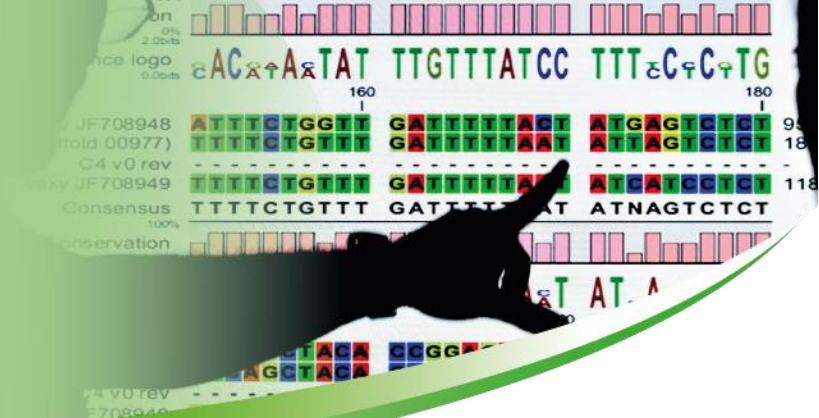
INITIATIVE ON
Genebanks

CGIAR



Digital sequence information is changing the way genetic resources are used in agricultural research and development: implications for new benefit-sharing norms

A discussion paper from CGIAR for consideration by delegates to the 15th Session of the Conference of the Parties to the Convention on Biological Diversity

INITIATIVE ON
Genebanks

CGIAR

Digital sequence information[1] is changing the way genetic resources are used in agricultural research and development: implications for new benefit-sharing norms

A discussion paper from CGIAR for consideration by delegates to the 15th Session of the Conference of the Parties to the Convention on Biological Diversity [i]

Contents

[1] This paper does not address the unresolved issue of the definition of the term Digital Sequence Information (DSI). As used in this submission, the term refers to sequences derived from DNA or RNA and includes short and long reads and all derived molecular markers such as Single Nucleotide Polymorphisms (SNPs), Presence Absence variations, insertions and deletions (InDels), chromosomal translocations and rearrangements detected through sequencing and Simple Sequence Repeats (SSRs), together with their epigenetic status and associated annotations setting out current knowledge of the functions of different parts of the genome. This corresponds to the narrowest potential definition of DSI as set out by the CBD Secretariat (2020a), and comprises a subset of potential broader definitions. Comments on this paper may be sent by email to m.halewood@cgiar.org

Executive Summary

Digital sequence information (DSI) is playing an increasingly important role in CGIAR's conservation and use of genetic resources for food and agriculture, thereby enhancing CGIAR's ability to create benefits for resource-poor farmers and consumers in developing countries. It is clear that DSI will continue to make positive contributions to all three objectives of the Convention for Biological Diversity (CBD): the conservation of biological diversity, the sustainable use of its components, and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources.

DSI is being used to assess genetic diversity of *ex situ* collections to identify *in situ* genetic resources that are not included in *ex situ* collections, and quantify the relationships between accessions within and among collections. This information is essential for developing more effective *ex situ* and *in situ* conservation strategies. The more knowledge we gain of biodiversity at the genome sequence level, the better the first objective of the CBD will be served.

DSI, coupled with phenotypic and other data, such as eco-geographic data, is being used to identify materials that are well adapted to different, and changing, agro-ecological conditions. Integrated into animal and crop breeding programs, DSI is increasingly useful for achieving targeted, efficient uses of genetic diversity in sustainable agriculture, more effectively and efficiently than is possible with traditional approaches. Use of DSI is essential for meeting the UN Sustainable Development Goals, in particular the first three goals, i.e., ending poverty, ending hunger, and ensuring good health and wellbeing of human beings.

The use of DSI to develop commercially viable products is predicated on the development and use of appropriate DSI-based methods and tools such as molecular markers, gene editing, algorithms, and information technology (IT) tools. Development of effective tools and technologies itself requires access and the extensive use of genetic resources, thus generating a two-step process: first, a diverse set of genetic resources are used to develop the required tools; then those tools are applied to a second set of genetic resources to develop commercial products.

The two-step process inherent in the use of DSI generates a new landscape for sharing the benefits arising from the use of genetic resources. The first step (developing the tools) requires the use of a wide range of genetic resources, most of which are not used in the second step (applying the tools to create commercial products). Hence, most genetic resources used and the countries they come from are entirely unrelated to the final commercial products. Consequently, it appears that multilateral approaches to sharing benefits from the use of DSI are the most practical, reflective of reality, and equitable since they could (and should) recognize the contribution of multiple countries and institutions to tools development, and not only the countries of origin of the genetic resources that are eventually integrated in products at the end of the research and development chain.

There are several ways of developing multilateral systems. Our experience in managing agricultural research projects that are highly reliant on genetic resources and DSI suggests that the broadest, flattest

possible multilateral system, with fewest exceptions, thresholds, and choices to be made by both users and providers would be the most efficient. Such a system should 'delink' access from benefit-sharing, and also delink use of DSI in particular commercialized products from benefit-sharing. The simplest way to minimize transaction costs would be for Contracting Parties to make initial payments based on formula to be agreed by the Conference of the Parties to the CBD, with the option to collect from users under their jurisdiction. Alternative approaches that require payments directly from individual commercializers, and/or linked to sales of products derived from accessed DSI, would likely have higher transaction costs for generators, providers, and users of DSI, as well as those organizations responsible for the administration of those systems. These transaction costs could have negative impacts on agricultural research and development that depends upon DSI, to the disadvantage of all countries. Complementary forms of monetary benefit-sharing could include payments from organizations providing DSI and analytical services on a commercial basis, and sales of reagents.

New norms and processes adopted under the CBD must be 'in sync' with access and benefit-sharing under the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty).

Ideally, for genetic resources for food and agriculture *both* DSI and genetic resources should be available under harmonized multilaterally oriented terms and conditions. Given the existence, already, of the Plant Treaty's multilateral system of access and benefit-sharing, such harmonization should be relatively easy to achieve with respect to plant genetic resources for food and agriculture. It may be more difficult to achieve with respect to other genetic resources for food and agriculture (e.g., livestock, fish, trees and crops and forages other than those listed in Annex 1 of the Plant Treaty).

Newly adopted benefit-sharing norms should also promote non-monetary benefit-sharing, particularly capacity strengthening, including access and use of IT analytical tools for DSI, to close the gap in abilities between developed and developing countries to generate, use and benefit from DSI in the future.

1. Introduction

This paper analyses the ways in which CGIAR Centers use digital sequence information (DSI) in their efforts to conserve and sustainably utilize the world's most important crop and livestock genetic diversity. The paper then reflects on which of the benefit-sharing options currently under consideration by the Contracting Parties to the CBD[2] (and the versions of those options that must be considered by the Governing Body of the Plant Treaty and the UN FAO Commission on Genetic Resources for Food and

[2] Co-leads' report on the work of the Informal Co-chairs' Advisory Group on digital sequence information on genetic resources since the fourth meeting of the Open-Ended Working Group on the Post-2020 Global Biodiversity Framework: note by the co-leads of the Informal Co-Chairs' Advisory Group on digital sequence information on genetic resources. https://www.cbd.int/doc/c/0c79/5954/8ec6714d513ecbd570c0b062/wg2020-05-inf-01-en.pdf

Agriculture) would provide effective policy support for the continued use of DSI in agricultural research and development in the future [3].

1.1. CGIAR

CGIAR has evolved into a new global research and innovation partnership addressing the evolving global context, which demands a systems transformation approach[4]. Its mission is to deliver science and innovation that advance the transformation of food, land, and water systems in a climate crisis. It works in five impact areas: nutrition, health and food security; poverty reduction, livelihoods and jobs; gender equality, youth and social inclusion; climate adaptation and mitigation; and environmental health and biodiversity. The last of these areas directly supports the CBD. The primary focus is the poor and needy in developing countries, scaling research and innovation through capacity development and policy, delivered through regional and global initiatives in close collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations and the private sector. Hence the *modus operandi* of CGIAR, seeking to bring benefits to developing countries where benefits are most needed, meshes closely with the concepts of fair and equitable benefit-sharing set out in the CBD.

Given the nature of our mission, the networks we operate under and our *modus operandi*, our experiences of using DSI are most relevant to agricultural biological diversity[5] and the use of genetic resources for food and agriculture. Most of the work that CGIAR Centers are doing in the generation and use of DSI is related to the conservation and use of intraspecific genetic diversity, and sharing of benefits related to those uses. Much of our work is focussed on major staple crops, forages and agroforestry species. Some CGIAR Centers also generate and use DSI from livestock, fish, neglected crops, known and novel pathogens and soil biota. We shall consider lessons related to the use of DSI from across the spectrum of our work in these areas.

[3] Some of these themes were considered in CGIAR's 2017 submission to the Secretary of the CBD on DSI entitled "Potential implications of the use of digital sequence information on genetic resources for the three objectives of the CBD: a submission from CGIAR" (available at https://www.cbd.int/abs/DSI-views/CGIAR-DSI-en.pdf), and later, in a journal article (Halewood *et al*, 2017) based on that submission.

[4] https://www.cgiar.org/how-we-work/strategy/

[5] From Conference of the Parties to the CBD, decision V/5, Appendix: *Agricultural biological diversity: review of phase I of the programme of work and adoption of a multi-year work programme*. CBD: Montreal, 2000. "Agricultural biodiversity includes all components of biological diversity that constitute the agricultural ecosystems: the variety and variability of animals, plants and micro-organisms, at the genetic, species and ecosystem levels, which are necessary to sustain key functions of the agro-ecosystem, its structure and processes."

2. DSI and the conservation of biological diversity

In 2010, the Conference of the Parties to the CBD adopted the Aichi Targets[6]. Aichi Target 13 states: "By 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of their wild relatives, including other socio-economically as well as culturally valuable species, is maintained, and strategies have been developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity." Unfortunately, the deadline for meeting the Aichi targets has passed and they have not been met (CBD Secretariat, 2020b), and new targets are being set through the Post-2020 Global Biodiversity Framework.

The failure to meet Aichi Target 13 highlights the difficulty and challenges to identifying and quantifying the genetic diversity that needs to be maintained. The target will be more achievable when whole *ex situ* collections and representative samples of *in situ* diversity are sequenced and, equally importantly, appropriate standards for processing and interpreting sequence data are better established. DSI will help strengthen the metrics, once it can be obtained on a large scale, for describing diversity and analysing the extent of diversity *in situ*. This will aid long-term conservation of maximum diversity so that all elements of the diversity present in any one crop plant or livestock species are conserved long-term for humankind. Thus, DSI is a powerful, and the best, means to assist genebank managers, national agricultural organizations, companies, indigenous communities and researchers to quantify the diversity present, *in situ* and *ex situ,* for the major crops and their wild relatives, (semi-)domesticated animal species, and other components of agricultural biodiversity.

As a part of a recent system level review (Hawtin *et al* 2020), a 30-year vision for CGIAR genebanks (Sackville Hamilton 2020) was developed highlighting the critical role of appropriately processed DSI for the rational, efficient and effective conservation and use of biological diversity in *ex situ* collections. Its role in promoting the use of diversity is based largely on developing a catalogue of functionally relevant genetic variants, which can be used to optimise the composition of active *ex situ* collections. Such a vision is both technically feasible and effective.

[6] Decision X/2 of the 10th meeting of the Conference of the Parties to the Convention on Biological Diversity, Nagoya, Japan, 18–29 October 2010. See pp 111-123 of the report on the 10th meeting at https://www.cbd.int/doc/meetings/cop/cop-10/official/cop-10-27-en.pdf

2.1. Intraspecific diversity among accessions[7]

Example: Asian rice

A ground-breaking assessment of genomic variation in Asian rice, based on whole-genome sequences of purified samples of 3,010 diverse accessions (Weng *et al.* 2018), highlighted the enormity of the diversity present within species and the need for further sequencing. The dataset revealed 27 million Single Nucleotide Polymorphisms (SNPs), 2.4 million small length polymorphisms and 90,000 large structural variations, with each accession having 1,000 unique SNPs that were not present in any of the other accessions. The abundance of variants was found to follow "Zipf's law"[8], indicating that most variants are rare. Sample-size trend analysis suggested that if the sample size had been increased to 100,000 accessions, the total number of SNPs discovered would have been doubled, and each accession would still have 100 unique SNPs.

One of the key lessons of the above analysis was that genome sequences must be processed to identify and catalogue functionally relevant variants. Many sequence variants have no known effect on the regulation of genes or the resulting proteins. As such they seem to contribute little to functional diversity and hence to rational decisions on conserving diversity. The analysis revealed around 50,000 genes, each represented by on average ten functionally distinct variants. Using this to reduce the raw DSI dataset to the gene variants of each accession, Zipf's law was again found to apply. In this case, each accession had on average 300 unique variants that were not present in any of the other 3,010 accessions, highlighting the near impossibility of conserving all diversity. Thus, even with this exceptional dataset, we have barely begun to quantify the diversity of rice. Nevertheless, it should be noted that focusing specifically on functionally distinct variants transforms the task of optimising conservation based on DSI into an achievable as well as an effective task.

[7] In *ex situ* collections, an accession is a unit of management of genetic material, managed with the intent to preserve its genetic composition as similar as possible to that of the material originally acquired from a provider. Two samples of identical genetic material received on different occasions would normally be managed as different accessions. One accession may comprise a single genome or a genetically heterogeneous population.

[8] Zipf's law describes an empirically observed negative relationship, linear on a log-log scale, between the abundance of a particular variant and the number of different variants observed to be present at that abundance. Originally formulated to describe the use of words in a language, by which a few words are used commonly whilst most words are used rarely, it has been found to describe abundance-rarity in many situations, including sports, arts and evidently biological diversity. https://en.wikipedia.org/wiki/Zipf%27s_law

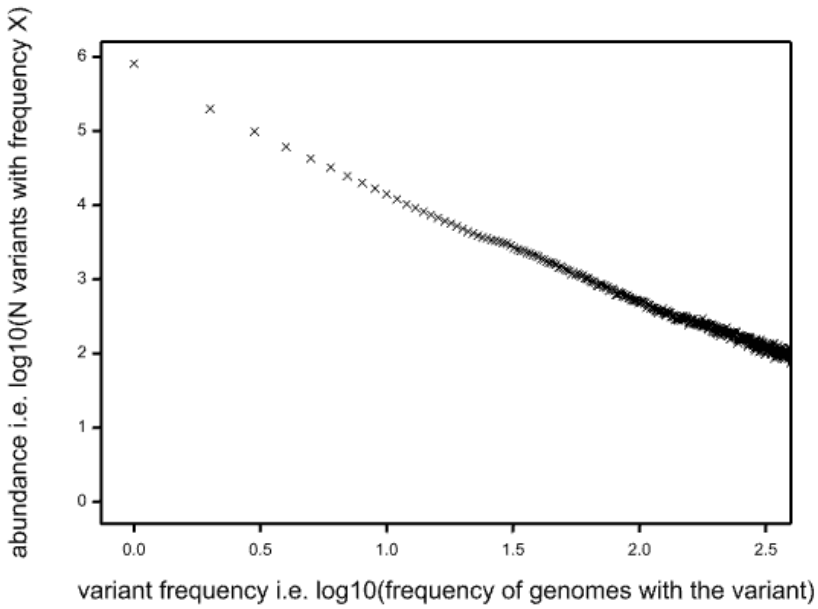


Fig. 1. Variant frequency, or proportion of 3,010 rice genomes that contain a given functionally distinct variant (X axis) vs the abundance or number of variants among that are present at that frequency (Y axis).

