

# Using Genomic Sequence Information to Increase Conservation and Sustainable Use of Crop Diversity and Benefit-Sharing

Michael Halewood,<sup>1</sup> Isabel Lopez Noriega,<sup>1</sup> Dave Ellis,<sup>2</sup> Carolina Roa,<sup>3</sup>  
Mathieu Rouard,<sup>4</sup> and Ruaraidh Sackville Hamilton<sup>5</sup>

This article describes how CGIAR centers and partners are using genomic sequence information to promote the conservation and sustainable use of crop genetic diversity, and to generate and share benefits derived from those uses. The article highlights combined institutional, and benefit-sharing-related challenges that need to be addressed to support expanded use of digital sequence information in agricultural research and development.

**Keywords:** genomic sequence information, conservation, sustainable use, benefit-sharing

## Introduction

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; they can aid in taxonomic classification, identification of unique genes, and gene combinations that encode valuable traits for sustainable agriculture under changing climates. While whole-genome sequences are increasingly available as a result of new-generation technologies,<sup>1,2</sup> our collective capacity to actually analyze and benefit from that data is lagging behind.<sup>3</sup> Currently, genetic markers—minuscule segments of DNA scattered throughout the whole genome—are successfully deployed to genotype individuals and, with some success, to identify individuals in breeding programs that contain traits of agronomic importance. Genotyping is a powerful tool to help identify gaps in collections of plant genetic resources for food and agriculture (PGRFA).<sup>4,5</sup> Unique genetic fingerprints of existing crop varieties can provide a baseline to assess losses of farmers' varieties, concomitant priorities for *in situ* conservation, and reintroduction into local production systems of "lost" farmers' varieties that have been conserved *ex situ*.

Plant varieties (including farmers' varieties, bred varieties, and wild ancestors) represent combinations of genetic sequences that underpin the traits of each particular variety in

interaction with the environment where the variety grows. While a variety may be unique as a whole, individual genetic sequences (e.g., coding an early-maturing or late-maturing trait) may be expressed in the same manner in many different varieties. The reverse can also occur: the same trait, for example, 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.

While a powerful tool, genetic sequence data cannot be used in isolation from other technologies, including phenotyping (i.e., the study of the observable traits of organisms). While genome sequencing and genetic fingerprinting may help to distinguish "what is the same" and "what is different" genetically, morphological (phenotypic) data are needed to fully interpret the digital sequence. Most traits, particularly those related to abiotic stresses, are under complex genetic control involving multiple forms of multiple genes interacting in networks.<sup>6–8</sup> 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

<sup>1</sup>Bioversity International, Rome, Italy.

<sup>2</sup>International Potato Center, Lima, Peru.

<sup>3</sup>Centro Internacional de Agricultura Tropical, Cali, Colombia.

<sup>4</sup>Bioversity International, Montpellier, France.

<sup>5</sup>International Rice Research Institute, Los Baños, Philippines.

in relation to the development of drought; and difficult-to-measure attributes of photosynthetic, respiratory, and other biochemical and physiological capacities of the plant.

This article describes how CGIAR centers and partners are using genomic sequence information to conserve and sustainably use crop genetic diversity, and to generate and share benefits associated with those uses. The article also includes reflections from CGIAR centers concerning technical and policy challenges that need to be addressed for the enhanced use of genomic sequence information in the future.<sup>9,10</sup>

CGIAR's mission is to advance agricultural science and innovation to enable poor people to better nourish their families, and improve productivity and resilience so they can share economic growth and manage natural resources in the face of climatic changes and other challenges. Fifteen CGIAR centers carry out research for development under this mission in close collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations, and the private sector. Most of CGIAR's (CGIAR, formerly the Consultative Group for International Agricultural Research) work is focused on developing countries and regions. Eleven CGIAR centers host international PGRFA collections under the framework of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). These centers use genomic sequence information to improve their conservation and sustainable use of crop genetic diversity. Some CGIAR centers also generate and use fish, livestock, and microbial genomic data. CGIAR centers are involved, either as leaders or members of research consortia, in whole- or high-density genome sequencing and genotyping of a number of crop plants, including banana,<sup>11</sup> cassava,<sup>12</sup> chickpea,<sup>13</sup> cowpea, groundnut,<sup>14</sup> millets,<sup>15</sup> maize, pigeon-pea,<sup>16</sup> potato,<sup>17</sup> rice,<sup>18</sup> sorghum, sweetpotato, wheat, and yam.<sup>19</sup>

### Using Genomic Sequence Information for Crop Diversity Conservation

Managers of *ex situ* collections have long sought to define the diversity of crops they hold, to know better what they are conserving and to identify gaps in their collections. Morphological descriptors have been developed and used to characterize *ex situ* collections, but these descriptors have limited value in assessing closely related individuals and in defining the range of genetic diversity that exists within and between landraces. Characterization of diversity found *in situ* remains an elusive goal. However, it will be achievable when whole *ex situ* collections and representative samples of *in situ* diversity can be sequenced.<sup>20</sup> DNA sequence data, once obtained on a sufficiently large scale, will be the best tool for describing and analyzing 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 are conserved long-term for humankind.

