access peer reviewed scientific journal published by Oxford University Press. The journal publishes two yearly special issues, one dedicated to biological databases published since January 1993, and the other on biological web servers published since July 2003. The current 2017 Nucleic Acids Research Database Issue, is the 24th annual collection of bioinformatic databases on various areas of molecular biology. It describes both newly created databases and updates on the databases that have been previously described.

- **Fairsharing** https://fairsharing.org/databases/

As of June 2017, BioSharing, is a searchable portal of three linked registries covering standards, databases, and data policies in the life sciences, broadly encompassing the biological, environmental and biomedical sciences. A product of The Research Data Alliance (RDA), which was launched as a community-driven organization in 2013 by the European Commission, the United States Government's National Science Foundation and the National Institute of Standards and Technology, and the Australian Government's Department of Innovation, with the goal of building the social and technical infrastructure to enable open sharing of data.

- **Ark DB (**Roslin Institute) http://www.ed.ac.uk/roslin/facilities-resources/bioinformatics

A generic, species-independent database built to capture the state of published information on genome mapping in a given species. It stores details of references, markers and loci and genetic linkage and cytogenetic maps which can be drawn using the online map-drawing application. Data from linkage maps held within the ArkDB system can be drawn alongside their corresponding genome sequence maps (extracted from ENSEMBL).

Another example, the **SNPchiMp**, is a MySQL database linked to an open access web-based interface. This tool combines many different sources of information that otherwise would be time consuming to obtain and difficult to integrate.

Currently, six commercial whole-genome SNP chips are available for cattle genotyping, produced by two different genotyping platforms. Technical issues need to be addressed to combine data that originates from the different platforms. Features of SNPchiMp include the following functions: 1) referencing the SNP mapping information to the latest genome assembly, 2) extraction of information contained in dbSNP for SNPs present in all commercially available bovine chips, and 3) identification of SNPs in common between two or more bovine chips (e.g. for SNP imputation from lower to higher density). This allows easy integration and standardization.

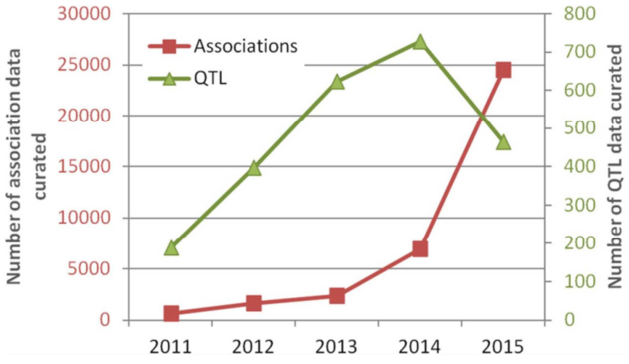
2.4. Scope of information and its use

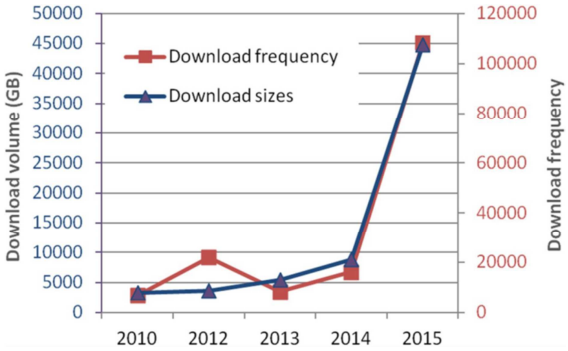
The amount of data inserted in databases is growing incredibly fast. As an example, since 19th June, 2017, till 4th September 2017, in the QTL db animalgenome.org, the number of QTLs/associations for cattle increased from 98 090 (773 journals) to 99 652 (799 journals). With chickens, the increase has been from 6 791 QTLs/262 publications to 7 812/273 publications, and with in pigs the increase from 17 955 QTLs/576 to 25 610 QTLs in 593 publications.

In comparison, in 2007, when the AnimalQTLdb was designed to house all publicly available QTL data on livestock animal species, the number of entries was: 630 for cattle, 657 for chicken and 1287 for pig QTLs. Overtime, the database tools were added to link the QTL data to other types of genomic information.

