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

Seventeenth Regular Session

Rome, 18–22 February 2019

SUBMISSIONS BY MEMBERS AND OBSERVERS ON "DIGITAL SEQUENCE INFORMATION" ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE

TABLE OF CONTENTS

I. INTRODUCTION

1. The Commission on Genetic Resources for Food and Agriculture (Commission), at its last session, established a new work stream on "digital sequence information" (DSI) on genetic resources for food and agriculture (GRFA) and requested the Secretariat to prepare an exploratory fact-finding scoping study on DSI on GRFA.[1] The Commission requested the Secretary to invite Members to submit information on the use of DSI on GRFA and potential implications for the conservation and sustainable use of GRFA, including exchange, access and the fair and equitable sharing of the benefits arising from their use, and to compile and submit this information to the Executive Secretary of the Convention on Biological Diversity (CBD), as a timely contribution to the process set by Decision XIII/16 of the Conference of the Parties to the CBD.[2]

2. This document contains all submissions received from Members and observers in response to the Circular State Letter of 22 May 2017, for information of the Commission. As requested by the Commission, all submissions were compiled and submitted to the Executive Secretary of the CBD. The submissions have also been made available on the website of the CBD[3] and are reflected in the CBD document *Synthesis of views and information on the potential implications of the use of digital sequence information on genetic resources for the three objectives of the Convention and the objective of the Nagoya Protocol*.[4]

II. SUBMISSIONS BY COUNTRIES

A. BRAZIL

1. DEFINITION AND SCOPE

Digital Sequence Information (DSI) is data information contained in a digital file with a precise order of nucleotides or amino acids. The nucleotide sequence is the main structure of nucleic acid molecules, DNA and RNA, whose main function is the storage and transmission of genetic information.

DSIs can be obtained by sequencing the DNA of organisms already existing in nature, through inference by reverse translation of amino acid sequences or fabricated through simulations and computer programs. In the first two cases they are called natural sequences (NS) and in the third case synthetic sequences (SS).

Public DSI information has been used to boost innovation through basic research, foster scientific collaboration and generate scientific knowledge. One of the most iconic examples is the National Center for Biotechnology Information (NIH-NCBI - https://www.ncbi.nlm.nih.gov/) Genbank database where researchers can currently download about 30 terabytes of data in a single day to make scientific and innovation discoveries – a resource which has been available since 1982. It was designed to provide and encourage scientific community access to DNA sequence information without imposing restrictions on the use or distribution of data. Another one is the European Molecular Biology Laboratory (EMBL). In both the genetic data of all living organisms can be deposited and freely accessed by different research groups, from all over the world.

The scientific community has recognized the benefits of open data, as this information serves as a source of both original and supplemental data for exploration and validation of findings and to make significant discoveries (https://ncbiinsights.ncbi.nlm.nih.gov/2013/09/16/ncbis-open-data-a-source-of-experimental-data-for-important-discoveries/). These international data repositories are fundamental to worldwide research, development and innovation (RDI) because they avoid duplication of efforts and expenditures and allow a much faster advancement of human knowledge. The exchange of digital sequence information is, therefore, essential to achieve important and highly relevant results for humanity. One significant example, in the medical area, is the influenza case (Dormitze,P.R., 2014). In agriculture, DSIs exchange and access might provide a fast response to pests that can devastate thousands of hectares, especially when compared to traditional breeding systems (without the use of

[1] CGRFA-16/17/Report Rev.1, paragraph 86.
[2] CGRFA-16/17/Report Rev.1, paragraph 88.
[3] https://www.cbd.int/abs/dsi-gr/ahteg.shtml
[4] CBD/DSI/AHTEG/2018/1/2.

molecular approach). It can also lead to a significant yield increase production in smaller areas. All those positive results lead to the achievement of the sustainable development goals (SDG).

2. DSI, CONVENTION ON BIOLOGICAL DIVERSITY, NAGOYA PROTOCOL AND FAO COMMISSION ON GENETIC RESOURCES

DSI discussion have started from considerations on Synthetic Biology (Synbio) at Twelfth Conference of the Parties to the Convention on Biological Diversity (CBD/COP12). COP13, serving as the meeting of the Parties to the Nagoya Protocol on Access and Benefit-Sharing (MOP2), adopted decisions on DSI on genetic resources in which they recognized the need for a coordinated and non-duplicative approach on this matter under the Convention and the Nagoya Protocol by decisions XIII/16 and NP2/14.

The FAO Commission of Genetic Resources understood the potential impact of DSI for the conservation and use of genetic resources for food and agriculture and therefore, decided to ask Members to submit information of "DSI on GRFA" and potential implications for the conservation and sustainable use, for the Secretariat to compile and submit this information to the Executive Secretary of the CDB. The Commission also requested the Commission Working Groups and the seven experts nominated for microorganism and invertebrate genetic resources to review and provide inputs to the draft exploratory fact-finding scoping study elaborated by the new work stream stablished at the last meeting, on "digital sequence information on GRFA" prior to its submission to the Commission for consideration at its next session - FAO, "CGRFA-16/17/Report" and "Follow-up to the 16th Regular Session" – FAO C/CBD-7, 22 May 2017.

3. TRACEABILITY, ACCESS AND BENEFIT SHARING

In general, international public databases of digital sequence, such as those cited above (NCBI and EMBL) require the provision of several information of the submitted sequences. In most cases traceability data is required but it is not mandatory as it must be to safeguard the objectives of ABS international instruments. It is therefore important to understand that the traceability of public databases sequences is feasible, but not yet mandatory and to a large extent, available at the species level and subject to the information provided by the scientists.

Some difficulties for traceability could occur: genomes are composed by repetitive regions and those regions are even conserved between species from different kingdoms, which the vast majority is not endemic to a single country. In this way, identical DNA sequences can exist in several organisms originated from different regions of the world. Changes in taxonomic status are still common, with many papers being published every day (e.g., Maronna et al, 2016; Amirahmadi et al, 2016; Uribe et al, 2017; Eyun, 2017). However, on smaller databases, such as the CBOL (www.barcodeoflife.org), whose objective is to identify species and it is required the deposit of a specie voucher in an indexed herbarium/ museum (DNA barcode, Hartvig et al, 2015) traceability is essential. The Barcodes are short nucleotide sequences of a standard genetic locus for use in species identification and taxonomic reviews. It also has application to identify regulated species, including invasive species and to test the purity and identity of biological products. However, it is of the common knowledge that part of DNA research, which seeks to identify molecular markers associated with defense and resistance (abiotic, biotic) genes or adaptation do not need, a priori, the correct species identification.

The difficulty in tracking digital sequences becomes more evident in the case of the microorganisms, since these organisms are characterized by their great diversity, cosmopolitan distribution and fast adaptability to different conditions by evolutionary processes.

For provider countries and countries of origin of genetic resources, traceability is crucial for the full implementation of benefit sharing in the context of the utilization of digital sequence information. Traceability is also fundamental for countries that utilize this information, since a consolidated legal framework is considered favorable by institutions that study the possibility of investing in research in their domestic contexts.

Despite those difficulties it is important to ensure the traceability of sequences submitted in databases. Thus, it is necessary to: (i) establish mechanisms to ensure that the genetic databases require standardized information necessary to the traceability of submitted sequences; (ii) establish differentiated treatment for cases of sequences submitted in genetic databases prior to the definition of this mechanism or whose traceability is not yet possible.

Access and Benefit-sharing has been established under article 10 of the International Treaty on Plant Genetic Resources for Food and Agriculture and is at the core of the work of the Commission on Genetic Resources for Food and Agriculture as it has agreed on the importance of considering access and benefit-sharing in relation to all components of biodiversity for food and agriculture.

Considering the amount of sequence information available in online databases combined with the dramatic reduction of costs and difficulty for DNA sequencing, in the near future the use of a physical sample of genetic material will be unnecessary. Thus, the benefit sharing mechanism should be functional and article 15 of the CBD and article 10 of ITPGRFA are applicable even when the information on genetic resources is digitally obtained. Otherwise it would compromise the achievement of the general purpose of the Convention on Biological Diversity and the International Plant Treaty and the observance its principles, especially in megadiverse and developing countries.

4. POTENTIAL IMPLICATIONS OF DSI - THE BRAZILIAN EXPERIENCE

4.1. Conservation and sustainable use

Brazil is considered one of the most important megadiverse countries on the planet (Forzza et al., 2012). With the most diverse flora in the world, which represents 15% to 20% of the world´s plant diversity. The Brazilian flora include 46,442 species from which 4,753 of Algae, 33,052 of Angiosperms, 1,559 of Bryophytes, 5,722 of Fungi, 30 of Gymnosperms and 1,326 of Ferns and Lyophytes. Despite this, Brazil is far from knowing all the components of the Brazilian flora, given its complexity and dimension.

The relationship between conservation and the List of Brazilian Flora Species is exemplified when, in 2010, the country published the Catalog of Plants and Fungi of Brazil and launched the first online version of the List of Species of the Brazilian Flora, meeting Target 1 of the Global Strategy for Plant Conservation (GSPC-CBD). This botanical milestone was only achieved due to the commitment of more than 400 Brazilian and foreign taxonomists who worked on a platform where information about our flora was included and disseminated in real time. The "Brazilian List", as it was popularly known, closed in November 2015 with the publication of five papers and their respective databases dealing with the different groups of fungi and plants (Brazilian Flora 2020)

It is important to emphasize that the list of species of “Flora do Brasil” was evaluated based on studies of molecular phylogeny and taxonomic reviews. Other search results for "Legume Phylogeny Working Group-LPWG" (2017) on the known non-monophyly of the traditionally recognized subfamily Caesalpinioideae, now recognize not only three but six robustly supported monophyletic subfamilies. This new qualification uses as structure, a more comprehensive phylogenetic analyzes of the legumes up to the information based on sequences of plastid genes, and include almost the complete sampling of genera and ca. 20% (3696) of the known species.

The above example illustrates the impact of molecular biology in conservation studies. This reality is straight related to cost reduction of DNA sequences generation, associated to the high quality and robustness of the analyzed data, providing greater efficiency in germplasm characterization, when compared to morphological traits. In addition, DNA data has been used to boost conservation studies in the following areas: (1) inbreeding depression, when selfing or crossing among relatives are more frequent than random crossing, resulting in high levels of endogamy; (2) deleterious mutations accumulation, which can radically reduce regeneration; (3) lost of genetic variation in small populations, when fragmentation (among other anthropomorphic actions) leads to a significant reduction in the genetic pool, causing a significant reduction in genetic variability; (4) adaptation to captive or isolation and its effect on reintroduction; (5) outbreeding depression, when the progeny shows lower fitness when compared to the parents in consequence of the high genetic distance between both parents; (6) fragmentation of populations and reduction in migration, when fragmentation causes isolation and prevents gene flow; and (7) taxonomic uncertainties and introgression (Frankham, 1995). To better understand those areas the free data access is very important to avoid duplication of efforts and to improve statistical robustness of analysis by considering data generated from multiple studies.

Regarding *ex situ* conservation, molecular characterization allows the generation of DNA data on a large scale, providing not only the optimization of conservation by enabling the elimination of duplicates, but also helping in the evaluation of the genetic representativeness of the germplasm banks, directing the needs of new introductions (e.g., Pessoa-Filho et al., 2007). Also allows comparison of representativeness among different germplasm banks, making it possible to identify access that may be

at risk for being conserved in only one place. The DNA-based study greatly assists the formation of nuclear collections by allowing an accurate analysis of the collection's genetic diversity. It also assists in the formation of thematic nuclear collections by identifying genotypes resistant to biotic and abiotic factors. Still, it allows the differentiation of cultivars and the identification of landraces.