Since the 1980s, multiple molecular (DNA) markers and associated techniques have been developed. Some of them, suitable for research on population diversity, evolution, gene flow, and inheritance of traits, are increasingly used to better understand the diversity conserved in international collections and facilitate researchers and breeders' work by identifying potential sources of useful traits (Table 1). CGIAR has embraced these new technologies to better characterize crop diversity, and understand the relationships among conserved

accessions. Furthermore, CGIAR takes advantage of these technologies to identify gaps in the international "in trust" PGRFA collections they host, identify materials for collection, and identify materials for reintroduction into *in situ* conservation and use programs.<sup>4,21,22</sup> For example, the International Potato Center (CIP) recently genotyped their entire *ex situ* collection of cultivated landrace potato and sweetpotato, laying the foundation for assessing the diversity present in such collection (Fig. 1).<sup>23</sup> Among other things, this information has been useful for identifying what materials can/should be restored to Peruvian indigenous communities. 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 now plant fewer than 10 varieties. One challenge associated with potato restoration work is that CIP scientists and local farmers do not know what diversity existed 50 or 100 years ago ... or even what diversity exists today in the *ex situ* gene bank. Genotyping the international potato collection hosted by CIP has helped to reveal the relatedness of individuals within the collection and can be used to identify novel alleles that are not conserved *ex situ* when similar genotyping is done *in situ*.

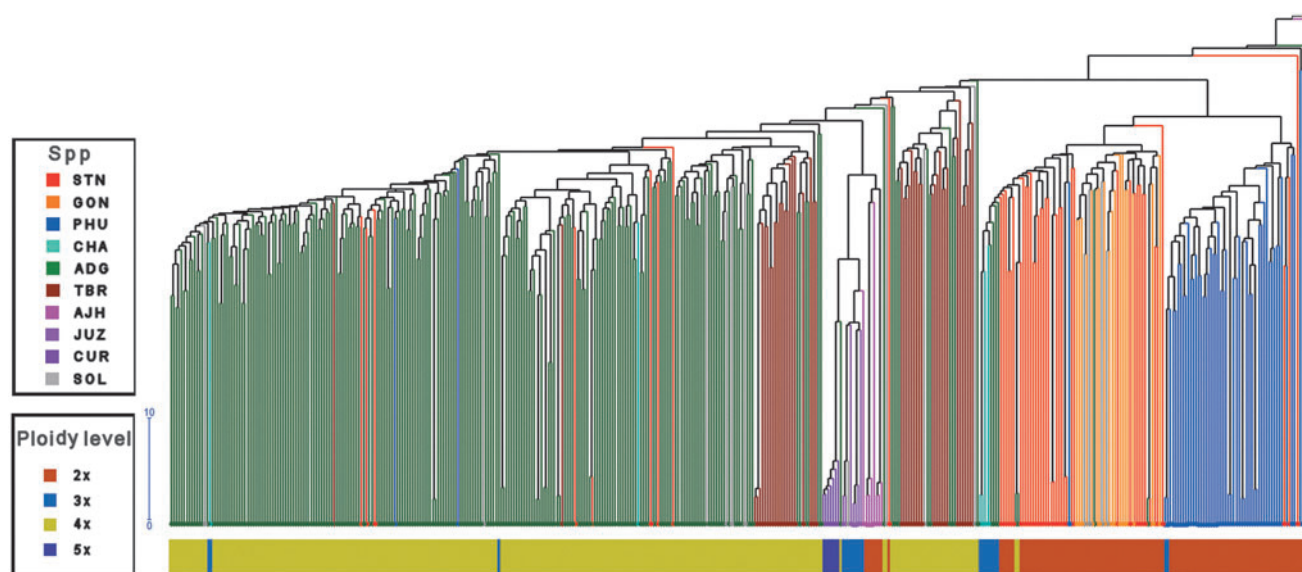
It is beyond the capacity of plant breeders and other users to sample and use all the genetic diversity of particular crops contained in the CGIAR collections. To help users find what they need, CGIAR centers' gene banks develop smaller, core sets that represent most of the diversity in the overall collection, or portions of the collection.<sup>24</sup> For example, CIMMYT (International Maize and Wheat Improvement Center, CIMMYT is a Spanish acronym) recently developed a core set representing the diversity of Mexican wheat landraces. Introduced into the Americas between the 16th and 18th centuries, wheat became gradually adapted to local environments, giving rise to Mexican wheat landraces known as "creole wheats." As such, they should have useful genetic variation for heat and drought stress tolerance. Scientists from CIMMYT and Mexico's Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) and the Punjab Biodiversity Board (India) carried out a study to: (1) characterize the collection of Mexican wheat landraces conserved at the CIMMYT gene bank; and (2) develop a core set using multiple variables. Eight thousand four hundred sixteen wheat landraces representing a range of Mexican agro-ecologies were characterized using genetic markers (DArTseq) and phenotypically for yield potential, drought and heat tolerance, and yellow rust resistance to identify a core set capturing 89% of the rare alleles present in the complete set (Fig. 2).<sup>25</sup>