A second key lesson was to highlight the need for more high-quality reference genome sequences. Most of the 3,010 rice genome sequences were assembled by cross-referencing to high-quality reference genomes, as is standard for low-cost high-throughput next-generation sequencing (NGS). For 453 accessions coverage was sufficiently high for *de-novo* sequence assembly, enabling a more in-depth analysis. Of the ~23,000 gene families present in at least one of these 453 accessions (defined as a "pan-genome"), little more than half were present in all of them (the "core genome"). This again emphasises the huge magnitude of within-species diversity, even in domesticated species such as rice which have been through the "genetic bottleneck" of domestication.

The creation of a pan-genome by combining multiple genomes of a species has also proven to have high potential value in other species such as tomato for the discovery of new genes possibly involved in some traits (Gao et al 2019).

A third key lesson is the critical importance of open and easy access to DSI datasets. If the data remains private, difficult to access or only in raw, unprocessed form, usage will be restricted to only those generating data, or with the resources to analyze it for their own use. To enable easy access to the 3K RG variant data by breeders and researchers, IRRI developed the SNP-Seek database and API (https://snp-seek.irri.org) (Mansueto *et al.*, 2017). SNP-Seek uses an open-source development model with CHADO relational database components combined with a non-relational datastore to handle genomic diversity data. SNP-seek is accessed on average by 50 users per day to undertake queries of SNP diversity, display of regions on the genome, determination of haplotypes, extraction of data, and others. Impact of GSI

usage is evident by the number of studies citing the database and 3K RGP data that are leading to improved use of the genetic stocks and creation of pre-breeding products.

Example: chickpea

Likewise, sequencing of 3,366 chickpea genomes including 195 wild species accessions (Varshney et al 2022) could identify genes that show signatures of selection during domestication, migration and improvement. Superior haplotypes for improvement-related traits in landraces that can be introgressed into elite breeding lines through haplotype-based breeding were identified. A detailed comparison of cultivated (2,258) and wild species, *C. reticulatum* (22) accessions with a coverage of greater than 10× revealed structural variations, including insertions (139,483), deletions (47,882), inversions (61,171), intra-chromosomal translocations (417) and inter-chromosomal translocations (2,410) in cultivated and 287,854 insertions, 67,351 deletions, 58,070 inversions, 446 intra-chromosomal translocations and 2,066 inter-chromosomal translocations among *C. reticulatum* accessions as compared to the reference genome (Varshney et al 2013).

The sheer magnitude of diversity is so great that it generates methodological problems using NGS, as the absence of a gene from the reference genome hampers the quantification of diversity in genomes that do have that gene. This in turn leads to ascertainment bias (where increasing differences between the reference genome and a test variety increase uncertainty over the genome of the test variety) and underestimation of the diversity of genetic resources that are genetically most different from the reference. This has been a particular problem for rice, as Nipponbare (the temperate japonica Japanese variety that for two decades has provided the standard reference genome sequence) has proved to be highly atypical, especially of the tropical indica varieties grown in most developing countries. The result has been growing awareness of the need for more high-quality genome sequences selected from across the full spectrum of diversity to provide a more appropriate choice of reference genomes for each test variety. For Asian rice, 15 such high-quality genome sequences have now been established, corresponding to the 15 clades previously identified (Zhou *et al* 2020), enabling renewed progress in quantifying rice diversity.

Alternatively, where reference genomes are not available for resequencing, a number of NGS based genotyping approaches do not use any individual reference in the identification of specific variation within a sample, rather these approaches compare sequences among the group of samples being interrogated, lowering the risk and impact of ascertainment bias.

Example: potatoes

The CIP Genebank has been collaborating with Peruvian farmers for decades, developing activities such as repatriation, characterization and conservation of potato landraces. A recent study developed by CIP estimated the genetic diversity maintained *in situ* by farmers from four communities in the Sierra de Lima and ten communities in Pasco in Peru through SNP analysis, and compared it with the molecular database of the *ex situ* potato collection of the genebank held at CIP. A total of 1,075 traditional potato varieties were collected from the fourteen participating communities in collaboration with farmers (one per community) through the support of the Association of Guardians of the Native Potato of Peru AGUAPAN and the NGO Grupo Yanapai. DNA was extracted from tuber sprout tissues and genotyping was performed with Illumina Infinitum SolCAP-V4 potato SNP microarrays and the use of GenomeStudio software. After filtering out the markers with low quality and reproducibility, 2,759 common markers for all the samples were used in a multivariate clustering analysis with the support of R software. From the comparison of the genetic relationships between the CIP collection and the local varieties conserved in *in situ*, six taxonomic groups (Hawkes, 1990 classification) and four ploidy levels (2x, 3x, 4x and 5x) were identified in the landraces conserved *in situ*. The genetic composition between the communities of Lima and Pasco was significantly different and the communities of Lima presented diversity values higher than those of Pasco, with the community of Huancachi in Lima presenting the highest value (He = 0.263) and the community from Paucartambo in Pasco the lowest value (He = 0.190). From the comparison in pairs of the genetic distances between communities, a lower number of significant distances was observed between the majority of the communities of Pasco, which could presume the existence of a greater exchange of local varieties among the farmers that comprise them. It was possible to identify 88 genetically-unique local varieties that were not represented in the CIP collection. In addition, it was identified that the species most frequently found in these communities are *S. stenotonum* subsp. *goniocalix* and *S. tuberosum* subsp. *andigena*. Through collaboration with Peruvian farmers, new genetic diversity was added to the CIP Genebank to preserve this diversity in perpetuity. Farmers in turn benefit from *ex situ* conservation by accessing the collection through CIP's repatriation program which gives farmers the opportunity to request traditional varieties they have lost, or which might be better adapted to changing climatic conditions in the Andes.

2.2. Intraspecific diversity within accessions

With the exception of accessions maintained as clones, accessions conserved *ex situ* are almost invariably populations of distinct genomes, even for crop species that are naturally self-pollinating such as wheat and rice. This presents an additional technical challenge to effective conservation *ex situ*.

Example: sorghum, pearl millet, and pigeonpea

A study conducted by the ICRISAT Genebank (Allan et al 2020), aiming to assess genotypic (DArTSeq) and phenotypic diversity of geographically representative diverse sorghum, pearl millet, and pigeonpea landraces. Thirty-six accessions originating from diverse parts of the world from each of the three crops were analysed, being 25 individual plants per accession for pearl millet and 15 individual plants per accessions for sorghum and pigeonpea. Within-accession distances based on phenotypic data varied from 0.038 to 0.141, 0.145 to 0.271, and 0.071 to 0.410 for sorghum, pearl millet, and pigeonpea, respectively. A total of 45,249 SNPs in pearl millet, 19,052 SNPs in sorghum, and 8,211 SNPs in pigeonpea were used for the analysis of the molecular data after filtering. Pairwise MRD within each accession was averaged, thus the overall mean genetic distance within each accession varied from 0.031 to 0.342 for sorghum, 0.181 to 0.300 for pearl millet, and 0.040 to 0.393 for pigeonpea. These results show that the within-accession diversity is highly variable among accessions and crops. Enormous variability was observed within and among landraces of sorghum, pigeonpea, and pearl millet. Molecular variance within accessions was observed to be low in sorghum (26.3%), highest in pearl millet (80.2%), while pigeonpea showing an intermediate within-accession variance of 57.0%. The level of heterogeneity and diversity in landraces are crop-specific and associated with their mode of fertilization. Based on the level of diversity within each accession of different crops, appropriate conservation and regeneration strategy should be followed to conserve the genetic integrity and diversity of landraces.

2.3. Quantifying and conserving livestock diversity

More than 8,800 livestock breeds have been recorded globally by the FAO Commission on Genetic Resources for Food and Agriculture, representing a valuable resource and a high biodiversity at the genetic level. Precise quantitative assessment of that biodiversity is needed to support international agreements that recognize the importance of its conservation, such as the 2020 Aichi Biodiversity Targets set by the CBD and SDGs 14 and 15 on protecting, restoring and promoting sustainable use. In the absence of appropriate DSI production and use, breed attributes and genes that are potentially beneficial in the future may be lost. Instead, some breeds may be condemned to extinction and in the process, some of the good genes that they may have possessed disappear with them.

NGS techniques have evolved rapidly, allowing unbiased and comprehensive analysis of genetic diversity parameters, providing individuals´ entire genetic information of neutral loci and coding regions, thus, demographic parameters related to selective or adaptive processes can be estimated more accurately. This offers new possibilities in genetic diversity analysis: detection Copy Number Variants (CNV), discovery of new causal mutations associated with traits of interest in livestock, identification of rare genetic variants with minor allele frequencies, etc.

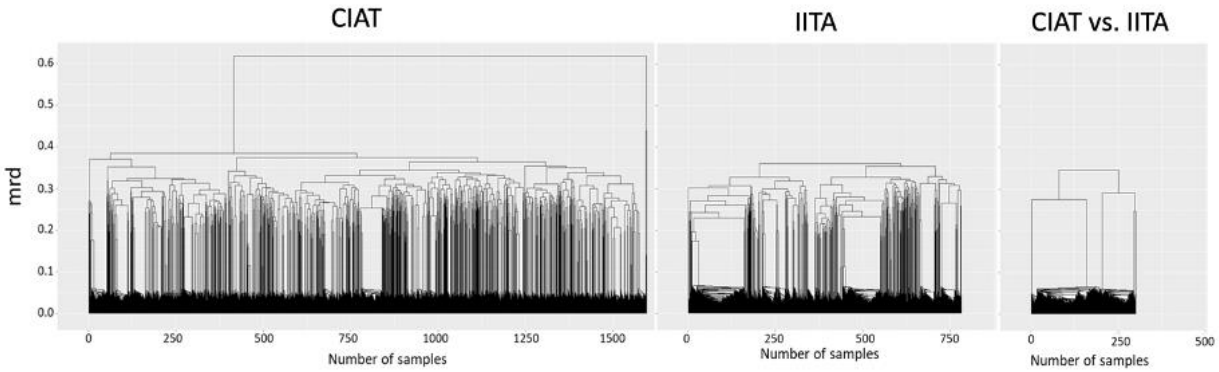
2.4. Enhancing cost-effectiveness of *ex situ* conservation

DSI is also showing the theoretical potential to enable optimisation of the cost-effectiveness of *ex situ* conservation. With a catalogue of functionally distinct variants, the composition of a collection can be altered to maximise the functional diversity that can be conserved within a fixed budget, by eliminating,

or placing into long term storage accessions that bring little unique diversity to the collection and creating space to bring in different accessions with greater diversity. Addressing the challenge of limited budgets for ex situ conservation has been a priority for many years, but without DSI it is essentially unsolvable. Again, DSI offers new opportunities to support the first objective of the CBD.

Example: cassava

The IITA and the ex-CIAT genebank of the Alliance of Bioversity International & CIAT are collaborating to jointly genotype their cassava collections on the DArTseq genotyping platform. The following figure displays the preliminary results for a subset of CIAT and IITA accessions that were genotyped first. The dendrograms shown are based on quantifying the similarity among genome-wide SNP profiles using modified Roger's distance (mrd) estimates, followed by hierarchically clustering those accessions that showed pairwise mrd < 0.05 estimate(s) with one or more accessions. Initial genetic redundancy estimates within collections are 32% (CIAT) and 20% (IITA), with the potential overlap between the two collections being relatively minor (4%).



- Three sub-matrices with cases of mrd < 0.05 within/between centers
- Hierarchical cluster analysis
- Counted number of clusters and number of samples within clusters

Description	CIAT	IITA	CIAT/IITA
No. samples	4,376	3,443	7,821
Samples with <0.05 mrd	1,597	782	301
Number of clusters	214	82	11
Approx. genetic redundancy	**32%**	**20%**	**4%**

Given the high heterozygosity levels of cassava, even a single (selfing) meiosis of any accession would be expected to produce a progeny with greater genetic-distance estimates among individuals than the distances among the putative duplicates in the figure above (mrd $\leq$ 0.05). This makes the identification of cassava duplicates easier than for self-pollinating crops with low heterozygosity levels. Before archiving genetically redundant cassava accessions, DNA samples of potential duplicates will be extracted and genotyped again to exclude any potential human error. In addition, genetically-redundant accessions may be grown side-by-side in the field to compare phenotypes and detect any (epigenetic) differences between potential duplicates which are only visible at the phenotypic level. Genotyping collections of highly heterozygote, clonally propagated crops, which are expensive to conserve, is a useful tool to identify a genetically non-redundant subset of accessions for prioritizing and enhancing the cost effectiveness of *ex situ* conservation efforts.

Example: species-specific diagnostic SNPs in rice

Taxonomic misclassification (misidentification or misnaming) and mislabelling during routine genebank operations are some of the major sources of errors at every genebank, which restrict effective use of germplasm for the correct purpose. To minimize those types of errors in the rice collection at AfricaRice in the future, the group identified 332 diagnostic SNPs out of 31,739 genome-wide SNPs generated via DArTseq technology, which requires conversion to Kompetitive allele-specific PCR (KASP™) (https://www.biosearchtech.com/). Since KASP genotyping can be carried out in 96, 384 and 1536-well PCR plates, genotyping cost is determined based on the PCR reaction volume. A 5 µL reaction volume in a 384-well PCR plate is much cheaper than 15 µL reactions in 96-well PCR plates. For such purpose, we converted 210 out of 332 diagnostic SNPs to KASP assays using the LGC Biosearch Technologies service lab, Hoddesdon, UK. One hundred and forty-six of the 210 KASP SNPs produced working assays. In the first stage of validation using 82 accessions of *Oryza glaberrima* (18), *O. barthii* (18), *O. longistaminata* (10), *O. sativa* Group indica (21) and *O. sativa* Group japonica (15), 86 of the 146 SNPs (59%) were found to be diagnostic as expected. The remaining 41% of the converted SNPs were either monomorphic (17 SNPs) or did not show clear haplotype patterns per species and subspecies to serve as diagnostic markers. A final validation was undertaken of 60 of the 86 diagnostic SNPs using 565 accessions of *O. barthii* (80 samples), *O. glaberrima* (130), *O. longistaminata* (77), *O. sativa* Group indica (141) and *O. sativa* Group japonica (137). Based on the final validation genotype data, three alternative panels of 10, 24 and 40 SNP were recommended to provide flexible, rapid turnaround and cost-effective means for rectifying errors during germplasm collection, acquisition or routine genebank operations, which ultimately facilitate germplasm curation and management.

