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[i]

Summary

The Secretary of the CBD called on relevant organizations and stakeholders "to submit views and relevant information on any potential implications of the use of digital sequence information on genetic resources for the three objectives of the Convention". CGIAR conducts strategic research for agricultural development ensuring food security with a mission to benefit small holder farmers in developing countries. CGIAR experience to date confirms that digital genomic sequence data[1] can play important roles in the management and sustainable use of biological diversity and in the sharing of benefits associated with the use of that diversity. With respect to conservation, digital genomic sequence data has been used to assess genetic diversity of *ex situ* collections and to identify unique germplasm in farmers' fields which is not included in *ex situ* collections; this baseline information is essential for developing more effective *ex situ* and *in situ* conservation strategies. Concerning sustainable use, genomic sequence information, coupled with phenotypic and other data, can be used to identify genotypes that are well adapted to different, and changing, agro-ecological conditions. Integrated into crop breeding programs, genomic sequence information is increasingly useful for achieving targeted, efficient uses of genetic diversity in sustainable agriculture. The most important benefit to be shared from the use of genomic sequence information in agricultural research and development and plant breeding is improved food and livelihood security. Other non-monetary benefits associated with the use by CGIAR Centres of genomic sequence information are farmers' improved access to technologies, enhanced institutional capacities of developing country research organizations, shared research results, and local and regional economic development. Monetary benefits linked to Centres' uses of PGRFA are largely under the multilateral system of access and benefit-sharing of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). The multilateral system regulates access to material genetic resources, and not to genetic sequence information. One option currently under

[1] 'Genomics sequence' refers to sequences derived from DNA and RNA and include short and long reads and all derived molecular markers such as Single Nucleotide Polymorphisms (SNPs)

consideration for revising the multilateral system – introduction of a subscription system – could have the effect of dissolving the distinction between access to and use of material genetic resources and genomic information, since benefit sharing would be based on total seed sales which would in turn reflect the benefits to the commercial user of accessing and using both genetic resources and genomic sequence data.

Technological capacities to generate genomic sequence data, currently known as Next-Generation Sequencing Technologies, have accelerated faster than capacities to enable practical use of this information. Relatively small investments in the initial generation of genomic sequences, must then be coupled with significantly larger investments to comparatively analyse genomic sequences, to link genetic variability to useful phenotypic traits or performance, to 'optimize' those traits, and ultimately, to develop new crop varieties for release and use in farmers' fields.

CGIAR's experiences generating and using genomic sequences is still relatively new, although for analysis of germplasm collections we are further ahead. We anticipate that genomic sequence information will play an increasingly important role in CGIAR genetic resources conservation and breeding programs, and in turn will create benefits for resource poor farmers in developing countries. CGIAR underscores the importance of capacity building for developing country research and development organizations to generate and use genomic sequence information as part of their own conservation and crop improvement programs, and to be able to participate on equal footing in internationally coordinated and funded research and development programs. As part of its mission, CGIAR seeks to enable national partners in developing countries to take advantage of these and other potentially revolutionary and rapidly evolving technologies, to enhance food stability and security and close potential technological gaps. To that end, CGIAR centres are providing training and technology transfer for scientists in developing countries so that the impact and advantages from digital sequence data can benefit all.

1. Introduction

This report is being submitted by CGIAR in response to an invitation issued by the Secretary of the CBD calling on "relevant organizations and stakeholders to submit views and relevant information on any potential implications of the use of digital sequence information on genetic resources for the three objectives of the Convention".

CGIAR

CGIAR is a global research partnership for a food-secure future. CGIAR science is dedicated to reducing poverty, enhancing food and nutrition security, and improving natural resources and ecosystem services. Its research is carried out by 15 CGIAR Centres in close collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations and the private sector. Our mission is 'to advance agricultural science and innovation to enable poor people, especially women, to better nourish their families, and improve productivity and resilience so they can share in economic growth and manage natural resources in the face of climate

change and other challenges.' The primary geographical focus of CGIAR research and development is developing countries and regions.

Given the nature of our mission, the networks we operate under and our modus operandi, our experiences of using genetic sequence information will be most relevant to the conservation and sustainable use of agricultural biological diversity and benefit sharing derived from the use of genetic resources for food and agriculture. While some CGIAR centres do have some experience generating and using animal and fish genetic sequence information, for the purposes of this submission, we will focus primarily on the implications of generating and using sequence information derived from plant genetic resources for the three CBD objectives.

The objectives of the Convention on Biological Diversity

The CBD has three objectives as defined in Article 1:

(1) the conservation of biological diversity
(2) the sustainable use of its components
(3) the fair and equitable sharing of the benefits arising out of the use of genetic resources.

After this introductory section, the rest of the report is divided into three separate sections concerning the contributions that genetic sequence information can and has made to each of these objectives.

Agricultural biological diversity and genetic resources for food and agriculture

According to the Conference of the Parties to the Convention on Biological Diversity (CBD) "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]

Virtually all of the work that CGIAR centres are doing in the generation and use of genetic sequence data is related to the conservation and sustainable use of instraspecific genetic diversity, and sharing of benefits related to those uses.

What is genetic sequence information and why (in very general terms) is it useful?

The DNA of every living organism on earth encodes the basic building blocks of its life. DNA sequences are unique from one organism to the next and these sequences can aid in taxonomic classification, identification of unique genes, and can help identify gene combinations that encode traits valuable for

---

[2] Conference of the Parties to the Convention on Biological Diversity. 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.

sustainable production of food in a changing environment. While whole genome sequences are increasingly available, the use of this information to create benefits still lags well behind the technology to generate the sequence data. Currently, genetic markers, minuscule segments of DNA scattered throughout the whole genome, are being successfully deployed to genotype individuals and, with some success, to identify individuals in breeding programs that contain traits of agronomic importance. In genetic resources conservation programs, genotyping is a powerful tool to help identify gaps in collections and in the future to ensure valuable unique genetic diversity is safely conserved for future and more effective use. For *in situ* conservation programs and indigenous communities, unique genetic fingerprints of existing crop varieties can be used to establish baseline information for future work looking at the conservation, or loss, of farmer-held varieties as well as to document the impact on diversity in farmer fields with different intervention methodologies, including the reintroduction of locally extinct native varieties conserved *ex situ*.

Varieties (farmers' varieties or bred varieties or wild ancestors) represent combinations of genetic sequences that underpin the traits of each particular variety in interaction with the environment where the variety is being grown. While a variety may be unique as a whole, individual genetic sequences (e.g. coding an early maturing or late maturing variety) may be expressed in the same manner in many different varieties. The reverse can also occur: the same trait, e.g. early maturing, can also be under the control of different sequences in different varieties. The expression of any given gene can be widely influenced by environment. The uniqueness of a variety comes from the combination of those genetic sequences and the environment's influence. The combination in any particular variety is the result of hundreds and thousands of years of random, environmental, farmer and breeder selection.

Digital sequence information can enable:

- Genebank managers to accurately assess and quantify the level of variation among individuals in a single seed pack, or between genebank accessions. This can ensure quality control, as well as proper maintenance, distribution, and use of genebank collections;
- Crop scientists to obtain the necessary genetic baseline information to compare *ex situ* genotypes with diversity maintained by farmers in *in situ* conditions, to improve understanding of the diversity present in these cultivated crops and domesticated animals overall, and to determine what is present, and missing, from the genebank or in community management landscapes;
- Breeders to understand and use the diversity that exists to develop more efficient breeding strategies to reach genetic gain objectives;
- Member countries and local peoples and indigenous communities, potentially, to be able to accurately and holistically quantify and fingerprint the diversity they hold *in situ*, identifying the diversity that is most at risk as well as diversity that is unique or being used by the community, or in individual farmers' fields.
- Governments to partner with farmers and other natural resource managers for developing conservation priorities for maintaining diversity and monitoring *in situ* populations at the gene

level, as well as for understanding how populations respond to changes in temperature, water, fertilizer, nutrients, management, etc.

While a powerful tool, genetic sequence data is not a panacea nor can it be used in isolation from other technologies. Although it is becoming increasingly inexpensive to obtain large amounts of data, the tools to assemble these data in a broad-based useable form that can drive genetic gain[3] in most crop plants, livestock and fish are still lacking, in both the global north and south. CGIAR Centres are involved at various levels in whole or high density genome sequencing and genotyping of crop plants (including banana, cassava, chickpea, cowpea, groundnut, millets, maize, pigeon-pea, potato, rice, sorghum, sweetpotato, wheat and yam), as leaders and as members of broader consortia.

While genome sequencing and genetic fingerprinting may help to distinguish "what is the same" and "what is different" genetically, a major bottleneck in the development of these tools is a lack of morphological (phenotypic) data which compliment, and are needed to fully interpret, the digital sequence data. Most traits are under complex genetic control involving multiple forms of multiple genes interacting in networks. Therefore, a cause and effect scenario for a given single gene to a given trait often cannot be readily understood. A crop's ability to tolerate drought, for example, depends on the anatomy and architecture of roots, leaves and stems, the rate of progress through the life cycle in relation to the development of drought, and difficult-to-measure attributes of photosynthetic, respiratory and other biochemical and physiological capacities of the plant (e.g. dropping of leaves). None of these mechanisms for drought tolerance is simple and none would likely be caused by a single gene. The situation is the same in animals: Key traits in livestock such as weaning weight, tolerance to heat and resistance to diseases depend on multiple characteristics regulated by myriad genes. CGIAR Centres and partners are gaining experience using high throughput phenotyping. However, this is expensive and requires a great deal of investment in expertise and resources.

In summary, DNA sequence information by itself is currently of limited value, though it has considerable incremental value as tools and methods are being developed for its more efficient, targeted use. Whole genome sequences can provide valuable information. However, DNA markers are presently much more widely used for genotyping and in attempts to understand the genetic basis of traits. These markers increasingly help to accelerate breeding and genetic gains worldwide. Also in developing countries, many breeding programs have begun to use these approaches to enable rapid selection of varieties and breeds with traits important to them.

---

[3] Genetic gain refers to the amount of increase in performance of traits of interest of selected individuals subject to genetic improvement programs, between the original generation and the next generation, when they are compared in the same environment.

2. The contribution of digital genetic sequence information to the conservation of biological diversity

The main obligations of Contracting Parties with respect to conservation are set out in CBD Article 8 ('In situ conservation'), Article 9 ('Ex situ conservation') and Article 7 ('Identification and Monitoring related to conservation and sustainable use').[4]

In 2010, the Conference of the Parties to the CBD adopted the Aichi Targets. Aichi Target 13 states: "By 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of 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."

One major challenge linked to the realization of Aichi Target 13 is to identify and quantify the genetic diversity that needs to be maintained. Managers of *ex situ* collections have long sought to define the diversity of crops they hold so that important gaps (missing diversity) in the collections could be identified and safely conserved. Characterization of diversity found *in situ,* both for crops and livestock, has also been one of the main goals of a range of actors involved in the conservation of agricultural diversity. This remains an elusive goal despite the work that has been done in developing DNA sequencing methodologies. However, it will be achievable when whole *ex situ* collections and representative samples of *in situ* diversity can be sequenced. In the past, morphological descriptors have been used to characterize *ex situ* collections and distinguish the genetic differences from one accession to another. This system works and is still the routine way to classify *ex situ* collections. The same approach has been applied to study crop and animal populations *in situ*. However, morphological descriptors have limited value in assessing closely related, yet genetically distinct, individuals and in defining the range of genetic diversity that exists within and between landraces of particular crops or domesticated animals. Defining genetic diversity using morphological descriptors is further confounded by the wide environmental plasticity of these morphological characters. DNA sequence data will be the best tools available, once they can be obtained on a sufficiently large scale, for describing diversity and analysing the extent of diversity *in situ* and to 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, DNA sequence data is a powerful, and perhaps the only, means to assist genebank managers, national agricultural organizations, companies, indigenous communities and researchers to fully achieve this Achi Target and quantify the diversity present, *in situ* and *ex situ,* for the major crops and domesticated animal species.

