Scope and roles of the CGIAR genebanks: 2030 vision

Necessity is the mother of invention; Diversity is its father

Contents

1. Executive summary

This document presents needs and opportunities for deep innovation in the CGIAR genebanks, in line with the ongoing change process towards "One CGIAR". It is based on a vision of what genebanks should be in 30 years, and on the changes that are achievable and should be achieved within the next 10 years. It considers two broad sets of questions: how to empower genebanks to conserve and deliver the right resources; and how to ensure that they conserve and deliver those resources in the right way.

Among the key points made are:

- Demand for genebank materials by breeders is likely to decrease at least for the more advanced crops, as they will rely more on information on genomes and their phenotypic consequences to select only the accessions they need. In contrast, demand for materials by other researchers is likely to increase with the increasing importance and ease of establishing genotype-phenotype associations.
- Other future changes are more speculative. Greater nationalism may mean less transboundary access to national genebanks; access to *in situ* materials may be increasingly restricted; there may be significant "privatization" of collections as private sector involvement in crop improvement accelerates; some national genebanks may have their funding and scope slashed in the future as short-term development goals are prioritised over sustainability. These changes will influence the future and role of CGIAR banks, for example to conserve legacy materials.
- Genebank managers should:

    - assign current, future, and heritage values to each accession, enabling clearer separation between the short-term and long-term functions of genebanks;
    - tailor the management regime of each accession to suit its current and future value.
- Once the current value of an accession for a given purpose can be estimated with reasonable confidence, the funding model may be modified:
    - The user pays for services provided directly to them; and
    - The endowment fund of the Global Crop Diversity Trust pays for services to future generations.
- Genebanks may have a role in actively creating new genetic resources that will help current users explore and exploit new opportunities for rapid advance.
- An opportunity exists for fundamental modernisation of process and materials management for consistently high quality and high throughput at minimum cost.
- Investment is needed in research in seed longevity (especially of wild and recalcitrant species) and cryopreservation, both of which are critical to improving efficiency and effectiveness of *ex situ* conservation protocols and genebank management.
- An opportunity exists to restructure CGIAR genebanks. Details of restructuring will need to be harmonized with the emerging "One CGIAR" change process, but common features of the options presented include:
    - A small number (1-3) of multi-crop central repositories;
    - A larger number of regeneration hubs and distribution hubs;
    - Services provided to, or outsourced from, others as appropriate;
    - More close engagement with users and other stakeholders.
- Fully benefitting from the digital age will demand that genebanks:
    - Develop a digital catalogue of functional genetic variants;
    - Implement a "smart phenotyping" strategy to maximise the return on investment in phenotyping;
    - Continue to contribute to development of a "digital genebank" as a global information resource for users needing access to information about accessions.
- CGIAR should further increase engagement in relevant fora on phytosanitary measures and genetic resources standards and policy.

2. Introduction

The United Nations Sustainable Development Goals envision the complete eradication of poverty and hunger by 2030, ensuring health and wellbeing for all, and with life on land protected, restored and sustainably used[1]. It won't be easy. We know the challenges of climate change, environmental degradation, population growth, and pandemics affecting humans, crops and farm animals alike. We know the risks of failing to meet these challenges. We know that increasing environmental unpredictability and socio-economic instability make the challenges ever harder to meet. We cannot begin to guess what new challenges might come tomorrow. And the rate of change is increasing.

The only way to meet these evolving challenges is through continuous innovation: "[f]uture technologies and systemic innovation are critical for the profound transformation the food system needs" [2]. We are in a vicious, unsustainable cycle of ever faster change necessitating ever faster innovation. The need for rapid innovation has never been greater. But that will not be enough. To

---

[1] Sustainable Development Goals 1, 2, 3 and 15.
https://www.un.org/development/desa/disabilities/envision2030.html

[2] Herrero M *et al* 2020. Perspective: Innovation can accelerate the transition towards a sustainable food system. Nature Food, 1, 266-272 https://doi.org/10.1038/s43016-020-0074-1.

achieve the goals of eradicating poverty and hunger, we must go beyond the process of innovation and the commercial exploitation of inventions: we must also have mechanisms in place to ensure that the benefits of innovation reach the right people. The approach may be disputed[3], but the need for change is universally accepted.