Genomic sequence information is very useful in creating these core sets, ensuring representation of collections' genetic diversity in the core set. Genomic sequence information also provides guidance for breeders to make use of these core sets in effective, efficient ways. Some CGIAR centers, in collaboration with local and national organizations, have applied genomic tools to genetically characterize landrace diversity found *in situ* with the aim to promote conservation of landraces and traditional varieties through their cultivation and commercialization by farmers. For instance, the Heirloom Rice Project, coordinated by the Philippine Department of Agriculture and the International Rice Research Institute (IRRI), has genotyped 124

TABLE 1. MOLECULAR MARKERS FOR DIVERSITY CHARACTERIZATION AND BREEDING APPLICATIONS

Molecular marker	Basis of polymorphism	Level of polymorphism	Suitable for	Advantages	Disadvantages
RFLPs	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 <i>a priori</i> knowledge of studied DNA sequence and a large sample size; technically demanding; difficult to automate; limited coverage of genome (low copy coding region); rarely used now
RAPDs	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	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 and plant resources	Large numbers of markers can be generated	Low reproducibility
SSRs	Simple sequence repeats in tandem from 1–6 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 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	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 and 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	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	Sequences of short regions (50–150 bases) flanking each and all restriction sites for a given endonuclease	High (uncovers 100 s–1000 s polymorphic genetic markers in a 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)

AFLP, amplified fragment length polymorphism; DARTseq, diversity arrays technology sequencing; GBS, genotyping by sequencing; RADseq, restriction site-associated DNA sequencing; RAPD, randomly amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

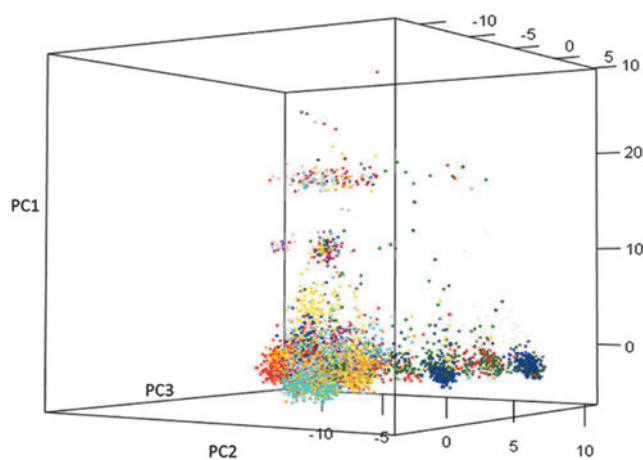


**FIG. 1.** Dendrogram of the CIP diversity reference core set of landrace cultivated potato (*Papa nativa*) based on genetic data from the SolCAP 12K SNP array. *Colored lines* in the dendrogram denote different potato species. *Bar* below the dendrogram denotes ploidy level (CIP data, unpublished). CIP, International Potato Center; SNP, single nucleotide polymorphism.

traditional varieties of rice to show how related each variety is to others from the same location or from other provinces. This work will increase farmers' knowledge of their heirloom rice diversity and their market opportunities.<sup>26</sup>

### Using Genomic Sequence Information for Sustainable Use

Conventional plant breeding relies on natural or induced genetic variation combined with efficient selection of favorable genetic combinations and evaluation of phenotypes to identify variants of interest for desirable traits. The use of genomic information can shorten these processes and make