---

[4] Annex 1 provides a list of the components of biological diversity that should be monitored, from landscape level, to species level to genetic level. Paragraph three states "Described genomes and genes of social, scientific or economic importance."

Since its inception, CGIAR has invested in conserving agricultural biological diversity *ex situ* and sustainably using this resource to fulfil its mission. Lack of knowledge of the genetic diversity has been a rate-limiting bottleneck and therefore CGIAR has embraced new technologies in order to better characterise crop diversity, understand the relationships among conserved accessions, and identify and fill gaps in the global *ex situ* collections CGIAR Centres maintain for the international community. As examples, Box 1 presents the experience of CIP in the genetic characterization of a set of potato landraces from the *ex situ* genebank. Studies such as these can show the relatedness of individuals in *ex situ* collections and can be used in the identification of novel alleles not conserved *ex situ* when similar genotyping is done *in situ*. Box 2 presents the genetic structure of the accessions maintained in IITA cassava *ex situ* collection.

Box 1: CIP genotypes accessions to understand the genetic structure of *ex situ* sweet potato collection

In Peru, farmers traditionally cultivated 20-40 different landraces of potatoes as a form of insurance; by planting such diversity, some landraces will produce a crop even in bad years, sustaining farmers until the following harvest. Over the past several decades, the planting of many different potato landrace varieties by some communities has gradually decreased and many families are now planting fewer than ten varieties. A program by the CIP genebank has been operating to give back, or repatriate, landrace varieties collected from these areas and thus contribute to building back-up systems whereby indigenous communities can continue traditional farming practices where diversity plays a key role in long-term sustainability. One challenge associated with this restoration work is that we do not know what diversity existed 50 or 100 years ago or even what diversity exists today in the *ex situ* genebank. Thus, the CIP genebank has recently genotyped their entire collection of cultivated landrace potato and sweetpotato, laying the foundation for assessing the diversity present in the *ex situ* collection

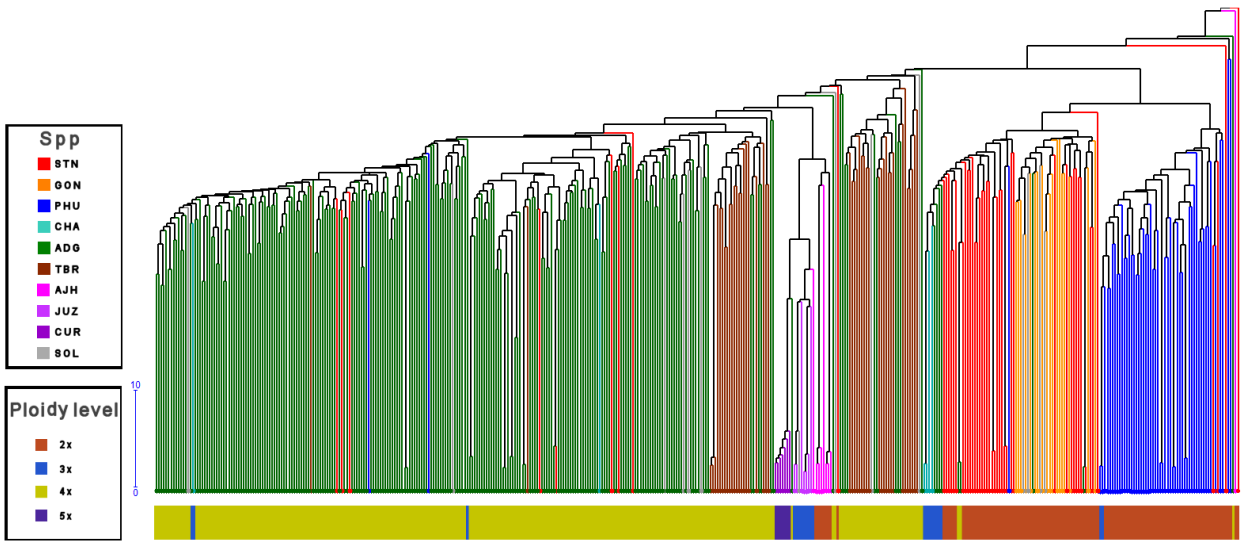
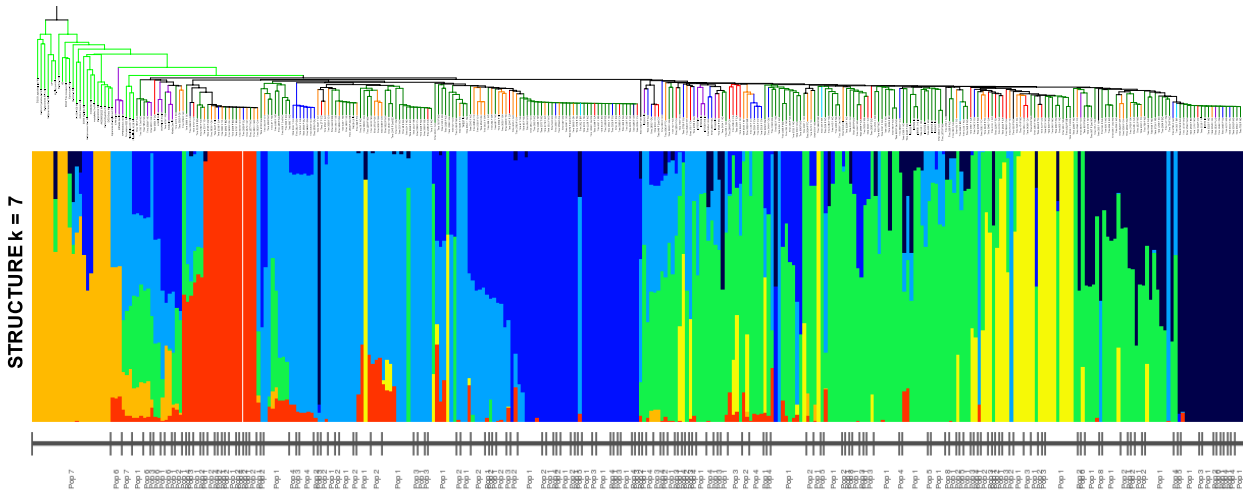


Figure 1: Dendogram of the CIP diversity reference core set of landrace cultivated potato (papa nativa) based on genetic data from the SolCAP 12K SNP array. Colored lines in the dendogram denote different potato species (based on Hawkes 1990) which show clear grouping by species. Bar below the dendogram denotes ploidy level. Data such as these help define diversity in the collection and identify individual accessions of interest for more detailed study. (CIP data, unpublished).

Box 2: Understanding the structure and genetic diversity of international cassava collection hosted by IITA



Cladogram analysis of 340 samples from 319 cassava accessions (25 Wild, 281 landraces of IITA diverse *ex-situ*/field core collection and 14 cultivated) genotyped with 38K DArTseq SNP markers showing available genetic diversity and relationship between accessions with origins from different regions of the West Africa (IITA, unpublished). Different colours indicate countries from which originally collected.

CGIAR Centres seek to maximise the use of their collections by characterizing plant genetic resources, and sharing information about the germplasm they maintain. However, just the identity of potential traits of interest is often not enough as the large size of collections make incorporation and use of these traits in improvement programs nearly impossible. Creating smaller subsets, or core reference sets, can help breeding programs to make use of such traits, as is the case with Mexican wheat landraces presented in Box 3, ICRISAT collections in Box 4, and African rice in Box 5. Genetic information helps create these core sets in a sensible way, ensuring representation of collections' diversity in the core set, and provides guidance for breeders to make use of the sets in a more effective way.

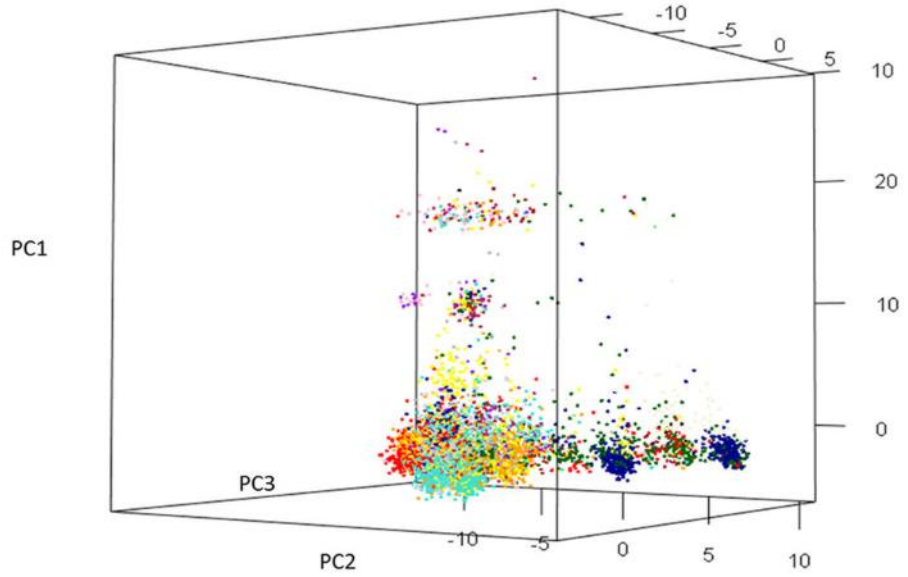
Box 3: Unlocking the genetic diversity of creole wheats

(Abstracted from Vikram et al. 2016)

Mexican wheat landraces, also known as 'creole wheats', were brought to the Americas from the 16th through 18th centuries and gradually became adapted to the local environments including many heat- and drought-stressed regions. As such, they should have useful genetic variation for stress tolerance.

The introduction of the genetic diversity of these creole wheats into breeding pipelines has potential for developing the next generation of wheat varieties. With this objective in mind, a team of scientists from CIMMYT, the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) in Mexico and the Punjab Biodiversity Board (India) carried out a study to: (1) characterize the collection of Mexican wheat landraces conserved in the CIMMYT germplasm bank; and (2) develop a core reference set using multiple variables. A core reference set is a subset of a genetic resources collection representing the diveristy present in the whole collection but small enough for breeders to evaluate for interesting traits. Core reference sets have been established in the past based on one variable, for example genotypic data or phenotype measures or geographical distribution. Simultaneous use of multiple types of variables (genotype, phenotype, geography, etc.) provide a robust diversity estimate for its application in plant breeding. As a result, 8,416 wheat landraces representing a range of Mexican agro-ecologies were characterized by genetic markers (DarTseq) and also phenotypically for yield potential, drought and heat tolerance, and yellow rust resistance to identify a core reference set that can represent this important variation. This core reference set capture 89% of the rare alleles present in the complete set.