Regardless of how we adapt agriculture to meet our current and future needs, genebanks are an essential element of the change process. They are the enablers of sustainable agricultural innovation, providing the genetic resources without which farmers and other stakeholders in agriculture could not meet current or future challenges. Their essential tasks are to conserve the right genetic resources in the right way, to match resources to the needs of users, and to deliver the right resources to each user. They must undertake these tasks as effectively and efficiently as possible and in compliance with applicable national legislation and international agreements.

As the challenges grow, we have both the necessity and the opportunity for deep innovation in the genebank community to enhance its contribution to progress. This document explores the framework and elements of modernising genebanks to be more efficient and effective. It begins by setting out the contribution of plant genetic resources and breeding to agricultural development, the essential starting point for mapping out the future of genebanks.

3. The evolving story of crop evolution

The genome[4] of a plant is unimaginably complex. The potential number of combinations of all the functional variants of all the genes in rice (the only crop to date for which we have sufficient information to begin to estimate this) exceeds the total number of atoms in the known universe. Further complexity arises from epigenetic variation, i.e. heritable variation in the structure (rather than base sequence) of DNA that affects gene expression. It is absolutely impossible to conserve, or discover the value of, all potential combinations, even for a single crop. At the same time, there is immense potential for innovation through developing novel combinations that have not previously existed.

The products of the genes in a genome work together in a network of interdependent metabolic pathways. As a result, the effect of one gene depends on other genes in the same genome. Many potential combinations of functional genetic variants are unproductive or even lethal. On the other hand, certain combinations of functional genetic variants function as “co-adapted gene complexes”, which must be inherited together if the offspring of a cross between parents bringing about such combinations are to be as productive as their parents.

Over ten millennia of agriculture, farmers and breeders have selected plants with combinations of genes that result in crops yielding orders of magnitude more than their wild relatives – the fastest evolutionary innovation in the entire history of life. Modern breeders typically work with a limited diversity of painstakingly chosen potential parents, enabling reliable progress that does not risk breaking up the superior combinations of genes. Crosses between more genetically distant parents bring the potential for much greater stepwise progress and become essential when breeding objectives change (such as when a new disease appears) or when an existing breeding programme stagnates for lack of diversity to work with[5]. However, such “wide crosses” bring the risk (or even the

---

[3] E.g. https://www.globalagriculture.org/report-topics/industrial-agriculture-and-small-scale-farming/industrial-agriculture-and-small-scale-farming.html promotes small-scale farming over “industrial agriculture”.

[4] The word “genome” is used in various senses. For the purposes of this document, the genome is defined as the genetic material (DNA in plants) of an individual. It includes both nuclear and cytoplasmic (mitochondrial and chloroplast) DNA.

[5] The amount of additive genetic variation for a trait under selection in a breeding population (σ_a) is one of the determinants of the annual rate of genetic gain $G = (\sigma_a ir)/L$, where *i*=selection intensity, *r*=selection accuracy

near certainty in crops such as maize or in interspecific crosses between crops and their wild relatives) of breaking up the desirable combinations of genes. The greater the genetic distance between parents, the greater the potential for improvement but the greater also the likelihood of reversing earlier gains. In times of dire need for reliable progress, most breeding programmes cannot afford this gamble, which is becoming the domain of a shrinking set of specialist, largely public-sector, upstream research and pre-breeding groups.