**FIG. 2.** Three-dimensional PCA graph showing the distribution of Mexican wheat landrace groups based on genetic markers. There were a total of 15 groups—each represented by a different color—that correspond to different Mexican states. PC1, PC2, and PC3 contribute 10.5%, 8.2% and 6.9% of the total variation, respectively. Source: Vikram et al.<sup>25</sup> PCA, principal component analysis.

them more precise. The use of molecular markers (Table 1) has allowed CGIAR scientists to identify genes that control important traits. For example, IIRI researchers identified a 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.<sup>27</sup> Researchers from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) have identified the molecular markers for quantitative trait loci (QTL) influencing grain iron and zinc content in sorghum, with the long-range potential impact of reducing malnutrition in sorghum-producing and -consuming countries.<sup>28</sup> Researchers from the Centro Internacional de Agricultura Tropical (CIAT) and the International Institute of Tropical Agriculture (IITA) have identified QTL associated with resistance to cassava green mite, cassava mosaic disease, cassava brown streak disease, high provitamin A, and high dry matter content in storage roots.<sup>29,30</sup> CIMMYT researchers have developed and deployed breeder-ready production markers for provitamin A content, maize streak virus resistance, maize lethal necrosis (MLN) resistance, and high haploid induction rate.

Digital genetic sequence data obtained from high-throughput genotyping techniques have dramatically enhanced CGIAR centers' ability to screen highly diverse germplasm from international collections at a large scale to identify allelic variation linked to agronomic traits. Genome-wide association studies conducted by CGIAR researchers for rice,<sup>31,32</sup> maize,<sup>33,34</sup> peanut,<sup>35</sup> sorghum,<sup>36</sup> banana,<sup>37</sup> and cassava<sup>38,39</sup> have led to the identification of useful diversity and molecular markers for a wide range of traits. For example, ICRISAT identified genetic markers associated with 36 agronomically important traits in peanut gene bank accessions, opening new prospects for future breeding programs.<sup>35</sup>

CGIAR centers are also involved in using genetic information to conserve and enhance underutilized plant species. The

World Agroforestry Centre (ICRAF) is a member of the African Orphan Crops Consortium (AOCC), which seeks to improve the livelihoods of smallholder farmers by improving 101 orphan crops, including trees, and strengthening associated value chains. As part of this work, AOCC is sequencing whole genomes and developing single nucleotide polymorphism (SNP) panels. Through the African Plant Breeding Academy, AOCC is training ~120 African plant breeders in national agricultural research programs to use genomics tools to take advantage of the genomic sequence data that is being generated.

CGIAR centers and partners are also using genetic sequence information to 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. For example, Bioversity International's Seeds for Needs programme uses a combination of participatory approaches, genomics, and quantitative genetics to trace the genetic basis of smallholder farmer preferences of durum wheat traits in Ethiopia. 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. One hundred twenty-four putative QTL affecting farmer-preferred 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. The combination of participatory variety selection and modern plant breeding can speed up the genetic gain in breeding that targets smallholder farming systems and also lead to improved varieties more closely addressing smallholder farmers' needs.<sup>40</sup>

### *Genome engineering and synthetic biology*

Since the 1970s, the focus of so-called "genetic modification" has been to target the insertion of new genetic material into the genomes of organisms and to control their expression. A distinct new generation of technologies, commonly known as genomic editing, is now available to induce mutations at preselected loci to disrupt the function of one or more specific genes; to alter existing sequences to reproduce ancient alleles or to introduce novel alleles; and to introduce new genetic material into specific loci or regions of the genome. Genome editing could vastly reduce the time to create new varieties by facilitating the direct modification of (un)favorable alleles in germplasm, thus reducing the number of breeding cycles required. Equally importantly, genome editing now offers a fast and economical method to validate the function of native genes, thus speeding up conventional marker-assisted breeding.

CIMMYT is using the CRISPR/Cas9 to edit genes for stress tolerance and quality traits in maize and MLN, a disease prevalent in East Africa that poses a significant threat to food security in that region. CIMMYT will gene-edit wheat for durable rust resistance and improved metal ions (zinc and

iron). CIAT started genome-editing research in 2014 with rice, and has since expanded its work to cassava and beans. CIAT's work on rice focuses on developing resistance to viruses and bacteria, improving nutritional quality and attaining hybrid seeds; its work on cassava seeks to improve starch quality and generate resistance to herbicides. The objective of CIAT's work on beans is to increase nutritional quality and early detection of pathogens for diagnostic purposes. Other CGIAR centers are also include gene-editing techniques as part of their crop improvement programs.