Figure 2: Three-dimensional Principal Component Analysis (PCA) graph showing the distribution of Mexican wheat landrace groups based on genetic markers.



There were a total of 15 groups that correspond to different Mexican states. 1 = Yellow (MEXICO, PUEBLA), 2 = Light blue (MEXICO, QUERETARO), 3 = Dark blue (CHIHUAHUA, OAXACA), 4 = Orange (MEXICO, PUEBLA, QUERETARO, HIDALGO), 5 = Light green (DURANGO), 6 = Dark green (CHIHUAHUA 95.5), 7 = Pink (OAXACA, TLAXCALA, TOLUCA, PUEBLA), 8 = Purple (OAXACA), 9 = Turquoise (MEXICO), 10 = Brown (MEXICO, MICHOACAN), 11 = Red (COAHUILA), 12 = Gray (TLAXCALA, MEXICO, MICHOACAN), 13 = Maroon (MICHOACAN), 14 = Beige (CHIHUAHUA 95.5), 15 = Black (GUANAJUATO). The PC1, PC2 and PC3 contribute 10.5%, 8.2% and 6.9% of the total variation respectively.

Box 4: Creating 'mini-core' collections to enhance utilization of germplasm for crop improvement

Germplasm diversity is basic to crop improvement programs. However much of the germplasm in genebanks has not yet to be used in crop improvement programs. Greater use of germplasm in crop

improvement programs is needed for sustained and enhanced agricultural production for food security. The ICRISAT Genebank conserves over 125,000 accessions of six mandate crops and five small millets from 144 countries. The main reason for low use of germplasm is lack of information on traits of economic importance and the large size of the collections. To enhance utilization of germplasm in crop improvement programs, representative core collections (10% of entire collection) were developed using data on quantitative and qualitative traits in chickpea, pigeonpea, groundnut, sorghum, pearl millet, finger millet, foxtail millet, Proso millet, barnyard millet, kodo millet and little millet. Unfortunately, the number of accessions in these core reference sets still remains too large for its meaningful evaluation in breeding programs. To overcome this, ICRISAT scientists (Upadhyaya and Ortiz, 2001) developed the mini-core (10% of the core collection or 1% of the entire collection) concept and proposed a two-stage strategy using qualitative and quantitative trait data of the mini-core collection. Extensive multi-disciplinary evaluation of mini-core collections has identified new sources of variation for multiple traits including tolerance to biotic and abiotic stresses, and for nutritional and agronomic traits. These mini-core reference sets have been distributed to breeders in 36 countries for use in improvement programs. Sequencing these mini-core collections would be of great value to determine sequence variation associated with traits and help identify the most useful germplasm lines for use as parents in breeding programs.

Box 5: The genetic variation and population structure of *Oryza glaberrima* and the development of a mini-core reference set using DArTseq

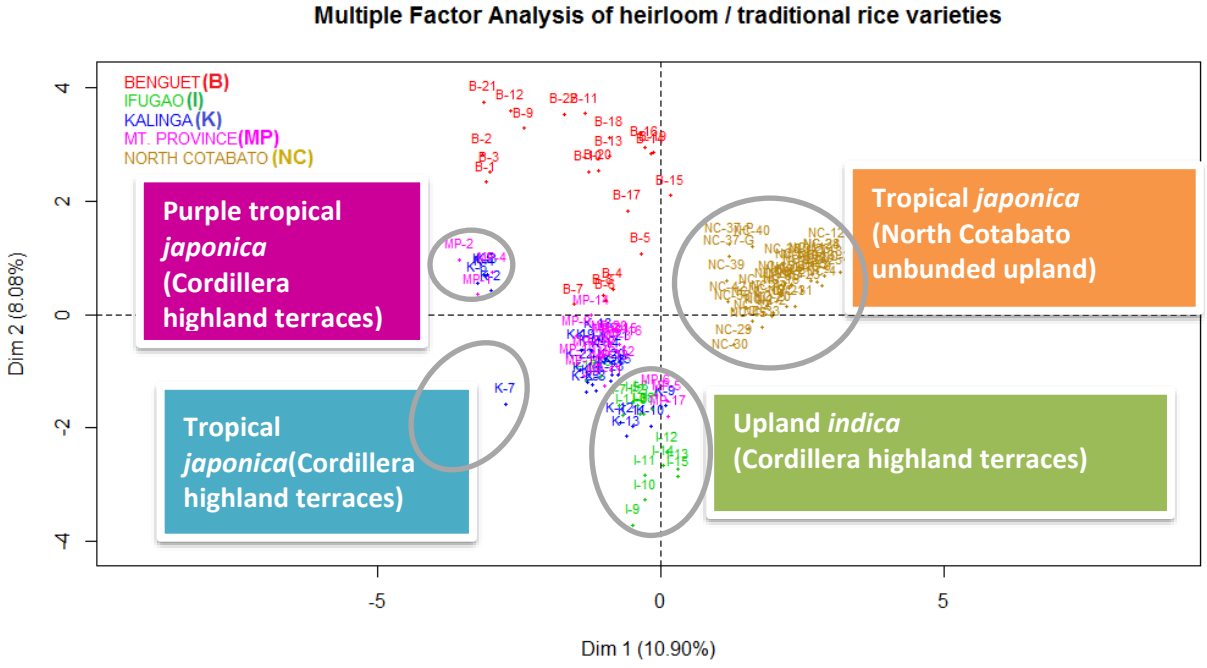
The AfricaRice genebank holds two cultivated species (*Oryza sativa* and *O. glaberrima*) and five African wild species (*O. longistaminata*, *O. barthii*, *O. punctata*, *O. brachyantha* and *O. eichingeri*), most of which originated in Africa. These rice accessions have adaptive or protective mechanisms for different abiotic and biotic stresses, but are generally characterized by a wide range of undesirable agronomic traits. To combine traits of economic importance from both the Asian rice (*O. sativa*) and African rice, interspecific breeding programs were initiated by AfricaRice in the early 1990s, which resulted in the development and release of several improved interspecific varieties for the different ecologies in Africa. Such successful interspecific hybridization between the two cultivated rice species, released with a trade mark of New Rice for Africa (NERICA), clearly demonstrated the usefulness of the African rice germplasm in developing modern improved varieties that combined the high yield potential from the *O. sativa* parents and the adaptability to different abiotic and biotic stresses from the *O. glaberrima* parents. The genetic analysis of the AfricaRice collection has enabled the development of core and mini-core reference sets of 1,330 and 300 accessions, respectively. The core and mini-core sets accounted for ~ 61 and 14%, respectively, of the whole collection, and represent 97-99% of the SNP polymorphism and nearly all allele and genotype frequencies observed in the whole *O. glaberrima* collection.

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 centres, in collaboration with local and national organizations, have applied genomic tools to genetically characterize landrace diversity found in *in situ*. A good example is the Heirloom Rice Project in the Philippines, which is presented in Box 6. For domestic animals, a number of studies, such as the one presented in Box 7 concerning the Bactrian camel, have made use of genetic sequence data to

understand the genetic structure of different populations of domestic races and breeds, and the relationship among populations, in this way providing useful information for designing conservation strategies and prioritizing conservation actions based on genetic erosion in certain populations, breeds or species.

Box 6: Applying genetic characterization to *in situ* conservation and use of Heirloom rice in the Philippines

The Heirloom Rice Project, which started in 2014, is supported by the Department of Agriculture of the Philippines and IRRI. Its aims are to enhance productivity, enrich the legacy of heirloom varieties and empower communities growing heirloom varieties in marginal rice-based ecosystems in the Philippines. Some heirloom rice varieties have exceptional cooking quality, flavor, aroma, texture, color, and nutritional value; some are also resilient, showing high levels of resistance to diseases and tolerance of environmental stresses. However there are challenges for farmers to grow and market these varieties, including lack of access to quality seed, low yield, etc. Some heirloom varieties are gradually disappearing. Market and product development, alongside maintaining biodiversity in the region, are crucial for farmers to continue growing these threatened rice varieties. By characterizing existing landraces, scientists have applied DNA analysis to understand relationships between and among varieties and to demonstrate the presence of novel genes associated with important traits such as resistance to pests and diseases that can protect these unique landraces against the stresses. This work will increase farmers' knowledge of their heirloom rice diversity, and their market opportunities.
In this project, farmers and other stakeholders recognized the importance of maintaining these unique varieties for generations to come, therefore, *ex-situ* conservation maintained under a "black-box" agreement at the national (Philippine Rice Research Institute Gene Bank) and international gene banks (IRRI International Rice Gene Bank) is in progress.



Multiple Factor Analysis (MFA) done in FactoMineR (R package) showing relationships among 119 rice

varieties grown *in-situ* by farmers based on the following groups of variables: (1) Philippine Region = Province **SOURCE** [categorical: Cordillera Administrative Region (CAR, irrigated highland terraces) = Benguet (B), Ifugao (I), Kalinga (K), Mountain Province (MP), and Mindanao (Region 12, unbunded upland) = North Cotabato (NC)]; (2) **Genotype**, 558 SNPs (categorical: A,T,C,G); (3) **Morpho-agronomic traits**, 8 categorical + 4 numerical; (4) **Grain quality**, 1 categorical + 21 numerical; (5) **Disease reaction** (BB) [categorical: Resistant (R), moderately resistant (MR), moderately susceptible (MS), susceptible (S)].

Box 7: Understanding genetic diversity of Bactrian Camel for conservation purposes

With their tolerance to cold, drought, and high altitudes, Bactrian camels (Camelus bactrianus) have been particularly appreciated in the steppes and mountains of Central Asia and could be a source of potentially useful traits in breeding programmes, but their number has been decreasing rapidly in recent years. Strategic conservation and breeding strategies are hindered by, among other factors, the very limited knowledge about the genetic diversity of Bactrian camels and the relationship among existing populations. With research partners from China and Mongolia, scientists in ILRI applied genetic analysis using microsatellites markers to characterize populations of Bactrian camels in these two countries. The study revealed significant differences among Chinese and Mongolian populations, showed gene flow within the target populations (possibly associated with trading along the Silk Road and transhumance) and confirmed that Bactrian camels from China and Mongolia are genetically distant and should be considered as distinct populations in conservation and breeding programmes (Jianlin et al, 2004).

3. The contribution of genetic sequence information to the sustainable use of biological diversity

The obligations of Contracting Parties on the sustainable use of biological diversity are set out in Article 10 (Sustainable Use) and article 7 (Identification and Monitoring related to conservation and sustainable use).

Natural genetic variation has permitted humans to select plants, animals and microorganisms for the past 13,000 years, and through the application of varied breeding techniques and methodologies, humans have been able to change the genotype of these resources and select for differential expression of traits (the phenotype) to obtain animal races, plant varieties or cultivars and strains of microorganisms responding to our emerging and changing needs.

Within the agricultural sector, genomic information contributes to the sustainable use of biological diversity in the context of plant and animal breeding. In a conventional sense, breeding relies on natural or induced genetic variation combined with efficient selection of favourable genetic combinations and evaluation of phenotypes to identify variants of interest for desirable traits. Conventional breeding can and has been enhanced, shortened and made more precise with the use of genomic information.