Microorganisms are an essential part of the biodiversity in nature and one of the greatest genetic reservoirs on Earth. These microscopic organisms are major players in the cycles of energy and matter, acting as protagonists of the biogeochemical cycles and the essential biochemical reactions that sustain the entire life in the planet. Understanding the role of microorganisms in the environment has provided conditions for the development of new biotechnological applications and the establishment of food security policies, sustainable agriculture and industrial development programs. However, scientists have only begun to understand this vast world, since a small fraction of the microbial diversity is known and/or can be cultured under lab conditions (Stewart 2012, Garza & Dutilh 2015). This fraction has been kept in culture collections, representing an important genetic resource for the countries, providing stocks of material for use in programs of interest to society. In Brazil, microorganisms have received more attention in the last decades, mainly due to their increasing importance for agriculture, food and environment, either in the form of bio-products or in different industrial and biotechnological processes.

Because of the advances in molecular biology and DNA sequencing techniques in the last decades, DSI of microorganisms available in different on-line databases grew at the same rate as sequencing costs were reduced. The availability of this information and free access to these sequences have been providing conditions for a huge scientific step in phylogenetic studies, biological characterization and ecology of microbial species. For most of the microorganisms, and despite of the importance of the morphological characterization, identification at the specific level depends exclusively on gene sequencing and phylogenetic analysis. The precise or correct identification of microbial strains deposited in culture collections is essential for conservation purposes and for the material exchange among institutions aiming its use in research programs of interest to society. It is noteworthy that advances in metagenomics have created the possibility of obtaining complete or nearly complete sequences from uncultured microorganisms, which can be collected directly from any environment, providing an important field for the study of biology, ecology, evolution and functional structure of microbial communities (Garza & Dutilh, 2015).

With regard to *in situ* conservation, the use of DNA sequences helps to determine priority conservation areas by studying the genetic diversity of species in different populations. It assists in evolution studies, taxonomy and biogeography. With the use of DNA sequences it is possible to identify populations with greater genetic diversity, populations with different gene pools, processes of evolution and adaptation. These molecular tools can aid in species identification (e.g., Nithaniyal et al, 2014) and the origin of the sample when the sequences obtained by different groups from different countries are made available in public databases and public articles. Another study case was identification of geographic origin of timber where researchers ensured that illegally harvested timber was traced (Degen et al, 2013). The mahogany logging in Brazil is illegal, but it is not in other countries of the Amazon region.

A new concept in conservation genetics is environmental DNA (eDNA). This makes reference to all genetic material that can be extracted from a bulk of environmental sample as soil, water and even air. This approach allows a significant and unprecedented ability to identify a range of species that are present in different areas and biomes and conduct analyzes for conservation, management and research. This kind of study has a high potential for helping in estimating population size, genetic diversity and also allow genomic analyzes, including biomolecules, such as eRNA and ePTNs (Barnes and Turners, 2016).

Knowledge about biodiversity is the first step towards conservation as well as a source of immense potential for economic use. To achieve these goals, the Brazilian Ministry of Environment's - Plants for the Future - initiative sought to promote the sustainable use of native species of Brazilian flora of current and potential economic value, used locally and regionally. Although today we have a considerable number of native species domesticated, or in process of domestication, the use of genetic resources of native species in commercial scale is still embryonic (Coradin et al.,2011). The use of the information already collected and published could help Brazil and other countries to reach a better nutritional status, also prompted by the SDGs (www.mma.gov.br/publicacoes/biodiversidade).

4.2. Breeding and exchange

The drastic reduction of arable areas coupled with extreme climatic events and the significant increase in the demand for food raise the awareness and dedication of genetic improvement in the development of cultivars/ breeds resistant to different biotic and abiotic factors such as pests and restriction of water use, respectively (Kage et al, 2016). The advance in DNA sequencing, with the new generation sequencing allows to integrate information of metabolomics, proteomics, transcriptome, together with mapping association studies, facilitating the identification of genes allelic variants, accelerating the breeding process through the development of functional molecular markers (FMMs) in plants or genomic selection in livestock.

The green revolution was achieved due to the development of dwarf and semi-dwarf cultivars of rice and wheat. Mutation points identified in two genes were identified in wheat, *Rht-B1b* on chromosome 4B and *Rht-D1b* on chromosome 4D. Markers for allele-specific PCR were developed and applied in selection of wheat genotypes with this characteristic of interest (Ellis et al, 2002). Deletion-based markers were developed for Dwarf8 gene coding region in maize for identification and selection of plant height and flowering time (Thornsberry et al 2001). There are hundreds of examples like these that make it clear that the use of DNA sequences is a reality for genetic breeding, since the research and development of a product goes far beyond the simple identification of a gene, but of its variables among different accesses. Each one of those accesses) may be originated in a different country. Most of the times scientific advances, specially for food and agriculture, are not due to the identification of a "magic gene", that occur in only one species. If it occurs, it is the exception. Most of the times, scientific advances as those mentioned above, are due to a massive characterization of lots of different accesses in order to identify differences in genotypes, including known hundreds if not thousands of genes, most of them common in many different species and available in public databases.

The identification of genes and their functions, and the availability of this information as DSI, has been expanding the field of applied research in areas of interest for food production and bio-industry. There are several applications of the microbial genome sequences in industrial processes (bioenergy, biofuels, processed foods and derivatives), on the improvement of microbial strains or in the development of clones with specific genes of interest (Fávaro, 2012), and in plant transformation (Shelton et al. Al, 2002, Lemaux, 2008).

The same is true for animal genetic resources (AnGR) that are represented by all species (exotic and native, terrestrial or aquatic) with direct or potential use for food and agriculture. From a specific gene banking point of view, there is a growing trend to make germplasm (semen and embryos), genomic, and phenotypic information associated with the sample available to the public and private sector. In 2014, Brazil, Canada and the US launched efforts to incorporate genomic information into their joint database Animal-GRIN (http://nrrc.ars. usda.gov/A-GRIN/ or http://aleloanimal.cenargen.embrapa.br/) and to make it publically available through the internet (Paiva et al., 2016). This approach reduces duplication of genotyping efforts and leverages investments already made in genotyping. Sample acquisition coupled with digital curation of genomic and phenotypic information represents new areas of repository utilization.

From a general point of view, public DSI databases have played and will continue to play a significant part for FAO's "Global Plan of Action" and "State of the World's Animal Genetic Resources" publications and policy directions. Without those data bases it is not going to be possible, for example, to expand current knowledge about origin and diversity of all livestock and to fulfill the Strategic Priority 1: characterization, inventory and monitoring of trends as associated risks to AnGR.

Brazil is recognized to be one of the most important players in animal protein production and exportation (USDA-FAS, 2017). For example, the country is the biggest broiler meat producer and the second largest beef producer in the world. To keep up with those standards, Brazil and other countries are sharing genetic information to improve the breadth and reach of conservation and breeding programs (e,g., Bovine, Ovine and Swine Genomic Consortia). Impacts of controlling the use of genetic information obtained from digital sources and imposing benefit sharing obligations for non-commercial research are by far the most difficult to predict, monitor and manage. Therefore, operational procedures to obtain prior informed consent and monitoring the use of DSI should be facilitated and the most user friendly as possible.

Impacts of controlling the use of genetic information obtained from digital sources for development of new breeds with potential use for food and agriculture could specially impact the use of aquatic species with potential for aquaculture, as most of these are currently undergoing domestication. In this sector, many production and supply chains are under development, with high demands for new and innovative technologies.

Regarding exchange of genetic resources, the presence of public DSI data bases is essential to speed up germplasm movement and manage countries food security. For example, it is crucial to have preventive breeding actions, especially for resistance to economically important diseases/pathogens that are not present in a particular country or region. The growing frequency of extreme climate events is expanding disease zones very rapidly (mainly South to North). Mitigation of those events will just be feasible if DSI information and germplasm are available among countries to allow swift actions to be set in place before outbreaks can destroy entire production and supply chains. A Brazilian long-term experience, the Embrapa-Labex USA, a 20 years partnership and collaboration with USDA/Agriculture Research Service is proving to be very successful in this regard.

5. RELATED BRAZILIAN LEGISLATION

Brazilian Law 13,123/2015 defines genetic resources (or genetic heritage) as the genetic information from plants, animals, and microbial species, or any other species, including substances originating from the metabolism of these living organisms (Article 2, I). Furthermore, Decree No. 8,772/2016, in what concerns registration in the National System for Genetic Heritage - SisGen states:

"*Art. 22. For the **registration** of access to genetic heritage or associated traditional knowledge, the natural or legal person must complete the electronic form of SisGen that will require: ...*

F) identification of the genetic heritage at the strictest possible taxonomic level or associated traditional knowledge, as the case may be, in particular:

*1. the origin of the genetic heritage, including georeferenced coordinates in the degree, minute and second format, from the place of in situ production, even if they have been obtained from ex situ or **in silico** sources; and...."*

…….

"§ 1 When it is not possible to identify the geo-referenced coordinate of the in situ sampling place, which is dealt with in item 1 of item "f" of item II of the caput, and only in cases in which the obtaining of the genetic patrimony occurred prior to the entry into force of Law 13,123, of 2015, the origin May be informed on the basis of the most specific geographic location possible by one of the following means:

*I - identification of the ex situ source of the genetic heritage, with the information Constants in the deposit register, when it comes from ex situ collection; or II - Identification of the genetic heritage origin database with the information In the deposit record, when it comes from an **in silico** database."*

Thus, the regulation of Law 13.123,2015 provides that research utilizing genetic information obtained *in silico* is to be carried out freely, and that registration is required only at the time of publication of the results, or upon application for a patent, or before introduction of a product on the market. The Brazilian legislation also brings facilitated mechanisms for access to genetic resources for food and agriculture, special considerations for scientific research for food and agriculture; economic exploitation as the point of incidence of benefit sharing obligation and the establishment of a benefit-sharing fund.

6 REFERENCES

Barnes MA, Turner CR. The ecology of envireonmental DNA and implications for conservation genetics. Conservation Genetics, 17(1): 1-17, 2016.

Brazilian Flora (Flora do Brasil) 2020. Jardim Botânico do Rio de Janeiro. Disponível em: < http://floradobrasil.jbrj.gov.br/ >. Acesso em: 27 Ago. 2017.

Degen B; Ward SE, Lemes MR, Navarro C, Cavers S, Sebbenn AM. Verifying the geographic origin of mahogany (Swietenia macrophylla King) with DNA-fingerprints. Forensic Science International: Genetics, 7(1):55-62, 2013.

Dormitzer,P.R. Rapid production of synthetic influenza vaccines. Curr. Top. Microbiol. Immunol. 2015;386:237-73.

Ellis M, Spielmeyer W, Gale K, et al. (2002). ‘‘Perfect’’ markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. Theoret Appl Genet, 105, 1038–42

Fávaro LCL. 2012. Melhoramento genético de microrganismos para produção de biocombustíveis e químicos renováveis: integração de abordagens de biologia molecular, biologia sistêmica, biologia sintética e engenharia metabólica. Agroenergia em Revista 5: 32-35.

Forzza, R. C. et al. 2012.New Brazilian Floristic List Highlights Conservation Challenges. Bioscience 62:39-45.

Frankham R. Conservation Genetics. Annual Reviews in Genetics, 29:305-27, 1995.

Garza DR, Dutilh BE. 2015. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems Cell. Mol. Life Sci. 72:4287-4308.

Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I. The Use of DNA Barcoding in Identification and Conservation of Rosewood (Dalbergia spp.). Lumbsch HT, ed. PLoS ONE. 2015;10(9):e0138231. doi:10.1371/journal.pone.0138231.

Kage U, Kumar A, Dhokane D, Karre S and Kushalappa AC. Functional molecular markers for crop improvement. Critical Reviews in Biotechnology 36(5), 917-930, 2016

Legume Phylogeny Working Group (LPWG). 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66: 44–77.