### *Genomic information systems for sustainable use of plant genetic resources*

Extensive genotypic data, linked to measured traits, allow germplasm repositories to be searched for materials containing desired genetic elements or trait characteristics for direct use in production or in breeding programs. CGIAR centers have been involved in developing information systems for a number of crops, including the Rice SNP-Seek Database<sup>41</sup> and the Musa Germplasm Information System<sup>42</sup> (Fig. 3).

CGIAR centers share information on genetic resources held in trust through their own databases and through the global online gateway Genesys.<sup>a</sup> CGIAR is participating in the creation of the Global Information System on PGRFA (GLIS<sup>b</sup>) under the framework of the ITPGRFA. CGIAR gene banks have started to issue digital object identifiers (DOI) for all genetic resources they maintain in their international PGRFA collections. These DOIs will provide a means of ensuring that research results, including genomic research, are associated with and traceable to the germplasm held in trust.

### **Using Genomic Sequence Information to Generate and Share Benefits**

In recent years, there has been a surge in interest, on the part of intergovernmental bodies dealing with the conservation and use of biological diversity, concerning the governance of genomic sequence information. The conference of the parties of the Convention on Biological Diversity (CBD) and the meeting of the parties of its Nagoya Protocol, the governing body of the ITPGRFA, and the FAO Commission on Genetic Resources for Food and Agriculture have all launched processes to gather information about how genomic sequence data are used, to set the stage for future discussions of governance options.<sup>c</sup> A growing number of developing countries are urging these bodies to require benefit-sharing from commercial users of genomic sequence data. CGIAR has made submissions to these bodies concerning CGIAR centers' uses of genomic sequence information.<sup>43</sup>

### *Nonmonetary benefit-sharing*

Improved plant varieties—including the increasing proportion of those whose development depends in some way on the use of genomic sequence data—are used by farmers the world over, generating household income and food security; at a larger scale, they contribute to national and regional economic development. Under the CBD, Nagoya

<sup>a</sup><https://www.genesys-pgr.org/>

<sup>b</sup><https://ssl.fao.org/glis>

<sup>c</sup><https://www.cbd.int/abs/dsi-gr.shtml>





Protocol, and ITPGRFA, income generated by farmers from the use of improved resources and local, national, and regional economic development are considered nonmonetary benefits. Monetary benefit-sharing is more narrowly construed to refer to payments from organizations who commercialize new genetic resource products to the providers of the genetic resources from which the products are derived.

CGIAR pursues its mission by working through partnerships to create nonmonetary benefits for farmers and national programs in developing countries. Perhaps the most important overall benefit from CGIAR's use of genomic sequence information is "food and livelihood security benefits" (one of the nonmonetary benefits listed in the Annex to the Nagoya Protocol). In the previous sections, we have provided examples where genomic information played a critically important role in the discovery of genes that contributes to the development of crop varieties that are more productive, nutritious, disease-resistant, less reliant on chemical inputs, and adapted to changing climatic conditions. Given the relatively recent emergence of next-generation sequencing technologies, 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 clear that use and reliance of genomics data in crop improvement programs will increase over time.

One challenge related to the evolution of new genomic-sequencing technologies will be to ensure they are used for sustainable development, generating benefits for resource-poor farmers in developing countries. There is a risk that new technologies/methods may disproportionately benefit developed countries, and perpetuate rather than diminish the existing North/South technology divide. To ensure next-generation sequencing technologies benefit the Global South, national agricultural research organizations of developing countries must be engaged as equal partners in identifying challenges to be addressed through genomics-based research.<sup>44</sup> They must also be partners in the actual conservation, research, and development work.<sup>10</sup> In the following paragraphs we provide examples of how CGIAR centers partner with, train, and transfer technology to agricultural research organizations in developing countries.

*Addressing climate change challenges through co-generation and transfer of technologies for genetic characterization and crop improvement.* CIMMYT, Bioversity, International Centre for Agricultural Research in the Dry Areas (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 climatic changes. These efforts are taking place under four projects funded by the ITPGRFA Benefit Sharing Fund (BSF). 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.<sup>45</sup>

*Increasing African breeding programs' capacities to incorporate genomic tools in sweetpotato improvement.* CIP is a partner in the Genomic Tools for Sweetpotato Improvement Project that aims to develop modern genomic, genetic, and bioinformatics tools to facilitate crop improvement and improve genetic gains in sweetpotato. Sweetpotato is an important

food security and cash crop with potential to alleviate hunger, vitamin A deficiency, and poverty in Sub-Saharan Africa. It is mainly grown by poor women farmers in small plots. The project supports capacity development. Traditional and web-based training workshops and seminars are organized for partner organizations in Africa to facilitate access to and use of molecular markers in breeding (<https://sweetpotatogenomics.cals.ncsu.edu/>).<sup>f</sup>