Economists estimate that the current rate of genetic gain in improving varieties and animal breeds must be doubled to keep up with population and income growth. There are significant scientific challenges that must be overcome to fully develop the technologies that will allow us to reach the target rate of genetic gain. To obtain meaningful information and knowledge about genetic resources we need to

learn how to manage and interpret huge quantities of data. This includes the analysis of data to understand how differences in genomic sequences lead to differences in adaptive potential as well as productive value to the farmer. This is a daunting scientific challenge that must not be underestimated. The function and action of each gene differs depending on the organ of the plant or animal, the age or developmental stage of the organ, the life stage of the crop or animal, from sowing/birth to the consumer's plate, and the value to the consumer results from the integration of all these lifetime effects. Moreover, the effect, impact and expression of each and every gene depend on a myriad of complex, interacting factors, including the full composition of the genome, abiotic and biotic factors, and agronomic practices. Therefore, selecting the correct genetic combinations to satisfy farmers, target markets and consumers is a colossal task. Developing countries have generally not benefitted to the same extent as developed countries from genetic gains (in the form of new varieties, races and breeds) made through conventional breeding. Levels and rates of adoption of new improved varieties in Sub-Saharan Africa are lower and slower than in Europe and North America. It is hoped that new sequencing and phenotyping capacities coupled with new plant and animal breeding techniques can accelerate the genetic gains made available to farmers in the developing world and help to close this gap between developed and developing countries. We will revisit this theme below in section 4 of this document, regarding benefit sharing. Meanwhile the remainder of this section is dedicated to a technical discussion of how genomic sequence information can contribute to the sustainable use of biological diversity.

Molecular markers as breeding catalysers

As part of many new breeding techniques, breeders are increasingly using molecular markers – fragments of DNA that vary due to a change in certain locations in the genome - as 'signposts' to identify particular alleles (forms or variants of a gene) of interest in breeds or accessions, to select breeds or lines as parents for crossing, to select offspring carrying favourable or deleterious alleles in segregating populations, and to perform genomic prediction (Collard & Mackill, 2008; McCouch et al., 2012). Thus, molecular markers can be used individually to follow individual variants of genes (alleles) with known effects on traits in segregating populations. This approach is often used for traits with simple genetic inheritance such as some disease-resistance traits. Alternatively, markers covering the entire genome can be used collectively to predict trait values to select promising progenies in a breeding program, an approach referred to as "genomic selection" (GS).

Since the 1980s, multiple molecular (DNA) markers and associated techniques have been developed; some of them are more suitable for research on population diversity, evolution, gene flow, and inheritance of traits while others are better suited for breeding programs (see Table 1 below). Molecular markers are based on two types of DNA polymorphism (or sequence variations): Single-Nucleotide Polymorphisms (SNP), which are variations in a single nucleotide at a specific position in the genome, and Presence-Absence Variation (PAV), which are insertions or deletions of entire stretches of nucleotides. The platforms used to detect and classify these DNA polymorphisms have evolved rapidly over the past decades and are transitioning from the detection of DNA polymorphisms via electrophoresis (e.g., RFLP, RAPD, AFLP, SSRs) to polymorphism detection on next-generation

sequencing (NGS) platforms, which determine the exact sequence of thousands to millions of given lengths of DNA in parallel. Thus, genotyping – the identification and analysis of different molecular markers (polymorphism) in individuals within and among populations – has increased in sensitivity, accuracy and speed. While Simple Sequence Repeats (SSR) have been the principal electrophoretic marker technique, this is rapidly being replaced by Single Nucleotide Polymorphisms (SNPs) as the molecular marker of choice with the ability to generate tens of thousands of genetic markers at low cost (see Table 1 below).

NGS-based marker methods — such as Genotyping-by-Sequencing (GBS), Diversity Arrays Technology Sequencing (DArTseq), and Restriction Site-Associated DNA Sequencing (RADseq) — all use NGS platforms to sequence defined subsets of DNA fragments derived from individual genomes. The sequencing of entire genomes is becoming more affordable and robust as a result of rapidly improving sequencing platforms, and although they have fallen short of their potential to broadly aid crop improvement programs, they have a huge potential for the future. Combining NGS with SNP genotyping is proving to be a relatively low cost and rapid tool to genotype breeding populations and apply genome-wide associated selection (GWAS) and genomic selection (GS) for the identification of traits regulated by multiple genes which are difficult to identify through conventional techniques due to complex genetic inheritance patterns. The understanding of the inheritance of such traits is further complicated by considerable environmental influence on factors such as yield and abiotic stresses in plants, growth and fertility in animals, and complex and large genomes in crops such as wheat and potato.

Table 1. Various types of molecular markers for diversity and breeding applications

Aspect					
Molecular marker	**Basis of polymorphism**	**Level of polymorphism**	**Suitable for**	**Advantages**	**Disadvantages**
RFLPs *Restriction Fragment Length Polymorphisms*	Different sizes of alleles associated with restriction fragments generated by enzymes (endonucleases)	Medium	Genetics, e.g., to find where a specific gene lies on a chromosome; gene flow; phylogenetic studies	First applied DNA marker for genotyping; useful in construction of genetic linkage maps	Requires *a priori* knowledge of studied DNA sequence and a large sample size; technically and time wise high demanding; difficult to automate; limited coverage of genome (low copy coding region); rarely used now

RAPDs *Randomly Amplified Polymorphic DNAs*	Different sizes of alleles based on length of short primers complementary to randomly targeted DNA in multiple locations	High	Diversity , e.g., closely related species; gene mapping	Cheap; technically and time wise low demanding; produces large no. of bands that can be characterized individually	Low reproducibility; mainly dominant; difficult to analyze; difficult to automate; cross-study comparisons are difficult
AFLPs *Amplified Fragment Length Polymorphisms*	Differences in length of selectively amplified restriction fragments generated by endonucleases	High	Diversity and genetics, e.g., population structure studies; evaluation and characterization of animal & plant resources	Large numbers of markers can be generated	Low reproducibility
SSRs *Simple Sequence Repeats*	Simple sequence repeats in tandem from 1 to 6 nucleotides in length	High	Diversity, genetics and breeding, e.g., to distinguish closely related genotypes (population studies); linkage disequilibrium studies (i.e., association of a disease-causing locus and a marker)	Highly informative (large no. of alleles, high heterozygosity); codominant; Easy to isolate; low ascertainment bias	High mutation rate; complex mutation behavior; not abundant enough; difficult to automate; cross-study comparisons require special preparation
SNPs *Single Nucleotide Polymorphisms*	Single nucleotide mutation at a specific place (locus) in a DNA sequence	High	Diversity, genetics and breeding, e.g., genetic variation in different species and breeds	Low mutation rate; high abundance; easy to type; high potential for automation; cross-study comparisons easy	Substantial heterogeneity rate among sites; expensive to isolate; low information content of a single SNP

GBS *Genotyping by Sequencing*	Sequences of the ends of all resulting DNA restriction fragments produced by a frequent cutter enzyme; generates large no. of SNPs	High	Genetic map construction; SNP genotyping in a variety of species and populations useful for breeding, plant genetics and germplasm characterization	Useful for high diversity & large genome species; cost effective for genomic-assisted breeding; high automation; technically easier to use and less demanding than RADseq	Management and analysis of large amount of data; proprietary technology
DArTseq *Diversity Arrays Technology Sequencing*	Works on a genome complexity reduction concept – selection of genome with predominantly active genes (target low copy sequences)	High	High resolution mapping and detailed genetic dissection of traits; phylogeography (in animals); genetic relatedness of species; species origin studies	Reduction complexity methods are simple and cheaper than other GBS-based methods; high reproducibility; high heterozygotes representation	Management and analysis of large amount of data; single source for proprietary technology
RADseq *Restriction Site Associated DNA Sequencing*	Sequences of short regions (50-150 bases) flanking each and all restrictions sites for a given endonuclease	High (uncovers 100s to 1000s polymorphic genetic markers in subset of genome)	Population differentiation and selection studies; phylogeography; ecological and evolutionary genomics; linkage mapping	Relatively low cost (greater no. of samples), and simple; greater coverage per locus; no prior genomic information required	Bias due to allele dropout, PCR duplicates and variance in depth of coverage among loci (all of the former vary according the RADseq method used)

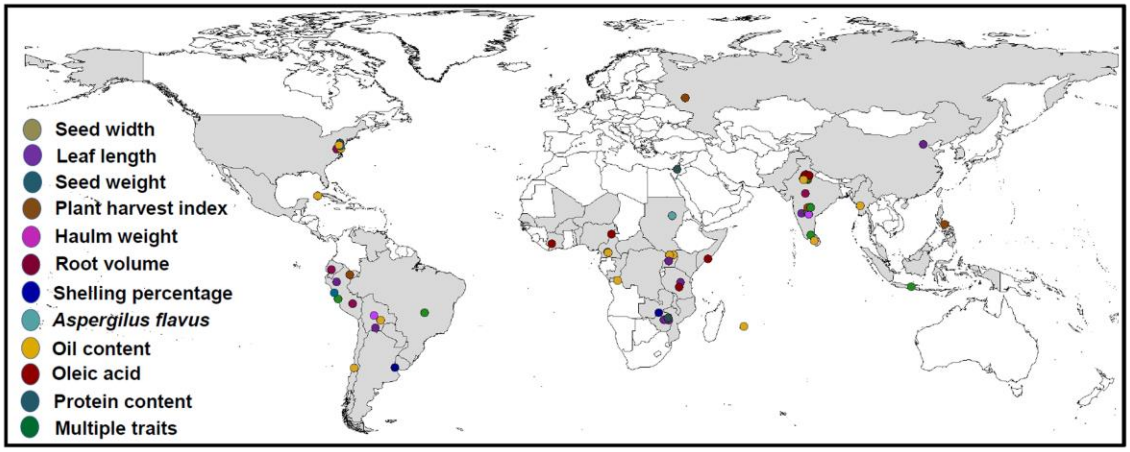
The use of molecular markers within CGIAR has allowed researchers to identify genes that control important traits (Box 5). For example, IRRI researchers identified an anaerobic germination gene that enhances rice germination under anaerobic conditions. Tolerance of anaerobic soil during germination enables uniform germination and seedling establishment under submergence, and is a key trait for the development of tropical direct-seeded rice, which represents a means of intensification and economization of rice production. ICRISAT researchers have identified the molecular markers for quantitative trait loci influencing grain iron and zinc content in sorghum, with the long-range potential impact of reducing malnutrition in sorghum producing and consuming countries. CIAT and IITA researchers have identified quantitative trait loci associated with resistance to cassava green mite,

cassava mosaic disease, cassava brown streak disease, high pro-vitamin A and high dry matter content in storage roots. CIMMYT researchers have developed and deployed breeder-ready production markers for pro-vitamin A content, maize streak virus (MSV) resistance, maize lethal necrosis (MLN) resistance, and high haploid induction rate. CIAT's work on forages has led to the identification of the genomic region associated with apomixis in *Urochloa* species, which facilitated acceleration of breeding cycles and therefore more rapid progress developing higher quality grasses. Box 8 presents cases in which genome-wide association studies have been carried out to identify genetic variants associated with relevant traits. Boxes 9 and 10 refer to advances in the use of genomic information and tools to breeding minor crops and fish respectively.

Box 8: Examples of breakthroughs for crop improvement by CGIAR Centres using genome-wide association studies.