Advances in genomics and informatics are transforming this breeding model. Breeders can now choose parents and select progeny based directly on genotype rather than phenotype, which can be much faster and cheaper. Selections based on phenotype only are often ineffective because of the low heritability of the desired traits, and their high dependence on the environment in which they are assessed and the genetic background of the desired genes. Phenotyping is also usually expensive and slow, as plants may have to be grown for one or even more cropping cycles before the traits can be measured. In contrast, provided there is good understanding of the genetic control of the trait(s) under selection and good markers[6] for the functional genetic variants desired, the required genotypes can be selected with 100% certainty at the seedling stage. Alternatively, for traits with complex or uncertain genetic control, genomic selection may be used. The prerequisites in this case are an effective intelligent algorithm for selection of variants across the whole genome and a good training population for the algorithm to learn from.

In addition, through gene editing it is now possible to add, delete or change a single gene in a genome. Where a causal relationship has been established between a single gene and a high-value trait, this makes it possible for breeders to add a high-value gene into a high-value genome in a single step, enabling large improvements without the risk of breaking up desirable gene combinations. In conjunction with synthetic biology, it will eventually become possible for breeders to even edit the gene without accessing physical material in genebanks.

However, gene editing is an effective breeding tool only after research to determine the sequence and function of the "best" functional genetic variant. This research will rely on continued access to physical genetic resources for the foreseeable future. Hence, it is likely that the direct clients of the physical genetic resources will increasingly become the upstream research community, while breeders will come to rely more on such researchers and the knowledge they create. Recalling the high degree of shared synteny between species (i.e. the preserved co-localization of genes on chromosomes of different species) and other forms of homologous relationships, knowledge about the functions of genes in one species may be derived from research on other species, which will be of particular importance for minor crops.

Moreover, gene editing typically[7] addresses only one gene at a time, about 0.002% of the functioning genome. As such, although for the purpose of introducing single genes it is much faster and more

---

and *L*=number of years per cycle (see e.g. Cobb *et al*. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. 2019. *Theor Appl Genet* 132, 627–645. https://doi.org/10.1007/s00122-019-03317-0).

[6] Good markers are often close to (and therefore genetically linked with), but not part of, the gene controlling the trait. The further the marker is from the gene, the less tightly it is linked, and therefore the more dependent on the specific materials being bred.

[7] Although techniques are being introduced to edit several genes simultaneously: e.g. see Campa CC, Weisbach NR, Santinha AJ et al. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat Methods 16, 887–893 (2019). https://doi.org/10.1038/s41592-019-0508-6

precise than conventional backcrossing[8], it continues the slow-but-sure approach of modern breeding. It still leaves a glaring need to explore the much greater potential of recombining widely different genomes, if we are indeed to accelerate progress to the required level.

In summary, technological opportunities promise gains in efficiency and effectiveness. The client base for physical material is shifting more from breeders towards upstream researchers. There is a rapidly emerging client base for "digital genebanks", i.e. for comprehensive online searchable repositories of information on genetic resources, as users increasingly require access to digital information associated with accessions as much as to the physical material. It is clear that genebanks will need to evolve: not only to improve how they work and catch up with the advanced state of breeding (particularly for the most advanced crops, although under-utilized and less intensively bred crops will follow), but also to accommodate a changing role.

4. Co-evolution of genetic resources policy

Changing approaches and technologies for breeding crops have attracted scrutiny in international policy fora for decades, and continue to do so[9]. The basic concern centres on a "north-south" divide: high-income countries and advanced organizations (the "global north") are the primary beneficiaries of technological innovation, but they rely on genetic resources that originate to a large extent in the global south. The divide is viewed as existing not only at national level (high-income vs low-income countries) but also at individual and organizational levels (big corporations vs traditional farmers).

A recurring theme has been the need to implement a Global System on Plant Genetic Resources for Food and Agriculture (e.g. http://www.fao.org/agriculture/crops/thematic-sitemap/theme/seeds-pgr/gpa-old/gsystem/en), a system that itself needs continual revision. Now the primary international policy instrument in this field is the International Treaty on Plant Genetic Resources for Food and Agriculture ("Plant Treaty": http://www.fao.org/plant-treaty/en/), which came into force in 2004 and has a current membership of 146 countries plus the EU. Its centrepiece, with the CGIAR genebanks as its backbone, is a multilateral system ("MLS") of access and benefit-sharing ("ABS"), by which all member countries agree the same set of conditions for providing facilitated access to genetic resources based on fair and equitable sharing of the resulting benefits.