Example: Use of Primordial Germ Cells (PGSC) cuts down the cost of African poultry conservation

Recent advances in genome editing methods like CRISPR/Cas9 allow precise heritable modification in the genome sequence information of a germline cell. These advances have attracted interest in using this technology for cost effective conservation. Thanks to DSI, the Centre for Tropical Livestock Genetics and Health (CTLGH) has developed an innovative technology for conservation of tropical poultry genetic resources using primordial germ cells (PGCs), and their resuscitation using the surrogacy technology for large-scale dissemination of elite tropical poultry genetic resources. This technology enables the maintenance of millions of animals stored in the form of cells in test tube, thus replacing the high cost of keeping and maintaining live animals with the low cost of test tubes in liquid nitrogen This is allowing the Kenyan Agricultural and Livestock Research Organization (KALRO) in collaboration with the International Livestock Research institute (ILRI) to easily conserve and restore Kenyan ecotypes of indigenous chicken and the improved lines.

As part of projects that aim to contribute to the conservation of genetically rich landraces and traditional varieties through their cultivation and commercialization by farmers, some CGIAR Centers, in collaboration with local and national organizations, have applied genomic tools to genetically characterize landrace diversity found *in situ*.

Example: cultivated bananas from Papua New Guinea

The Alliance of Bioversity International and CIAT in collaboration with National Agricultural Research Institute (NARI) collected plants grown in Papua New Guinea (Sardos et al, 2017), a center of diversity and domestication for banana. This project resulted in interaction with farmers and selection of plants that either looked new or were variants of known genotypes. Genomic characterization was conducted to confirm their unique genetic patterns and 40 new accessions were safeguarded in the Musa Germplasm Transit Centre (ITC). Related DSI datasets were made available as part as a bigger dataset to the whole community (Rouard et al, 2021, Sardos et al 2022)

3. DSI and the sustainable use of the components of biological diversity

In this section we address how DSI has fundamentally changed the traditional approach to the use of genetic resources to develop new products. The implications of these new approaches for benefit-sharing will be addressed in the next section.

Over ten millennia of farming, varieties and breeds have been improved based on phenotypic selection of parents and offspring that resulted in the alteration of their genomes. A high-yielding, low-quality variety might be crossed with a low-yielding, highly nutritious variety with the objective of developing a new high-yielding, highly-nutritious variety. The same principle can be applied for example to develop varieties with less reliance on chemical inputs through better climate adaptation, better pest and disease resistance, or greater tolerance of drought, flooding, or polluted soils.

The process of developing superior new products in this way is slow, costly, inefficient and potentially unreliable because phenotype varies with age, environment, organ and pre-conditioning, and because phenotypic traits can be a function of multiple interacting genes resulting in complex and potentially non-obvious inheritance[9]. Prior to the 21st century, this approach was sufficient to meet our needs because of the relatively slow pace of change in human society; indeed it was the basis of the massive civilisations that have appeared since agriculture started. However, with the current planet crisis caused by unprecedented rates of change in population size and climate, we have to change more rapidly than is possible with the traditional approach. Continuing to rely on this traditional approach would result in failure to meet SDGs 1 to 3.

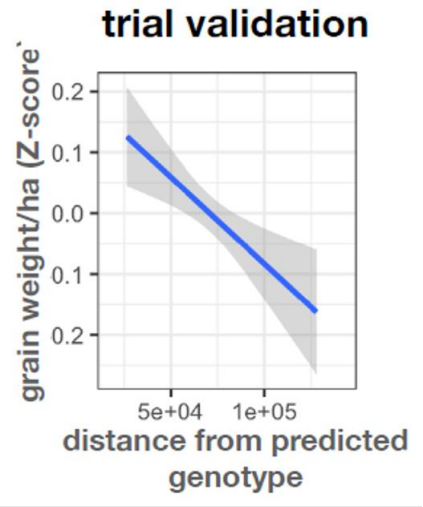
[9] A classic example is the shape of a chicken's comb. When a chicken with a "rose" comb is crossed with a chicken with a "pea" comb, the offspring have a "walnut" comb that is entirely unlike either of the parents' combs (Miko 2008). Another example is the difference between spring and winter cereal crops: crossing two spring barley varieties can result in winter barley progeny if the two parents have complementary vernalisation genes (Takahashi, R. and Yasuda, S. 1971).

In contrast, genotype is fixed and can be determined at birth without having to wait months or years for the trait to be expressed in an appropriate organ and developmental stage of the organism grown in an appropriate set of environments. Hence, given sufficient knowledge of the genetic control of the traits of interest, new products can be developed faster, more efficiently and more effectively by basing selection directly on genotype rather than indirectly through the resulting phenotype. The benefits are greatest for traits that are difficult, expensive or slow to measure, show complex inheritance, are expressed late in the life of the organism, have low heritability, are the result of recessive genes, or are traits of response to environment requiring assessment in multiple environments.

Example: adapting maize to climate change

It is exceptionally difficult to improve adaptation to climate based on phenotypic selection, as climate is a complex composite of numerous variables, and response to climate must be measured by comparing performance across numerous environments.

CIMMYT and collaborators have tested the potential for adapting maize to new climates using genotype-based prediction. They established a "training panel" of landrace varieties of maize that had been collected from precisely known locations across a range of countries. They DNA-fingerprinted the training panel and quantified the relationship between the genotype and the climate at its location of origin. Using this information, they predicted the best synthetic genotype for a new test location. To check their prediction, they sourced data on the genotype and phenotypic performance in the test location of progeny related to the training set along with unrelated entries. The genetic distance of each tested entry to the predicted best synthetic genotype was calculated. As shown in the figure below, the further away from the predicted genotype an individual gets, the lower its relative performance within the specific trial location and environment, confirming the merit of adopting a DSI-based approach to climate adaptation.



3.1. DSI-based knowledge and tools

Effectively basing selection on genotype is predicated on understanding the complex dependence of phenotype on genotype. Hence the application of DSI adds a new initial step in the research-for-development pathway which includes 1) the creation of knowledge about the genome, and 2) the development of DSI-based tools (such as genetic markers, DNA constructs, or algorithms). Only then can the tools be applied to create commercially viable products.

3.1.1. Marker-assisted selection

The simplest and oldest approach based on DNA analysis dates back to the 1980s, with the discovery of RFLPs, which enabled genetic variants to be revealed at random positions spread across the genome. RFLPs have since been replaced with other genotyping technologies such as SSRs, AFLPs, RAPDs, SNPs, DArTs and GBS, but the principle remains the same. The genetic variants, while typically not part of a gene, may be physically close to, and thus tend to be inherited with, a gene controlling a trait of interest. Such a genetic variant can be an indicator, or marker, of the likely presence of a trait.

Traditionally, the marker development process starts with selecting two parents based on their phenotypes, one with and the other without a particular trait of interest. By creating large numbers of progeny of the two parents over several generations, and recording their phenotype and their genotype, associations can be established between the trait and DNA variants. Variants that are associated with the trait are then candidates for use as a marker for selecting progeny with the trait.

The original approach to marker-assisted selection has major limitations. It is slow and expensive and has to be repeated for each trait of interest. The association between the marker and the trait is a correlation, which does not necessarily imply causation, and the correlation is prone to breakdown in other genetic resources. Even where the association is causal, the effect of a gene depends on the composition of the rest of the genome. For both reasons, such markers are typically effective only with the specific, small breeding populations, for which they were developed, and must be re-developed for different breeding populations. The weaker the genetic association, the greater the functional interactions between genes and the narrower the range of efficacy. Further, since the approach targets a single gene-trait combination, it is most useful for transferring single genes that have major effects on high-value traits into an existing elite variety.

3.1.2. Building knowledge

The fact that marker-assisted selection has become so pervasive in crop improvement despite its limitations is testament to its superiority over basing selection on phenotype. Massive research efforts have been devoted to overcoming the limitations, by building greater knowledge of important genomes and the functional relationship between genotype and phenotype in order to develop tools better and faster, to enable genotype-based selection of parents as well as offspring. Much of the knowledge has

been built by the upstream research community independently of the downstream product development community, and as such is often based on genetic resources that are not relevant to or incorporated into commercial products. Moreover, some of the methods of knowledge-building intrinsically rely on the use of large numbers of genetic resources from many different countries of origin, most of which are used solely for the purpose of building knowledge and are never incorporated into commercial products.

DNA sequencing is crucial for building knowledge of the dependence of phenotype on genotype. All CGIAR Centers have started sequencing genomes, even for taxa that are hard to sequence because they are outbreeders (e.g., maize and pearl millet) or have polyploid genomes (e.g., potato, banana, bread wheat, finger millet and peanut).

Sequencing a genome accurately from scratch is extremely time consuming and expensive. Although time and costs have been reduced dramatically in recent years, it remains prohibitively expensive to undertake on a large scale. However, a single genome that has been sequenced with high accuracy and reliability may be used as a "reference genome" for low-cost high-throughput resequencing ("next generation sequencing", or NGS) of many similar genomes. This low-cost resequencing data then becomes a persistent resource; new data can be added to incrementally over time and re-used by multiple researchers simultaneously around the world for a wide variety of purposes and traits. For genetically diverse species, a single reference genome is not sufficient for reliable NGS of the whole species. For example, a set of 16 high-quality reference genomes has been developed for Asian cultivated rice *Oryza sativa* (Zhou et al 2020) and has enhanced the NGS sequencing of 3,010 accessions (Weng *et al* 2018). More and more genomes are being sequenced for other crops and animals but at variable scales and pace (Table 1). When genome position or order of variants is not needed, *de novo* generation of polymorphic data through comparison of NGS-based data within a set of entries of interest can be conducted.

Crops (Annex 1)	**Reference / high quality genomes**	**Draft assembled genomes**	**Resequencing (R) & NGS Genotyping (G)**	**Total *ex Situ* accessions**
Banana	5	9	300 (R) + 600 (G)	*7,000*
Barley	1	19	22,000 (R)	371,304
Bean	3	5	?	165,418
Breadfruit	0	1	-	388
Cassava	2	6	16,600 (G)	16,841
Chickpea	4	4	3,366 (R)	91,654
Coconut	2	1	-	256
Cowpea	1	6	1,300 (G)	56,723
Faba bean	1	0	250 (G)	33,707
Finger Millet	1	2	314 (G)	31,229
Groundnut	2	5	384 (R) + 812 (G)	61,143
Lentil	1	0	-	42,713
Maize	9	60-80	26,825 (G)	218,522
Pea	1	4	-	67,743
Pearl Millet	1	6	36 (G)	51,624
Pigeonpea	1	2	292 (R) + 663 (G)	28,868

Crops (Annex 1)	Reference / high quality genomes	Draft assembled genomes	Resequencing (R) & NGS Genotyping (G)	Total *ex Situ* accessions
Potato	2	25	4,000+	52,155
Rice (Asian)	15	600	10,000 (R) +	481,554
Rice African	1	-	246 (R) + 9380 (G)	11,270
Sorghum	2	7-16	176 (R)	172,928
Sweet potato	3	4	5,771 (G)	19,148
Wheat	3	10	80,000+ (G)	659,801
Yams	2	5	1,500 (G)	7,008
Forages				
Buffelgrass	0	0	205	-
Napier grass	1	0	350	
Rhodes grass	0	0	104	
Urochloa brizantha	1	0	254	
Animals	Reference / high quality genomes	Draft assembled genomes	Resequencing (R) & NGS Genotyping (G)	Populations
African cattle	1	-	220+ (R)	180 breeds
Nile Tilapia	1	-	100 (G)	-

Table 1. Progress in sequencing, including reference genomes, draft *de novo* genomes, low-coverage whole genome resequencing, NGS genotyping for Annex 1 crops and animals in the CGIAR (Sources: NCBI, PUBMED, FAO VIEWS, CGIAR) as of November 2022. Statistics are exhaustive and can be an underestimation of DSI produced worldwide.

The utilisation of such genetic resources in this way generates enormous indirect benefits through the knowledge generation they facilitate, even if they are never used directly for the creation of commercial products.

Data on genome sequences is of value only once it is "decoded" to identify polymorphisms and their associations with phenotypic data, to understand their organizational structures and their evolutionary changes (comparative and/or structural genomics) and to elaborate how the genes and intergenic regions of the genome contribute to different biological processes ("functional genomics"). The combination of this knowledge with the underlying sequence data is the body of information referred to as DSI. Much of our knowledge of how the genome and genes are organised, how their expression is regulated and ultimately how the genome determines phenotype, is based on studies with model species such as the weedy flowering species *Arabidopsis thaliana*, first sequenced in 2000 (Arabidopsis Genome Initiative, 2000). Moreover, because of the observed synteny between the genomes of different taxa, knowledge of DNA sequences and genes and their functions in *Arabidopsis* and other model species is often directly transferable (Salse *et al* 2002, Shultz *et al* 2007).

Genetic resources of these model species are rarely used directly to create commercial products, but by making it possible to identify genes and their functions within DSI, they have completely transformed our ability to utilise other genetic resources and hence have created huge benefits.

Ideally, a genetic marker for a trait should not simply be physically close to the DNA sequence that actually determines the trait, but should actually be based on that sequence. Selections based on such ideal markers will allow reliable improvement based on these functionally important alleles. However, finding which portion of the genome actually determines a trait is a major challenge and typically involves seeking convergent evidence of multiple kinds. This can include for example, studies of developmental genetics, physiology and biochemistry, analysis of genetic and environmental effects on gene expression (transcriptomics and proteomics), knowledge of gene function in related species, and mapping studies. Again, this involves utilizing a potentially large number of genetic resources from a wide range of species, genera and kingdoms for the purpose of building knowledge, not directly for creating commercial products.

Genome Wide Association Studies (GWAS)

With a DSI dataset comprising whole-genome genotyping data for a large number of genomes, GWAS becomes possible for the fast detection of markers for marker-assisted selection. By evaluating the same set of genomes phenotypically, correlations can be determined between genotype and phenotype. The same DSI dataset can be reused for multiple traits in multiple environments. Hence GWAS is becoming a widespread tool for the initial identification of chromosomal regions and likely genes controlling target traits, adding major value to phenotyping trials where the DSI dataset is available: what had previously been a simple trial to evaluate phenotypic diversity now serves to identify markers associated with traits.

In addition, where the DSI dataset is based on genome sequences that have been resolved into genes, the technique allows the preliminary identification of candidate genes controlling traits. Moreover, it is relatively straightforward to train scientists to undertake GWAS for their own phenotypic data, making the technology highly scalable and transferable. This has resulted in a massive increase in demand for genetic resources that have been sequenced, especially for crops such as rice, sorghum and chickpea where large numbers of accessions have been genotyped or sequenced and access is provided not only to the material but also to the associated DSI and analytical skills.

As with so much of the knowledge relating to DSI, the genetic resources used for GWAS are not necessarily utilised directly in the development of commercial products. The benefits realised from utilising them for GWAS lie in the resulting knowledge of genotype-phenotype relationship, leading to the development and use of markers for more effective application of commercial products derived from other genetic resources.