Newly released QTL/association data are also exported to data alliances (Ensembl, NCBI Entrez GeneDB, UCSC). Users can employ tools on these respective sites for information mining in relation to animal QTL and association data.

According to Zhi-Liang et al. (2015), in recent years, the Animal QTL Database (QTLdb; http://www.animalgenome.org/QTLdb) has undergone dramatic growth as regards new data curated, data downloads and new functions and tools. The development efforts were focused on coping with challenges arising from rapid growth of newly published data and end users' data demands, and a need to optimize data retrieval and analysis to facilitate users' research. The Google Scholar indicated that the Animal QTLdb has been cited in the literature over 1010 times in last 10 years. The graphs below, ilustrate well the state of content of these databases as well as their use (Zhi-Liang et al., 2015).





2.5. The submission and use of data on example of the GenBank

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up dated and comprehensive DNA sequence information available. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights for all or for a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore, cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

GenBank Submission

GenBank accepts mRNA or genomic sequence data directly determined by the submitter. The submission must include information about the source organism and annotation provided by the submitter. Data are submitted together with the respective scientific publication (i.e. with full publication data-all authors, title, journal, volume, pages and date).

Some authors have expressed concern that the appearance of their data in GenBank prior to publication will compromise their work. Accordingly, GenBank will, upon request, withhold release of new submissions for a specified period of time. In order to prevent the delay in the appearance of published sequence data, GenBank urges authors to inform them of the appearance of the published data. Since 1982 to the present, the number of bases in GenBank has doubled approximately every 18 months, now containing more than 200 M GenBank DNA sequences and 487 M whole-genome sequences.

The statistics (number of bases and the number of sequence records) is available at https://www.ncbi.nlm.nih.gov/genbank/statistics/

3. USE IN DEVELOPMENT

Use of digital sequence information in animal breeding and in the management of animal genetic resources is extensive. First, we briefly describe the main types of digital sequence information. Second, we present practical applications using digital sequencing information.

3.1 Relevant types of digital sequence information

From the 1970s and onward, the era of molecular genetics provided new opportunities for the use of DNA markers. Single Nucleotide Polymorphisms (SNPs) and whole genome sequence information are the main and most relevant types of digital sequence information

(Valdani et al., 2016; Oldenbroek, 2017). Former types of genetic markers, such as microsatellites nowadays have limited value, but microsatellites are still used in genetic diversity studies or parentage testing.

- ***Single Nucleotide Polymorphisms (SNP)***

For many species low density and/or high density SNP panels are now available for research and breeding

- ***Whole genome sequencing***

Due to growing interest in human genome resequencing, a new generation of sequencing technologies emerged. These next-generation sequencing (NGS) technologies are able to generate DNA sequence data at low cost and at a rate much faster than that of traditional technologies. With NGS technologies it is possible to resequence entire genomes or sample entire transcriptomes more efficiently and economically than in past, and in greater depth than ever before. This makes it possible to sequence hundreds or even thousands of related genomes, and to determine the genetic basis of trait variation and adaptation. Farm animal species genomes have been sequenced and annotated for many species during the past decade, including for chicken, dog, cattle, horse, pig, sheep and rabbit.

3.2. Practical applications

- ***Insight in the origin and domestication of farm animal species***

Genetic and genomic data provide a powerful resource for answering questions on the origin of diversity. A large number of diversity studies have allowed the reconstruction of the domestication, migration, selection, and adaptation history of most farm animal species. In particular for animals, studies of mitochondrial DNA genomes, which are maternally inherited, have helped to gain more insight in the origin and domestication of the species. The availability of high density SNP chips and full sequence information has further increased the understanding of domestication processes, breed characteristics, and introgression, signatures of selection and adaptation mechanisms.

- ***Analysis of within and between breed genetic diversity***

In the past, effective population sizes and inbreeding levels of AnGR, could be estimated only on the basis of pedigree data and/or using a limited number of markers. With the availability of sets of high density SNP markers or even whole genome sequence data, the variety of alleles, haplotypes and genotypes can be assessed more precisely. Analysis of dense markers will give information about the level of heterozygosity, genetic diversity and signatures of selection.