*Tools, training, and joint research using maize and wheat sequence information in Latin America.* The CIMMYT-led "MasAgro Biodiversidad" promotes use of wheat genetic resources through the creation of a platform linking germplasm, data, software tools and services, and capacity-building. It has developed free distance-learning modules focusing on (1) the theory of genotypic data, and technologies to generate and analyze it, and (2) practical use of an Android application to record phenotypic data of maize and wheat varieties in the field<sup>g</sup>. In 2016 and 2017, MasAgro Biodiversidad supported 13 projects led by Mexican scientists/organizations focusing on characterization of yellow rust resistance of Mexican wheat accessions, genomic characterization of maize in University of Guadalajara gene bank, informatics models to select gene bank accessions based on allele frequency, genetic resources for forage maize in highland Mexico, mapping traits associated with highland adaptation in Central Mexico, nutritional quality of maize, identification of molecular markers associated with expansion (popping quality) in popcorn maize, and searching for sources of resistance to Karnal bunt disease in wheat. Forty-two students have pursued PhD, MSc, or BSc thesis projects within MasAgro Biodiversidad (Kevin Pixley, pers. comm.).

Shared research results based on genetic resources are another critically important benefit. CGIAR's default policy is to make all research results, including data, freely available, as set out in the CGIAR Principles for Management of Intellectual Assets in 2012 and open access policy in 2013. The Intellectual Assets Principles allow CGIAR centers 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.<sup>46,47</sup>

### *Monetary benefit-sharing*

The ITPGRFA establishes the monetary benefit-sharing rules that apply to most of the CGIAR centers' research and development work that involved accessing, conserving, improving, and distributing PGRFA. By extension, recipients of PGRFA from CGIAR centers are also bound by the terms of the ITPGRFA's multilateral system of access and benefit-sharing. Accordingly, commercializers of new PGRFA products derived from germplasm they receive from CGIAR centers are obliged to pay 0.77% of sales to the ITPGRFA's BSF if those products are not made available for use by others for further research or breeding.

The ITPGRFA (like the CBD and the Nagoya Protocol) applies in scope to material genetic resources and not to

<sup>f</sup><https://sweetpotatogenomics.cals.ncsu.edu/>

<sup>g</sup><http://seedsofdiscovery.org/>

genomic sequence information per se, so it does not require benefit-sharing from users of genomic sequence data alone.

The terms of access and benefit-sharing under the Treaty are currently under review with the objective of increasing the flow of payments from users to the BSF, and possibly increase the scope of materials in the multilateral system. The review process was launched because, at the time, no mandatory payments have been made by commercial users. Very recently, the BSF received its first payment pursuant to the mandatory benefit-sharing conditions. Generally speaking, it appears that those users who might be obliged to make payments—for example, those whose product development and delivery strategy includes patenting new PGRFA products—have chosen not to obtain materials from the multilateral system. Some delegations have stated they will only endorse a revised multilateral system if it includes benefit-sharing from commercial users of genomic sequence information.

One option that is being considered is to create a subscription system, whereby PGRFA users could choose to subscribe to the multilateral system for 10 years, during which they would make annual payments to the BSF based on all their seed sales, regardless of whether or not all of their commercialized PGRFA products actually incorporate materials from the multilateral system of access and benefit-sharing. The subscription system could be the key to resolving tensions concerning genomic data. Payments from subscribers could be construed to reflect the commercial value of their own use of both genetic material and sequence data. In this way, the governance of both, both genetic materials and genomic data could be addressed under the same policy framework. Of course, all regions will need to agree on the rate of payments.

Another option that 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 ITPGRA benefit-sharing fund, based on total seed sales within their jurisdictions. In return, all users in those countries would enjoy facilitated access to genetic resources in the multilateral system and related genomic information. In this scenario, similar to the subscription option described earlier, 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 benefit-sharing fund would presumably reflect the value of access to the genetic resource or gene sequence data, or both. It is hoped that these kinds of simplified approaches can help avoid disincentives rooted in unresolved controversies over benefit-sharing, and instead encourage proactive use of new-generation technologies in sustainable, equitable agricultural research and development.