As with traditional marker development, GWAS is based only on observed correlations between genotype and phenotype and do not necessarily reveal causal relationships. Potential markers need at least to be validated in target breeding materials. As with other approaches to identifying correlations between genotype and phenotype, additional research is needed to reveal causal relationships.

Example: GWAS in support of wheat selection and breeding for developing resistance to wheat blast

Wheat blast is a devastating fungal disease caused by the fungus *Magnaporthe oryza* pathotype *triticum* (MoT). It was first reported on wheat (*Triticum aestivum L.*) in 1985 in Paraná, Brazil and has since spread throughout many of the important wheat-producing areas of South America. In 2016, wheat blast spread to Bangladesh, which suffered a severe outbreak. Currently this disease threatens wheat production in tropical areas in South America and South Asia. Directly striking the wheat ear, wheat blast can shrivel and deform the grain in less than a week from the first symptoms, leaving farmers no time to act. There are no widely available resistant varieties, and fungicides are expensive and provide only a partial defense. They are also often hard to obtain or use in the regions where the blast occurs, and must be applied well before any symptoms appear — a prohibitive expense for many farmers. Against this background, CIMMYT scientists have made use of genomic tools to identify sources of blast resistance. In 2017-2020, multi-environment genome-wide association mapping for blast resistance enabled identification of consistent blast associated genotyping-by-sequencing markers and provided valuable insights into the genetic architecture of resistance against different isolates of MoT in Bolivia and Bangladesh. As part of the efforts, 4,143 lines were fingerprinted to identify the presence of *Aegilops ventricose* 2NS translocation segment, a segment that has been widely utilized in wheat breeding for disease-resistance and which has been recently associated with resistance to wheat blast. The results have provided insights into the increasing frequency of the segment in wheat lines and have emphasized the urgent need to identify novel sources of blast resistance beyond the 2NS translocation segment.

Example: African rice

AfricaRice and partners evaluated amylose content (AC) of 1,020 *Oryza glaberrima* accessions conserved at the AfricaRice genebank. AC is one of the important quality traits that influence texture and taste of cooked rice, but the genetic control of AC in *O. glaberrima* still remains poorly understood. The study reported high variation in AC that ranged from 15.1 to 29.6%, with an overall mean of 26.2%, which is higher than the values reported in *O. sativa*. Nearly 25% of the *O. glaberrima* accessions had AC values greater than the highest value (28%) reported in indica and japonica. A GWAS was conducted to map the genomic regions associated with AC in a panel of 386 *O. glaberrima* accessions that represented the phenotypic variation for amylose content. The panel was genotyped with 31,739 single nucleotide polymorphism (SNP) markers using DArTseq, of which 4,135 SNPs were polymorphic. Using a threshold of $p < 10^{-5}$ and weighted mixed linear model, three genomic regions that individually explained up to 12.1% of the phenotypic variation of AC in *O. glaberrima* were identified. The significant GWAS hits were validated using a biparental interspecific population derived from a cross between *O. glaberrima* and *O. sativa*, but the genomic regions identified in the two populations were different. One of the points in the GWAS panel was a very narrow genetic variation among the *O. glaberrima* accessions. To identify the most genetically diverse set of *O. glaberrima* collections for future use in breeding and gene discovery (including AC), the entire O. glaberrima collections conserved at the AfricaRice genebank were genotyped and a core collection of 350 accessions that captured 97% of the molecular variation observed in the whole set was developed. This entire collection for quality and nutritional traits and the core collection for diverse agronomic and biotic traits are currently being evaluated.

Example: GWAS for the identification of new sources of resistance to *Striga hermonthica*

A recent GWA study in sorghum in collaboration with Kenyatta University (Kenya) made use of a subset of a reference panel to understand pre-attachment resistance of sorghum to *Striga*. The study not only identified markers and genes associated with the trait but also reported new genotypes from the reference set that had higher levels of resistance than the checks (Mallu et al. 2022).

Fine mapping

By genotyping or sequencing more genomes at high density, correlations between genotype and phenotype can be progressively refined to a smaller part of the genome. This is known as “fine mapping”. If the genetic data include DNA sequences that have been resolved into genes, the genes in that part of the genome may be considered as candidate genes for controlling the trait of interest. It can be an important tool in the long process of determining the functional dependence of phenotype on genotype, as it reduces the number of candidate genes that need to be validated.

Example: maize lethal necrosis (MLN)

At CIMMYT, MLN was fine mapped with large multiple F2 populations. MLN is a viral disease with a devastating effect on maize production. Screening thousands of global maize lines revealed KS23-5 and KS23-6 as two of the most promising donors of MLN resistance alleles. Both linkage mapping and association mapping approaches were used to discover and validate genomic regions associated with MLN resistance. Selective genotyping of resistant and susceptible individuals within large F_2 populations coupled with GWAS identified a major effect QTL (*qMLN06_157*) on chromosome 6 for MLN disease severity and AUDPC values in multiple F_2 populations involving one of the KS23 lines as a parent. The major effect QTL (*qMLN06_157*) is recessively inherited and explained 55 to 70% of the phenotypic variation with an approximate 6 Mb confidence interval. Linkage mapping in three F_3 populations and three F_2 populations involving KS23-5 or KS23-6 as one of the parents confirmed the presence of the major effect QTL on chromosome 6, demonstrating the efficacy of the KS23 allele at *qMLN06.157* in varying populations. This QTL was further fine mapped with large F2 populations and flanking markers were identified. Based on the consistent and effective resistance afforded by the KS23 allele at *qMLN06.157*, the allele has been used extensively in both marker-assisted forward breeding and marker-assisted backcrossing schemes, to improve MLN resistance of breeding populations and key lines for Eastern Africa. Until now >130 elite lines in Eastern Africa are converted into MLN resistance through MABC and released as public goods to be used in breeding program (Murithi et al 2021; Awata et al., 2021).

Example: maize streak virus

Maize streak virus (MSV) disease is a devastating disease in the Sub-Saharan Africa (SSA), which causes significant yield loss in maize. MSV disease outbreaks often coincide with drought periods or irregular early rains exacerbating crop failures and often resulting in complete crop losses. Studies have reported losses to the tune of US$120-480 million per year due to MSV in Africa in terms of lost income and higher maize prices, and indicated that at least half of such loss could be potentially recovered with the effective control of MSV. Resistance to MSV was mapped to a major QTL (*Msv1*) on chromosome1 (bin 1.05) that is germplasm and environment independent and to several minor loci elsewhere in the genome. Initial mapping for response to MSV in a linkage mapping population identified *Msv1* in a 5.9 cM interval, which explained approximately 67% of phenotypic variance. Fine mapping of a large effect QTL in the presence of other minor effect loci requires some special considerations. A variant of the F_3- based QTL Isogenic Recombinant (QIR) approach was designed for fine mapping, in which the genomic regions harboring minor effect loci were selectively homogenized in one generation as against the NIL (near-isogenic line) approach, which seeks to homogenize the entire background except the target locus through repeated backcrosses. Seed-DNA based genotyping enabled non-destructive and efficient screening of a large number (4,725) of F_2 seeds for the flanking markers of the *Msv1* and other QTL intervals for identifying QIRs. The *Msv1* interval was saturated with additional markers and seven recombination exchange points were identified, enabling delimitation of *Msv1* to a smaller interval of 0.87 cM. A parallel GWAS undertaken in the DTMA panel that broadly represents tropical and subtropical cultivated maize genetic diversity, and genotyped at high density of ~1 million GBS SNPs identified an overlapping region as the fine mapped interval of *Msv1*, and helped optimize the haplotype for selection. The most significant SNP was located in the 10^{th} exon of a U-box domain containing tyrosine kinase family protein coding gene. Though a number of protein kinases have been established as disease resistance genes and U-box domains that are part of E3 ligases have been demonstrated to be associated with plant defense mechanism through ubiquitination-mediated protein degradation, further experiments will be needed to unambiguously identify the gene underlying the QTL. The Kompetitive Allele-specific PCR (KASP) assays developed for the production haplotype (Nair et al., 2015) have been used in more than 100,000 breeding lines since the last seven years in the relevant breeding pipelines in SSA for forward breeding. This helped to increase the frequency of favorable alleles for Msv1 in the early stages of selection in the breeding pipeline.

Example: chickpea and drought

Chickpea production is vulnerable to drought stress. Identifying the genetic components underlying drought adaptation is crucial for enhancing chickpea productivity under water stress conditions. A chickpea landrace, ICC 1882 of Indian origin was identified as tolerant to drought in the 1980s. Using a RIL population a 'QTL-hotspot', a genomic region controlling chickpea growth with positive consequences on crop production under drought. Following various strategies and 1911 BC_6F_2 plants, Barmukh et al (2022) were able to fine map and identify a nonsynonymous substitution in the transcription factor *CaTIFY4b* that regulates seed weight and organ size in chickpea. It is suggested that allelic variation in 'QTL-hotspot' improves pre-anthesis water use, transpiration efficiency, root architecture and canopy development, thus enabling high-yield performance under terminal drought conditions. While another germplasm line, ICC 4958 also drought tolerant and collected from India, has the drought tolerance associated with its larger root length and volume.

Proof of gene function

As indicated above, proving that a putative candidate gene indeed has an effect on a trait has been a very challenging task, often involving convergent evidence of different kinds. The definitive final proof is to modify just the one gene within a genome. However, this is also challenging, as the close linkage between adjacent genes in a genome makes it very difficult by conventional breeding to change only one gene without also changing its neighbours. Nevertheless, it has been a key challenge to address, because it results in "perfect" markers for selection, which enables parents as well as offspring to be selected on the basis of their genotype. Options to overcome the difficulties posed by linkage between genes and modify a single gene at a time include chemical- or radiation-induced mutants or GM methodologies. However, both these approaches present other challenges.

Gene editing is rapidly becoming the preferred tool of choice for proof of gene function, as it is simpler, more efficient, faster and more reliable than other approaches. Once a candidate gene has been identified, it can be edited directly or transferred to another genome by gene editing, changing the genome sequence in a precisely controlled manner at precisely the correct position. The technology can be applied to any part of the gene, including not only the coding portion of a gene but also the regulatory portion and other non-coding parts of the gene. This can provide unequivocal proof of the function of one gene or several genes within a year, rather than the many years required with other technologies. In fact, the major use of gene editing is now in the realm of developing DSI tools (validating gene function and assessing phenotype in a range of genetic backgrounds) rather than developing commercial products.

Example: gene editing for gene validation

IITA has developed a robust CRISPR/Cas9 mediated genome editing tool for banana, testing it by knocking out a specific gene (Phytoene dehydrogenase PDS) in a banana cultivar 'Sukali Ndiizi' and a plantain cultivar 'Gonja Manjaya' (Ntui et al 2020). To demonstrate the reliability of the method, they knocked out the gene in 77 plants of the banana and 17 of the plantain cultivar by altering two different targeted parts of the gene. The gene-edited plants showed albino and variegated phenotypes as expected, indicating mutations at the targeted sites had indeed disrupted the function of PDS gene.

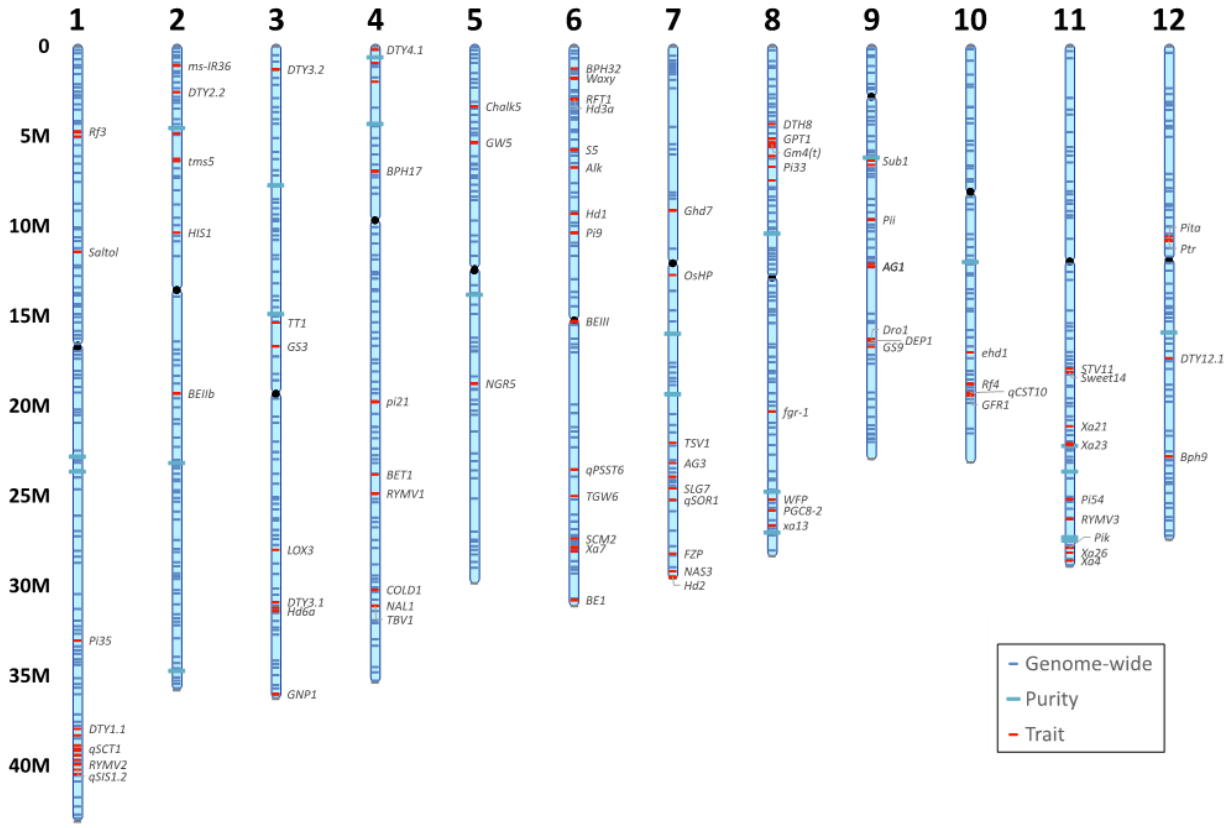
3.1.3. Improving tools

Improving markers

Differences in the DNA sequences of the allelic variants of a gene (including not only the coding portion of the gene but also the regulatory portion and other non-coding parts of the gene) controlling a trait of interest can be expressed as one or more Single Nucleotide Polymorphisms (SNPs) (e.g., Liu *et al* 2012), i.e., as short lengths of DNA that differ by a single nucleotide. DSI resources such as genome sequences can be used to inform the discovery and evaluation of new candidate markers. The dense nature of genome sequencing data provides ample source of SNP and other genomic polymorphisms, which can be evaluated for their suitability as markers, particularly when genome sequencing data is compared across a range of specified diversity. For example, trait markers targeting a specific desirable allele of a gene can be assessed for accuracy (Platten *et al.* 2019), a task for which genome sequencing data on diverse gene bank accessions is remarkably useful. Likewise, genome-wide fingerprinting panels can be designed by considering allele frequencies across the diversity observed in breeding cohorts.