Since 2014 the Nagoya Protocol[10], with a current membership of 123 countries plus the EU, has provided a new international ABS regime to govern cases (countries, materials and purposes) that are outside the scope of the Plant Treaty.

However, on one side, concern is growing that these benefits are in fact not yet being fairly or equitably shared; that while the global north continues to benefit, the global south is not being recompensed for the genetic resources it has made available to it; and that while the rights of big corporations to profit from their inventions are well protected, farmers' rights are not being recognized. On the other side, concern is similarly increasing that the lack of progress on these policy issues is hindering the innovation that is so vital to the future of this planet.

---

[8] Backcrossing: starting with a cross between an elite variety and a donor variety, each generation is created by crossing the prior generation to the elite variety, resulting in a genome that comes almost entirely from the elite variety, with just a small portion from the donor.

[9] Roa C, Sackville Hamilton R, Wenzl P, Powell W. 2016. Plant Genetic Resources: Needs, Rights, and Opportunities. Trends in Plant Science: Science & Society. 23 (8), 633-636. https://doi.org/10.1016/j.tplants.2016.06.002

[10] In full, the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity: https://www.cbd.int/abs/

In addition, in recent years, awareness has grown that the value of genetic resources lies increasingly in the digital information generated from those resources, rather than in the material resources *per se*. This has exacerbated the concerns of the global south, partly because the ability to derive advantage from the information is even more dependent on advanced technical capabilities and therefore constitutes an even greater north-south divide; and partly because existing international instruments only address use of the material, not use of the information. The disputes show few signs of abating. The need to keep rethinking the global PGRFA system continues.

Such uncertainties make it difficult to predict the future. There has been a tendency, which may continue, for many countries to become increasingly protective of their own genetic resources. This increasing nationalism may lead to less transboundary access to material conserved in national genebanks, and increasingly restricted access to *in situ* materials. Accelerating involvement of the private sector in crop improvement, in parallel with emerging regulations on benefit-sharing, may lead to significant development of “privatized” collections. The increasing difficulty and urgency of the immediate challenges facing this world may lead to short-term development goals being prioritised over sustainability; this may in turn lead to some national genebanks may have their funding and scope slashed in the future. These changes will influence the future and role of CGIAR banks, for example to conserve legacy materials.

CGIAR has had, and must continue to have, a critical role to play in this continuing evolution of policy. Uniquely among international treaties (which are legally binding agreements among governments), the CGIAR Centres with genebanks have legally binding agreements with the Governing Body of the Plant Treaty, recognizing the special history and situation of the CGIAR genebanks and their mandate to manage genetic resources for the benefit of the international community. CGIAR is thus ideally positioned to address the need to ensure that the benefits arising from the use of genetic resources do indeed reach the global south fairly and equitably: since CGIAR is funded primarily by high-income countries and mandated to help developing countries through building capacity and transferring technologies to them, the need fits perfectly with CGIAR’s mandate and mission.

5. The genebank of the future

The ideal future genebank will be part of a rational, efficient and effective system in which genebanks work in close partnerships with each other, and in harmony with the scientific and the policy dimensions of research-for-development, ensuring that the benefits of innovation reach those who most need them.

It will meet the needs of immediate users more accurately and efficiently than at present, including those of a burgeoning community of researchers needing both material and in-depth genomic information. It will achieve this by using a wide range of technological advances in genomics, phenotyping, informatics, bioinformatics, and process and materials management, and by interacting more closely with the user community, to ensure that it conserves the right genetic resources in the right way, and closely matches resources to the needs of users.