An increasing amount of genomic data can be expected to support improved management of diversity in selected populations and ex-situ collections. Genomic data is being used to determine molecular coancestry, which is a more accurate indicator for inbreeding compared to pedigree based coancestry. The effectiveness of strategies to maintain within breed genetic diversity and to control the genetic background of a breed can be improved when genomic data is used. Relationship estimates on the basis of whole genome sequence data are significantly different in comparison to estimates on the basis of pedigree data.

- ***Marker assisted selection***

Methods of Marker-Assisted Selection (MAS) became operational with emergence of the first DNA-based genetic markers in late 1970. Molecular markers can be used to identify genes or genomic regions that control traits of interest.
Combining MAS with traditional/conventional selection methods allows for more precise selection, as well as improving selection response. The use of MAS has potential if the markers are highly correlated with the desired phenotype to enhance efficiency and power of breeding strategies.

- ***QTL mapping***

Traits can be controlled by one or few genes (qualitative traits) or complex quantitative trait loci (QTL), e.g. milk yield and growth rate, where expression of traits involves many genes e.g. milk yield and growth rate. Most of the genetic traits of farm animals are the result of quantitative variation. Locating this loci is called Quantitative trait locus (QTL) which is a small segment of DNA that has large effect on the trait. A substantial numbers of QTL and marker-phenotype associations have been detected, and also causative mutations. QTL (quantitative trait loci) mapping has been used to determine the genetic bases of complex traits. The phenotyping of large numbers of genotypes makes possible the identification of trait-associated genomic regions and marker-based selection. To date, thousands of QTLs have been reported. However, the identification of the underlying causative mutations remain challenging. There is a long history of research on the use of genetic markers to identify quantitative trait loci and their use in marker-assisted selection, but with limited implementation in practical breeding programs. Currently, the high-density SNP data provides new opportunities to detect QTL and to better understand the genetic architecture of quantitative traits.

- ***Genome Wide Association Studies (GWAS)***

Genome-wide association (GWA) studies use a quantitative genetic approach to find genetic associations between genotype and phenotype in a population of individuals of unknown relatedness to identify genetic loci contributing to such a phenotype. Genome-wide association analysis (GWAA) provides a new approach for high resolution genetic analysis, thanks to the development of large panels of SNPs and the development of cost-effective methods for large-scale SNP genotyping and analysis. Next-generation high-throughput DNA sequencing technologies and the completion of high-quality reference genome sequences have enabled the development of sequencing-based genotyping and genome-wide association studies (GWAS).

- ***Genomic selection***

The availability of high-density SNP genotyping, combined with novel statistical methods for the use of this data to estimate breeding values, has resulted in extensive application of genomic selection or whole-genome selection. Genomic selection is a form of marker-assisted selection where a very large number of genetic markers are used covering the whole genome. The large number of markers is obtained by chips using Single Nucleotide Polymorphims (SNP's), a point mutation of a single nucleotide. The genomic selection is based on the analysis of 10.000 up to 800.000 SNP's. This high number of genetic markers is used as input in a genomic prediction formula that predicts the breeding value of an animal. Genomic selection is fundamentally distinct from marker assisted selection in that all available genetic markers are fitted simultaneously to develop a prediction model utilizing

phenotypic and genotypic data collected from a training populations. These models are then used to predict genomic breeding value of progeny in future generations. It has had an enormous impact on the livestock sector, starting with the dairy cattle sector. The key advantage of genomic selection is increased genetic gain through shortening of the generation interval, but investments are also high.
When costs of genotyping drop, this will result in more marker based genetic evaluations. Future focus could result in more f inclusion of causal variants in the genetic model, instead of genome wide genomic selection only.

- ***Proteomics and metabolomics***

The assessment of RNA (transcriptomics), protein (proteomics) and metabolite (metabolomics) levels can deliver information on genes in the target region associated with mRNA, protein or metabolite shift linked to the trait of interest. It is now possible to generate omics datasets for many species, and, although the high costs of metabolomics limits direct application in breeding, developments in omics technology are helping to elucidate the biological processes that determine gene effects.