Another important improvement in the tools for marker-assisted selection is the development of customisable assays such as sequence-based assay sets and the SNP chip. SNP chips or arrays enable multiple SNPs to be tested on a single chip, enabling multiple genes to be probed in a single assay. Based on synthesis of knowledge from the various sources described in the previous section, potentially derived from a wide range of genetic resources, chips, and sequence-based systems can be designed to probe anything from a few SNPs to tens of thousands of SNPs and are now available for many crops (e.g., Singh *et al* 2020). Designing chips to probe genic SNPs overcomes many of the limitations of traditional marker assisted selection, including enabling the ability to select for multiple traits simultaneously.

Example: Marker systems for public-sector rice breeding programs



Data from the 3000 rice genomes and other similar projects, together with targeted resequencing of elite germplasm from IRRI, AfricaRice, BRRI and other public-sector institutes, have been used to design and refine marker systems in use for public-sector rice breeding efforts. These marker systems include a comprehensive set of over 260 high-value trait markers designed to be accurate across the entire diversity of *Oryza sativa* germplasm, and implemented at publicly-accessible service providers such as Intertek. These trait markers allow public-sector rice breeders to make reliable, cheap, high-throughput selections for over 100 major genes contributing to a wide range of traits including major diseases (blast, bacterial blight, rice yellow mottle virus and many others), abiotic stresses (drought, salinity, submergence, heat etc.), maturity, grain quality and agronomic performance.

In addition to the trait-specific markers, a genome-wide panel of markers has been designed, informative on the elite indica material that forms the basis of nearly all breeding programs in developing countries. This set of ~1000 markers (the 1k RiCA panel, Arbalaez *et al.* 2019) enables genome-wide fingerprinting of elite breeding programs, and forms the basis of genomic predictions in public-sector rice breeding programs.

Both of these marker systems were developed for public-sector use by leveraging major DSI resources such as the 3000 rice genomes project, which have been made available through public repositories such as SNP-seek. Ongoing efforts to sequence elite breeding core panels and key trait donor lines are contributing to continued refinement and improvement of these public resources. Improved elite varieties developed through these resources are already being tested in target countries in Africa, South Asia and South-East Asia.

IRRI has also developed SNP panels to correctly identify different species and sub-groups within the cultivated species.

Example: sorghum, groundnut, pigeonpea and finger millet

ICRISAT in collaboration with CGIAR's Excellence in Breeding (EiB) initiative and national partners, have developed trait marker and mid-density panels (MDP) for sorghum, groundnut, pigeonpea and finger millet. The trait marker panels contain markers associated with traits of interest that can be used for early generation selection. The MDPs contain combinations of genome-wide markers as well as trait markers and can be used for genomic selection. These panels have been made available through the EiB tools page: https://excellenceinbreeding.org/toolbox

Genomic selection

Bypassing the need to generate trait-specific knowledge about the dependence of phenotype on genotype, genomic selection is an entirely different approach. It uses advanced algorithms and can include machine and deep learning techniques applied to a "training population" of genetic resources to guide progress based on the genome. Selection of an appropriate training population is key to the success of the approach. Including genetic resources that will knowingly not become part of the commercial product can be critical in guiding the machine learning algorithms to determine what to discard.

Example: genomic predictions in rice

Prediction-based breeding has been implemented in IRRI's mainstream rice breeding programs for a number of years. Genomic selection allows breeders to reduce cycle times by selecting and recycling promising lines as parents with minimal or in advance of field testing, and to increase selection pressure by scaling up the program, growing a larger cohort of progeny and only testing those predicted by the genome-wide fingerprint to have superior performance.

To further improve the accuracy and usefulness of genomic predictions, breeding programs are actively engaged in whole-genome resequencing of elite core panels and cohorts. This allows continued assessment of marker panels for informativeness, and where necessary the refinement of these panels to maintain their utility as the composition of breeding cohorts changes. It also enables better design of training cohorts, and is a fantastic resource for imputation of high-density haplotypes based on mid-density genotyping data.

Example: genomic selection in maize

Genomic selection (GS) is successfully and routinely applied in CIMMYT's tropical maize breeding program. GS is applied at two stages in the maize breeding pipeline (i) at F3 cycle with the aim to accumulate maximum favorable alleles, and other at (ii) early-stage testing for yield with the aim to predict the performance of hybrid or GCA of a line. GS at F3 stage is applied on population developed by using two parents or multiple parents. Biparental based rapid cycle genomic selection (RCGS) was routinely used in maize since 2012 in Eastern Africa and South Asia breeding pipelines with the aim to increase the grain yield under drought, low N and heat stress management. With biparental based RCGS, the average gain from GS per cycle across populations was 0.086 t/ha. Hybrids derived from RCGS produced 7.3% (0.176 t/ha) higher grain yield than those developed through the conventional pedigree breeding method. Even though biparental based RCGS is effective and developed several high yielding elite lines and hybrids, it incorporates maximum number of alleles from the selected two parents only. On the other hand, to increase the number of favorable alleles to the maximum extent, multi-parents based RCGS strategy was applied. Since 2015, both in Latin America and south Asia CIMMYT's breeding pipeline, multiple parent-based RCGS was applied and demonstrated higher genetic gain compared to traditional recurrent selection with reduced resources (Zhang et al., 2017; Das et al., 2021). Multi-parent based RCGS is useful for high complex traits like grain yield under drought stress, heat stress, water logging tolerance and low soil N stress.

There are three product profiles in CIMMYT's East and Southern Africa breeding pipeline, for which GS is routinely applied at early stage of testing for yield under optimum, drought and low N stress conditions. Prediction by using historical data alone yielded relatively low accuracy for GY and other agronomic traits. However, we showed that by combining only 10% to 30% of the lines from the year of testing led to significant increase in the predicting accuracy. CIMMYT Africa's breeding program was implementing GS at the early stage of testing using a test-half-predict-all strategy (Beyene et al. 2019, 2021). With the use of historical data, we reduce the training population size from 50% to 10-30% to achieve the same or even higher level of accuracy. This saves the costs associated with testcross formation and multi-location evaluation, and also helps to reduce the breeding cycle time.

3.2. DSI and the replication of units of heredity

Units of heredity are normally transmitted across generations through the replication of DNA or RNA; by binary fission in prokaryotes and by mitosis and meiosis in eukaryotes, subject to variations caused by mutation and recombination. DNA or RNA in the offspring is merely a copy of, not materially the same as, that in the parent(s). Only the abstract concept of the nucleotide sequence persists across generations. This has led to the concept of DNA as information (Willis 2016), in which the subject matter of inheritance is in fact the information contained within the nucleotide sequence, and the DNA or RNA is merely the ephemeral material carrier of the information. Progeny "incorporate" genetic material from their parent(s) only in the sense that their functional units of heredity have the same information content.

Genetic modification and gene editing remove this dependence on meiosis and mitosis, through enabling *in vitro* replication of DNA and RNA and its subsequent horizontal transfer into a recipient genome.

Even with these tools, DNA and RNA remain the carriers of the information content of DNA, as the information is replicated solely by replicating DNA molecules. Synthetic biology also removes this dependence: where the information content of DNA has been replicated digitally, synthetic biology enables this digital information to be restored to its biological form by resynthesizing DNA with the defined sequence. It further enables novel sequences to be developed and tested. Hence the use of genetic resources can now involve a combination of *in vivo*, *in vitro* and *in silico* replication of units of heredity, with a combination of vertical and horizontal gene transfer, and even the creation of novel units of heredity through *in silico* mutation and recombination.

3.3. DSI and the identification of alternative genetic resources

Whilst the concept of a "country of origin" may in some cases be reasonable at the whole genome level, it is less applicable at the level of individual genes. Many gene variants can be found in unrelated germplasm from multiple countries of origin. Searching for alternative sources of the same high-value genes becomes routine through the use of DSI. Hence, once a user has built knowledge of phenotype-genotype dependences using genetic resources acquired under one agreement, it is then relatively easy (at least within the constraint that the effect of a gene depends on the rest of the genome) to use that knowledge to develop commercial products from genetic resources acquired under a separate agreement. Similarly, if one user builds knowledge of phenotype-genotype dependences using genetic resources acquired under one agreement and then makes the associated DSI and knowledge freely available, it can be relatively easy for other users to exploit that knowledge to develop commercial products from genetic resources acquired under a separate agreement.

The potential to utilize unrelated genetic resources in the two steps (knowledge creation and product development) clearly has significant implications for benefit sharing, which are addressed in the next section.

4. DSI and the fair and equitable sharing of the benefits arising out of the use of genetic resources

As an international organization dedicated to creating benefits specifically for developing countries, CGIAR's primary concern is that the policy environment should facilitate the creation of those benefits. To this end the benefit-sharing regime needs to be efficient and inclusive.

4.1. Monetary benefit-sharing

In the following paragraphs we comment on some of the leading options for monetary benefit-sharing from DSI users currently under consideration by the Informal Advisory Committee (IAG) to the Co-Chairs

on DSI, OEWG-5, and ultimately, the CBD-COP.[10] Our comments are based on our expertise and experiences using DSI in agricultural research and development, as highlighted in Sections 2 and 3 above.

4.1.1. The need for a multilateral solution

As shown in previous sections, DSI fundamentally changes the approach to creating commercially viable products derived from genetic resources for food and agriculture. The changes argue strongly for a multilateral approach to benefit-sharing. In particular:

- The benefits of DSI arise largely from the use of genetic resources to create DSI-based knowledge and tools, and the development of these tools relies on many genetic resources from many countries. Once a tool has been developed, it may be used to incorporate high-value genetic material originating from just a small number of countries, perhaps even only one. It does not appear to be fair nor equitable to share benefits solely with the few countries that provided the genetic material that finally appears in the product. Benefits must be shared with all the many countries that provided material used for the development of the DSI-based tools.

- DSI-based tools and technologies enable *in vitro* and *in silico* replication of DNA sequences followed by horizontal gene transfer to reintroduce them into a material product. It does not seem reasonable to base monetary benefit-sharing solely on the narrow concept of *in vivo* replication and vertical gene transfer through the natural process of cell division.

- DSI enable developers to focus on single genes, where a single specific gene sequence can be found in numerous different varieties from numerous different "countries of origin". This argues in favour of multilateral benefit-sharing as the alternative is a major loophole: a developer can identify a gene from one variety under an ABS agreement with the applicable country of origin, sequence it, find the exact same gene sequence in a different variety that is not subject to any ABS agreement, and use that second variety for further product development free of any benefit-sharing requirements.

An international system regulating access and benefit-sharing for DSI on a bilateral basis would be disastrous for agricultural research and development. Indeed, the negative impact of bilateral ABS regulation for Plant Genetic Resources for Food and Agriculture (PGRFA) was already evident to the negotiators of the Plant Treaty when they agreed, 21 years ago, to create the multilateral system of access and benefit-sharing for genetic resources of crops and forages upon which the international community is interdependent, and which are important for food security. The rationale for multilateralism is even stronger with respect to DSI derived from PGRFA (NB we suggest below that multilateral benefit-sharing for DSI related PGRFA could/should be layered into the Plant Treaty's multilateral system). Given the manner in which DSI is generated and stored, we are even more interdependent with respect to how we

[10] Co-Leads' Report on the Work of the Informal Co-Chairs' Advisory Group on Digital Sequence Information on Genetic Resources, CBD/WG2020/3/INF/8, https://www.cbd.int/doc/c/079d/1142/339a68fee2d22e95fb2b1c4c/wg2020-03-inf-08-en.pdf

access, use and rely on that information. It would be impossible to negotiate all the bilateral agreements that would be required to assemble the data sets used in research activities described in Sections 2 and 3. And it would be even more impossibly complex to manage those data sets with each datum being potentially subject to different terms and conditions of use and benefit-sharing. As highlighted in sections 2 and 3 above, these uses of DSI derived from GRFA are critically important for food security. CGIAR very much appreciates that the negotiators under the aegis of the CBD (and the Plant Treaty) are generally in agreement that a global system of bilateral regulation of DSI is untenable, and that a multilateral approach is more responsive to the reality of the way DSI is generated, shared and used.

There are of course different ways to develop multilateral systems, with different combinations of elements. Our experience suggests that the broadest, flattest possible multilateral system, with fewest exceptions, thresholds, and choices to be made by both users and providers would be the most efficient. Such a system could/would 'delink' access from benefit-sharing, and also delink use in particular commercialized products from benefit-sharing. The system should apply to both material genetic resources and to DSI under the same terms and conditions.

By far the simplest multilateral approach, as far as the generation, sharing and using of DSI is concerned, would be to have Contracting Parties undertake to make annual payments to a multilateral benefit-sharing fund based on a percentage of overall sales, within their jurisdiction, of classes of products that are generally developed through use, in some substantial way, of genetic resources and/or DSI. To work, it would be essential for negotiators to agree upon what classes of products, in which sectors, should attract such payment obligations, as well as the formula to be used for calculating amounts to be paid. (This has already been addressed, for example, under the Plant Treaty's multilateral system of benefit-sharing, where it was decided that sales of seed on the open market are what attract payment obligations; not any other uses further up, or further down, the research, development, use and commercialization chain.) Pursuant to this option, it could be left open to each contracting party if they would require the commercializing entities within their jurisdiction to 'pay them back'. This approach would not require any tracking or tracing of uses of genetic materials or DSI, would guarantee predictable flow of income to the benefit-sharing fund, and relieve commercial users of GR and DSI from having to make difficult decisions about whether or not to seek materials or DSI from a system that link payments to access and use of GRs and DSI. Since payments would be made by contracting parties, there would be no need to monitor if commercializing entities that should be declaring and making payments actually did so. No centralized system for monitoring and enforcement would be necessary. All companies would be on the same level playing field. The rate of payment would be set to reflect benefit-sharing with respect to both genetic resources and DSI. This approach is not explicitly included in the range of options considered in the Options paper or Analytical Framework.

A variation on this approach, would be to require payments to a centralized benefit-sharing directly from commercializing entities based on sales of designated classes of products (as reflected in Option 6 in the Analytical Framework). While this approach relieves the contracting party from being involved, it potentially carries more transaction costs as a result. Without contracting parties as intermediaries, all

commercializers would need to make declarations and payments to a centralized international authority; and the centralized international authority would need to monitor commercializing entities worldwide to ensure they are making required payments.