Digital genetic sequence data obtained from high-throughput genotyping techniques have drastically modified our ability to screen highly diverse germplasm at large scale in global genebanks to identify allele variation linked to agronomic traits. Genome-wide association studies conducted by CGIAR researchers for rice (McNally et al, 2009; Zhao et al, 2011), maize (Yang et al, 2011; Romero Navarro et al, 2017), wheat (Ogbonnaya et al, 2007), peanut (Pandey et al, 2014), sorghum (Upadhyaya et al, 2016), banana (Sardos et al, 2016) and cassava (Wolfe et al, 2016) have led to the identification of useful diversity and molecular markers for a wide range of traits.

As an example, a study carried out on a reference set of accessions of the ICRISAT genebank identified genetic markers associated with 36 agronomically important traits in peanut, opening new prospects for future breeding programs. The following figure is extracted from that study.



Source: Pandey et al. 2014

A high-resolution, open-access research platform to facilitate genome-wide association mapping in rice was recently opened as described in McCouch et al, 2016.

Box 9: The African Orphan Crops Consortium

The African Orphan Crops Consortium (AOCC), a consortium involving ICRAF and various other organizations, was established in 2014-15 with the objective of improving livelihoods of smallholder farmers by improving access, quality and markets for the selected 101 orphan crops. AOCC represents a wide spectrum of players, both public (government, non-government, international, intergovernmental organizations) and private (food products, commodity and technology companies) who are stakeholders in the whole ecosystem of orphan crops. AOCC is developing genomics resources by whole genome sequencing and diversity sequencing using the latest sequencing and genotyping technologies. This is going to benefit millions of small holder farmers who are the major custodians of this treasure. The sequencing primarily leads to developing SNP panels for these crops, 50 of which are perennial trees. These developments are expected to feed into the national breeding programs through the involvement of New Partnership for Africa's Development (NEPAD), which is the African Union's development arm for Pan-Africa. The African Plant Breeding Academy (AfPBA), a capacity development initiative started under AOCC, is training ~120 African plant breeders in the use of genomics tools under breeding programs of national agricultural research systems.

Box 10: Applying genomics to fish improvement

Since 1988, WorldFish has used selective breeding to develop and manage the fast-growing Genetically Improved Farmed Tilapia (GIFT) strain. The strain has been disseminated to at least 16 countries, mostly in the developing world, and is grown by millions of small-scale fish farmers for food, income and nutrition across the globe.
Use of genomic selection tools, which enable the selection of animals based on genetic markers, will allow WorldFish to expand its GIFT research beyond a growth-only focus and introduce selection for characteristics that are otherwise difficult to measure, such as resilience and feed efficiency.
As a preliminary stage, WorldFish scientists and partners from the Roslin Institute and other research organizations have identified 13 215 SNP markers and 12 490 silicoDArT (dominant) markers from broodstock of two selective breeding programs (Genetically Improved Farmed Tilapia (GIFT) strain from Malaysia and the Abbassa strain from Egypt) and have started to test these markers for rapid genomic screening and use in Tilapia breeding programs.

Genome engineering and synthetic biology

Late in the 1970s, researchers documented yeast and bacteria integrating exogenous DNA into their genomes randomly. Since then the focus of so-called 'genome engineering' has been to target the insertion of new genetic material into the genomes of organisms and to control their expression. Animals, microorganisms and plants have been genetically engineered at experimental and commercial levels. Genetically engineered bacteria and fungi are used for a growing number of applications in bioremediation, food, textile and paper industries, generation of pharmaceutical and health products and in agriculture. The first genetically modified crops to be commercialized were tomatoes with extended shelf life (1994), insect resistant potatoes (1995), herbicide (glyphosate) resistant soy (1996) and virus resistant papayas (1998). Twenty one years after the commercialization of the first genetically modified crop, they are now grown on 185.1million hectares by approximately 8 million farmers in 26 countries (ISAA, 2016).

In the past decade or so, emerging technologies such as programmable nucleases, e.g. zinc finger nucleases and RNA-guided Cas9 from bacterial CRISPR systems, have enabled so-called precision genome engineering (or genome editing): the induction of targeted modifications to the genome, its contexts (e.g. epigenetic marks) or its outputs (e.g. transcripts) (Schiml & Puchta, 2016; Petolino et al., 2016). Targeted genome modifications include the induction of mutations at pre-selected loci to disrupt the function of one or more specific genes; the editing of existing sequences to reproduce ancient alleles or to introduce novel alleles; and the introduction of new genetic material into specific loci or regions of the genome. It is also possible to change DNA modifications, such as methylation, in order to modulate gene expression. When coupled with the ability to chemically synthesize DNA molecules at ever diminishing costs, genome engineering may enable multiple novel variations to be designed and tested at any desired genetic locus, including in multifactorial combinations (Puchta, 2017).

Although technologies and processes such as genome editing and synthetic biology[5] are still at relatively early stages of development and difficult to execute, they are developing quickly. They have the potential to vastly reduce the time taken for knowledge generated in the laboratory to transition into marketable products by allowing the direct modification of unfavourable to favourable alleles in agronomically valuable germplasm or breeds, thus reducing the number of breeding cycles required. Ultimately, if these technologies continue to develop and are widely available, it may be possible for farmers to request that a targeted set of changes be made to a highly valued cultivar as part of the breeding process. In response, a new genetic trait or combination of genetic characteristics could be rapidly designed and introduced into a cultivar to improve its resilience to stress, nutritional quality or architectural characteristics, making it a better fit for either the traditional cropping system or the modern agricultural landscape. Box 11 presents two cases in which gene editing has been incorporated in plant breeding in CGIAR Centres.

Box 11: Exmples of CGIAR Centres' genome editing

Gene editing for rice, cassava and bean improvement in CIAT research programmes

The International Center for Tropical Agriculture (CIAT) started to work on genome editing in 2014 with rice, and since then it has also tried CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats generating the endonuclease Cas9, cutting at specific sites in the DNA guided by tailored RNA) in cassava and beans, at experimental levels (laboratory and confined trials) up to this date. CIAT's research focus is geared towards achieving resistance to viruses and bacteria, improving nutritional quality and attaining hybrid seed in rice; improving starch quality and generating resistance to

[5] For the purposes of this submission we consider synthetic biology as the introduction of large sets of genes encoding complete biochemical pathways or modifying or creating new artificial organisms, enabled by a combination of powerful computer science and engineering technologies.

herbicides in cassava; increasing nutritional quality of beans, and early detection of pathogens for diagnostic purposes. Working on a proof of concept for the CRISPR/Cas9 system in rice, CIAT researchers induced a mutation in a gene regulating the formation of leaf rachis and flower carpels, obtaining leaves with drooping effect and male-induced flowers. The methodology is being evaluated to eliminate antibiotic resistance genes as selectable markers from transgenic rice plants expressing higher content of zinc and iron (Valdés, 2016; Russell, 2015).

Gene Editing at CIMMYT: Disease Resistance and Grain Quality

CIMMYT is in the process of editing genes using the CRISPR/Cas9 system for stress tolerance and quality traits in maize and wheat. Maize Lethal Necrosis (MLN), a disease prevalent in East Africa that is caused by a combination of two viruses, poses a significant threat to food security in that region. CIMMYT, in collaboration with DuPont Pioneer (now DowDuPont), has identified a strong source of MLN resistance and is close to isolating the responsible gene. Most of the hybrids in Africa are generated from combining three parents. The MLN resistance CIMMYT has identified is recessive, which means each of the parent lines in hybrids would need to be modified to confer a resistant phenotype on the hybrids. Conventional backcrossing is a time-consuming and resource-intensive process. In addition, it is nearly impossible to recreate the original makeup of the hybrid parents as the residual genome of the donor, resistant parent causes variable drag on grain yield. CIMMYT and DowDuPont will edit the susceptible gene to its resistant form directly in the parents of the susceptible commercial African hybrids. This will not only save years of time but also eliminate any chance of yield drag by precisely modifying a single locus.

In wheat, CIMMYT plans to focus its gene editing efforts on creating additional variation for durable rust resistance, the preferred mode of resistance by the breeders, and on improving available metal ions (zinc and iron) through downregulation of phytate in the grain

Promoting sustainable use through complementary uses of gene sequence information and farmers' knowledge

Formal sector breeding is only one component of sustainably using genetic resources of crops and domesticated animals.

Genetic sequence information can make useful contributions to farmer-led improvement programs, as a means of completing farmers' knowledge and adding to the knowledge base for local level management and future enhancement efforts. Box 12 presents an example from Bioversity International. Genetic sequencing has also proven to be a useful and reliable tool for measuring levels of adoption of improved varieties by farmers, as showcased in Box 13.

Box 12: Combining participatory farmer variety evaluation and selection with genotyping: wheat landraces in Ethiopia

In a recent study carried out by Bioversity International under the Seeds for Needs programme, a combination of participatory approaches, genomics, and quantitative genetics was used to trace the genetic basis of smallholder farmer preferences of durum wheat traits in Ethiopia. Two smallholder communities evaluated 400 Ethiopian wheat varieties, mostly landraces, for traits of local interest in two

locations in the Ethiopian highlands. For each wheat variety, farmers provided quantitative evaluations of their preference for flowering time, spike morphology, tillering capacity, and overall quality. Ten agronomic and phenology traits were simultaneously measured on the same varieties, providing the means to compare them with farmer traits. The durum wheat varieties were genotyped for more than 80,000 SNP markers, and the resulting data were used in a genome wide association study that resulted in a molecular dissection of smallholder farmers' choice criteria. 124 putative quantitative trait loci (QTL) affecting farmer traits and 30 putative QTL affecting metric traits were found. The study showed that smallholder farmers' traditional knowledge can be associated with QTL for desired phenotypes. These results demonstrate that it is feasible and appropriate to involve farming communities to directly evaluate broad collections of genotypes using a selected set of previously agreed summary traits. The combination of participatory variety selection and modern plant breeding can not only speed up the genetic gain in breeding targeting smallholder farming systems but also lead to improved varieties more closely addressing smallholder farmers' needs (Kidane et al. 2017)

Box 13: Assessing adoption of improved cassava varieties by farmers in Colombia through the use of SNPs

To estimate the adoption of improved varieties of cassava –a vegetatively propagated root crop cultivated by small-holder farmers and consumed as a staple food in many countries, including Colombia – CIAT characterized the genotype of 436 samples of cassava stems representing varieties identified/named by farmers from 217 households. Using an array of 93 Single Nucleotide Polymorphisms (SNPs) the genetic materials were compared with mostly Latin American accessions held at CIAT's genebank to determine their genetic relationship. Among the results, genomic information revealed misclassification of grown improved varieties *vs* landraces (37 improved and 180 landraces according to farmers' self-identification *vs* 20 improved and 197 landraces through SNPs) and generated a better estimate of the land area under cultivation of improved varieties (close to 13% by DNA fingerprinting *vs* 24% according to farmers' account). As the application of appropriate agronomic practices is variety related, misidentification has potentially a negative effect on the production and yield attained with the grown materials.Extension services appeared to be a decisive factor in the adoption of improved varieties as locations where farmers' identification of improved varieties and prevalence of them coincided with DNA information tended to be exposed to active extension services. Finally, 60 unique cassava cultivars not present in CIAT's genebank were identified, showing the genetic diversity of the crop in Latin America, and informing strategies for *ex situ* and *in situ* conservation and sustainable use (Floro IV, in press).