- ***Phenomics***

'Phenomics' has been proposed as a novel discipline in biology, involving the gathering of high-dimensional phenotypic data at multiple levels of organization to progress towards the full characterization of the complete set of phenotypes of a genome, in analogy with whole genome sequencing (Dhondt et al., 2013).

- ***Landscape genomics***

Landscape genomics is an approach combining genetic marker or genomic data with GIS data and may be used to improve *in situ* conservation strategies.

- ***Identification of genetic defects***

One of the first applications of genetic markers was the discovery of the genetic basis and development of genetic tests for single gene defects. Today, many genetic defects and disorders have DNA tests available. There has been a lot of research on developing genetic markers for monogenic recessive genetic defects and the genes themselves that are present in all species and are responsible for genetic defects.

- ***Maximizing genetic progress while maintaining genetic variability***

In order to manage inbreeding in livestock breeding population's, different methods were developed based in order to manage inbreeding while increasing genetic gain. One of them is optimum contribution selection, where inbreeding is limited to a specific level and the rate of gain maximized for that specific level of inbreeding. These optimal selection principles have been shown to maximize genetic gain at lower rates of inbreeding so that selection response can be maintained over the long term. DNA marker based estimation of breeding values is used in many breeding programmes, and it has been suggested that genomic estimation of breeding value could also be used as a tool to reduce inbreeding. It was proved that estimation of genetic relationship between any two individuals in the population is much more accurate using SNPs than pedigree based methods.

- ***Authenticity of products***

DNA markers can be used to proof the authenticity of products and whether a product is a product of a certain breed.

4. CONSERVATION

4.1 History, characterization and breed distinction

Since the domestication of farm animals, humans were making breeding decisions. Different breeds of livestock were first developed on the basis of phenotypical traits only. Together with animal/environment interactions, this resulted in the diversity of AnGR known today (Andersson 2001). Distinguishing between breeds by phenotype has always been important. Phenotype often is regarded as trademark for a breed such as curly hair in the Mangalica pig breed.

Characterization of animals by recording of production traits and progeny testing is a much newer approach . The International Committee for Animal Recording (ICAR) started its work in 1951 with dairy cattle (http://www.icar.org/). Without recording and evaluation of productivity, no successful management of breeding programs is possible (FAO 2010). With the emerging of DNA technologies from 1970 onward, selection intensity was further increased. One of the results of intensive selection based on reliable data is the impressive breeding progress achieved over the last 60 years represented by specialized international breeds. Less intensively selected and managed breeds were gradually marginalized or even lost completely.

4.2 *In situ* conservation, control of inbreeding, development of mating plans

According to FAO (2013), the conservation of AnGR *in situ* is the most effective and versatile method. *In situ* conservation enables the ongoing adaptation of animals to their production environment, the traditional utilization and/or the development of new uses and products and raises public awareness on the issue of endangered breeds.

Intensive selection together with the development of biotechnologies in reproduction, like artificial insemination, leads to an accelerated inbreeding rate. In small closed populations, like most endangered breeds, the inbreeding rate as well as the loss of genetic diversity through genetic drift is an important issue.

Until recently, many *in vivo* conservation programmes for endangered AnGR relied on mating plans requiring or not requiring pedigree information to reduce inbreeding (FAO 2013). Even if there are reliable pedigrees of sufficient depth for calculations for a breed, the relationship of founder animals remains unclear (Binder 2016). Additionally, there is almost no information on relationship between breeds available in pedigrees.

By genomic analysis, especially when using the SNP technology for the first time, information about breed history, about genetic diversity within and between breeds or populations, and about allocation of individuals to a breed or population is available for landrace populations without pedigree information or records (Fernãndez & Bennewitz 2017). Mating plans based on actual relationship can be determined for each individual and carriers of rare alleles can be identified and used preferentially.

Another advantage of genomic methods is the evaluation of long-term *in situ* conservation programmes.
The drawbacks are the limited availability of data for landrace breeds and the price. Both issues currently are rapidly improving with increasing use of the technology.