Another, considerably more complicated, multilateral approach would be to provide open access to DSI and genetic resources, but require users to agree to standardized monetary benefit-sharing terms when they access genetic resources or DSI from an online data base. Here again, one payment option could be that users agree to make payments, over a period of time to be agreed by contracting parties negotiating this new arrangement, on sales of prescribed classes of products commercialized by their own organization (regardless of whether or not they used access genetic resources or DSI in the development of the products of those classes that they sold). The down-side of this approach is it requires companies to make difficult decisions about whether or not to use the multilateral system, or continue to seek materials and or DSI from outside the system, risking slowing essential progress in order to avoid making payments. As a result, it would be impossible to ascertain how much income will actually flow to the benefit-sharing fund. And it would require all those entities that do access DSI materials through the related multilateral system to declare sales of relevant products and make payments to a centralized entity, and it would require investment in systems and transaction costs for entering into contractually binding undertakings by relevant users when they access materials and or DSI from online data bases, and for monitoring and enforcing payments. This option is very similar to 2.2 in the Analytical Framework, but expanded to include MAT for access to genetic resources and/or DSI rather than just DSI.

An additional possibility that could complement payments pursuant to multilateral options above would be to require payments from DSI service providers (for example, companies that sell sequencing equipment and reagents, or that charge for sequencing and analytical services). This option (which corresponds to 3.2a in the Analytical Framework) has the advantage of not affecting access to or use of genetic materials or DSI. Such approach would not be the first example of taxation on value creation on digital economy[11]. However, further investigation would be required to define its feasibility (voluntary micro levy, tax enforced by national legislations) and to assess its potential for monetary benefit-sharing based on market size.

4.1.2. The relationship between a multilateral system for DSI and sovereign rights of control of genetic resources and related DSI

One of the primary motivations, from the perspective of research and development organizations, to support multilateral approaches for the governance of DSI (and for genetic resources), is the guarantee of low-transaction-cost access to a wide range of data and genetic materials, and that over time, more and more data and materials will become available. However, for clarity, it is important to underscore that creation of a multilateral system for benefit sharing for DSI does not mean that all data everywhere will suddenly be made available upon request. Presumably privately held, confidentially managed, DSI (like their associated genetic resources) would only be made available through open access infrastructure

[11] Fair Taxation of the Digital Economy (https://taxation-customs.ec.europa.eu/fair-taxation-digital-economy_en)

subject to approval of the organizations, individuals, communities that manage or create them, subject to their existing rights of control (including potentially, national ABS laws that extend to DSI). By contrast, contracting parties would be expected to take measures to maintain open access status of data bases they support and maintain. And they would be expected to create incentives – along with the international benefit sharing fund -- for natural and legal persons (including companies, individuals, indigenous peoples, local communities) that manage genetic resources and generate related DSI to deposit that DSI on open access servers.

Indeed, this is how the Plant Treaty's multilateral system for access and benefit-sharing works. PGRFA of 64 explicitly listed crops and forages that is 'under the management and control of contracting parties and in the public domain' is considered to be automatically included upon a countries' ratification. However, pursuant to that same formula, PGRFA that are under management and control of natural and legal persons (including companies, individuals, indigenous peoples, etc) are not automatically included in the multilateral system. Contracting parties agree to take measures to encourage them to voluntarily make their materials available through the multilateral system. Furthermore, while some contracting parties are voluntarily including PGRFA of additional crops and forages, beyond the 64 explicitly listed, they are under no obligation to do so. During the 6 years of negotiations to enhance the functioning of the multilateral system from 2016-2019, contracting parties discussed the option of expanding the scope to *all P*GRFA. In that context, some contracting parties argued that with such a potentially open-ended expansion of scope of the multilateral system, they would want to have flexibility to exclude some culturally important, indigenous crops from the scope of the system. With suspension of the negotiations, these issues were not finally resolved, but they do provide some insight into the range of possibilities for developing formulas concerning what is in, and what is out, of a multilateral system of access and benefit-sharing.

4.1.3. The need to maintain harmony with the Plant Treaty

Any new benefit-sharing norms for DSI developed under the aegis of the CBD must take into account the fact that there are already other international access and benefit-sharing systems developed under the aegis of other UN bodies. The most important such agreement for CGIAR is the Plant Treaty. The Plant Treaty has already created a multilateral system of access and benefit-sharing for PGRFA. It is reasonable, and most efficient to take advantage of that multilateral system, building on it, make such adaptations as are necessary to govern benefit-sharing from use of DSI derived from PGRFA in the multilateral system.

The 9th session of the Governing Body of the Plant Treaty in September 2022 agreed to launch a process to enhance the functioning of the multilateral system of access and benefit-sharing. Among other things on the table for consideration during the course of those negotiations, are new obligations for monetary benefit-sharing from commercial uses of DSI.

The most logical and feasible approaches to enhancing the Plant Treaty's multilateral system, including integrating benefit sharing related to DSI, is very much in line with our stated preferences (above) concerning how benefit-sharing for DSI and genetic resources should be addressed under the CBD.

Again, it seems the best way to both accumulate regular payments to the benefit-sharing fund, and minimize transaction costs, would be to have contracting parties undertake to make annual payments to the Treaty's benefit-sharing fund based on seed sales within their jurisdictions, using a fixed royalty rate that reflects the value of access to, and use of, both PGRFA and DSI. Contracting parties could then, if they wished, recoup a portion, or all of the payments they made, from commercial users in their jurisdictions. Our 'fall back' position is the same as described above in the context of negotiations under the aegis of the CBD, i.e., to have commercializing users make the payments directly to the benefit-sharing fund, either as a direct consequence of their country ratifying the Plant Treaty members state, or because they agree to standardized, international, mutually agreed terms. The governing body could also consider additional modalities, such as payments from agencies who sequence/analyze DSI on a fee-for-service basis.

It should be noted that the international benefit-sharing fund disperses funds for priority areas of work, and for various groups of stakeholder/beneficiaries, as established by the Governing Body, through a process of competitive grants. This process appears to work very well. It provides an example for how funds collected through multilateral mechanisms under the CBD can also be allocated.

Ideally, the multilateral systems adopted under both the CBD and the Plant Treaty will be very similar, if not identical, thereby minimizing, or erasing, all uncertainties and transaction costs for conservers, users of genetic resources and DSI.

4.1.4. The desirability of a harmonized approach for accessing both DSI and genetic materials

The foregoing discussion about the relationship with the Plant Treaty brings us to a broader discussion about the relationship between new norms to regulate DSI, and *any* norms to regulate access and benefit -sharing for genetic materials. Under the Plant Treaty, because there is already a multilateral system for genetic materials, it seems likely/logical/most efficient that new norms for DSI would be integrated/build into that same multilateral system. Therefore, both PGRFA materials and related DSI would be regulated under harmonized conditions under a single system. It is hard to imagine it working any other way – i.e., complying with one system for accessing materials, and another system for accessing related DSI -- without creating enormous uncertainty, and creating extraordinary challenges/transaction costs for agricultural research activities such as we described in sections 2 and 3; activities that depend upon access to, and use of, an extraordinarily diverse range of *both* genetic materials and DSI. In the absence of an international legal instrument like the Plant Treaty establishing multilateral approaches for other GRFA (i.e., livestock, aquatic, soil GRs and crops, forages and trees that are not included in Annex 1), the way forward for creating harmonized conditions for both genetic materials and DSI is less clear. In this context, it is worth underscoring that the Nagoya Protocol Article 8(c) recognizes the contracting parties may consider the importance of genetic resources for food and agriculture and their special role for food security. Perhaps adoption of a multilateral system for DSI would be an occasion/stimulus for contracting parties to consider harmonized multilateral approaches for regulating access to GRFA at the same time.

4.1.5. Regarding hybrid options

There has been a lot of discussion of possible hybrid approaches to access and benefit-sharing for DSI. Here we discuss two models or archetypes.

De facto hybridization, because a new multilateral system would not be all inclusive: The discussion above concerning 'the relationship between a multilateral system for DSI and sovereign rights of control of genetic resources and related DSI' underscored that a multilateral approach to access and benefit-sharing for DSI would almost certainly not be expected to include all data from all potential providers. By corollary, a multilateral system would leave space for a range of actors (e.g., IPLCs, companies) to exert various forms of control over DSI that is not included in the multilateral system, either because it is controlled, owned, managed by natural and legal persons whose data would not be affected by creation of a multilateral system, or because it was derived from species or genera that are not included in the multilateral system. Perhaps more to the point, many actors already have the right, pursuant to existing national ABS laws, to deny access to genetic resources from which DSI could be derived in the future. While these controls may be of limited utility in requiring benefit-sharing from DSI that is 'already out there', or derived from GRs that are already widely disbursed, they should be effective going forward, when the actors concerned are approached by agencies seeking to access genetic resources or DSI, including instances where the DSI concerned would be destined for inclusion in the multilateral system by the access seeker. Parties with a recognised right of control over the genetic resource or data could refuse, or agree to provide access to it only upon mutually agreed terms.

Hybridization as a component of a larger harmonized system including both multilateral and bilateral elements. Presumably, here too, not all DSI would be included, *ab initio*, in a newly minted DSI multilateral system, as discussed immediately above. Here, 'hybrid' refers to different possible approaches to benefit-sharing derived from the use of DSI that is already included, or available through, an open access/multilaterally oriented digital infrastructure. The benefit-sharing could be achieved through either a) multilateral means in some cases (i.e., payments to an international fund) or b) bilaterally in other cases (subject to MATs with the providers of the data that was accessed from open access/multilaterally organized digital infrastructure). The bilateral rules would be triggered when DSI from only a few associated genetic resources of known country of origin are used in the development of a commercialized product. Clearly more consideration is necessary to see if such a system is technically feasible. However, we note there that this approach does not appear to respond to the three ways DSI is changing the way genetic resources can be used to create commercial products through agricultural research and development, and breeding in particular (as noted above): i) that DSI from a wide range of genetic resources is used in developing tools in early stages of research that are incorporated in final products; ii) that DSI-based tools and technologies enable in vitro and in silico replication of DNA sequences followed by horizontal gene transfer to reintroduce them into a material products and iii) DSI enable developers to focus on single genes, where a single specific gene sequence can be found in numerous different varieties from numerous different "countries of origin". Users could therefore avoid bilaterally oriented

benefit-sharing obligations by accessing the same sequence from other materials available from multilateral or unregulated sources.

Another option for 'building in' some form of bilateral benefit-sharing, would be to set aside a proportion of funds from a multilateral benefit-sharing fund to be allocated to countries which are the source of materials whose associated DSI is available. In this context it also seems just and equitable to set aside a significant proportion of the multilateral funds for indigenous and local peoples, including farmers.

4.2. Non-monetary benefit-sharing

Regardless of the monetary benefit-sharing formula that is eventually adopted, a revised, enhanced multilateral system should include considerably more focus on promoting non-monetary benefit-sharing, including capacity building to close capacity gaps among nations and regions.

In the report of the Informal Co-Chairs Advisory Group (IAG)[12] CGIAR notes that policy option 4 (enhanced technological and scientific collaboration and capacity-building) was the only option on which there was consensus within the group to recommend further consideration, and was widely assessed as scoring highly on most of the criteria for an appropriate ABS regime for DSI. We appreciate that about 40% of IAG members noted it should be considered a complementary component to monetary benefit-sharing. CGIAR is ideally positioned to help build relevant technological capacity in developing countries.

Promoting the accrual of non-monetary benefits (in the lexicon of the Plant Treaty and Nagoya Protocol) to developing countries through the conservation and sustainable use of genetic resources has always been, and remains, at the very core of CGIAR's mission, independently of specific ABS agreements setting out CGIAR's legal obligations to providers. CGIAR follows a hybrid bilateral and multilateral approach: it develops bilateral partnerships tailored to the needs of individual developing countries while transferring materials knowledge and skills multilaterally across the full spectrum of partners from all developing countries. The non-monetary benefits generated and shared by the CGIAR Centers and Initiatives include all the modalities recognized by the Plant Treaty and most of those recognized by the Nagoya Protocol.

Impact studies indicate that developing countries have benefited economically by billions of dollars per year from the CGIAR. DSI, while not changing the principle of non-monetary benefit-sharing by CGIAR, enables CGIAR to bring these benefits to developing countries more efficiently and effectively.

4.2.1. Examples of non-monetary benefit-sharing in the CGIAR

Here we give some representative examples of non-monetary benefit-sharing in relation to DSI mediated by CGIAR. The examples are cross-referenced to modalities for sharing benefits recognised by the Plant Treaty (Article 13 paragraphs 1 and 2.a)-2.c) and by the Nagoya Protocol (Annex, paragraph 2 clauses (a)-(q); https://www.cbd.int/abs/text/articles/?sec=abs-37)

[12] https://www.cbd.int/doc/c/c064/37f6/d5024789093ef19bf5f84519/wg2020-05-03-en.pdf

Non-monetary benefit-sharing example 1: Access to genotyping services

Benefit-sharing modalities: Plant Treaty 2.a)-2.c); Nagoya (a),(b),(d),(e),(g),(h),(j),(k),(m)

In order to facilitate low cost and high throughput genotyping for CGIAR Centers and partner institutes including national agricultural research organizations, the High Throughput Genotyping Project (HTPG) was established with a primary objective, to broker access to a shared SIngle plex SNP genotyping platform at an accessible cost ($1.50 to $2.00 per sample for DNA extraction and 10 SNP markers). Under subsequent evolution with the Excellence in Breeding Platform and Breeding Resources Initiative of the CGIAR more than 40 public institutions in the global south have access to the platform, which has persisted beyond the expiration of the original funding cycle. The Low-Density Genotyping platform offers SNPs genotyping services in rice, wheat, maize, millets, legumes, clonally propagates species and other agriculturally important commodities, which are readily deployable in the breeding programmes.

Non-monetary benefit-sharing example 2: Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub

Benefit-sharing modalities: Plant Treaty 2.a)-2.c); Nagoya (a),(b),(d),(e),(g),(h),(j),(k),(m)

The hub is a shared agricultural research and biosciences platform located at ILRI in Kenya. It provides world class laboratories to African and international scientists conducting research on African agricultural challenges. It was established as part of the New Partnership for Africa's Development (NEPAD). The Hub mobilizes biosciences for Africa's development; it enables research, capacity building and product incubation, conducted by scientists in Africa and for Africa, and empowers African institutions to harness innovations for regional impact.

The BecA-ILRI has various technology platforms. The Genomics Platform includes high output sequencing machines that provide cutting-edge genome sequencing and analysis technology that can be used in pathogen detection, diagnostics, and risk assessment, gene expression analysis, metagenomics, and to determine genetic diversity in crops and animals. The SegoliP platform includes Sequencing, Genotyping, and Oligonucleotide Services (SEGOLIP), and are available to Hub users and other scientists and institutions located anywhere in the world.

Capacity development includes: Annual hands-on training workshops in key skill areas; research placements enabling early career scientists to spend up to six months at the Hub; and institutional support to key laboratories and individuals in the region. Over 1000 scientists from more than 20 countries benefited from short skill-enhancement training courses between 2015 and 2017. As part of network building, the hub has established 13 communities of practice, linking African scientists of different disciplines with similar interests, some of them related to genomics.

