

Views and other information on “Digital sequence information (DSI)” on PGRFA from Italy

During its Seventh Session, held in Kigali, Rwanda, from 30 October to 3 November 2017, the Governing Body invited Contracting Parties for submissions on “Digital Sequence Information” (DSI) on Plant Genetic Resources for Food and Agriculture (PGRFA), including terminology used in this area, actors involved, the types and extent of uses of “DSI” on PGRFA, as well as on relevance of “DSI” on PGRFA for food security and nutrition, in order to facilitate consideration by the Governing Body, at its Eighth Session, of the potential implications of the use of “DSI” on PGRFA for the objectives of the International Treaty, including exchange, access and the fair and equitable sharing of the benefits arising from their use.

Digital Sequence Information (DSI) is at the forefront on the sustainable use of plant genetic resources for food and agriculture (PGRFA). It has implications in several areas on which the International Treaty acts such as characterization, conservation, breeding and genetic improvement and identification. This terminology designs several kinds of data that can be stored in databases in the form of electronic information. Mainly it refers to DNA information in the form of sequence of nucleotides (A, T, C, G). The genetic information can include also protein information such as aminoacid sequences, RNA data related to the transcriptome, epigenetic signatures, metabolite and phenotypic data. All this information can be transmitted and exchanged electronically without the possession of the biological material.

There are several debates about the terminology used, if DSI can substitute the possession of the biological genetic material and how regulate the exchange of this kind of information.

Terminology

Several terms have been used to address this area that include: “genetic sequence data”, “genetic sequence information”, “genetic information”, “dematerialized genetic resources”, “in silico utilization”. Concerning the terminology, for Italy, Digital Sequence Information (DSI) does not reflect the most important trait that characterizes this kind of data. DSI are several forms of “genetic” data such as DNA or RNA sequence, protein information such as aminoacid sequence, epigenetic signature, metabolic or phenotypic data. So, the term “genetic” should be included in the terminology. Moreover, the term DSI is rarely found in scientific literature. Among the different terms proposed “Genetic Sequence Data” (GSD) appears the most appropriate.

Actors

Many actors play important roles in GSD on PGRFA. Governments and public institutions play a central role in the maintenance of about 1700 public databases so far recorded. Maintaining these resources is highly costly and massive investments are needed to uphold them active and freely available.

Scientific journals and granting bodies have a central role in keeping freely available and open access GSD. The most important scientific journals require the submission of GSD in a public database as a condition of publication. Even granting bodies start to put requirements to make public GSD arising from work completely or partially funded through them.

Private sector as well private acting public institutions may have interest in generating and using GSD such as establishing intellectual property (IP). The public at a large may have interest in GSD not primarily including income or IP.

Relevance of GSD for food security and nutrition and for the three objectives of the International Treaty on PGRFA.

GSD are playing an increasing important role for food security and nutrition. There are many uses of GSD that affect characterization, conservation and use in breeding and genetic improvement of crops. Management of PGRFA collections, for example, relies on genotyping with molecular markers (e.g. SNPs) that allows the correct estimation of genetic diversity, the redundancies within and among collections and can inform germplasm curators about the best actions for maintenance and use of PGRFA. Identification of important genes controlling agronomic traits important for food security and nutrition throughout association studies (e.g. GWAS, biparental genetic mapping) requires the use of massive genotyping and phenotyping data. Then, Marker Assisted Breeding (MAB) and Genomic Selections (GS) strategies involve large use of GSD data such as whole genome SNP data. Eventually, application of New Breeding Techniques (NBT) such as the emerging genome editing CRISPR-Cas9 technique require, as well, GSD for RNA guide (gRNA) design and off target checking.

Nevertheless, GSD cannot be substitutive of the genetic material. They are not a genetic resource.

In the International Treaty on Plant Genetic Resources for Food and Agriculture “genetic resources” and “genetic material” are defined in Article 2 as follow:

“Plant genetic resources for food and agriculture”: “any genetic material of plant origin of actual or potential value for food and agriculture”.

“Genetic material”: “any material of plant origin, including reproductive and vegetative propagating material, containing functional units of heredity”.

These definitions state clearly that the genetic resources must contain “functional units of heredity”. Genetic sequence data by definition do not contain “functional units of heredity” like DNA so, they are not genetic resources nor genetic material.

Previsions about using this kind of data to reconstruct the biological material from which it was obtained, as reported in the Background Study n.68 (Heinemann et al 2018), are quite speculative and premature. The examples reported about the reconstruction of a functional virus from its genomic sequence (Wimmer et al, 2006) as a proof of interchangeability is not very pertinent. Viruses have a small genome (up to few kb) and are something in between life and non-life. They have not a cellular machinery, they recruit the host cellular machinery to complete their cycle. Reconstructing from a scratch a functional eukaryotic, or even a prokaryotic organism, starting from GSD would be quite challenging and, at the current state of the art, not achievable. The background study n.68 itself reports in the conclusion alerts about future projections of GSD on GRFA such as: *“...some of the projections of future value may fail to be realized, or even later viewed as hype”* highlighting that some of prospectations are not achievable at the current state of the art.