## Acknowledgments

We thank Matthew Morell, Marianne Bänziger, Etienne Duveiller, Hari Upadhyaya, Kevin Pixley, Theresa Gresl, Marie-Noëlle Ndjiondjop, Michael Abberton, Peter Wenzl, Alice Muchugi, Elise Perset, Prasad Hendre, Marianne Yazbek, Hugo Campos, Carlo Fadda, and Rajneesh Paliwal for the useful discussions and input. This article is a redacted version of the much-longer submission from CGIAR to the Ad

Hoc Technical Experts' Committee on Digital Sequence Information (DSI) under the CBD.<sup>43</sup>

This work has been supported by the CGIAR Genebank Platform.<sup>48</sup>

## Author Disclosure Statement

No conflicting financial interests exist.

## References

1. Michael TP, Jackson S. The First 50 Plant Genomes. *Plant Genome* 2013;6.
2. Michael TP, VanBuren R. Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol* 2015;24:71–81.
3. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: Higher than you think! *Genome Biol* 2011;12:125.
4. McCouch SR, McNally KL, Wang W, Hamilton RS. Genomics of gene banks: A case study in rice. *Am J Bot* 2012;99:407–423.
5. Kilian B, Graner A. NGS technologies for analyzing germplasm diversity in genebanks. *Brief Funct Genomics* 2012;11:38–50.
6. Alcázar R, García AV, Parker JE, Reymond M. Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. *Proc Natl Acad Sci U S A* 2009;106:334–339.
7. Gutierrez-Gonzalez JJ, Wu X, Gillman JD, et al. Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. *BMC Plant Biol* 2010;10:105.
8. Waters AJ, Makarevitch I, Noshay J, et al. Natural variation for gene expression responses to abiotic stress in maize. *Plant J Cell Mol Biol* 2017;89:706–717.
9. Marden E. U.S. ratification of Plant Treaty: Benefit sharing ambiguity for plant genomics researchers does not change. *Genet Resour Crop Evol* 2017;1–7.
10. Halewood M, Chiurugwi T, Sackville Hamilton R, et al. Plant genetic resources for food and agriculture: Opportunities and challenges emerging from the science and information technology revolution. *New Phytol* 2018;217:1407–1419.
11. D'Hont A, Denoeud F, Aury J-M, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012;488:213–217.
12. Wang W, Feng B, Xiao J, et al. Cassava genome from a wild ancestor to cultivated varieties. *Nat Commun* 2014;5:5110.
13. Varshney RK, Song C, Saxena RK, et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol* 2013;31:240–246.
14. Chen X, Li H, Pandey MK, et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc Natl Acad Sci U S A* 2016;113:6785–6790.
15. Varshney RK, Shi C, Thudi M, et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat Biotechnol* 2017;35:969–976.
16. Varshney RK, Chen W, Li Y, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* 2012;30:83–89.
17. Consortium TPGS. Genome sequence and analysis of the tuber crop potato. *Nature* 2011;475:189–195.
18. Li J-Y, Wang J, Zeigler RS. The 3,000 rice genomes project: New opportunities and challenges for future rice research. *GigaScience* 2014;3:8.
19. Tamiru M, Natsume S, Takagi H, et al. Genome sequencing of the staple food crop white Guinea yam enables the de-