Creating better varieties and animal breeds faster will not, on its own, constitute sustainable use of biological diversity or contribute to sustainable development. Improved varieties and breeds must be incorporated into farming systems that are economically, socially and environmentally sustainable. This will require associated advances in, and contributions from, agronomy, water and soil science, landscape ecology, market analysis, value chain development, social science, extension services and integrated seed supply systems.

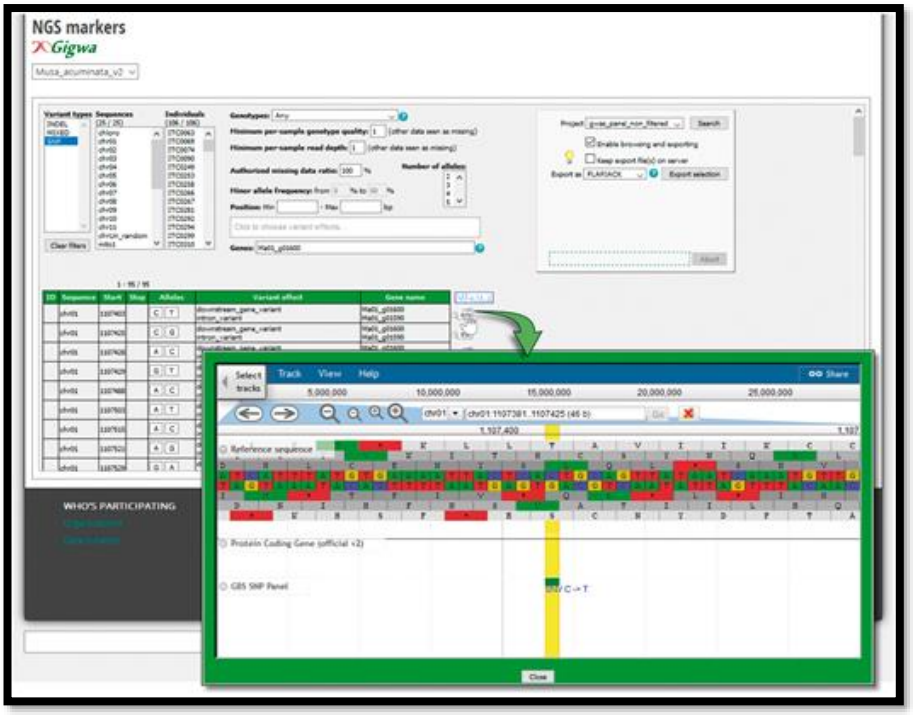
Information systems for sustainable use

Extensive genotyping, linked to measured traits, allows germplasm repositories to be searched for those materials containing desired genetic elements or trait characteristics for use either directly in production or to be introduced in programs to develop of new varieties. The development of such information systems is critical to facilitating the use of genetic sequence information, along with other relevant information, to contribute ultimately to the sustainable use of biological diversity. Such information systems are being developed inside and outside CGIAR for many crops (e.g. Rice, Maize and Banana). Two examples are provided in Box 14.

Box 14: Examples of information systems maintained by CGIAR centres that provide access to data.

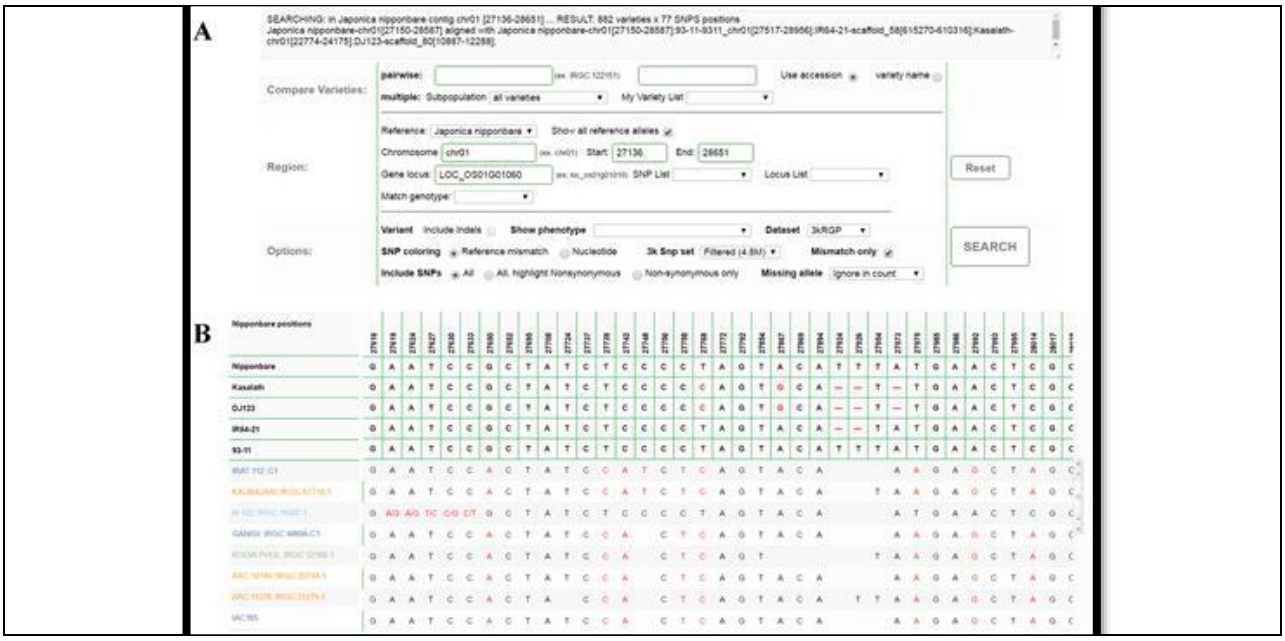
The Musa Germplasm Information System

The figure below is a screenshot of the Musa Germplasm Information System (https://www.crop-diversity.org/mgis/). The system allows users to explore allelic gene diversity in a panel of banana (Musa spp.) accessions. This tool and related methods are described in Ruas et al. 2017.



The Rice SNP-Seek Database

The following is a screenshot of the Rice SNP-Seek Database(http://snp-seek.irri.org) that enables queries associated with 3024 sequenced rice accessions. This is described in Mansueto et al, Nucleic acid research 2016.



4. Genetic sequence information and the fair and equitable sharing of the benefits arising out of the use of GRFA

Benefit-sharing under the CBD is addressed directly and indirectly in Articles 8j (benefit sharing from use of traditional knowledge), 12 (research and training), 15 (access to genetic resources), 16 (technology transfer), 17 (Exchange of Information), 18 (Technical and Scientific Information), 19 (Handling of Biotechnology and Sharing of Benefits), 20 (Financial Resources) and 21 (Financial Mechanism).

Concerns about low levels of benefits sharing led to the negotiation of the Nagoya Protocol on access and benefit sharing. Annex 1 of the Nagoya Protocol includes the following non-inclusive list of potential monetary and non-monetary benefits that could be shared with Parties that are the providers of genetic resources and traditional knowledge (Box 15).

Box 15: Nagoya Protocol, Annex 1, Monetary and Non-monetary benefits

1.Monetary benefits may include but not be limited to:
(a) Access fees/fee per sample collected or otherwise acquired;
(b) Up-front payments;
(c) Milestone payments;
(d) Payment of royalties;
(e) License fees in case of commercialization;
(f) Special fees to be paid to trust funds supporting conservation and sustainable use of biodiversity;
(g) Salaries and preferential terms where mutually agreed;
(h) Research funding;
(i) Joint ventures;
(j) Joint ownership of relevant intellectual property rights.

2. Non-monetary benefits may include, but are not limited to:
(a) Sharing of research and development results;
(b) Collaboration, cooperation and contribution in scientific research and development programmes, particularly biotechnological research activities, where possible in the Party providing genetic resources;
(c) Participation in product development;
(d) Collaboration, cooperation and contribution in education and training;
(e) Admittance to ex situ facilities of genetic resources and to databases;
(f) Transfer to the provider of the genetic resources of knowledge and technology under fair and most favourable terms, including on concessional and preferential terms where agreed, in particular, knowledge and technology that make use of genetic resources, including biotechnology, or that are relevant to the conservation and sustainable utilization of biological diversity;
(g) Strengthening capacities for technology transfer;
(h) Institutional capacity-building;
(i) Human and material resources to strengthen the capacities for the administration and enforcement of access regulations;
(j) Training related to genetic resources with the full participation of countries providing genetic resources, and where possible, in such countries;
(k) Access to scientific information relevant to conservation and sustainable use of biological diversity, including biological inventories and taxonomic studies;
(l) Contributions to the local economy;
(m) Research directed towards priority needs, such as health and food security, taking into account domestic uses of genetic resources in the Party providing genetic resources;
(n) Institutional and professional relationships that can arise from an access and benefit-sharing agreement and subsequent collaborative activities;
(o) Food and livelihood security benefits;
(p) Social recognition;
(q) Joint ownership of relevant intellectual property rights.

Monetary benefit sharing

Improved plant varieties are the source of monetary benefits enjoyed by individual farmers the world over; at a larger scale, they contribute to national and regional economic growth. As stressed above, it is anticipated that genomic sequence information will play an increasingly important role in future crop development efforts and by extension in monetary benefits for farmers, communities, companies and countries.

The ITPGRFA establishes the monetary benefit sharing rules that apply to most of the research and development work in which CGIAR engages including accessing, conserving, improving and distributing PGRFA to downstream recipients. By extension, recipients of PGRFA from CGIAR Centres are also bound by the monetary benefit sharing conditions of the ITPGRFA. In short, the ITPGRFA requires users to make financial payments of 0.7% of sales of seeds of commercialized plant varieties that incorporate materials accessed from the multilateral system, if they are not made available for use by others for further research or breeding. The payments are made to an international benefit sharing fund (BSF) and distributed following decisions by the Governing Body of the ITPGRFA.

The ITPGRFA (like the CBD and the Nagoya Protocol) applies in scope to material genetic resources and not to genomic sequence information *per se*. On one hand, genomic sequence information adds value to PGRFA collections and makes them more desirable to use. If those collections are under the multilateral system, then ultimately monetary benefit sharing conditions will apply to the plant varieties that are eventually developed and commercialized that incorporate genetic materials from the multilateral system, including those hosted and improved by CGIAR. On the other hand, it is also possible that information about interesting or valuable genetic sequences of materials inside the MLS, could give users information they need to identify the same sequences in genetic material from outside the multilateral system, thereby allowing them to circumvent the monetary benefit sharing provisions that would otherwise apply.

The terms of access and benefit-sharing under the Treaty are currently under review with the objective of increasing a) the flow of payments from users to the BSF, and b) the range of materials included in the MLS. One option that is being considered is creating a subscription system, whereby PGRFA users would agree to subscribe to the MLS for a fixed number of years (e.g., 10) during which they would make payments to the BSF based on all their seed sales, regardless of whether or not their commercialized PGRFA products incorporate materials from the MLS. Adoption of this system would have the effect of dissolving the distinction between material genetic resources and digital sequence information as far as monetary benefit sharing is concerned. It would not matter if a user accessed and used data or genetic material as long as payments based on sales are made to the system.

Additional reasons for supporting a subscription approach are rooted in challenges related to attribution of contributions of genetic sequence information (and material genetic resources). We highlight here three examples of scenarios where such challenges arise.