Lemaux, P. 2008. Genetically Engineered Plants and Foods: A Scientist’s Analysis of the Issues (Part I). Annual Review Plant Biology 59:771-812

NCBI > National Center for Biotechnology Information <https://www.ncbi.nlm.nih.gov> . Acesso em Ago 2017

Coradin, L; Siminski, A; Reis, A (2011). Espécies nativas da flora brasileira de valor econômico atual ou pontencial: plantas para o futuro – Região Sul. / Ministério do Meio Ambiente. Secretaria de Biodiversidade – Brasília, DF: MMA, 934p.

Nithaniyal S, Newmaster SG, Ragupathy S, Krishnamoorthy D, Vassou SL, Parani M. “DNA Barcode Authentication of Wood Samples of Threatened and Commercial Timber Trees within the Tropical Dry Evergreen Forest of India.” Ed. Arndt von Haeseler. PLoS ONE 9.9 (2014): e107669. PMC. Web. 25 Aug. 2017.

Paiva, S. R., McManus, C. M., & Blackburn, H. (2016). Conservation of animal genetic resources–A new tact. Livestock Science, 193, 32-38.

Pessoa-Filho, M., Beló, A., Alcochete, A. A., Rangel, P. H., & Ferreira, M. E. (2007). A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. BMC plant biology, 7(1), 23.

Shelton AM, Zhao JZ, Roush RT. (2002). Economic, Ecological, Food Safety, and Social Consequences of the Deployment of Bt Transgenic Plants. Annual Review of Entomology 47:845–81

Stewart EJ. 2012. Growing Unculturable Bacteria Journal of Bacteriology 194(16): 4151–4160.

The Angiosperm Phylogeny Group . (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants : APG IV. Bot. J. Linn. Soc. 181: 1-20

Thornsberry JM, Goodman MM, Doebley J, et al. (2001). Dwarf polymorphisms associate with variation in flowering time. Nat. Genet, 28, 286–9.

Vieira, RF.; Camillo, J.; Coradin, L. (2016). Espécies nativas da flora brasileira de valor econômico atual ou potencial: Plantas para o Futuro - Região Centro-Oeste / Ministério do Meio Ambiente. Secretaria de Biodiversidade – Brasília, DF: MMA, 1.160p.

Science Policy Team of the Royal Society of Biology, (2017) - response to consultation by DEFRA about DSI, July.

FAO, Sixteenth Regular Session of the Commission on Genetic Resources for Food and Agriculture, CGRFA-16/17/Report. Rome, 30 January – 3 February, 2017.

B. CANADA

During its Sixteenth Regular Session held 30 January – 3 February 2017, the Commission on Genetic Resources for Food and Agriculture established a new work stream on "digital sequence information on genetic resources for food and agriculture" (Report, para.86). The Commission requested its Secretary (para.88) to invite Members to submit information on the use of "digital sequence information on genetic resources for food and agriculture" and potential implications for their conservation and sustainable use, including exchange, access and the fair and equitable sharing of the benefits arising from their use, and to compile and submit this information to the Executive Secretary of the Convention on Biological Diversity (CBD), as a timely contribution to the process set by decision CBD COP XIII/16.

The present document includes Canada's information on the use and potential implications of "digital sequence information" on genetic resources for food and agriculture.

Introduction

Agriculture, fisheries and forestry, as well as food and non-food product value chains, depend on living organisms; this includes pollinators, naturally-occurring organisms that support ecosystem services (e.g. soil microorganisms), and natural enemies that can be used as biocontrol agents to manage pests. Biodiversity is the pool of resources upon which production systems are based. A vast pool of genetic diversity has been developed under domestication and is in urgent need of effective conservation and sustainable use strategies around the world. Therefore, efforts to understand, preserve and utilize biological diversity for agriculture, forestry and fisheries have been and continue to be essential to bio-based industries and to sustain the enormous improvements science has brought to production, processing, conservation efforts and utilization of genetic resources.

The study of biodiversity has become central to Canada's research activities as science started uncovering the important interconnections between biodiversity, productivity and resiliency in agriculture, fisheries and forestry. The study of the biological diversity of crops, animals, forests, wild relatives, beneficial species, pests, invasive and alien species, and native species enables advances in areas such as genetic improvement of forest populations and breeds of farm animals, improvement of pest and nutrient management methods, development of new crops or crop attributes, threat identification and risk mitigation, and identification and management of invasive alien species. It supports resiliency of productive systems by enabling the development of beneficial management practices to manage the effects of intensive management practices on populations of beneficial organisms and the ecosystem services they provide.

Furthermore, in an increasingly globalized world, the movement of pests and diseases is a constant threat to production and could result in catastrophic losses of major economic importance for the agriculture, fisheries and forestry sector, as well as catastrophic damage to biodiversity. Therefore, the capacity to identify and characterize organisms that may threaten its ecosystems and biological resource base is vital.

The Government of Canada has committed to openness, transparency and information sharing through the Open Government Partnership and the Global Open Data for Agriculture and Nutrition initiative (GODAN - http://www.godan.info/). Since joining the Open Government Partnership, a multilateral initiative to foster greater transparency and accountability, Canada developed three national action plans aimed at enhanced access to information and expanded open data, among other things. The Ministry of Environment and Climate Change Canada is the federal lead on the Open Science element of Canada's 2016-2018 Action Plan on Open Government. Sharing data related to genetic sequences is not new – the GenBank initiative (https://www.ncbi.nlm.nih.gov/genbank/) has made genetic sequence data available worldwide since 1982. Full sequences of genotypes of various crops (wheat in 2012, soybean in 2010 and others) and annotations of animal genomes carried out by various countries (http://www.ensembl.org/info/about/index.html) have also been published and made available online.

As a federal government department, the Ministry of Agriculture and Agri-Food Canada (AAFC) is also committed to providing access to data produced, collected and used to the Canadian public, researchers and industry communities. AAFC is a contributor to the Government of Canada's Open Data Portal and more recently, in May 2016, joined the GODAN initiative, which supports global efforts to make nutritionally relevant data publically available for use worldwide. Joining GODAN makes AAFC's data

accessible to an even wider audience and helps strengthen ties with key partner countries and organizations.

Given the importance to agriculture of understanding the various organisms at play, their life cycles and the conditions they need in order to thrive, the study of agrobiodiversity and bio-resources has always been a priority research area supported by Agriculture and Agri-Food Canada. Its scientific collections, as well as the biosystematic study of species important for agriculture, are as old as research in this department (128 years). Over the years, AAFC has built a critical mass of knowledge, expertise and science capacity to study taxa of importance for agriculture. AAFC shares information on its holdings of plant and animal genetic resources collections using Internet websites and other tools. Biodiversity in agricultural systems can be studied at the various levels of organization, and all these levels are important: alleles, genes, populations, species, communities in an agricultural landscape and the agro-ecosystem as a whole. This ranges from genetic variation that provides the basis for valuable traits in a crop or affects the behaviour of a pest species, to populations of soil microorganisms and their importance for nutrient or carbon cycling or productivity of agricultural crops dependant on pollinator populations.

Forests are the most diverse ecosystems on land, containing the majority of the world's terrestrial species. Timber, pulpwood, firewood, cash crops, fish and medicinal plants from forests provide livelihoods for hundreds of millions of people worldwide. Unlike agricultural plants, forest trees are largely undomesticated and highly heterozygous, due to their outcrossing breeding systems.

In Canada, genetic sequence data is used to generate molecular markers which are important tools for forest trees and associated forest species (e.g. caribou) contributing to areas such as: (1) the assembly of breeding populations in newly developed and advanced breeding programs, and (2) species conservation activities such as the proper delineation of species taxonomy for management issues associated with conservation. Integrative approaches are being implemented within Canada to address the impacts of abiotic and biotic stressors. An example of this is where future climate modeling has been very successful in uncovering potential threats of declines in genetic diversity and the distribution of forest tree species, so that timely precautions to conserve the species can be undertaken.

Canada has developed the National Forest Information System (https://ca.nfis.org/index_eng.html) in 2000 as a collaborative effort with Natural Resources Canada, Canadian Forest Service and the Canadian Council of Forest Ministers to provide access to the most current, consistent and reliable forest resources information, integrating information across jurisdictional boundaries. This system contains data that allows for an accurate picture of Canadian forest practices and forest biodiversity, including assessing genetic conservation requirements of native tree species of Canada, information concerning threats to genetic diversity (e.g. invasive alien species), species biology and ecology (https://pfc.cfsnet.nfis.org/CAFGRIS/home.jsp).

Nonetheless, characterising biodiversity can be challenging. For example, the soil microbiome is composed of microscopic organisms that are in many cases very challenging to isolate, making downstream analysis complex. Other examples include the identification of very similar species or detection of alien invasive species over a vast territory. Genomics has provided very powerful solutions, which allow the identification of specimens without having to isolate them. This highlights the importance of genetic sequences for biodiversity in bio-based resource sector science.

Terminology

Canada is not comfortable with the term “digital sequence information” to characterize this data. First, the sequences in question are genetic sequences, not (for example) mathematical sequences or sequences of events. Therefore, to avoid confusion, the word “genetic” should be part of the term identifying this issue.

Second, the fact that genetic sequences can be conveyed digitally is of secondary importance; the “digital” aspect only refers to a particular mode of storage or mode of transmission. Genetic sequences listed on a sheet of paper would still be genetic sequences. Therefore “digital” should be removed from the term.

Third, the term “information” is not appropriate. “Data” would be more accurate, because data can be codified and is transmissible, which is the case for genetic sequences, but not always the case for information.

Therefore, Canada would prefer to use the term "genetic sequence data" (GSD), which we shall use in the rest of this document.

Genetic Sequence Data are not Genetic Resources

It has been proposed that GSD should be treated as if they were genetic resources. Canada does not agree; we are of the view that GSD are not genetic resources.

The definition of genetic resources in three legally-binding international instruments makes that clear.

In the Convention on Biological Diversity, Article 2:

> *"Genetic resources"* means genetic material of actual or potential value
>
> *"Genetic material"* means any material of plant, animal, microbial or other origin containing functional units of heredity.

In the International Treaty on Plant Genetic Resources for Food and Agriculture, Article 2:

> "Plant genetic resources for food and agriculture" means any genetic material of plant origin of actual or potential value for food and agriculture.
>
> "Genetic material" means any material of plant origin, including reproductive and vegetative propagating material, containing functional units of heredity.

In the Nagoya Protocol on Access to Genetic Resources and Benefit-Sharing, Article 2:

> The terms defined in Article 2 of the Convention shall apply to this Protocol.

These definitions are consistent and aligned. They state clearly that genetic resources "contain" "functional units of heredity" like DNA. The functional units of heredity are not themselves genetic resources, and even less so the base sequences within these functional units. Genetic sequence data, by their nature, do not and cannot contain functional units of heredity like DNA. Genetic sequence data do not accord with the legal definition of genetic resources in these international instruments. They therefore provide no legal basis to treat GSD like genetic resources.

There are also risks incurred in treating GSD as genetic resources. Genetic sequence data are not permanent and would change over the years; the sequences will be modified by punctual, random mutations. It is also evident in other articles of these instruments that they were not intended to relate to genetic sequence data. For example, provisions in these international instruments to conserve genetic resources certainly do not apply to strands of DNA.

Examples of Research Using Openly Available Genetic Sequence Data

In this section, we provide examples of research done in Canada that require simple and ready access to genetic sequence data of Canadian and international origin. They will, in particular, demonstrate that the use of GSD is not limited to substituting for genetic resources. We will also describe how access to GSD is important for these projects to support the objectives of the CBD. This research contributes to implement Articles 7, 10, 12 and 17 of the Convention on Biological Diversity, and contributes to meeting reporting requirements as defined in Article 26. It also supports the delivery of one of the 2020 Biodiversity Goals for Canada: "By 2020, Canadians have easy access to adequate and relevant information about biodiversity and ecosystem services to support conservation planning and decision-making", and in particular Target 14: "By 2020, the science base for biodiversity is enhanced and knowledge of biodiversity is better integrated and more accessible." (http://www.biodivcanada.ca/default.asp?lang=En&n=9B5793F6-1&offset=3#target_14);

1. *Study of soil and water microbiome: the EcoBiomics project*

Agriculture and Agri-Food Canada is leading a federal interdepartmental Genomics Research and Development Initiative (GRDI) that aims to develop new knowledge to improve water quality and soil health by comprehensively characterizing aquatic microbiomes, soil microbiomes, and invertebrate zoobiomes, and testing hypotheses to enhance environmental monitoring, assessment and remediation activities. It also aims at establishing comprehensive biodiversity baselines for assessing future changes to water and soil biodiversity at important long-term environmental monitoring sites in Canada.