- velopment of a molecular marker for sex determination. *BMC Biol* 2017;15:86.
20. McCouch S, Baute GJ, Bradeen J, et al. Agriculture: Feeding the future. *Nature* 2013;499:23–24.
  21. Upadhyaya HD, Gowda CLL, Buhariwalla HK, Crouch JH. Efficient use of crop germplasm resources: Identifying useful germplasm for crop improvement through core and mini-core collections and molecular marker approaches. *Plant Genet Resour* 2006;4:25–35.
  22. Christelová P, Langhe ED, Hříbová E, et al. Molecular and cytological characterization of the global *Musa* germplasm collection provides insights into the treasure of banana diversity. *Biodivers Conserv* 2017;26:801–824.
  23. Ellis D, Chavez O, Coombs J, et al. Genetic identity in genebanks: Application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome* 2018;61:523–537.
  24. Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK. Accessing genetic diversity for crop improvement. *Curr Opin Plant Biol* 2010;13:167–173.
  25. Vikram P, Franco J, Burgueño-Ferreira J, et al. Unlocking the genetic diversity of Creole wheats. *Sci Rep* 2016;6:23092.
  26. IRRI—The Heirloom Rice Project [Internet]. IRRI. [cited April 24, 2018]. Available from: <http://irri.org/resources/publications/brochures/the-heirloom-rice-project> Accessed October 1, 2018
  27. Septiningsih EM, Ignacio JCI, Sendon PMD, Sanchez DL, Ismail AM, Mackill DJ. QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. *TAG Theor Appl Genet Theor Angew Genet* 2013;126:1357–1366.
  28. Kotla A, Phuke R, Hariprasanna K, et al. Identification of QTLs and underlying candidate genes controlling grain Fe and Zn Concentration in Sorghum [*Sorghum bicolor* (L.) Moench]. 2016. In: Plant and Animal Genome Conference XXIV, January 08 - 13, 2016, San Diego, CA. Available from [http://oar.icrisat.org/9408/1/PAG%202016\\_Identification%20of%20QTLs%20and%20underlying%20candidate%20genes%20controlling%20grain.pdf](http://oar.icrisat.org/9408/1/PAG%202016_Identification%20of%20QTLs%20and%20underlying%20candidate%20genes%20controlling%20grain.pdf) Accessed October 1, 2018.
  29. Nzuki I, Katari MS, Bredeson JV, et al. QTL Mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*. *Front Plant Sci* 2017;8:1168.
  30. Welsch R, Arango J, Bär C, et al. Provitamin A Accumulation in Cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* 2010;22:3348–3356.
  31. McNally KL, Childs KL, Bohnert R, et al. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci* 2009;106:12273–12278.
  32. Zhao K, Tung C-W, Eizenga GC, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2011;2:467.
  33. Yang X, Gao S, Xu S, et al. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breed* 2011;28:511–526.
  34. Navarro JAR, Willcox M, Burgueño J, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet* 2017;49:476–480.
  35. Pandey A, Misra P, Alok A, et al. Genome-wide identification and expression analysis of Homeodomain Leucine Zipper Subfamily IV (HDZ IV) Gene Family from *Musa accuminata*. *Front Plant Sci* 2016;7:20.
  36. Upadhyaya HD, Bajaj D, Narnoliya L, et al. Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. *Front Plant Sci* 2016;7:302.
  37. Sardos J, Rouard M, Hueber Y, et al. 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* 2016;11:e0154448.
  38. Wolfe MD, Rabbi IY, Egesi C, et al. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 2016;9. DOI:10.3835/plantgenome2015.11.0118
  39. Rabbi IY, Udoh LI, Wolfe M, et al. Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *Plant Genome* 2017;10. DOI: 10.3835/plantgenome2016.09.0094
  40. Kidane YG, Mancini C, Mengistu DK, et al. Genome wide association study to identify the genetic base of smallholder farmer preferences of durum wheat traits. *Front Plant Sci* 2017;8:1230.
  41. Mansueto L, Fuentes RR, Borja FN, et al. Rice SNP-seek database update: New SNPs, indels, and queries. *Nucleic Acids Res* 2017;45(D1):D1075–D1081.
  42. Ruas M, Guignon V, Sempere G, et al. MGIS: Managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. *Database* 2017;2017. DOI:10.1093/database/bax046
  43. Halewood M, Lopez Noriega I, Ellis D, Roa C, Rouard M, Sackville Hamilton N. 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 (CBD). 2017. Available from: <https://www.cbd.int/abs/DSI-views/CGIAR-DSI-en.pdf> Accessed October 1, 2018.
  44. A welcome framework for research in Africa. *Nature* 2018; 556:274.
  45. bsf-third-cycle | International Treaty on Plant Genetic Resources for Food and Agriculture | Food and Agriculture Organization of the United Nations [Internet]. [cited April 24, 2018]. Available from: [www.fao.org/plant-treaty/areas-of-work/benefit-sharing-fund/projects-funded/bsf-third-cycle/en/](http://www.fao.org/plant-treaty/areas-of-work/benefit-sharing-fund/projects-funded/bsf-third-cycle/en/) Accessed October 1, 2018.
  46. Office CSM. CGIAR Principles on the Management of Intellectual Assets. 2012 [cited April 24, 2018]; Available from: <https://cgspace.cgiar.org/handle/10947/4486> Accessed October 1, 2018.
  47. Office CC. CGIAR Open Access and Data Management Policy (approved by Consortium Board October 2013). 2013 [cited April 24, 2018]; Available from: <https://cgspace.cgiar.org/handle/10947/2875> Accessed October 1, 2018.
  48. CGIAR Genebank Platform. Available from: <https://www.cgiar.org/research/program-platform/genebank-platform/>

Address correspondence to:

Michael Halewood, PhD  
Genetic Resources Policies and Institutions Group  
Biodiversity International  
Via dei Tre Denari, 472/a  
00057 Maccarese (Fiumicino)  
Rome  
Italy

E-mail: m.halewood@cgiar.org