4.3 *Ex situ* collections, sampling strategies, evaluation of collections

Ex situ cryo-conservation in established genebanks is the most efficient tool to complement *in vivo* conservation programmes. Which kind of material to collect and the collected amount are decided based on the goals of the collection. They can be roughly divided into a backup function and/or ongoing use of material in conservation breeding programmes (Woolliams et al. 2008). The collection should aim at containing the complete genetic variability of the stored breeds.

Sampling decisions should follow genetic criteria to maximize efficiency and to reduce cost (Toro & Mäki-Tanila 2008). As mentioned above, decisions based on pedigree analysis alone are not always possible in landrace breeds and relationships between breeds are not well documented.

For a comprehensive sampling strategy, genomic analysis of the active breeding populations and the content of the genebank has to be done to identify potential donors and gaps in the collection. Relying on these data the genebank is able to optimize the supply of reproductive material according to the demand of the mating programme.

Another opportunity provided by genomic data analysis is identification and allocation of very old and poorly documented samples.

5. RELEVANCE FOR ABS

There are a number of approaches regarding how to address DSI implications for the access and benefit sharing of genetic resources.

The background study, prepared over 13 years ago, (CBD, 2005) noted that ”genomes and proteomes may extend beyond individual lands or territories, the jurisdictions of individual states, regions, population groups and ultimately generations”. Therefore the study proposed that genomes and proteomes could usefully be seen as “global public goods”.

For others, DNA sequence generated from genetic resources might require the same ABS procedures as genetic resources themselves. Therefore, using sequence from the public database if it was obtained from an improperly acquired sample and if national legislation covers intangible genetic information may be considered as a form of biopiracy (Bagley, 2015).

The course of debate on this subject at COP/MOP2 in December 2016, indicates that developing countries are likely to incorporate to national legislation or into MAT contracts, conditions regarding generating, using and publishing of DSI data.

However, it might be difficult to implement restrictions on already available DSI in light of current practices of DSI data holders, where DNA, RNA and amino acid sequence data stored in various types of databases/databanks are considered in the public domain and their unrestricted use is taken for granted by the research community.

Moreover, it may be extremely challenging to establish operational ABS arrangements and restrictions on the use of DSI stored in publicly available databases; the costs might be high.

Once the terminology around DSI has become clear, a discussion on the nature of DSI, which will be undertaken by the Parties to Convention and the Nagoya Protocol may address the question if DSI should be treated as genetic resource are or due to its non-material nature, not as a genetic resource, and therefore, not within the scope of the Convention or Protocol. Posturing so far suggests that it will be difficult to reach agreement on whether DSI should be treated as genetic resources are, and that potential long-term adverse impacts of treating information as a genetic resource will not be fully considered. This could affect advancing research related to agriculture and food security with long-term adverse impacts to food insecure regions.

The most important question is the long-term outcome of restricting use of DSI, both for research (especially taxonomy) and its use for development, especially in sectors addressing human/animal health and food production. It is very important issue for DSI on GRFA when use of DSI is crucial for development of animal and plant breeding and for supporting sustainable, environmentally friendly food production.
It should also be stressed that the Convention on Biological Diversity calls for facilitated access and transfer of technologies that are relevant to the conservation and sustainable use of biological diversity or make use of genetic resources. Restrictions on the use of DSI would run counter to counter to these provisions.

The restricted use of DSI will also limit non-monetary benefits resulting from for international cooperation in taxonomy, conservation genomics, ability to better address human health issues (e.g. rare and not sufficiently studied diseases) or provide solutions for adaptation to climate change that are enabled by genomics and DSI.

As was concluded in the study by Oldham (2009), the investments and international collaboration that exists in genome sequencing and genomic research are valuable in themselves in terms of knowledge and technology transfer and capacity-building.

6. References

Andersson L., 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature reviews. Genetics* 2.2 pp. 130 - 138.

Bagley M. A., 2015. Digital DNA: The Nagoya Protocol, Intellectual Property Treaties, And Synthetic Biology. Wilson Center, USA (*Synthetic Biology Project*) http://www.synbioproject.org/site/assets/files/1395/digital_dna_final.pdf

Binder S., 2016. Genetic variability in three highly endangered cattle breeds in Austria. *Master Thesis*. University of Natural Resources and Life Sciences. Vienna.