The Secretariat of the International Treaty of Plant Genetic Resources for Food and Agriculture (ITPGRFA) commissioned a scoping study to highlight how the emerging sectors of DSI, synthetic biology and genomic affect the Treaty and specifically its access and benefit sharing protocols (Welch et al 2017). The scoping

study highlighted that the genomic advancement *“raises significant challenges to the ABS logic of identification, and the different expectations of monitoring that are currently in the Treaty framework”*. The author reported that *“monitoring DSI exchange is a challenging prospect”*. Moreover, they noted that the same sequence may occur in multiple organisms making difficult to identify the source of the gene sequence. In addition, an individual sequence may have value as part of a group of sequences from diverse sources in order to associate a particular trait to the causal gene (e.g. through GWAS). All these aspects further challenge the ABS principles and the Multilateral System. As genomic and *“gene editing techniques become more accessible and democratized, the Treaty will need to keep pace with the rapidly expanding ecosystem”* stated the authors in the scoping study. They suggest, as well, a wide range of benefits related to GSD, most of which can be categorized as non-monetary benefits, such as facilitated access to PGRFA within the MLS, exchange of information, capacity-building, and access to and transfer of technology.

Considering GSD as if they were genetic resources or genetic material can have deleterious consequences in term of scientific progress and innovation at large and on PGFRA in particular. This would hamper the free sharing of information that are crucial for identification, characterization, conservation and use PGRFA threaten their efficient conservation and sustainable use.

Constrains to manage, generate and analyze GSD exist in developing countries hampering access, exchange and benefit sharing. These constrains are infrastructural, the so called “digital divide”, but also cultural, a mixture of social, psychological, economic and research capabilities and skills, (Bezuidenhout et al 2017). Nevertheless, these constrains cannot be the reason to restrict the free use of GSD. The main goal in the future is to remove obstacles and constrains that exist and not to add further ones limiting the open access to GSD.

An example of the access and benefits sharing is well represented by the recent work done at the International Institute of Tropical Agriculture (IITA) in Nairobi, Kenia, aimed at inactivating the banana streak virus (eBSV), a major challenge in banana cultivation and breeding. Banana is a staple food for many African populations. Moreover, because of its deep sterility, the cultivated banana is a very vulnerable crop. In this work the IITA applied the emerging and innovative CRISPR-Cas9 technology to inactivate the virus (Tripathi et al 2019). This work is an example of the benefit sharing for developing countries of free GSD availability. RNA guides (gRNAs) were designed using virus strains sequence freely downloaded from NCBI database. The checking of potential off targets in the banana genome was done using the freely available whole genome sequence downloaded from the banana genome hub database (<http://banana-genome-hub.southgreen.fr>; D’Hont et al 2012). This is a clear example of access and fair and equitable benefit sharing of genetic information whose beneficiary are the developing countries. In fact, all the genetic sequence data, for both the viruses and for the banana genome, were obtained from laboratories distributed in developed countries and made freely available without any restriction or constrain. The beneficiary of the free access and use was a developing country. Restriction in usage of the sequence data would have hampered this important step for an important staple food for developing countries.

Italy in the last decades has played an important role in this arena generating and using freely available GSD. It took part to international consortia that obtain and release the whole genome sequences of important plants for food and agriculture such as wheat, grape, tomato, apple, peach, strawberry, citrus, potato and pear, in some cases coordinating the initiative (as for grapevine, apple, tomato and peach). Some whole genome sequences were made freely available to the scientific community prior the publication under the Fort Lauderdale agreement. It contributed as well to the built and the release of important genomic tools such as SNP arrays. Data and tools were made freely available to the whole

scientific community in the strong believe that public available data would have been beneficial for the progress of science and knowledge. Italy, as well, has had benefits of the freely available data. Management, characterization and innovative conservation strategies of Italian *ex situ* germplasm collections of PGRFA have been achieved thanks to the availability of freely available molecular markers (e.g. SSRs and SNP arrays). Genetic improvement of traits related to important agronomic traits such as disease and pest resistance (apple scab, grapevine downy and powdery mildew, wheat powdery mildew among others) and tolerance to climate constrains (e.g. drought) have been achieved or are underway thanks to the free availability of GSD.

In view of the increasing importance of GSD on PGRFA further in-depth studies highlighting their role on PGFRA are necessary. The studies should cover examples of uses of GSD on PGFRA addressing, identification, conservation, characterization, and sustainable use in breeding and genetic improvement in both developing and developed countries. These studies need to address, as well, economic evaluations of the consequences of putative restrictions in access and use of GSD on PGFRA for both developing and developed countries. It is important to perform cost benefit analyses that take in consideration both direct and indirect costs and benefits arising from different scenarios of restrictions in access and use of GSD on PGFRA. These analyses should include benefits and costs of capacity building and support in developing countries and least developed countries to allow them to generate, manage, analyze and use of GSD. This analysis would be a cornerstone to take informed decisions on this important issue that may affect the sustainable use of PGFRA and food and nutrition security of the whole planet.

References

Bezuidenhout ML et. al (2017). Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, 44(4), 2017, 464–475.

D'Hont A et al. (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* volume 488, 213–217.

Heinemann J, Coray DS, Thaler DS (2018). Exploratory fact-finding scoping study on “digital sequence information” on genetic resources for food and agriculture. BACKGROUND STUDY PAPER NO. 68 FAO

Tripathi JN et al (2019) CRISPR/Cas9 editing of endogenous banana streak virus in the B genome of *Musa spp.* overcomes a major challenge in banana breeding. *COMMUNICATIONS BIOLOGY* 2:46

Welch EW, Bagley M, Kuiken T, Louafi S (2017). Potential implications of new synthetic biology and genomic research trajectories on the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA or ‘Treaty’) FAO.

Wimmer, E (2006). The test-tube synthesis of a chemical called poliovirus. *EMBO Rep.*, 7: S3–S9.