One, many markers may help to define favourable regions of the genome which may contain multiple genes of importance. There may be multiple options for the specific molecular marker to use to define such a region of the genome. In this case, how it is possible to place a value on taking one option or another?

Two, a favourable sequence variation may be present in multiple genotypes, including genotypes from both inside and outside the multilateral system of access and benefit sharing. How can one ascribe or capture values to a sequence that is present in the public domain outside the provisions of the Treaty?

Three, complex traits are controlled by multiple genes that interact with each other in different ways in differing environments. The effect of one gene on an agronomic trait depends on what other genes are in the genome. The commercial value of a product depends on the specific combination of genes, and is not simply the sum of the effects of each gene. How can we disentangle the monetary benefits for one from the other?

These are just a few scenarios highlighting attribution challenges: challenges that could potentially be avoided through adoption of a subscription approach wherein monetary benefits are based on total

sales of crops, and do not require tracking and tracing of the use of particular genetic sequences and genetic sequence data in the development of particular new PGRFA products. Through this kind of simplified approach, it is hoped that possible, potential disincentives for using genomic sequence information in the future can be avoided, and a proactive approach to developing and using new technologies for agricultural research and development can be supported.

Another option which seems rational and straightforward from the point of view of research and development organizations conserving, accessing, distributing and using genetic resources -- but which is not currently getting much support from a number of Contracting Parties -- would be for Contracting Parties to agree to make payments themselves to the BSF, based on total seed sales within their jurisdictions. All users in those countries would enjoy facilitated access to genetic resources in the multilateral system. In this scenario, like the subscription option described above, the distinction between access and use of material genetic resources and genetic sequence data disappears for practical purposes, as do challenges linked to attribution, because payments to the BSF would presumably reflect the value to the commercializer of access to the genetic resource or gene sequence data or both.

Non-monetary benefit sharing

The primary *modus operandi* of CGIAR towards its mission is create non-monetary benefits for developing countries through collaboration with partners in those countries. Thus it is not surprising that its genomics related work gives rise to a number of the non-monetary benefits listed in Nagoya Protocol Annex 1, often in combinations, depending on the nature of the projects and partnerships involved.

Perhaps the most important overall benefit from CGIAR's use of genomic sequence information is 'food and livelihood security benefits' (Annex, Article 2(o) as reproduced in Box 15 above). In the previous section we highlighted a number of instances where genomic information has played or can play a critically important role in the discovery of genes and combinations of genes that can eventually contribute to the development of new crop varieties and animal breeds that are more productive, nutritious, disease resistant, less reliant on chemical inputs, and adapted to changing climate conditions, to name few. All of these are important for food security and farmers' livelihoods.

Given the relatively recent emergence of next generation sequencing technologies, and the time and effort needed to develop new plant varieties and animal breeds, there are still not many examples of fully developed research and development chains that start with the generation of raw sequence data and end with released cultivars and breeds adopted by farmers. However, it is clearly the case that new enhancement methods that rely on the use of genomics data will be increasingly common, contributing to the development of newly adapted and more productive crop varieties and animal breeds that are essential for sustainable agriculture. Box 16 presents a concrete case in which the use of genomic tools will lead to a product with the potential to have a huge impact in the health of millions of people in Sub-Saharan Africa.

Box 16: The promise of maize provitamin A biofortification through discovery of natural genetic variation

In a project funded by, among others, the U.S. Department of Agriculture, Harvest Plus, and the Borlaug Fellowship, researchers from CIMMYT and several other educational and research institutions have been involved in efforts to develop maize crops with enhanced levels of the precursors to vitamin A.[6] Maize is an important subsistence crop in sub-Saharan Africa, where vitamin A deficiency is common and can lead to blindness and increased susceptibility to infections. Researchers have been working to identify natural variation in the amount of carotenoids produced in kernels of maize through association analysis, linkage mapping, expression analysis and mutagenesis. Natural variation at a lycopene cyclase locus was shown to affect the flux down certain carotenoid pathways, further affecting development of provitamin A compounds. The selection of the alleles with molecular markers identified through sequence data will enable breeders to produce maize grain with enhanced vitamin A levels, potentially having a transformational effect to improve the nutritional status of millions in sub-Saharan Africa.

One of the challenges with the introduction of new genomic technologies is ensuring that they are used in ways that contribute to sustainable development and that generate benefits for developing countries and resource poor farmers in those developing countries. There is a risk that the introduction of these technologies may benefit developed countries more than developing countries, and perpetuate, rather than diminish, the technology divide between the global north and south. To ensure next generation sequencing technologies are deployed to benefit developing countries, it is important that developing countries' national agricultural research organizations are engaged as partners in identifying challenges and genomics-based research and development streams to meet those challenges, and are engaged in the actual research and development.

Technology transfer (Annex 1, Article 2(f)), partnerships and collaboration (Article 2.(b)), and training (Article 2(j)) are essential for ensuring that developing countries have the capacity to use, and benefit from these technologies. Boxes 17, 18, 19, 20 and 21 present cases in which partnerships, technology transfer, information sharing and training involving CGIAR centres are allowing research and other actors in developing countries to increase their capacities to efficiently use genomic resources.

Box 17: Addressing climate change challenges through co-generation and transfer of technologies for genetic characterization and crop improvement

CIMMYT, ICARDA and IRRI have partnered with national research organizations from 13 countries in Africa and South Asia to co-generate and share technologies for genetic characterization and marker-assisted improvement of wheat, barley and rice, focusing on traits and alleles that are important for the crops' adaptation to climate change. These efforts are taking place under four projects funded by

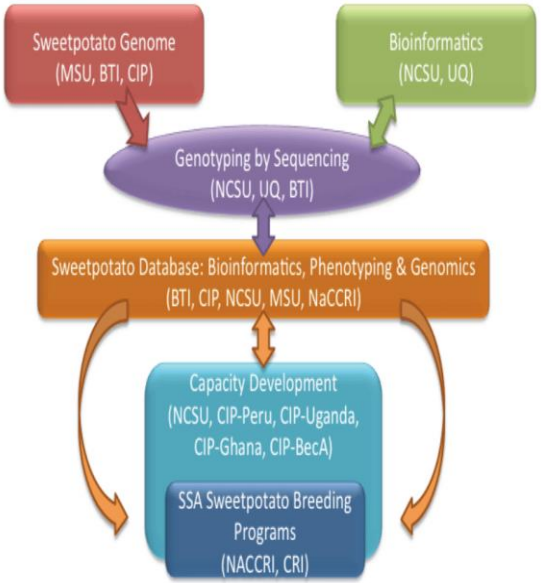
[6] Harjes et al. Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319:330-33, 18 January 2008.

the Benefit Sharing Fund of the ITPGRFA as part of the third round of projects supported by this fund. Training of plant researchers, breeders and informaticians from the countries involved and from other target countries on the use of genomic tools is a central part of the four projects.

Box 18: Increasing African breeding programmes' capacities to incorporate genomic tools in sweetpotato improvement

The Genomic Tools for Sweetpotato Project (GT4SP) is funded by The Bill and Melinda Gates Foundation and lead by North Carolina State University in partnership with the International Potato Center; The Boyce Thomson Institute at Cornell University; Michigan State University; the University of Queensland-Brisbane, Australia; the Uganda National Agricultural Research Organization, National Crops Resources Research Institute of Uganda; and the Ghana Council for Scientific and Industrial Research Crops Research Institute.

The project aims to develop modern genomic, genetic, and bioinformatics tools to facilitate crop improvement and improve genetic gains in sweetpotato, an important food security and cash crop with highly recognized potential to alleviate hunger, vitamin A deficiency, and poverty in Sub-Saharan Africa and predominantly grown in small plot holdings by poor women farmers. One of the project components (which are represented in the figure below) focuses on capacity development. Traditional and web-based training workshops and seminars are being organized for partner organizations in Africa to facilitate access to, and use of molecular markers in breeding.



Box 19: Partnerships and capacity development through CIMMYT-led 'Masagro Biodiversidad'

The CIMMYT-led 'MasAgro Biodiversidad' (Seeds of Discovery – SeeD) initiative has the goal of increasing effective and equitable use and benefit sharing from maize and wheat genetic resources

conserved in germplasm banks. A platform of publicly accessible germplasm, data, tools and services is being developed through public-private partnerships that contribute expertise and resources to the project (see http://seedsofdiscovery.org/es/catalogo/). Capacity development, including graduate student thesis projects, technical workshops, visiting scientist projects, and publicly available software tools, forms the cornerstone of a strategy to enhance and extend project impacts, and provide an equitable framework for scientific innovation and benefit sharing.

Beginning in 2016, public "calls for proposals" from Mexican scientists have invited requests to work with or in the MasAgro Biodiversidad project on specific, user-defined objectives to strengthen the users' programs. Eleven proposals were received in 2016, of which five were implemented. Topics were characterization of yellow rust resistance of Mexican wheat accessions; genomic characterization of maize accessions in the University of Guadalajara germplasm bank; development and validation of informatics models to select germplasm bank accessions based on allele frequency; genetic resources for forage maize for highland Mexico; and, analysis of pigmented maize landraces. The scientists were from three universities and the Mexican national agricultural research institute (INIFAP). Nine proposals were received in 2017, of which eight are being implemented. Topics include mapping traits associated with highland adaptation in Central Mexico; identification of genomic regions of Mexican popcorn race 'Palomero Toluqueño' associated with adaptation to highland environments; nutritional quality of maize; identification of heterotic patterns for maize lines of a mid-size private company; identification of molecular markers associated with expansion (popping quality) in popcorn maize; searching for sources of resistance to Karnal bunt disease in wheat; and two projects on genomic diversity of coffee. The scientists are from three universities, one national public research institution, one private seed company, and the Mexican national agricultural research institute (INIFAP). The two coffee projects are especially exciting because the Mexican scientists obtained funding and approached CIMMYT to assist them to implement a project similar to MasAgro Biodiversidad to help them characterize the diversity in the Mexican coffee germplasm collection.

In addition to the annual calls for proposals, seven institutions, UNAM (Mexico), INIFAP-Sinaloa(Mexico), CNRG (the Mexican Genetic Resources Center), CATIE (Costa Rica), ICARDA, IITA, and a Mexican private seed company, have made use of the genotyping services of the MasAgro Biodiversidad project to enhance their own research. Almost 300 scientists and technicians have participated in diverse technical workshops, and 34 students have pursued PhD, MSc or BSc thesis projects within MasAgro Biodiversidad.

Box 20: Examples of capacity building programmes across CGIAR Centres

Training Plant Breeders through the African Plant Breeding Academy at ICRAF-AOCC

AOCC and ICRAF along with University of California-Davis, USA established the African Plant Breeding Academy (AfPBA), which is training ~120 African plant breeders over a period of 5 years (2015-2020). Approximately 25 breeders are trained every year. To date ~50 breeders have been trained and training of 25 is currently going on. The program has an inclusive selection process which considers representation from diverse regions, gender, and various crops especially the orphan crops and trees. For more information, please visit http://africanorphancrops.org/africa-plant-breeding-academy/.