Non-monetary benefit-sharing example 3: ICRISAT's Center of Excellence in Genomic and Systemic Biology

Benefit-sharing modalities: Plant Treaty 2.c); Nagoya (d),(g),(h),(j)

The Center offers training and short courses to provide scientists/researchers updated information on application of novel genomics technology in crop research and breeding. These courses are specifically designed to address the requirements of users of the Center's high-throughput Genotyping Service Lab. This includes: experimental design, sample submission and data analysis; genomics applications covering from molecular markers and marker-based diversity assessment to gene/QTL/trait mapping and marker-assisted breeding; construction of genetic linkage maps; QTL mapping based on purpose-created populations; and association (LD) mapping using breeding lines or germplasm accessions are covered, as is use of decision support systems for plant breeding etc. Courses also provide an introduction to cluster analysis and construction of dendrograms, diversity analysis software and next generation sequencing and bioinformatics tools.

The Center has organized 15 training courses so far and has trained a total of 485 scientists (including 117 female scientists) coming from national agricultural research centers, universities and private entities in Asia and sub-Saharan Africa.

Non-monetary benefit-sharing example 4: Using DSI to build 'data bridges' among genebanks

Benefit-sharing modalities: Plant Treaty 2.a)-c); Nagoya (a),(b),(d),(h),(j),(k),(n)

Sharing germplasm between genebanks located in different countries is not straightforward due to phytosanitary considerations. Sharing DSI data, however, is easy from a technical point of view. DSI data generated with genotyping platforms that minimize ascertainment bias, shed light on the true genetic relationships among accessions. Provided a single genotyping platform is used for generating DSI data, gaps and genetically redundant accessions can be identified across multiple germplasm collections conserved by different genebanks. Such cross-genebank analyses would enhance the efficiency of the Global System by identifying/reducing the redundancy among genebank collections and encouraging decisions to collect new germplasm to be taken based on the *total* diversity conserved across multiple genebanks.

To initiate work towards this goal, CIMMYT and the Alliance of Bioversity & CIAT, both acting as 'hubs' for the DivSeek Initiative, co-organized a meeting of Latin American genebanks in November 2022. The meeting was sponsored by the CGIAR Genebank Initiative and followed three preceding remote meetings on DSI-related aspects. It included the genotyping of accessions from maize, bean, potato, cassava, and other germplasm collections managed by workshop participants using the DArTseq platform of CIMMYT's SAGA genotyping service. Workshop participants familiarized themselves with the genotyping platform, the data generated, and the algorithms and software available for analyzing and interpreting high-density molecular data. Once all genotyping data becomes available, follow-up remote workshops will be organized to co-analyze, for each crop, data from multiple collections as an initial step to compare the diversity conserved at national genebanks in the region. Based on these initial results, additional funds may be sought for more comprehensive cross-genebank diversity studies. The workshop, which also included capacity building on gap-analysis algorithms, has created a foundation for an active community of practice of Latin American genebanks, which is already planning to jointly address other genebank-related topics beyond DSI in the future.

Non-monetary benefit-sharing example 5: national genebank of Tunisia

Benefit-sharing modalities: Plant Treaty 2.a)-c); Nagoya (a),(b),(d),(h),(j),(k),(m),(n)

ICARDA and national Genebank of Tunisia have done a study to characterize for the first time the genetic diversity within and among the Tunisian durum wheat collection, look at potential mis-classification by linking off types of landraces to local modern durum wheat varieties and/or other landraces, and compare the Tunisian durum wheat landraces to landraces from the countries around the Mediterranean basin and West Asia region using high throughput DArTseq™ technology (Robbana et al. 2019). The study has helped the Tunisia Genebank ensuring efficient conservation and management of genetic resources identifying miss classifications that could be due mainly to misnaming of the landraces during the collecting missions and to possible mixtures in the landraces. The genebank in Tunisia will be using outcomes from this study to conserve the bulk seeds of each landrace instead of seeds of many individual lines constituting each landrace. This will reduce the cost of conservation and avoid conserving several copies of the same line. The study was also an opportunity for the Tunisia Genebank staff to get training on curating and using DSI in conservation and use of genetic resources. A similar study on Barley is undertaken between ICARDA Genebank and the Genebank in Morocco to identify duplicates and also unique barely accessions missing between the two genebanks.

Non-monetary benefit-sharing example 6: DSI data hosting, curation and accessibility

Benefit-sharing modalities: Plant Treaty 2.a)-c); Nagoya (a),(b),(d),(h),(j),(k),(n)

As indicated above, DSI only reaches its full usefulness when there is (A) a critical mass of data available, and (B) this data is curated, digested and presented in an accessible format to act as input to higher-level analyses. Public repositories serve several critical roles in this process. Firstly, they allow cumulative aggregation of data submitted from multiple researchers, in multiple countries – allowing the formation of a critical mass of data. Secondly, repositories typically go beyond simply storing raw data; in most cases they include value-added functionality related to various types of data querying, aggregation and summarizing, and higher-level analysis tools.

CGIAR, through its Open Access and Data Management Policy[13], is committed to the widespread dissemination of data and knowledge, including DSI. It is globally recognized that open access is a key concept in biology that drives research and innovation and foster international collaborations. When publishing scientific evidence related to our mission, CGIAR scientists are DSI providers and submit raw DSI datasets in INSDC databases (NCBI, EMBL-EBI and DDBJ) as expected by academic publishers and in line with current good practices in the scientific community. However, generating new data by itself is

[13] https://www.cgiar.org/how-we-work/accountability/open-access/faq/ https://www.cgiar.org/how-we-work/accountability/open-access/faq/

insufficient if it cannot be effectively disseminated, queried, visualized, and analyzed. To facilitate access and use of the data, in particular for countries that are not the bioinformatic capacities to handle such large datasets, CGIAR Centers develop and host repositories for DSI related to genebank and/or breeding material.

Data volumes, data backup, site design and maintenance for these repositories are massive challenges, and extremely expensive. They are well beyond the means of individual researchers or labs even in developed countries, let alone the global south. Yet by making these resources freely accessible, researchers in all countries, including the global south, can leverage these benefits without incurring the costs of their design and maintenance.

Examples of such repositories include:

The 3000 Rice genomes

IRRI has developed the Rice SNP-seek database, a database that allows quick retrieving of SNP alleles for all varieties in a given genome region, finding different alleles from predefined varieties and querying basic passport and morphological phenotypic information about sequenced rice lines (Alexandrov et al, 2015)

Digital Genebanks

The Alliance in collaboration with 30 national partners has implemented a digital ecosystem via MGIS (Ruas et al, 2013) to enable access to DSI-related banana ex situ genetic resources worldwide and for 30% of the International genebank collection. Moreover, with the launch of the Future Seeds, a global innovation hub for the conservation and use of crop diversity In Colombia, the ambition is to build a 'knowledge bank' through DSI, digital phenotyping, and information technologies for Cassava, beans, forages, and other crops[14].

Seeds of discovery (SeeD)

CIMMYT with MasAgro launched the Seeds of discovery project with the objective to comprehensively study maize genetic diversity by obtaining, processing, and analyzing the world's largest genotypic data set to help scientists identify new genes of interest for maize breeding programs. More than 2 billion genotypic data and more than 870,000 phenotypic data of maize field trails have been processed and uploaded to SeeD's database and repository making them available to the scientific community via the project website. DSI datasets are made accessible, through authenticated users, via the Germinate data repository developed at the JHI (Raubach et al, 2020).

[14] https://alliancebioversityciat.org/fr/node/18239

Breeding informatics

Breeding programs are key components of the CGIAR. Over the past years an ambitious modernization process has been proposed, that includes development and adoption of modern breeding informatics solutions. Such systems are designed DSI information useful for trait discovery and support genomic selection. One of them, Breedbase, co-developed between the Boyce Thompson institute, IITA, CIP, the Alliance and national partners in Africa. Cassavabase, the flagship Breedbase database, has accumulated an immense amount of cassava breeding data, consisting of information on more than 500,000 cassava accessions, characterized in over 4,000 trials, and nearly 35,000 genotyping experiments (Morales et al, 2022).

Without these repositories, researchers would be limited to just that data they themselves have generated (giving an inherent advantage to large, well-funded organisations) and/or those with the resources to aggregate and analyse the data at scale to extract meaningful information. Public repositories thus disproportionately benefit smaller and less-resourced actors.

Non-monetary benefit-sharing example 7: Improved varieties

Benefit-sharing modalities: Plant Treaty 1; Nagoya (a)-(d),(h),(k),(l),(m),(n),(o)

As shown above, DSI significantly enhances the breeding process. This work is resulting in improved rates of genetic gains in target environments of countries in the global south, resulting in superior cultivars developed in collaboration with, and publicly available in, those target countries. This has always been a core purpose of the CGIAR, and active involvement in generating, storing and leveraging DSI at CGIAR Centers is accelerating delivery of these new varieties, accelerating delivery of key benefits to developing countries.

4.2.2. Promoting capacity building and technology transfer as the component of a new DSI-related benefit-sharing norm

There is general agreement that capacity building, technology transfer and other forms of non-monetary benefit-sharing must be an important part of a new international system for the sharing of benefits arising from the use of DSI. However, there is not a concrete proposal on how such a system would look like in relation to non-monetary benefit-sharing.

Funds collected through a multilateral monetary benefit-sharing mechanism should be used to facilitate capacity building, technology transfer and information sharing for the benefit of research and development organizations, and indigenous and local peoples (including farmers) in developing countries.

However, the international benefit-sharing fund should not be the only source of support for such capacity building and technology transfer. Ideally, a multilateral system for DSI should have mechanisms for technology transfer and capacity building that operate on their own, independent from the availability of funds raised through monetary benefit-sharing mechanisms.

Another option (which could be combined with the previous one) would be that countries commit to put in place mechanisms to incentivize or oblige users of DSI (especially those in developed countries) to transfer capacities, technologies, information and other non-monetary benefits to stakeholders in developing countries. The CBD and the Plant Treaty already include this commitment in relation to technologies and capacities that contribute to the conservation and sustainable use of biodiversity (CBD) and PGRFA (Plant Treaty). However, advances in the implementation of these obligations have been modest; so, negotiators could consider adopting more concrete commitments this time. For example, countries could commit to dedicate an agreed percentage of international aid funds to initiatives that focus on increasing DSI capacities in developing countries' organizations, and they could disburse these funds only to research and development organizations that show long term partnerships with organizations in developing countries. Another example could be the requirement to collaborate with research and development organizations in developing countries as a pre-condition to obtain public funds for DSI related research from national or regional funding agencies

4.2.3. The need for indicators for monitoring non-monetary benefit-sharing

No matter what option for non-monetary benefit-sharing is adopted for DSI, an important step in its implementation will be the development of indicators for monitoring non-monetary benefit-sharing. Contracting parties and stakeholders (like CGIAR) can use those indicators for periodic reporting.

Indicators would have to be developed that reflect the reality that, in many cases, non-monetary benefit -sharing cannot be linked back to (and are not conditions required by) specific ABS agreements. Indeed, a lot of the valuable work to add value to GRFA (characterization, evaluation, even genetic manipulation) is conducted for the very purpose of making those resources more attractive to potential users/access seekers. Consider for example, the efforts/contributions of a CGIAR Center to manage and add value to an internationally available collection of crop genetic resources. The CGIAR-hosted international collections comprise more than 700,000 accessions of more than 20 crop species. Without information concerning each of those accessions, users would not know which to select and request. The collections would essentially be useless. Consequently, Centers spend millions of dollars annually generating accession level information about those collections. This includes passport information (which is often incomplete and needs time and effort to bring up to internationally agreed levels) phenotypic and DSI-based information (of the sort described in part 1 above), generating additional layers of information concerning climatic, soil and other environmental conditions that prevailed when/where the material was collected, evaluation of performance of accessions under various biotic and abiotic stresses, etc. All this information is put online, in publicly available accession level databases, for potential users to search

through, to identify which materials are useful for their needs. The Centers also spend millions of dollars annually ensuring that the materials they conserve and make available are free of quarantine pests and diseases. Indeed, one of the services that some CGIAR Centers provide is to receive samples of diseased clonally propagated crops, clean them, and send them back, disease free, to the original recipients (and others). Another bypasses phytosanitary concerns by providing access to samples of DNA extracted and purified from accessions. All of these services, value additions – non-monetary benefits – are generated and shared before anyone seeks access to those materials. And of course, these forms of non-monetary benefits are not generated and shared by CGIAR Centers alone: these efforts are undertaken by regional and national and even community genebanks around the world. All of which should be appreciated and reflected (and counted) with tailored non-monetary benefit-sharing indicators. Some indicators could include the following:

- numbers of accessions in internationally available collections;
- numbers of accessions characterized (with the possibility of specifying the kind of characterization, e.g., phenotypic, molecular, etc; numbers of accessions tested for/cleaned of quarantine pests and diseases);
- numbers of accessions evaluated for climatic or nutritional traits;
- Number of diverse sources identified for a given trait;
- Number of gene mapped in a given crop for diverse traits and made available for use through MAS.

In other cases, it would be (theoretically at least) possible to track the generation and sharing of non-monetary benefits with previously accessed genetic resources (and DSI) under ABS agreements. CGIAR genebanks and breeding programs send over 50,000 samples annually under Standard Material Transfer Agreements (SMTAs) of the Plant Treaty to users around the world, primarily public sector agricultural research organizations in the global south. The Centers' genebanks also transfer approximately 30,000 samples to their own Centers' breeding programs under the terms of the SMTA. In both cases (distributions within and outside the CGIAR Centers) the transferred genetic resources and related information are often used to develop improved breeding lines or selections, or identify genetic traits. The CGIAR Centers share these improved materials/non-monetary benefits freely with public research organizations all around the world.

It is a relatively easy matter for CGIAR breeders to keep track of improved lines, traits and varieties they have developed using PGRFA received under the SMTA. CGIAR would be happy to provide such statistics in the future if indicators and a reporting framework are established. However, the situation is very different with respect to organisations outside CGIAR to whom CGIAR transfers materials under the SMTA, and who exchange materials among themselves. There is no obligation on the part of recipients of materials under the SMTA to share such information bilaterally with the provider; and providers of materials under the SMTA undertake that the material is provided without an obligation to track the

received materials. On the other hand, recipients are obliged to share all non-confidential information that results from research and development carried out on the material received multilaterally, sharing it through the Global Information System (GLIS) of the Plant Treaty. While the SMTA does not specify details of how the information is to be shared through GLIS, the only universal mechanism currently available for doing so relies on identifying genetic resources using Digital Object Identifiers (DOIs) assigned by the Secretariat of the Plant Treaty[15]. To date over 1.3 million GLIS DOIs have been generated[16]. These GLIS DOIs then become the basis for sharing information and can be used to access information from any data sources that can resolve DOIs, ranging from Genesys[17] for accessions in *ex situ* collections to DSI databases and scientific publications. Without imposing new reporting requirements, which would involve new obligations to track and trace uses of materials, the only way to aggregate data from around the world concerning this form of non-monetary benefit-sharing is through the GLIS for genetic resources that have a GLIS DOI. The potential exists to build on GLIS technology to establish standards for sharing DSI that would support emerging ABS regimes. One alternative option would be for countries to include such information in their periodic reports to the UN FAO concerning the state of implementation of the Global Plans of Action for PGRFA, and this information could be then relied upon as part of global monitoring of the Post 2020 GBF. To do so, national focal points would need to interrogate research and development organisations to get at least minimal information confirming whether the materials they have developed and make publicly available are derived from materials accessed under some form of ABS agreements (including the SMTA)

CGIAR has recently developed a results dashboard, which includes aggregate information about germplasm distribution, partnerships, capacity building, [etc] from across the entire CGIAR Centers, departments and research and development initiatives. The dashboard currently does not include a filter for activities, partnerships that involve enhanced uses of genetic resources and DSI, but it could in the future, to be in line with indicators, and monitoring and reporting frameworks such a filter could eventually be introduced.