The EcoBiomics project directly supports the objective of identifying biological diversity for conservation. This project uses metagenomics approaches to profile microbial and invertebrate communities in varying habitats tied to water and soil including pristine natural areas, agricultural, forestry and fishery systems, oil sands, and lakes and rivers. The result of the sampling in these locations based on standardized molecular approaches will identify the biodiversity within a taxonomic, ecological and functional context. The ability to do this analysis is absolutely dependent on known genetic sequence data from sources such as GenBank to identify the taxa in the samples and potentially their role in the ecosystem being studied. This project may also contribute to the identification and naming of taxa that are not known to science and to share this knowledge openly. This biomonitoring is essential to understand taxa and communities that may need conserving and thus potential regulatory action. There is also a strong economic tie to the production systems involved as maintaining diversity is typically correlated to healthy systems, and thus production as well as monitoring for alien invasive species and functional changes that could damage productivity.

The objectives of EcoBiomics also directly support the sustainable use of the components of biological diversity, as the production systems rely on sustainable biodiversity-based ecosystem services to be productive. The biomonitoring will produce metrics to assess sustainability. It should also indicate what remediation in terms of soil or water diversity is necessary to allow ecosystems to return to "normal" function.

Fair and equitable benefits of this project can only be realized through sharing genomic sequence data. Without sharing, the reference sequences will be internalized and there will essentially be no libraries available to carry out identification and quality control. This would obstruct research on biodiversity, especially on microbial life, and have a ripple effect to industry and the production systems they serve. Projects like EcoBiomics would no longer be possible if access to reference libraries of genetic sequence data and associated annotations becomes complicated and expensive.

2. *Studying pollinator diversity and understanding threats to their populations*

At least 35% of global crop production, and the majority of fruits, depend on bees and other insects for pollination services. While most interest in insect pollinators has focused on commercial honey bees, there is growing understanding that native bees are also efficient and effective crop pollinators. Alarming declines in both commercial honey bees and wild bee populations have led to increased fears that current agricultural productivity may be unsustainable without concerted efforts to maintain and enhance both wild and cultivated bee populations. Agriculture and Agri-Food Canada conducts research focussing on documenting bee diversity in our agricultural landscapes and studying the impacts of land use, pesticides, and pathogens on bee populations. This work involves sampling large numbers of bees which all need to be identified to species level.

In order to efficiently and accurately identify species, we often sequence various genomic regions of these bees and compare our genetic sequences to those that have been made available in online databases through collaborative efforts of the scientific community for well over a decade. Having access to genetic sequences of bees from other countries is vital to be able to detect any newly introduced species that our samples may be the first to detect in Canada. We also make use of worldwide genetic sequences of pathogens known to negatively impact bees in other parts of the world, in order to screen bees sampled in Canada for these pathogens. Importantly, the genetic sequences from the thousands of bees and pathogens that we have produced are invaluable tools for other researchers worldwide.

3. *Understanding gene flow related to herbicide resistance*

Since the beginning of agriculture, controlling weeds has been a major concern for farmers. Initially these were managed through labour intensive hand weeding. Currently, over 450 weed species globally have evolved resistance to at least one, but often many, herbicides. Within Canada, just over 60 species are herbicide resistant and in the United States there are over 150, some of which are spreading or will spread into Canada. Furthermore, with the introduction of crops modified to have herbicide resistance, the genes that confer these traits could move into wild relatives, providing additional challenges for weed management on the farm. Herbicide resistant weeds, whether they evolved or received this resistance through gene flow, are a clear threat to agricultural production within Canada and globally. Within the *Brassicaceae*, it has been documented that transgenes for herbicide resistance can move from crops (e.g. canola, *Brassica napus*) into weeds (e.g. Bird rape, *Brassica rapa*).This creates herbicide

resistant weeds that are difficult to control. Further, if canola crops with different herbicide resistance genes are grown, it is possible for these traits to combine resulting in weeds with multiple resistances. As a result, it is important to assess the potential of transgenes escaping from *Brassicaceae* crops

The research conducted at Agriculture and Agri-Food Canada to address this threat relies at every stage on free and open access to genomic resources, including genetic sequence data. As a first step, we use freely available information on genes from species around the world to determine which weed species are most closely related to novel crops and, therefore, most likely to receive transgenes. This enables efficient allocation of resources to where the risk of gene flow is the highest. Second, we use similar information to develop molecular markers to detect hybrids between crops and their wild relatives, allowing determination of the rate at which this hybridization is likely to occur. This information is needed by the Plant Biosafety Assessment Office at the Canadian Food Inspection Agency to make informed, science-based decisions on the unconfined release of enhanced cultivars. Finally, the availability of accessible published genetic sequence data allows us to create the tools needed to dissect the consequences of hybridization so that an informed risk benefit analysis can be made including: 1) pinpointing the potential frequency of this hybridization in nature, 2) determining the portions of the genome that are most likely to be exchanged, and 3) quantifying how quickly a transgene may spread once introduced. For example, a glyphosate resistant (e.g. Roundup resistant) ecotype of the weed *Kochia scoparia*, has recently invaded Canada from the USA and the gene that confers this resistance is spreading through Canadian populations. This species has the potential to cause devastating losses in wheat and soybean. Its genome has just been made available allowing us unprecedented ability to understand how seed and pollen movement contribute to the spread of these genes through populations. This information will contribute to the development of more effective and sustainable weed management strategies for this species.

This work supports the mainstreaming of biodiversity in agricultural production, and the promotion of "awareness, use and sustainable use of agricultural ecosystems; improve its productivity and diversification; integrate positive incentives for biodiversity in field production and reduce agricultural pollution".

4. Protection of Canadian biodiversity by monitoring invasive alien species

Every year, new and emerging insect pests arrive in Canada and threaten Canadian biodiversity as well as production of agricultural crops. Chemical pesticides are often the primary short-term solution for crop protection against invasive insects. However, alternative pest management solutions are urgently needed as public concern over pesticide use continues to rise, and as many pesticide chemistries are phased out. Canada conducts research to identify organisms from all life stages through the development of new genomics tools for efficient detection of quarantine and invasive species; thus substantially bolstering Canada's operational strategy for both prevention and effective eradication or mitigation of new invaders.

To make this work possible, scientists use DNA barcodes to identify invasive and pest insects of importance to agriculture and forestry. Genetic sequences are a very important tool for taxonomists and work well in combination with other systems of characterization (morphology, host plant identities and biogeographic patterns). Genetic sequences allow scientists to rapidly distinguish species that are potentially invasive and harmful from those that are beneficial and part of natural ecosystems. Globally, this project is central for Canada's work on alien invasive species, as it provides a diagnostic tool for early detection (significantly lowering risk of establishment), surveillance and management of thousands of species, including alien invasive species.

A project funded by Genome Canada is involved in the protection of Canadian biodiversity through biosurveillance of Alien Forest Enemies (BioSAFE). It is led by researchers from Canadian universities in collaboration with the Canadian Food Inspection Agency. This national level project will enable forest health professionals to track and identify forest invasive pests and diseases using a genomics based approach. This project will enable the development of DNA-based diagnostic tests to identify and monitor these pathogens quickly (within hours) and accurately, which will then be used as a decision support tool to mitigate threats. The success of this project requires the availability of open access to GSD of known pathogens for identification.

5. Inventorying forest genetic diversity

Canada has a long history of forest research that generates genetic sequence data and utilizes open access GSD. Research has been conducted through diverse collaborative efforts involving the federal government (Natural Resources Canada, Canadian Forest Service), jurisdictional governments (e.g. British Columbia, Alberta, Ontario, Quebec and New Brunswick), academia (e.g. University of British Columbia, Laval University) and industry. This research has included assessing the genetic diversity in natural and managed forest tree populations and in identifying the geographic scale for capture of diversity, and for the conservation of rare forest tree species. Research of this type contributes directly to *ex situ* and *in situ* forest conservation activities and tree improvement programmes.

Improving the understanding of genetic diversity is recognized as important for the sustainable management of forest genetic resources. It is recognized that monitoring changes below the species level provides necessary information for ensuring that adaptive potential is maintained so that species can evolve in response to changing environmental conditions. Ensuring that species can respond to environmental change continues to be priority for much of the forest research conducted within Canada.

Other implications for the Government of Canada

Further to supporting the three objectives of the CBD, the above-mentioned research activities also contribute to support other Canadian international engagements such as:

- Meeting Canada's obligations regarding the stewardship of plant genetic resources for food and agriculture based on the *International Treaty on Plant Genetic Resources for Food and Agriculture* (http://www.planttreaty.org/). Art.13.2(a) identifies exchange of information as a benefit of the Multilateral System, including information on technologies, and results of technical, scientific and socio-economic research, including characterization. Genetic sequence data are a tool in the sustainable use of plant genetic resources, as per Article 6, paragraphs 6.2 (b), (c), (d), and (f) and in monitoring their degree of genetic variation as per Article 5.1(f);
- Implementation of NAPPO objectives. The North American Plant Protection Organization (NAPPO) was created under the International Plant Protection Convention (IPPC) to help achieve the goal of protecting the world's cultivated and natural plant resources from the spread and introduction of plant pests while minimizing interference with the international movement of goods and people. The regional mandate for NAPPO was formalized by Canada, the United States and Mexico in a Cooperative Agreement signed in 1976 at the Minister/Secretary of Agriculture level.
- Implementation of the World Trade Organization (WTO) Sanitary and Phytosanitary (SPS) Agreements. The World Trade Organization and its agreement on the application of the SPS Agreements came into force in 1995. The WTO-SPS Agreement recognizes the International Plant Protection Convention (IPPC) as the relevant international organisation for the elaboration of science-based International Standards for Phytosanitary Measures (ISPMs) to help ensure that phytosanitary measures are not used as unjustified barriers to trade. Regional Standards for Phytosanitary Measures (RSPMs) are often established before the international ones by organizations such as NAPPO or the European Plant Protection Organization (EPPO).
- Supporting the recommendations of the Organization for Economic Cooperation and Development (OECD) for Biological Resource Centres (BRCs) which specify that "for networking of BRCs to be truly successful, technology for co-ordinating and combining catalogues and databases to meet the requirements of science in the post-genomics era will have to be implemented." Informatics needs for Biological Resource Centres are specified in the OECD Best Practice Guidelines for Biological Resource Centres.
- Implementation of the DivSeek Initiative (http://www.divseek.org/), of which Agriculture and Agri-Food Canada is a partner. DivSeek brings together 69 founding organizations, including the Global Crop Diversity Trust, the Consortium of the Consultative Group on International Agricultural Research (CGIAR), the Beijing Genomics Institute and the Global Plant Council. Its aim is to "unlock crop diversity" by extracting genetic information from samples of crop varieties and wild relatives stored in gene banks. The overall aim is to accelerate development of "climate-ready, high-yielding and nutritious crops", which is a global good and benefits everyone.

Concluding Remarks

The various applications of genetic sequence data make important contributions to the utilization of genetic resources for food and agriculture, and to practical market-based or producer-oriented options that can enhance improvement of crops, farm animals, forests and fisheries, and facilitate their effective conservation.