Capacity building in Africa on Animal Quantitative Genetics and Genomics

Each year, the Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub organizes a course on Animal Quantitative Genetics and Genomics. The Hub seeks to strengthen the capacity of the African scientific community in the understanding and application of methods in quantitative genetics and genomics to support research that will improve agricultural products and enhance food security in the region. The 10 day training course on Animal Genetics and Genomics targets individuals from eastern and central Africa who are currently affiliated with a national research program or university in the target countries. It takes place on ILRI campus in Nairobi. For some participants, participation is subsidized. The course includes lectures and practical sessions in population genetics, mixed linear models, genetic markers, GWAS and genomic selection among other topics. Participants also have practical sessions on programming.

Free and accessible to all online learning platform to accelerate the development of improved maize and wheat varieties

An online learning platform created in partnership among the Seeds of Discovery (SeeD) Initiative at CIMMYT (in Mexico), the National Institute of Agricultural Botany (NIAB) (Cambridge, UK) and Diversity Arrays Technology Pty (Canberra, Australia) comprises distance learning practical and theory modules about how to enhance the use of genetic diversity in wheat and maize. The online modules are available free of charge and to anyone who wishes to access them. The first module focuses on the theory of genotypic data, its importance for genetic diversity, how it is used and the technologies used to generate and analyze the data. The second module focuses on the practical use of KDSmart, an Android based application to record information on physical traits (phenotypic data) of maize and wheat varieties in the field. The modules are complemented with videos that show how the modules may help prospective users to solve problems found in their research and that explain the aims and outputs of the Seeds of Discovery Initiative. The creation of the platform and modules was driven by the felt need of reaching out to a much larger number of researchers than through the limited spaces of face-to-face courses and workshops offered by SeeD on genetic diversity analysis, breeding and use of datasets and software tools. The modules are aimed at postgraduate students, researchers, crop breeders and university members. Initially developed in Spanish, to respond to the capacity building needs of users in Mexico and Latin America, it will be available in English to reach a wider spectrum of people interested in the characterization and use of genetic diversity.
More information at:
http://seedsofdiscovery.org/new-online-learning-platform-offers-capacity-development-for-all/

Box 21: Leveling the playing field on use of genomic digital information through CIMMYT's SAGA

The Genetic Analysis Service for Agriculture called SAGA (Spanish acronym) of the Seeds of Discovery Project (SeeD), funded by the Mexican Government and run by CIMMYT since 2011, has used DArTseq technology to characterize genetically 100% of the CIMMYT maize collection (approx. 29,000 accessions) and approximately 40% of the 150,000 accessions in CIMMYT's wheat collection. Public and private institutions such as UNAM (Mexico), INIFAP-Sinaloa (Mexico), CNRG (the Mexican Genetic Resources Center), CATIE (Costa Rica), ICARDA, IITA, and a Mexican private seed company have

availed the genotyping services of the SeeD project to enhance their own research. Many Mexican researchers work with SeeD to address their specific program needs (e.g. heat or drought tolerance, disease resistance) and opportunities (e.g. enhanced nutritional qualities, enhanced forage productivity) within their projects. In the area of strengthening the capacities of agricultural research in Mexico, several public institutions and private companies have benefited from the training of 267 participants attending 18 technical workshops. Students of different levels (16 PhD, 10 MSc and 15 BSc) have and are developing their thesis projects on topics relevant to agriculture with the support of SeeD.

Another challenge – much closer to the downstream end of the R&D chain – that is common to all crop and animal enhancement programs, (including those that integrate the use of genomic sequence information) is to ensure that improved varieties and breeds are actually used by farmers. If quality seed of improved varieties is ultimately not available for use by farmers, then all the upstream research efforts will be wasted and no one will benefit. The factors that can contribute to this situation – such as weak seed systems, lack of effective extension, underinvestment, poor market linkages etc. -- are beyond the scope of this paper. However, it is important to underscore that these very fundamental issues significantly impact the sharing of benefits that can be generated through the use of genetic sequence information in crop improvement.

Sharing research results is also a critically important benefit (Article 2(b)) as is access to *ex situ* genetic resources collections and related data (Article 2(e)). Of course, genetic sequence data itself is a research result, as are markers, information about QTLs, traits and the genomics platforms that are developed to facilitate other research.

CGIAR's policy is to treat all such research results as global public goods, making them freely available for use by others. CGIAR adopted a policy on Management of Intellectual Assets in 2012 and Open Access policy in 2013, both of which underscore this basic approach. The Intellectual Assets policy allows centres to restrict access to an intellectual asset if necessary for the further development or dissemination of that asset in furtherance of the CGIAR mission. However, even where such restrictions are justified, the policy specifies that the asset must be freely available to public organizations in developing countries for research and breeding.

CGIAR Centres share information on genetic resources in the international collections they host through their own databases and through GRIN Global and Genesys. One important initiative CGIAR is participating in is the creation of the so-called Global Information System on plant genetic resources for food and agriculture (GLIS) under the framework of the ITPGRFA. Eventually, under the GLIS, CGIAR genebanks will issue digital object identifiers (DoI) for all genetic resources they maintain in their international PGRFA collections. These DoI will provide a means of ensuring that research results, including genomic research, are associated with and traceable to the PGRFA that the Centres conserve and make available to the international community under the ITPGRFA framework.

Just as is the case of crops, it is expected that genomic sequence information will play an increasingly important role in future animal breeding/improvement efforts and the concomitant increase in benefits derived from livestock and aquaculture to be shared with farmers, communities, companies and countries. The patterns and direction of animal germplasm exchange are North-North, North-South and to some extent South-South, but with limited movements South-North. Breeding strategies targeting developing countries need to adapt to the particularities and needs of smallholder systems, where local breeds have the biggest potential to significantly contribute to rural development. In this scenario, non-monetary benefit sharing is expected to have a greater impact, particularly if it focuses on supporting communities that conserve specific breeds and agro-ecosystems and facilitates decentralized, community-based breeding programmes, tailored to specific communities (Köhler-Rollefson and Meyer. 2015; Marshall, K. et al. 2009).

Literature cited

Collard, B.C., Mackill, D.J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci*. 12; 363(1491):557.

Floro IV, V., Labarta, R., Becerra López-Lavalle, L. A., Martínez, J. M. and Ovalle, T. M. 2017. Household determinants of the adoption of improved cassava varieties using DNA fingerprinting to identify varieties in farmer fields: A case study in Colombia. *Journal of Agricultural Economics*, In Press.

Hawkes J. G. 1990. *The potato: evolution, biodiversity and genetic resources*. Belhaven Press, Oxford, UK.

ISAAA. 2016. *Global Status of Commercialized Biotech/GM Crops: 2016*. ISAAA Brief No. 52. ISAAA: Ithaca, NY.

Jianlin, H., Ochieng, J.W., Lkhagva, B. and Hanotte, O. 2004. Genetic diversity and relationship of domestic Bactrian camels (Camelus bactrianus) in China and Mongolia. *Journal of Camel Practice and Research* 11(2): 97.

Kidane, Y., Mancini, C., Mengistu, D., Frascaroli, E., Fadda, C., Pè, M. E., and Dell'Acqua, M. 2017. Genome Wide Association Study to Identify the Genetic Base of Smallholder Farmer Preferences of Durum Wheat Traits. *Front. Plant Sci*. 8:1230.

Köhler-Rollefson. I. and Meyer, H. 2015. *Access and Benefit-sharing of Animal Genetic Resources*. Bonn: The ABS Capacity Development Initiative. Ober-Ramstadt: The League of Pastoral People and Endogenous Livestock Development.

Marshall, K., Quiros-Campos, C., van der Werf, J.H.J., Kinghorn, B. 2009. Marker-based selection within small-holder production systems in developing countries. *Livestock Science*. 136 (1)

McCouch, S. et al. 2016. Open access resources for genome-wide association mapping in rice. *Nature communications*. 7

McCouch, S. McNally, K.L., Wang, W. and Sackville Hamilton, R. 2012. Genomics of gene banks: A case study in rice. *Am J Bot*., 99 (2):407

McNally, K.L. et al. 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A*.,106(30):12273

Ogbonnaya, F. C., Imtiaz, M., DePauw, R. M., 2007. Haplotype diversity of preharvest sprouting QTLs in wheat. *Genome*. 50(2):107.

Pandey, M. K. et al. 2014. Genomewide Association Studies for 50 Agronomic Traits in Peanut Using the 'Reference Set' Comprising 300 Genotypes from 48 Countries of the Semi-Arid Tropics of the World. *PLoS ONE*. 9(8)

Petolino, J. F., Srivastava, V., Daniell, H. 2016. Editing Plant Genomes: a new era of crop improvement. *Plant Biotechnology Journal*. 14 (2): 435

Puchta, H. 2017. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology*. 36:1–8

Romero Navarro et al. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature Genetics.* 49: 476

Ruas, M., et al. 2017. MGIS: managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. *Database*, Volume 2017

Russell, N. 2015. Genome Editing: as Easy, Useful and Safe as it sounds? At http://blog.ciat.cgiar.org/genome-editing-as-easy-useful-and-safe-as-it-sounds/ (visited on 21 September 2017)

Sardos, J. et al. 2016. A Genome-Wide Association Study on the Seedless Phenotype in Banana (*Musa*spp.) Reveals the Potential of a Selected Panel to Detect Candidate Genes in a Vegetatively Propagated Crop. *PLoS ONE*. 11(5)

Schiml, S. and Puchta, H. 2016. Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods*. 12:8

Upadhyaya, H. and Ortiz, R. 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *R. Theor Appl Genet*. 102: 1292.

Upadhyaya, H. et al. 2016. Association mapping of germinability and seedling vigor in sorghum under controlled low-temperature conditions. *Genome*. 59(2):137

Valdés, S., D. Marín, G. Delgado, M. Prías, M. Endo, T. Seiichi M. Ishitani, P. Chavarriaga, J. Tohme. *Editing the genome of rice (IR64) to remove selectable markers*. Ninth Latin American and Caribbean Agricultural and Forestry Biotechnology Meeting "IX Encuentro REDBIO 2016-PERU". June 27 - July 1, 2016. Lima- Perú

Vikram, P. *et al.* 2016. Unlocking the genetic diversity of Creole wheats. *Sci. Rep.* 6: 23092.

Wolfe, M. D. et al. 2016. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. *Plant Genome*. 9(2)

Yang, X. S. et al. 2011. Gene expression biomarkers provide sensitive indicators of in planta nitrogen status in maize. *Plant Physiol*. 157(4):1841-52

Zhao, K. et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa. Nat Commun.* 13; 2: 467.

[i] This submission was developed by a core team consisting of Michael Halewood (Bioversity), Isabel Lopez Noriega (Bioversity), Dave Ellis (CIP), Carolina Roa (CIAT), Mathieu Rouard (Bioversity), and Ruaraidh Sackville Hamilton (IRRI). Other contributors were Matthew Morell (IRRI), Marianne Banziger (CIMMYT), Etienne Duveiller (Africa Rice), Hari D Upadhyaya (ICRISAT), Kevin Pixley (CIMMYT), Theresa Gresl (CIMMYT), Marie-Noelle Ndjiondjop (Africa Rice), Michael Abberton (IITA), Peter Wenzl (CIAT), Alice Muchugi (ICRAF), Elise Perset (SMO), Prasad Hendre (ICRAF), Marianne Yazbek (ICARDA), Hugo Campos (CIP), Carlo Fadda (Bioversity) and Rajneesh Paliwal (IITA).