CBD, 2005. UNEP/CBD/WG-ABS/3/INF/4.Global status and trends in intellectual property claims: genomics, proteomics and biotechnology https://www.cbd.int/kb/record/meetingDocument/8667?RecordType=meetingDocument&Event=ABSWG-03

Casas, E., Kehrli & M.E.Jr. 2016. A review of selected genes with known effects on performance and helath of Cattle. *Frontiers in Veterinary Science*. Vol 3, No.113

Dhondt S., Wuyts N., Inze D. 2013. Cell to whole-plant phenotyping: the best is yet to come. *Trends Plant Sci.* 18 428–439. 10.1016/j.tplants.2013.04.008 [PubMed] [Cross Ref]

FAO**,** 2010. Breeding strategies for sustainable management of animal genetic resources*. FAO Animal Production and Health Guidelines*. No. 3. Rome. pp. 75 – 82.

FAO, 2013. In vivo *conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14*. Rome. pp. 150 – 152.

Fernãndez J. & Bennewitz J., 2017. Defining genetic diversity based on genomic tools. In: Oldenbroek, K. (Ed.) *Genomic management of animal genetic diversity*. Wageningen Academic Publishers. pp. 49 – 76.

Hu Zhi-Liang, Park C. A. & Reecy J. M. , 2015. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res*. 2015; 44(D1):D827-D833. doi:10.1093/nar/gkv1233 DOI: Nucleic Acids Res (2016) 44 (D1): D827-D833.

International Committee for Animal Recording (ICAR) http://www.icar.org/

Mauwissen, T.H.E., Hayes, B., Goddard, M. 2016. Genomic Selection: A paradigm shift in animal breeding. *Animal Frontiers*. Vol 6, No.1

Mauwissen, T.H.E. & Oldenbroek, K. 2017. Management of genetic diversity including genomic selection in small in vivo populations. In: Oldenbroek, K. (Ed.) *Genomic management of animal genetic diversity*. Wageningen. pp. 139 – 151.

Oldenbroek, K. (Ed.), 2017. *Genomic management of animal genetic diversity*. Wageningen Academic Publishers, ISBN 978-90-8686-297-9

Oldham, Paul D., 2009. Global Status and Trends in Intellectual Property Claims: Genomics, Proteomics and Biotechnology (January, 22 2009). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1331514

Sabir, J., Mutwakil M., El-Hanafy, A., Al-Hejin, A., Sadek. M.A., Abou-Alsoud, M., Qureshi, M., Saini, K., Ahmed, M. (2014). Applying molecular tools for improving livestock performance: From DNA markers to next generation sequencing technologies. *Journal of Food, Agriculture & Environment*. Vol.12 (2): 351-363.

Toro M. & Mäki-Tanila A., 2007. Establishing a conservation scheme. In: Oldenbroek, K. (Ed.) *Utilisation and conservation of farm animal genetic resources*. The Netherlands: Wageningen Academic Publishers, pp. 75 – 91.

Valdani Vicari & Associati (VVA); Arcadia International E.E.I.G.; Wageningen University (WUR) including Centre for Genetic Resource, the Netherlands (CGN), Plant Research International (PRI) and the socio-economics research institute (LEI); Fungal Biodiversity Centre of the Royal Academy of Arts and Science (CBS-KNAW); Information and Coordination Centre for Biological Diversity of the German Federal Office for Agriculture and Food (BLE-IBV). 2016. Preparatory action on EU plant and animal genetic resources. Final Report. Literature review. EUROPEAN COMMISSION. Directorate-General for Agriculture and Rural Development.

Williams, J.L. (2005). The use of marker assisted selection in animal breeding and biotechnology. *Rev Sci Tech*. 24 (1): 379-391

Woolliams J. A., Matika O. & Pattison J., 2008. Conservation of animal genetic resources: approaches and technologies for in situ and ex situ conservation. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, *42*,pp. 71 – 85. ://doi.org/10.1093/nar/gkv1233

Woolliams, J.A. & Oldenbroek, K. 2017. Genetic diversity issues in animal populations in the genomic era. In: Oldenbroek, K. (Ed.) *Genomic management of animal genetic diversity*. Wageningen Academic Publishers, pp. 13 – 45.