Restricting the dissemination of genetic sequence data would limit the benefits that can be generated from the use of genetic resources and jeopardize their efficient conservation. Sharing information on genetic sequences is the corner-stone of entire scientific disciplines such as phylogeny, phylogenetics, molecular biology, molecular genetics, and many more, contributing to the global understanding of life in all its forms. Understanding diversity within a species is needed to maintain the viability of ecosystems and ecosystem services. Genetic sequence data are not genetic resources and should not be treated as such.

Genetic sequence data has the potential to bring numerous benefits to Canadian and global society. In this document, we provided examples of research projects that directly support the objectives of the international agreements on biodiversity and genetic resources. Such projects require a flow of genetic sequence data among countries that is not complicated or expensive in order to advance these objectives.

C. ECUADOR

6. INFORMACIÓN DIGITAL SOBRE SECUENCIAS DE RECURSOS GENÉTICOS PARA LA ALIMENTACIÓN Y LA AGRICULTURA

Acción: Presentación de información sobre el uso de "información digital sobre secuencias de recursos genéticos para la alimentación y la agricultura" y sus posibles implicaciones para la conservación y utilización sostenible de los recursos genéticos para la alimentación y la agricultura, con inclusión del intercambio, el acceso y la distribución justa y equitativa de los beneficios derivados de su uso

Informe: En el Ecuador todavía no se está realizando el uso de información digital sobre secuencia de RGAA debido principalmente a dos elementos a tomar en cuenta:

1. Existe poca información generada sobre secuencias en las universidades e instituciones públicas y privadas
2. La información que se ha generado puede ser sujeta a temas de propiedad intelectual y a leyes nacionales y regionales sobre intercambio, acceso que necesita autorización previa de la Autoridad Nacional Competente. Inclusive las implicaciones todavía son impredecibles dado que el Ecuador hace pocos días ha ratificado el Protocolo de Nagoya y este instrumento está en proceso de implementación en el país.

D. GERMANY

Germany is of the view that the use and international exchange of digital sequence information (DSI) on genetic resources for food and agriculture should be considered in the overall context of conservation and sustainable use of GRFA as well as access and benefit sharing. In this context, Germany would like to highlight Articles 8b) and 8c) of the Nagoya Protocol which request Parties to pay due regard to cases of present or imminent emergencies that threaten or damage human, animal or plant health and to consider the importance of GRFA and their special role for food security.

International collaboration and common efforts in research and development are essential to achieve food security and sustainable development of agriculture worldwide in the context of poverty alleviation and climate change adaptation. In this regard, Germany wants to underline that it is important to maintain GRFA by widespread use and to support such use to achieve food security as envisaged in SDG 2.

Germany is of the view that access to DSI from public databases and publications does not constitute access to genetic resources in the meaning of the CBD and the Nagoya Protocol.

Including DSI in the concept and/or definition of genetic resources would result in insurmountable financial and administrative burden for both developing and developed countries and thus hamper research and development. This would be in contradiction with the objective of the Nagoya Protocol and the Convention itself.

Information exchange including digital sequence information is essential all over the world to strengthen efforts in conservation and sustainable use of GRFA, e.g. for,

- **identification** and **characterization** of GRFA, e.g. as an important step to prioritise target GRFA for *in situ*, on farm or *ex situ* conservation
- **conservation** of GRFA, e.g. by exploring population size and structure or to estimate relationships between populations, e.g. for classification of endangerment of species and to plan measures to minimise further genetic loss or to discover and monitor invasive alien species, e.g. invasive pest organisms in forestry and agriculture
- **breeding** and **genetic improvements**, e.g. by identifying key agronomic traits that can be useful for e.g. climate change adaptation of food crops (e.g. GENESYS and DIVSEEK, contributing to the Global Information System on Plant Genetic Resources according to Article 17 of the ITPGRFA), or for the further development of the aquaculture sector.

Apart from commercial considerations, the accessibility of reference data of DSI on GRFA is equally important for all countries in areas related to food and agriculture such as

- **food safety** (e.g. to test food products for the presence of contaminants, diseases or pests; in this context gene sequence information are an important tool to discover such contaminants etc.)
- **monitoring of plant and animal health** (e.g. reference data of DSI on GRFA facilitate urgently required tests)
- **quality controls of products in the food and feed sector** such as proof of authenticity and origin (e. g. reliable identification of fish species (aquagene.org database) that allows an unambiguous determination of fisheries products and obviates mislabelling and its negative consequences for consumers, traders and the environment)
- **testing of products** (based on genetic resources) to fulfil regulatory requirements (e.g. the EU Timber Regulation that aims to reduce illegal logging by ensuring that no illegal timber or timber products can be sold in the EU)
- **verification of the descent of certain breeding animals** (e.g. in order to implement the EU zootechnical legislation)

All abovementioned activities require continuous improvement or new development of methods, which are necessary worldwide. Internationally accessible databases including DSI on GRFA promote international research cooperation and global efforts in those fields. Therefore, Germany would like to stress that access to research information including DSI on GRFA constitutes a global benefit per se and is essential even for international research cooperation.

E. INDIA

Since the genetic sequence information/data has assumed global importance in view of the recent technological advances at faster pace, India is of the view that digital sequence information should be treated as " information associated" with the species. The benefits associated with it should be dealt at par with traditional knowledge associated with genetic resources within the scope of CBD.

In addition, the Protection of Plant Varieties and Farmers' Rights Authority is planning to generate digital sequence information of the registered varieties including registered farmers' varieties that will help in achieving the mandate under this provision.

F. UNITED STATES OF AMERICA

Terminology

The United States understands the term "digital sequence information on genetic resources for food and agriculture" to mean the genetic sequence data (GSD) that describe the order of nucleotides in DNA or RNA in genetic material of actual or potential value for food and agriculture. We therefore will use the term GSD instead of DSI in our response.

We note that GSD are neither genetic material nor a genetic resource. It is essential to maintain a conceptual and definitional distinction between genetic material itself and data describing that material.

Use of GSD describing GRFA

Researchers around the world use GSD describing GRFA as a tool to advance scientific study and technological innovation, with enormous benefits for people and nature. The use and exchange of GSD describing GRFA are critical for efforts to improve agricultural productivity and resilience, which are essential for ending hunger and poverty around the world and achieving shared development goals. Any move to restrict or impede access to and use of GSD could slow or halt such research, with serious consequences for the conservation and sustainable use of GRFA, economic development and food security.

Broadly speaking, researchers can use GSD describing GRFA to (1) define populations, (2) characterize genetic diversity within and between populations, and (3) better understand traits of interest; all of which can inform decisions related to the conservation, management and use of GRFA in a manner that promotes food security and economic development.

GSD can be used to define populations of interest based on genetic sequence signatures. Therefore, researchers can use GSD to distinguish between populations that may seem identical in appearance, behavior, and other properties but are genetically distinct. For example, crop genebanks look at differences in diagnostic nucleotide sequences to identify unknown samples, maintain these samples genetically true-to-type, and choose the best samples for specific research or breeding purposes. In aquaculture, researchers have used GSD to define populations released for commercial production, facilitating the identification and management of those fish populations.

In addition to defining specific distinct populations, researchers use GSD describing GRFA to establish the degree of genetic divergence and diversity within and between populations. As diversity within a population is essential for adaptation; researchers can use GSD to determine whether a population has the ability to adapt to environmental changes, contributing to the development of resilient production systems.

Knowledge of the genetic diversity within and between populations is also useful for maintaining genetic diversity and reducing inbreeding, both of which are important for selective breeding and conservation activities, such as cryopreservation. In aquaculture, researchers have used GSD to develop breeding programs for rainbow trout with increased yield and production efficiency. And in forestry, researchers use GSD to develop robust ex-situ collections of at-risk species by identifying distinct natural populations and those with increased diversity. In plant breeding, knowledge of genetic interrelationships of parental lines is a prerequisite for producing hybrids of maize, sorghum, sunflowers, and certain vegetable and fruit crops. Hybrid vigor in these crops results in yield and product quality that substantially exceed those of non-hybrid plants.

Researchers also use GSD describing GRFA to study traits of interest in agricultural products, such as disease resistance, product quality, production efficiency, and resilience to extreme conditions. GSD can be used to advance discovery and development of new crop and livestock varieties, with enhanced outcomes for food security. For example, researchers used GSD to identify genetic sequences in hybrid catfish associated with tolerance for low oxygen conditions. Similarly in cattle GSD are useful in determining adaptability to high altitudes and reduced pulmonary edema as well as increasing tolerance to high ambient temperatures and humidity. Researchers also use GSD to determine milk production potential in dairy cattle, which enables selective breeding for increased milk yield with less time and money required to assess genetic merit. Greater productivity and lower costs contribute to greater availability for consumers.

GSD is also being used extensively in all crop development, and especially for production of drought- and pest-resistant crops and crops altered for enhanced nutritional and economic value. Use of GSD has enabled researchers to rapidly identify markers for genes associated with drought tolerance in sorghum, maize, wheat and other crops. Scientists are also using GSD to breed beans that cook more quickly, which could reduce fuel use and therefore economic and environmental costs associated with this food staple. In summary, researchers can use GSD describing GRFA to accelerate the progress of genetically improving dietary staples, enabling yield and productivity gains that underpin food security by providing more abundant food to humanity.

The generation and exchange of data and information associated with GRFA are important modes of benefit-sharing. As part of research best-practices, GSD are openly available via international data

repositories such as GenBank and the International Nucleotide Sequence Database Collaboration, as well as in journals found in print and online. Additionally, scientists make large amounts of GSD freely available through sector and species specific online databases, such as Gramene (link: http://www.gramene.org/), GrainGenes (link: https://wheat.pw.usda.gov/GG3/), TreeGenes (https://dendrome.ucdavis.edu/index.php) and SoyBase (link: https://www.soybase.org/), to name a few.

These repositories and journals further engender collaboration by providing a free flow of GSD to both researchers and to the general public. This open access and collaboration are key benefits of the use of GSD. Regulations that would restrict or preclude access to and sharing of GSD would likely lead to a significant reduction in data sharing through these and other such mechanisms, stifling innovation, slowing agricultural research and development, and potentially impeding responses to crises affecting agricultural production, with negative impacts for food security.

III. SUBMISSIONS BY INTERNATINAL ORGANIZATIONS

A. AFRICAN CENTRE FOR BIODIVERSITY

We write to you from the African Centre for Biodiversity (ACB). The African Centre for Biodiversity (previously 'Biosafety') was established in 2003 and registered in 2004 in terms of the laws of the Republic of South Africa. ACB carries out research, analysis, capacity and movement building, and advocacy, and shares information to widen awareness and catalyse collective action and influence decision-making on issues of biosafety, agricultural biodiversity and farmer-managed seed systems (FMSS) in Africa. The ACB's work both informs and amplifies the voices of social movements fighting for food sovereignty in Africa.

We make these submissions based on our grave concerns about the implications stemming from the ability of corporate 'breeders' to use genetic engineering technologies such as Clustered regularly interspaced short palindromic repeats (CRISPR), which edit DNA, to access genetic sequence data pertaining to genetic resources, and convert this data back to DNA or RNA and to use it in living organisms, and the risks this poses to the conservation and use of biological diversity and access and benefit sharing international and national regimes and agreements.

Genetic sequence data can be accessed on the internet or in an email, this means that it may no longer be a need to access and exchange of the physical genetic resources/biological materials. Crop traits can be accessed in this way as well as genes that encode for active compound in medical plants, and be used in the manufacture of pharmaceuticals. This gives rise to the possibility that genetic resources can be accessed without prior informed consent and in the absence of a benefit sharing agreement. This would thus undermine several international agreements, including in particular the 3rd objective of the Convention on Biological Diversity (Convention), the Nagoya Protocol on Access and Benefit Sharing (Protocol) and the International Treaty on Plant Genetic Resources for Food and Agriculture (Treaty). Indeed we will go so far as to say that current benefit sharing regimes may be rendered redundant as we go into the future particularly those of the Protocol and Treaty. The rationale underpinning objective 3 of the Convention, and the central imperatives underpinning the Protocol and the Treaty to prevent biopiracy or the misappropriaton of genetic resources will be totally eroded.