5. Conclusions

DSI is increasingly central to CGIAR Centers' agrobiodiversity conservation, research and plant and animal breeding. CGIAR appreciates that recent technological advances in genome sequencing, gene synthesis and gene editing facilitate commercial exploitation of genetic resources without triggering monetary benefit-sharing conditions under the Plant Treaty and the Nagoya Protocol. As such, we appreciate that the international access and benefit-sharing policy framework needs to be updated. At the same time, the nature of DSI, including its easy reproducibility and global distribution require more than simply 'doubling down' on previous approaches to access and benefit-sharing. The advent of DSI has brought

[15] https://ssl.fao.org/glis/
[16] https://ssl.fao.org/glis/site/doiindex
[17] https://www.genesys-pgr.org/

the international research and policy making communities to a turning point, where we need to consider new approaches to access and benefit-sharing: approaches that both support the further expansion of open access architecture for agricultural research and development, and simultaneously ensure more monetary and non-monetary benefit-sharing.

Given the nature of DSI and the underlying genetic resources from which they are derived, and our own experiences as scientific research organizations, we agree with the majority of the IAG members that multilateral approaches, delinking access and benefit-sharing, are the most appropriate, to increase monetary benefit-sharing without interfering with the international open access research and development architecture. In addition, we urge the international community under the aegis of the CBD (and the Plant Treaty) to develop the broadest, flattest, most inclusive multilateral system possible for DSI. Access and benefit-sharing systems under both the Plant Treaty and Nagoya Protocol have insisted upon, and been structured around, the idea that benefit-sharing payments should come directly from commercial users. DSI has pushed challenges associated with this model to the limit, with the possibility of requiring tracking, tracing, reporting and monitoring systems that are entirely unsustainable. While we appreciate that it will not be popular among many delegations, in the interest of minimizing administrative burdens for conservers, providers, and users of DSI (and underlying GS) we suggest that contracting parties assume the initial role/responsibility for making requisite payments to a centralized, multilateral benefit-sharing fund, following whatever formula is agreed by the CBD/COP, and thereafter, collect payments from constituent users within their own borders ensuring the benefit reaches the Treaty, technology developing institute and the country whose genetic resources were used to develop the technology in a format agreeable to all three.

Equally importantly, new access and benefit-sharing norms should also support enhanced non-monetary benefit sharing, including, increased investments in capacity building to close north-south capacity gaps to generate, use and benefit from DSI. CGIAR already plays a significant sui generis role in technology transfer, training and capacity building for use of DSI in agricultural research and development in developing countries. The contributions of CGIAR and other organizations that play similar roles could be monitored and encouraged as non-monetary benefit-sharing under a new regime.

Very importantly, new norms for DSI under the aegis of the CBD must respect, and be 'in sync' with, approaches for access and benefit-sharing for both PGRFA and DSI under the Plant Treaty. Because the Plant Treaty has already created a multilateral system of access and benefit-sharing for plant genetic materials, it should be possible to develop harmonized conditions for associated DSI under the same legal framework. It is to be hoped that those conditions, and what is adopted under the CBD, will be very similar (if not identical), thereby reducing potential uncertainties and inefficiencies for users who are governed by measures implementing the CBD/Nagoya Protocol, and the Plant Treaty.

6. Literature cited

Allan, V., Vetriventhan, M., Senthil, R., Geetha, S., Deshpande, S., Rathore, A., Kumar, V., Singh, P., Reddymalla, S. and Azevedo, VCR. (2020) Genome-Wide DArTSeq Genotyping and Phenotypic Based Assessment of Within and Among Accessions Diversity and Effective Sample Size in the Diverse Sorghum, Pearl Millet, and Pigeonpea Landraces. Front. Plant Sci. 11:587426. doi: 10.3389/fpls.2020.587426

Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815 (2000). https://doi.org/10.1038/35048692

Arbalaez, JD., Dwiyanti, MS., Tandayu, E., Llantada, K., Jarana, A., Ignacio, JC., Platten, JD., Cobb, J., Rutkoski, JE., Thomson, MJ., Kretzschmar, T. 2019. 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. Rice 12, 55. https://doi.org/10.1186/s12284-019-0311-0

Awata, LAO., Ifie, BE., Danquah, E., Jumbo, MB., Suresh, LM., Gowda, M., Marchelo-Dragga, PW., Olsen, MS., Shorinola, O., Yao, NK., Boddupalli, PM and Tongoona, PB (2021) Introgression of Maize Lethal Necrosis Resistance Quantitative Trait Loci Into Susceptible Maize Populations and Validation of the Resistance Under Field Conditions in Naivasha, Kenya. *Front. Plant Sci.* 12:649308. doi: 10.3389/fpls.2021.649308

Barmukh, R. et al., 2022. Genetic variation in CaTIFY4b contributes to drought adaptation in chickpea. Plant Biotechnology Journal, 20: 1701–1715

Beyene, Y., Gowda, M., Olsen, M., Robbins, KR., Pérez-Rodríguez, P., Alvarado, G., Dreher, K., Gao, SY., Mugo, S., Prasanna, BM., and Crossa, J. (2019) Empirical Comparison of Tropical Maize Hybrids Selected Through Genomic and Phenotypic Selections. Front. Plant Sci. 10:1–11. doi:10.3389/fpls.2019.01502

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B. (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci., 55 (2015), pp. 154-163

CBD Secretariat (2020a) Digital Sequence Information on genetic resources: concept, scope and current use. CBD/DSI/AHTEG/2020/1/3. https://www.cbd.int/doc/c/fef9/2f90/70f037ccc5da885dfb293e88/dsi-ahteg-2020-01-03-en.pdf

CBD Secretariat (2020b) Global Biodiversity Outlook 5. Montreal. http://www.cbd.int/GBO5

Das, RR., M.T. Vinayan, B. ManishPatel, K. Ramesh Phagna, S.B. Singh, J.P. Shahi, Akashi Sarma, N.S. Barua, Raman Babu, K. Seetharam, A. Juan Burgueño, P.H. Zaidi (2021) Genetic gains with rapid-cycle genomic selection for combined drought and waterlogging tolerance in tropical maize (Zea mays L.) Plant Genome, 13, pp. 1-15.

Gao, L., Gonda, I., Sun, H. et al. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51, 1044–1051 (2019). https://doi.org/10.1038/s41588-019-0410-2

Halewood, M., Lopez Noriega, I., Ellis, D., Roa, C., Rouard, M., Sackville Hamilton, R., Morell, M., Bänziger, M., Duveiller, E., Upadhyaya, H.D., Pixley, K., Gresl, T., Ndjiondjop, M-N., Abberton, M., Wenzl, P., Muchugi, A., Perset, E., Hendre, P., Yazbek, M., Campos, H., Fadda, C., Paliwal, R. (2017) Potential implications of the use of digital sequence information on genetic resources for the three objectives of the Convention on Biological Diversity. A submission from CGIAR to the Secretary of the Convention on Biological Diversity. https://www.cbd.int/abs/DSI-views/CGIAR-DSI-en.pdf

Hawtin, G., Wood, M., Vermeulen, S. et al. 2020. CGIAR-Crop Trust System Level Review of Genebank Costs and Operations (GCO): Report October 2020. https://www.genebanks.org/wp-content/uploads/2020/11/GCO-Report_261020.pdf

Liu, J., Huang, S., Sun, M. et al. 2012. An improved allele-specific PCR primer design method for SNP marker analysis and its application. Plant Methods 8, 34. https://doi.org/10.1186/1746-4811-8-34

Mansueto, L., Fuentes, R.R., Borja, F.N., et al. (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucl Acids Res 45:D1075-D1081. DOI: 10.1093/nar/gkw1135.

Mallu, TS., Irafasha, G., Mutinda, S., Owuor, E., Githiri, SM., Odeny, DA., Runo, S. (2022) Mechanisms of pre-attachment Striga resistance in sorghum through genome-wide association studies. Mol Genet Genomics. 297:751-762. doi: 10.1007/s00438-022-01882-6.

Miko, I. (2008) Epistasis: Gene interaction and phenotype effects. *Nature Education* 1:197

Morales, N., Ogbonna, AC., Ellerbrock, BJ. et al. Breedbase: a digital ecosystem for modern plant breeding. G3 Genes|Genomes|Genetics 2022; doi: jkac078.

Murithi, A., Olsen, MS., Kwemoi, DB., Veronica, O., Ertiro, BT., Suresh, L. M., Beyene, Y., Das, B., Prasanna, BM., and Gowda, M. (2021) Discovery and Validation of a Recessively Inherited Major-Effect QTL Conferring Resistance to Maize Lethal Necrosis (MLN) Disease. *Front. Genet.* 12:767883. doi: 10.3389/fgene.2021.767883

Ntui, VO., Tripathi, JN., Tripathi, L. (2020) Robust CRISPR/Cas9 mediated genome editing tool for banana and plantain (*Musa* spp.), Current Plant Biology, 21, 100128, https://doi.org/10.1016/j.cpb.2019.100128.

Platten, JD., Cobb, JN., Zantua, RE. 2019. Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. PLoS One 14: e0210529. https://doi.org/10.1371/journal.pone.0210529

Robbana, C., Kehel, Z., Ben Naceur, M., Sansaloni, C., Bassi, F., Amri, A. Genome-Wide Genetic Diversity and Population Structure of Tunisian Durum Wheat Landraces Based on DArTseq Technology. *Int. J. Mol.*

Sci. **2019**, *20*, 1352.

Raubach, S., Kilian, B., Dreher, K. et al. From bits to bites: advancement of the germinate platform to support genetic resources collections and pre-breeding informatics for crop wild relatives. Crop Science 2020. doi:10.1002/csc2.20248.

Rouard, M., Sardos, J., Sempéré, G., Breton, C., Guignon, V., Van den Houwe, I., Carpentier, S., Roux, N. (2021) A digital catalog of high-density markers for banana germplasm collections. Plants People Planet, Online first paper (01 March 2021) 7 p. ISSN: 2572-2611

Ruas, M., Guignon, V., Sempere, G. et al. MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data. Database (Oxford) 2017; 2017. doi:10.1093/database/bax046.

Sackville Hamilton, R. (2020) The 30-Year Vision: The Future of CGIAR Genebanks. In "Crop Diversity for Challenging Times: the Role of Genebanks in Sustainable Development. Chatham House Dialogue", 21-23 September 2020. https://www.genebanks.org/news-activities/news/30-year-vision/

Salse, J., Piégu, B., Cooke, R., Delseny, M. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. Nucleic Acids Res. 2002 Jun 1;30(11):2316-28. doi: 10.1093/nar/30.11.2316.

Sardos, J., Paofa, J., Janssens, S., Sachter-Smith, G., Rauka, G., Roux, N. (2017) Banana collecting mission in the Autonomous Region of Bougainville (AROB), Papua New Guinea. [n.p.] 26 p.

Sardos, J., Breton, C., Perrier, X., Van den Houwe, I., Carpentier, S., Paofa, J., Rouard, M., Roux, N. (2022) Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. Frontiers in Plant Science 13:969220 ISSN: 1664-462X

Shultz, J.L., Ray, J.D. & Lightfoot, D.A. A sequence based synteny map between soybean and Arabidopsis thaliana. *BMC Genomics* 8, 8 (2007). https://doi.org/10.1186/1471-2164-8-8

Singh, S., Mahato, AK., Jayaswal, PK., Singh, N., Dheer, M., Goel, P., Raje, RS., Yasin, JK., Sreevathsa, R., Rai, V., Gaikwad, K., Singh, NK. (2020) A 62K genic-SNP chip array for genetic studies and breeding applications in pigeonpea (*Cajanus cajan* L. Millsp.). *Sci Rep*. 18;10(1):4960. doi: 10.1038/s41598-020-61889-0

Takahashi, R. and Yasuda, S. (1971) Genetics of earliness and growth habit in barley. In: R.A. Nilan (Ed.), Barley Genetics II. Proc 2nd International Barley Genetics Symposium, Washington State University Press, pp. 388–408.

Varshney, RK et al. (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat. Biotechnol. 31: 240–246

Varshney, RK, et al (2021) A chickpea genetic variation map based on the sequencing of 3,366 genomes. Nature 599: 622–627

Wang, W., Mauleon, R., Hu, Z. *et al.* (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, 557, 43-49. https://doi.org/10.1038/s41586-018-0063-9.

Willis, P.R. (2016) DNA as information. *Philosophical Transactions of the Royal Society A*, 374. https://doi.org/10.1098/rsta.2015.0417

Zhang, X., P. Pérez-Rodríguez, J. Burgueño, Olsen, M., Buckler, E., Atlin, G., Prasanna, BM., Vargas, M., Vicente, FS., Crossa, J. (2017) Rapid cycling genomic selection in a multi-parental tropical maize population G3-Genes Genomes Genet., 7 (2017), pp. 2315-2326

Zhou, Y., Chebotarov, D., Kudrna, D. *et al*. (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. *Scientific data*, 7, 113. https://doi.org/10.1038/s41597-020-0438-2

[i]The following people contributed to this discussion paper: Ruaraidh Sackville Hamilton (consultant, Genebank Initiative), Michael Halewood, Isabel Lopez Noriega and Mathieu Rouard (Alliance of Bioversity International and CIAT), Christian Tiambo (ILRI), Vania Azevedo (CIP), Ranjana Bhattacharjee (IITA), Wilmer Cuellar (Alliance of Bioversity International & CIAT), Andreas Gisel (IITA), Sarah Jane Hearne (CIMMYT), Bettina Heider (CIP), Prasad Hendre (ICRAF), Jan Kreuz (CIP), Alice Muchugi (ILRI), Okeyo Mwai (ILRI), Marie-Noelle Ndjiondjop (AfricaRice), Damaris Odeny (ICRISAT), Ekta Patel (ILRI), Elise Perset (CGIAR System Organization), John Platten (IRRI), Venuprasad Ramaiah (IRRI), Ana Carolina Roa (CGIAR System Organization), Kuldeep Singh (ICRISAT), Peter Wenzl (Alliance of Bioversity International & CIAT), Sonja Vermeulen (CGIAR System Organization), Xiaofei Zhang (Alliance of Bioversity International & CIAT)