Today, what are commonly referred to as the "Big Six" mega seed and agrochemical corporations - namely: BASF, Bayer, Dow, DuPont, Monsanto and Syngenta - together control 75% of the global agrochemical market, 63% of the commercial seed market and over 75% of all private sector research and development (R&D) in the sector.

However, three mega-mergers in input supply underway between ChemChina and Syngenta, Dow and DuPont and Monsanto and Bayer are set to entrench the existing global oligopoly built on a cartel-like technological platform in biotechnology traits, commercial seed and patented agrochemicals. These mergers are indicative of broader processes and the threats they pose to economic participation, social equity and ecological sustainability.

These multinational corporations are making major strides in the new genetic engineering technology that will supersede 'aging' technologies such as transgenic crops: CRISPR, genome editing technology and synthetic biology, which are cheaper and quicker to develop, and for now, unregulated.

Recommendations

1. Parties and all other relevant stakeholders and interested parties to the Convention, the Protocol and the Treaty must ensure that a legally binding decision is taken to unequivocally require that sequence data be considered equivalent to its physical biological counterparts.
2. Sequence data in this regard, must be broadly defined, because it is not necessary to synthesize an entire genome in order for the data to be useful and profitable. Individual genes synthesized from data and inserted into living organisms can be of enormous commercial value, particularly in industrial and medical applications. Sequence data includes DNA, RNA and amino acid sequences as well as accompanying characterization information. Parties and all other relevant stakeholders and interested parties to the Convention, the Protocol and the Treaty must thus pay attention to the scope of digital sequence data. The hereditary material of an organism is not just DNA but in some cases it is RNA. Since the complementarity between molecules, and their important functions, the sequences of both must be covered. The sequences of amino acids that

nucleotides encode are similarly valuable and can be used to replicate and modify natural compounds and in design of biological systems. These sequences too need to be addressed with respect to benefit sharing, see below.

3. Further, Parties and all other relevant stakeholders and interested parties to the Convention, the Protocol and the Treaty must agree to measures to be taken at the national level, to ensure that access and benefit sharing laws and agreements not only apply to the physical transfer of biological material but also to sequence data. In this regard, the scope as mentioned above in point 2, must be broad enough to ensure that all relevant data is caught in the benefit sharing net.

4. Parties and all other relevant stakeholders and interested parties the relevant international agreements must ensure that a legally binding decision is made to require that repositories of sequences and databases in turn require their users to up front agree to benefit sharing as a pre-condition to accessing any sequence information. In this regard, Parties and all other relevant stakeholders and interested parties to the relevant international agreements must be asked to elaborate rules and procedures to govern such user agreements.

5. Turning to the first and second objectives of the Convention, we make the following submissions:

   5.1 The gene foundry and synthesis equipment industries are largely unregulated, and this creates risks to biodiversity. Since synthesis equipment does not care what it is synthesizing, and companies are generally not under any legal obligation to consider the safety of the nucleotide sequences they are producing, much less the ecological impacts, invasive and harmful organisms may be produced and thereby pose a risk to biodiversity. This needs urgent attention by the Parties and all other relevant stakeholders and interested parties to the Convention, including to take into account emerging newer equipment such as “digital-to-biological-converters” that are smaller, faster, cheaper, more portable and able to synthesize ever longer sequences.

   5.2 The misuse of sequence data may also impact negatively on the conservation of biological diversity. The ability to synthesize organisms, or modify organisms with synthesized genes (creating a new sub-species/strain) presents a serious and novel challenge to efforts to curb the negative effects of invasive species. While traditional efforts to prevent invasive species have focused on physically preventing introductions-e.g. phytosanitary measures and border protection-the use of genetic data can leapfrog these controls, resulting in the introduction of invasive species from within a country rather than from outside its borders.

   5.3 A link may be drawn between sequences and sustainable use of biological resources. The unregulated and free access to sequences of genetic resources such as natural medicinal compounds and “climate smart genes” (sequences conferring drought or salt tolerance for example) may harm communities and decrease the perception of the importance of maintaining and developing the resiliency of the genetic resources in situ and within context specific and appropriate ecological environments.

   5.4 We make mention here that we do not accept that the recording of sequences in databases is sustainable use. The potential for the disruption or collapse of small farmers’ markets particularly in Africa, where 80% of the food is produced by small-scale farmers farming on average of 2 ha, engendered by unfettered and unrestrained use, is a real and grave danger. This will have severe negative impacts on the conservation and sustainable use of biological diversity by economically and culturally undermining indigenous peoples and local communities conserving and sustainably utilizing a wide pool of biodiversity.

We trust that these comments will contribute to the further work by Parties and other stakeholders towards realizing the goals and objectives of all three international agreements referred to in this submission.

B. CABI response on potential implications of the use of digital sequence information (DSI) from genetic resources for the three objectives of the Convention, and for the objective of the Nagoya Protocol.

At its thirteenth meeting, the Conference of the Parties to the Convention on Biological Diversity (CBD) considered the issue of digital sequence information on genetic resources and adopted decision XIII/16, in which it decided to consider, at its fourteenth meeting, any potential implications of the use of digital sequence information (DSI) on genetic resources for the objectives of the Convention on Biological Diversity.

Summary statement

CABI beleives that amendments to the Nagoya Protocol are **not** necessary in respect to Digital Sequence Information (DSI) as it is already captured in the definition of the genetic resources and genetic material covered by the Convention on Biological Diversity (CBD). "Genetic resources" means genetic material of actual or potential value. "Genetic material" means any material of plant, animal, microbial or other origin containing functional units of heredity. CBD 19 Mar 2010; https://www.cbd.int/doc/meetings/abs/abswg-09/information/abswg-09-inf-01-en.pdf. DSI describes the resource or material and in a functional form would be a "derivative" and its use and benefits can be treated at a country level at the point of access to genetic material. However, it would be helpful to have a common agreement on the generation of DSI and how it can be used in order not to impede innovation in the life sciences.

CABI position

- Generating and publishing sequence data is considered by CABI as the production of descriptive information on the organism and therefore not utilisation. As such, it is out of Nagoya Protocol regulatory scope.
- Publishing the sequence as electronic data is an act of sharing such descriptive information and thus meets any benefit-sharing commitment required from access to sequence the organism.
- DSI can be used at many non-exploitative levels: for example, it is used to confirm identification and in the CABI understanding this is an observation; in most cases the sequence is published.
- if DSI is used for financial benefit then this could be considered utilisation and the full benefit sharing aspect would be negotiated with the provider country as would be done for access to the organism
- The generation and use of DSI must be considered at the point of access and be expressed in the Mutually Agreed Terms (MAT) and presented in any Material Transfer Agreement (MTA) for clarification on what can and cannot be done regarding DSI.

CABI reasoning

The debate continues within the regulator and stakeholder communities, on whether access to Digital Sequence Information (DSI) should be treated in the same way as would accessing the genetic resource or material (organism) itself. It is obvious they are not the same thing as currently more can be done with the organism than with a partial sequence or even an entire genome. In essence, generation of a sequence requires the genetic resource, itself, to be accessed; the DSI is a product of that access, a "derivative". However, generating and publishing sequence data is considered by CABI as descriptive information and therefore not utilisation and, as such, it is out of regulatory scope. Publishing the sequence as electronic data is an act of sharing such descriptive information with the wider community – including the provider country. DSI can be used at many non-exploitative levels; for example, it is used to confirm identification and in the CABI understanding this is an observation; in most cases the sequence is published. European draft guidance indicates that taking the sequence information and using it to develop a product or tool is out of scope but we are aware that other countries are not of that opinion (this is still in debate at COP). However this does not negate the need for benefit sharing and in this case it is the actual publication of the DSI that shares the benefits of access to the genetic resources from which it was generated.

Selected DSI are becoming standard tools for identification and phylogenetic characterisation of species and populations: mitochondrial CO1 'barcodes' for animals; plastid matK and rbcL barcodes for

plants;16S ribosomal DNA for bacteria, and ITS for many eukaryotes (including Fungi), have become standard tools in modern taxonomy and identification, although the real power of this approach becomes most obvious when sequences from multiple individuals and sources (countries) is freely shared and duly acknowledged. In the future, whole genome sequencing will, without any doubt, have a similar impact on taxonomy. We argue strongly that DSI generated for taxonomic or descriptive purposes needs to be freely shared in the public domain to help address the taxonomic impediment that the CBD recognises, as well as to meet the needs of agriculture and other sciences.

Conclusion

Generation and use of DSI should be considered when accessing the genetic resource to the extent that benefit-sharing is required under the CBD. However, the generation and publishing of such data is not "utilisation" and should not trigger the Nagoya Protocol. Further, we suggest that such generation and publishing of DSI should be considered as part of a country's responsibilities under Article 7 (Identification and Monitoring) and 17 (Exchange of Information) of the CBD. As new technologies develop, DSI may have additional uses that could trigger benefit-sharing and this should be subject to equitable sharing of benefits. Requirements will vary from country to country but should include placing the DSI in the public domain and its subsequent use can be defined in standard Material Transfer Agreements. Likely benefits could include sharing the developed tool or enabling access to the generated product but in most cases the benefit is likely to be facilitating access to the published data and no more. It would be preferred that a single common global understanding was reached to ensure full compliance and thereby reduce confusion.

NOTE FOR SUBMISSION IN UK and to EC: In Europe there is an additional problem because the EU in their guidance documents are using the Frascati definition of research and development http://www.oecd-ilibrary.org/science-and-technology/frascati-manual-2002_9789264199040-en to help define "utilisation" and this includes the generation of knowledge. In this case generating DSI would consequently be "utilisation" and trigger benefit-sharing. However, the guidance documents go on to describe sequencing as a descriptive step or confirming identity which is NOT utilisation – there is still work needed to clarify.

C. INSTITUTE FOR AGRICULTURE AND TRADE POLICY

The Institute for Agriculture and Trade Policy (IATP) appreciates this opportunity to submit a short comment on an issue of critical importance before the Commission. We offer the following comment on digital sequence information to the Commission to assist it to "compile and submit this information to the Executive Secretary of the CBD [Convention on Biological Diversity], as a timely contribution to the process set by decision CBD COP XIII/16."[5] IATP is a non-profit, non-governmental organization (501.c3, in U.S. law) headquartered in Minneapolis, MN (U.S.) with offices in Washington, DC and Berlin, Germany. Since IATP's founding in 1986, we have participated in numerous meetings organized or co-organized by the Food and Agriculture Organization, including those of the Committee on World Food Security and the Codex Alimentarius Commission. IATP staffer Shiney Varghese and advisor Sophia Murphy have participated in the High-Level Panel of Experts advising the Committee on World Food Security.

Much of the discussion about governance of advanced techniques of genetic engineering, such as genome editing, in relation to the conservation and sustainable use of genetic resources, has concerned potential impacts of gene drives on biological diversity.[6] Part of that concern derives from the inability of those who would release gene drives to forecast the biological diversity impacts of suppressing one or more species and/or to reliably program the self-termination of the gene drive before biological resistance to the gene drive has formed.[7]

---

[5] Commission on Genetic Resources for Food and Agriculture, "Follow Up to the 16th Regular Session, C-CBD-7, 3. May 22, 2017, referencing https://www.cbd.int/doc/decisions/cop-13/cop-13-dec-16-en.pdf

[6] E.g. Nicole Gutzman et al, "CRISPR-based gene drive in agriculture will face technical and governance challenges," *EMBO Reports*, August 7, 2017. http://embor.embopress.org/content/early/2017/08/07/embr.201744661

[7] Robert L. Unckless, Andrew G. Clark, Phillipp W. Messer, "Evolution of resistance against CRISPR/Cas9 gene drive," *Genetics*, Early online December 10, 2016. as 10.1534/genetics.116.197285

Similarly, it is difficult to forecast the impact of the cross-border transfer or "trading" of digital sequence information derived directly or indirectly from genetic resources on the conservation and sustainable use of those resources, including fair and equitable Access and Benefit Sharing (ABS) resulting from the use of those resources, including their digital sequencing. Sequence data includes DNA, RNA, and/or amino acid sequences as well as the accompanying characterization information for those sequences. Current ABS governance in domestic regulation is implemented through a material transfer agreement (MTA) that requires signing of the agreement prior to transfer of the biological materials containing DNA, RNA or amino sequences and characterization information.

However, MTAs are already subject to abuse and circumvention.[8] The cross-border transfer of digital sequence data via the internet or even on a bootlegged thumb drive, which can be subsequently be synthesized as organisms either for non-commercial research or commercial purposes, will circumvent the current ABS and prior informed consent requirements of the CBD's Nagoya Protocol. The simplest form of circumvention is to claim that utilization of digital sequencing data is not "access" to genetic resources, and therefore no ABS for use of those data is required. A corollary argument for not recognizing and meeting ABS obligations is that to so will inhibit innovation.[9] Such circumvention, however commercially advantageous in the short term, will deprive CGRFA members of adequate resources to conserve *in situ* and *ex situ* the genetic resources upon which digital sequences to develop agricultural traits depend.

To prevent the potential misappropriation of genetic resources via circumvention in the digital sequencing of those resources, IATP urges CGRFA Members and the Commission Secretariat to respond to the CBD decision on digital sequence information[10] by informing the CBD Secretariat that the CGRFA Secretariat will consider doing the following:

1. Request that FAO Members inform the CBD Executive Secretary of both governmental and non-governmental digital sequence data bases in their jurisdictions of genetic resources that can be used for food and agriculture, per the Strategic Goals and Objectives of the CGRFA.[11]
2. Request that FAO members inform the CBD Executive Secretary of measures that govern access to and use of those data bases, both within their jurisdictions and across borders, per Goals 1 and 5.
3. Request that FAO Members inform the CBD Executive Secretary of voluntary guidelines, mandatory rules or proposed rules in their jurisdictions concerning digital sequence information of genetic resources for food and agriculture, including guidelines and rules pertaining to the cross-border transfer of such information.
4. Request that FAO members support the recommendation of the CGRFA ABS Expert Team "that the Commission convene an international workshop on Access and benefit-sharing for genetic resources for food and agriculture, to be jointly organized, as soon as practicable after the 16th Regular Session of the Commission, by the Secretariats of the Treaty and the Commission, possibly in collaboration with or supported by the Secretariat of the CBD."[12]
5. Request that FAO members support the convening of this international workshop with the collaboration of the Secretariat of the CBD.
6. Request that this workshop include a discussion of digital information sequence, as was begun by the ABS Expert Team at its September 2016 meeting.[13]
7. Request that the CBD Secretariat recommend to CBD Members that the next Meeting of the Parties of the Nagoya Protocol agree on a further specification of the term "utilization" for use in the ABS Clearing House and elsewhere in the CBD. Further specification, including the CBD's

[8] E.g. Tasnia Bubela, Jenilee Guebert and Amrita Mishra, "Use and Abuse of Material Transfer Agreements: Lessons in Proportionality from Research, Repositories and Litigation," *PLOS Biology*, Published online February 3, 2015. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4315468/

[9] Steve Burgess and Dominic Berry, "Regulating the Use of Genetic Sequence Data," *PLOS Synbio Community*, December 15, 2016. http://blogs.plos.org/synbio/2016/12/15/regulating-the-use-of-genetic-sequence-data/

[10] https://www.cbd.int/abs/dsi-gr.shtml

[11] http://www.fao.org/nr/cgrfa/cgrfa-vision/en/

[12] "Third Session of the Technical and Legal Experts on Access and Benefit Sharing," Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations, 13-15 September 2016, CGRFA/TTLE-ABS-3/16/Report, paragraph 39. http://www.fao.org/3/a-bp766e.pdf

[13] Ibid., paragraph 38.

understanding of the utilization of digital sequence information, will help the CBD respond to the CGRFA ABS Expert Team's view that "the Nagoya Protocol does not provide any specific guidance as to the nature or type of research and development activities covered by the term "utilization"[14] as they are applied generally, and more specifically to genetic resources for food and agriculture. The CBD's work on and definitions of the utilization of digital sequence information may assist the CGRFA in the international workshop recommended by the ABS Expert Team to further the CGRFA's Strategic Objectives and Goals.

8. If the next CBD Conference of the Parties decides that in principle digital sequence information is equivalent to biological material to enable implementation of the objectives of the CBD, the Secretary of the CGRFA should recommend to CGRFA Members that they invite the Secretary of the CBD to explain such a Decision and its possible follow-up at the international workshop, as recommended above. The CBD's ABS Expert Team should likewise be invited to meet with the CGRFA ABS Expert Team to discuss how a CGRFA Decision on digital sequence information may assist the realization of the CGRFA's Strategic Objectives and Goals.

The Institute for Agriculture and Trade Policy thanks the CGRFA for its consideration of these actions to inform the CBD of work by FAO members and the CGRFA on digital sequence information in mutual support of CBD and CGRFA goals and objectives.

D. THIRD WORLD NETWORK

Sequence information on genetic resources for food and agriculture (SI)[15] 1 has long-term implications across the mission, strategic goals and objectives of the Commission, including for the loss of genetic resources, food security, conservation, and sustainable use, as well the fair and equitable sharing of benefits derived from their use. For example, it might be argued that placement of SI in databases "in silico" constitutes a form of protection against loss of genetic diversity. Yet, at at the same time, undue reliance on SI as a conservation (or even use) mechanism may sap government resolve and even drain away what little resources that are available for in situ conservation, interrupting and devaluing the work of small farmers, pastoralists, fishers, indigenous peoples, and others who create, conserve, and develop GRFA.

Sequence information and fair and equitable sharing of benefits arising from the use of genetic resources

Most urgently at present, however, are questions regarding SI and the fair and equitable sharing of benefits arising from the use of genetic resources. Gene segments, genes and, indeed, entire organisms of high economic value (e.g. vaccine viruses) are now synthesized from SI that may have been exchanged via the internet or email, meaning that organisms and genetic variants can effectively cross borders without physical biological material changing hands. In the area of microbial genetic diversity, for example, in 2002, synthesis of poliovirus from data was considered a significant technological advance. Yet poliovirus is a mere ~7750 bases long. Fifteen years later, at the end of 2017, the Synthetic Yeast Genome Project anticipates creating an entire 12 million base yeast strain from sequence data. The length of the largest wholly synthetic genome will have grown in 15 years from 7,750 bases to 12 million – over 1500 times as long. Or course it is not necessary to synthesize an entire genome in order for SI to generate benefits. Individual genes synthesized from SI and inserted into living organisms can be of enormous value, particularly in industrial, agricultural and medical applications. For example, the gene(s) encoding a valuable industrial enzyme may be synthesized from SI and inserted into microbes for production in fermentation vats. Such uses of SI may be accomplished without accessing the microbe (or plant, animal, etc.) itself or obtaining prior informed consent (PIC) from the originators of the genetic resources and knowledge holders. Most benefit sharing agreements, policies, and laws are predicated on physical transfers of material and may not be applicable to SI in their current forms. This is a large problem for ensuring fairness and equity in use of GRFA and it is poised to continue growing as the cost of sequencing diminishes and tools for storage and manipulation of SI are further developed.

[14] Ibid., paragraph 16.

[15] We note decisions by the CBD and discussion at the Commission that terminology should be carefully reviewed and possibly modified. "SI" is used here in the interest of clarity.

Recommendations

1. In addressing the obvious threat that the combination of SI and synthesis technologies pose to fair and equitable sharing of benefits, the Commission should adopt the approach that sequence data be considered the equivalent of biological material. In other words, users of SI should, in general, be subject to the same benefit sharing obligations as users of the biological materials that are the source of that SI. Alternatively, some genetic resource providers may choose to make SI available without the underlying biological material and, in such instances, these providers should be fully enabled to ensure the application of obligations that will result in fair and equitable benefit sharing. Thus, when GRFA is sequenced – for example, a collection of the diversity of farmers' varieties of a particular crop species – if and when such information is shared and/or placed in databases, due consideration must be given, and steps taken, to ensure that users of that data are obligated to share benefits, and that Farmers' Rights and other rights of genetic resource providers are protected.

2. Importantly, providers, especially in developing countries, need to maintain awareness that transfers of non-reproductive materials – e.g. leaf matter or killed cell cultures – typically are potential transfers of SI (if the recipient extracts and sequences nucleic acids, at the time of transfer or if the samples are preserved at a future date). Thus, even transfers of "dead" biological materials can give rise to the generation of SI that may lead to the utilization of those genetic resources in biological systems again.

3. SI should be understood to include sequences of DNA, RNA in all their forms, as well as the sequences of amino acids and accompanying characterization information. Like DNA and RNA sequences, the sequences of the amino acids that nucleotides encode are valuable and can be used to replicate and modify natural compounds and in design of biological systems.

4. The Commission should explore how repositories of SI, such as the databases operated or provisioned by international agricultural research centres (IARCs), microbial collections, botanical gardens, etc, and databases such as Genbank and the European Nucleotide Archive, can require users to agree to benefit sharing as a precondition of access to SI. The Commission may, in collaboration other relevant intergovernmental bodies, consider the development of provisions for such user agreements (e.g. "click-wrap" terms and conditions) for databases, and develop recommendations on how databases should be required to implement them.

5. The Commission should study and consider developments in the gene foundry and synthesis equipment industries and their implications for SI and fair and equitable benefit sharing. The gene 18 foundry and synthesis equipment industries are largely unregulated, even to the extent of copies of some of the world's most dangerous pathogens being reproduced from SI, causing alarm in security circles. Put simply, synthesis equipment does not care what it is synthesizing, from neither safety nor fairness perspectives, and companies that commercially synthesize SI generally do not consider Commission-relevant obligations associated with the nucleotide sequences they are producing.

   Further, efforts are underway to create smaller, faster, cheaper, and more portable machines to synthesize ever-larger molecules from SI, including machines that synthesize double stranded DNA and that can be directed by e-mail. These so-called "digital-to-biological converters" aim to expeditiously and easily complete the loop from biological to SI and back to biological. These aim to be portable and easy to operate, broadening the possibilities to use SI to modify and recreate organisms – particularly microorganisms – anywhere.

6. Socio-economic and sustainable use impacts of the use of SI – e.g. on vanilla and vetiver farmers – have been described. The Commission will continually need to bear in mind that the recording of SI is not sustainable use per se, and that the potential disruption or collapse of small farmers' systems engendered by unrestrained use of SI could more broadly impact conservation and

sustainable use of genetic resources by economically undermining communities that conserve and sustainably utilize a wide variety of GRFA.

Thank you for the opportunity to make these comments. Third World Network looks forward to the result of this information gathering exercise and will continue to monitor the Commission's work on SI.

IV. SUBMISSIONS BY OTHERS

A. ASSOCIATE PROFESSOR JENS SUNDSTRÖM, FACULTY PROFESSOR PÄR INGVARSSON, DEPARTMENT OF PLANT BIOLOGY AT THE SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES IN UPPSALA: DIGITAL SEQUENCE INFORMATION (DSI); ITS USE AND IMPORTANCE FOR AGRICULTURAL RESEARCH.

Global research commons

Digital Sequence Information used for research and breeding are to a large extent stored at, and maintained by the global DNA database consortium GenBank/Embl/DDJB[1]. This is an open source format which builds on the principles that 1) results from publicly funded research should be published in open access, 2) that knowledge assets should be governed as common gods on a global scale, and 3) that data-sharing that contribute to global research commons is a legitimate form of benefit-sharing[1].

What is Digital Sequence Information?

The heritable material *i.e.* the DNA of any living organism is in principle built up by four different nucleotide bases (Adenine, Cytosine, Thymine and Guanine) that are held together by a sugar-phosphate backbone. The order in which the nucleotide bases appear on the DNA-strand can be determined using different sequencing techniques and the resulting sequence (ACTGCTT…etc.) can be stored in a digital form. In its most simplistic form the term Digital Sequence Information refer to such a sequence of nucleotide bases.

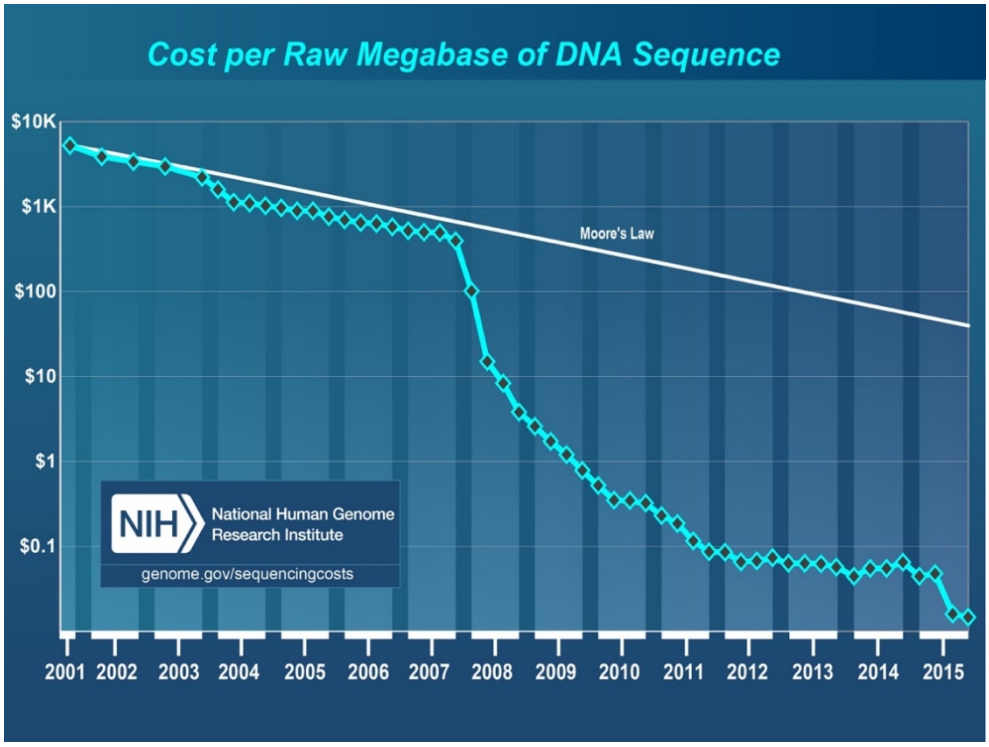
The global research community have built up databases of Digital Sequence Information that harbor information from many different organisms. The information can be in the form of **genomic DNA** that provide information on the different genes as well as the chromosomal regions in-between the genes. It can also be in the form of Expressed Sequence Tags (ESTs) or coding sequences, which correspond to the **RNA transcript** of a gene. Complementary to the nucleotide sequence of a gene, the translated **amino acid sequence** of the corresponding protein is also often deposited in the same database.

Digital Sequence Information can also be in the form of **Single Nucleotide Polymorphisms**, often abbreviated as SNPs. A SNP is a variation in a single nucleotide that occurs at a specific position in the genome. Each SNP is present to a variable degree in a population. For instance, at a specific position in the wheat genome the base C appear in most cultivars, whereas in a few cultivars the position is occupied by the base A. There is a SNP at this specific base position, and the two possible nucleotide variations – C or A – are said to be alleles for this base position. Information about SNPs can be directly or indirectly coupled to different traits and are, hence, widely used in plant breeding.

During recent years, a deeper understanding of the importance of how the DNA is packed into the cell nucleus has emerged. Epigenetic modifications, in form of *e.g.* **acetylation and methylation of the DNA** nucleotides, influence the three-dimensional structure of the DNA, which in turn may affect transcription and, hence, cellular functions. Digital information about acetylation and methylation patterns of the DNA are often deposited alongside with the DNA-sequence in the public databases.

Technology development

In 1977, Frederick Sanger and colleagues developed a method for DNA-sequencing; termed Sanger sequencing. Sanger sequencing gives relatively long reads (>500 nucleotides) with high accuracy, at a relatively high cost. During recent years technology development have drastically reduced the costs for sequencing, as illustrated by the graph provided by NIH[1].



In 2007, the first method for high-throughput sequencing was developed, also called next-generation sequencing. Several different technology platforms for high-throughput sequencing have since been developed. These methods typically give fairly short reads (around 100 nucleotides) that subsequently have to be assembled into longer continuous sequences. Using these modern methods, it is now possible to sequence the entire genome of a whole organism at a relative low cost. It is likely that all crops covered by the FAO international treaty on plant genetic recourses for food and agriculture will be sequenced within the next few years. In addition, transcriptomes (all RNA transcripts present in a sample) from many different crops, cultivars, strains or wild crop relatives are already published and made publicly available. In fact, sequencing of a transcriptome is now often considered a standard procedure, and may in many research areas only be included as part of a larger study, if published in high ranking journals.

Still, most sequencing techniques demand large sequencing facilities and require adequate computer capacity to handle the large data sets that are generated. Recently, portable real-time nanopore sequencing (RTnS) has become available. This technique gives long reads from a sequencing device in the size of a USB-stick, and offers opportunities to rapidly collect and analyze genomic data anywhere. It was used to monitor the Ebola outbreak in 2015[1] and has recently been used to analyze SNP variation among closely related plant species during ongoing field work[1].

As stated above, a publication in peer reviewed scientific journals often requires that digital sequence information is presented according to internationally agreed standards and is made public through *e.g.* the global DNA database consortium GenBank/Embl/DDJB. In many cases, species or genera specific databases are also built up that harbor genomic sequence along with transcriptomes, information about SNPs and other genetic tools that facilitate further research[1,1]. It is important to note that these databases often are global research commons. The databases are the results of joint efforts by the scientific community, and the open source format is essential for the curating of the databases.

Digital Sequence Information and the species concept

One of the lessons learned from evolutionary studies and the collective sequencing of different species is that many genes are shared even between distantly related species. This is not that surprising since all living organisms originate from a common ancestor. Sequencing of genomes from different species have shown that while there may be genes that are unique to a species, a majority of the genes have counterparts in other species and may not differ substantially in function or sequence composition. For instance, the genes that determine reproductive organ identity *i.e.* stamen- and carpel-identity are highly conserved within all flowering plants, and it is in principal possible to move those genes in-between species without changing either form or function of the flower. This implies that the species concept is

not relevant on the gene level. This has implications for the handling and the regulation of digital sequence information since it is not, by necessity, possible to assign a specific digital sequence to a specific species.

Examples

Digital Sequence Information is part of what in general terms are called Big Data. The digitalization of sequence data facilitates cross comparisons within and between species. Genetic diversity in e.g. the field of population genetics has increasingly become a tool for understanding basic biological processes and is equally important for plant breeding.

Since many genes are shared between different organisms, model systems are often employed to study basic biological functions. Generation of databases with digital sequence information, along with genetic tools and seed collections greatly facilitate the transfer of knowledge gained in a model system to crops covered by the FAO international treaty on plant genetic recourses for food and agriculture

Arabidopsis - a model system for plant breeding

Arabidopsis thaliana is a small weedy plant that belong to the Brassicaceae family. The genome of *Arabidopsis thaliana* was fully sequenced already in 2001[1] and it now serves as plant model species, not only for other Brassicaceae family members but for all seed plants. Digital Sequence Information is made available through the Arabidopsis Information Resource (TAIR). Arabidopsis is used to study almost all aspects of plant life such as development, metabolic pathways, disease resistance and adaptation. The knowledge gained using this model system and the Digital Sequence Information made available through TAIR have influenced breeding and helped to breed for a wide variety of traits such as *e.g.* salt tolerance in Rice[1], flowering time in Sugar beet[1] and resistance to clubroot disease in *Brassica napus*[1]. Important for the success of such breeding efforts, that in essence build on a comparative analysis between a crop and a model species, is of course that a database with Digital Sequence Information is available also for the crop species.

Digital sequence information in population genetics

Population genetics as a discipline dates back to the 1930s and has over the last century developed a rich theoretical understanding of how various evolutionary processes interact to shape genetic variation within and between species. Access to data from real populations has, however, always been scarce and largely limited a few special cases in model systems. The technological advances in DNA sequencing over the last few decades have yielded unprecedented levels of genomic information that can be applied to test and extend existing theories in population genetics on just about any organism. It is therefore now feasible to study entire genomes on population level scales with hundreds or thousands of samples. This has allowed for a much better understanding how genetic diversity varies across the genome of an organism and how this diversity is shaped by various evolutionary processes, such as natural selection, genetic drift and recombination[14]. Population-level genomic data from closely related species has also facilitated studies of speciation and how this process shape genome divergence[15]. Free access to Digital Sequence Information from published studies, such as whole-genome DNA sequences, through GenBank/EMBL/DDJP has been instrumental in many of these studies.

The use of Digital Sequence Information in genomic selection

The continuous decline in sequencing costs greatly facilitate the discovery of genome wide SNPs that can be used as markers for breeding purposes. In parallel high-throughput sequencing, breeders are increasingly using large scale phenotyping to couple those markers to desired traits. Hence, genome wide selection is a breeding method which use high-throughput methods for both phenotyping and genotyping. It has the advantage over traditional selection methods, since it in a cost-effective manner allows the breeder to simultaneously select for several complex traits. The availability and free access to curated reference genomes of via GenBank/EMBL/DDJP is very important for the implementation of Genome Wide selection, and often a prerequisite for the adoption of this powerful breeding technique in national or local breeding programs[16]

B. 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.

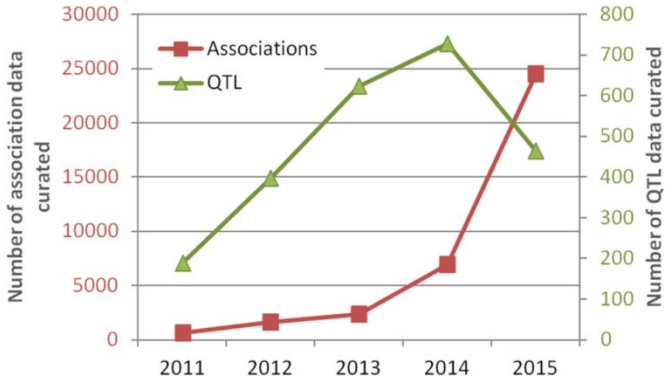
2.4. Scope of information and its use

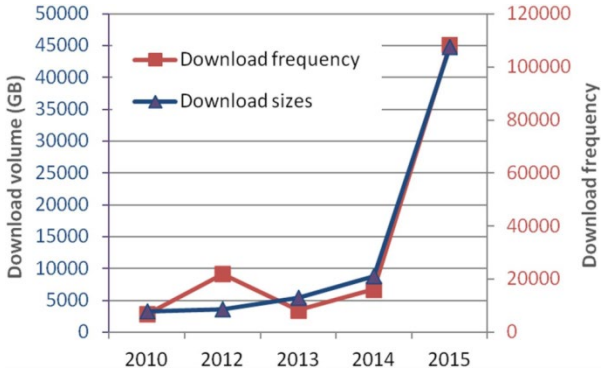
The amount of data inserted in databases is growing incredibly fast. As an example, since 19th June, 2017, till 4th September 2017, in the QTL db animalgenome.org, the number of QTLs/associations for cattle increased from 98 090 (773 journals) to 99 652 (799 journals). With chickens, the increase has 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.