The ability to ensure delivery of the right material to users will vastly increase the Return on Investment in genebanks, and it may even change the genebank’s funding model. Today, contrary to normal practice for other services, whereby users pay for services provided, genebanks pay to provide genetic resources to users. The justification is that users need to be encouraged to broaden the diversity of materials they use, and, as the genebank cannot know which accessions will actually help any given user, users will not pay for such services. Once genebanks start delivering well-targeted materials that meet users’ needs, the normal “user pays” funding model will work for genebanks as well, subject to the provisions of the Plant Treaty.

The introduction of a user pays model will also enable refinement of the role of the endowment fund of the Global Crop Diversity Trust. This fund provides guaranteed, long-term, ring-fenced funding for key genebanks, to ensure that long-term sustainability in support of future generations is not compromised by the vagaries of funding for short-term needs. If users pay for services to meet their immediate needs, the endowment fund can be focussed more specifically and precisely on those services that meet the needs of future generations.

Conserving and delivering the right resources

The emergence of low-cost, high-throughput genomic sequencing and genome annotation technologies fundamentally changes the ability of genebanks to conserve and deliver the right resources. The majority of genetic variants such as SNPs have little or no effect on phenotype, and knowledge of such variants has little value to genebanks and their users. The key is to identify which genetic variants have real-world effects. Whole genome sequences help reveal the functional variants, including for example through pangenomics analyses that enable the discovery of structural variations that cannot be seen with genotyping. However, having served their purpose of revealing the functional variants, the whole genome sequences themselves would not be an essential component of a catalogue of such variants.

The development of a comprehensive catalogue of the functionally significant genetic variants of each accession has become a feasible target. Even with as many as a million genomes per crop for 20 crops, with around 25,000-75,000 genes per crop genome, the data in the catalogue would require only about 2 terabytes of storage capacity. This is tiny relative to modern “big data” applications, and readily tractable. It is a game-changing contrast with the challenge to effective conservation and use caused by relying solely on the never-ending treadmill of phenotyping. For the first time ever, genebanks will be able to make reliable, objective decisions on what genetic resources to conserve and deliver.

Given the rate of progress to date, including automated algorithms for genome annotation, it should be possible to build an initial, reasonably comprehensive, multi-crop catalogue of functional genetic variants within 10-20 years. However, change is forever, and the catalogue would need to be progressively refined for decades after that.

In the ideal genebank of the future, every gene (including coding, non-coding and regulatory regions) within the crop genepool will be known, and every existing potentially functional variant of each of these. The majority of those variants will have their phenotypic effect either predicted or empirically demonstrated in at least one environment, genetic background and epigenetic status. The genebank’s catalogue will document the variant of every gene of at least one genome from every accession (in cases where accessions are genetically diverse). The genebank will also have access to the equivalent information for all relevant genetic resources held outside the genebank – critically important to enable the genebank and users alike to discover what the genebank has that users (and indeed other genebanks) do not already have.

Using such information on functional genetic variants, the future genebank will have estimated a “future value” for each accession. Specialist analysis will be needed to determine the financial implications of this concept. For the purpose of this document, we simply define the future value of an accession as the extent to which it provides options for addressing future challenges and opportunities that would be impossible with current varieties. This is a function of the number of functional genetic variants (including epigenetics factors, structural variants and transposable elements) that are present in the accession but that are either not known, or at risk of extinction, outside the genebank.

Varieties may also be conserved long-term because they are prized for their "heritage value" for a country or community. The importance of such material typically resides in the variety as a whole; that is, it is a function of the entire genome, rather than of specific rare genes or gene combinations. Currently, international genebanks conserve such varieties as a service to the country or community concerned and to the international community. This will need to continue unless there is significant investment in the capacity of developing countries to conserve their own heritage varieties safely and securely.

Accessions with high future value or heritage value will form the core future-oriented component of the collection, safely conserved for the long term, enabling adaptation to unknown future challenges, and enabling breeders to change breeding objectives as market needs change.

On the other hand, advanced algorithms, which are already transforming the ability of breeders to make the right breeding decisions faster, based on genome-wide selection, will be used by the genebank of the future and its users to explore, through digital evolution, the likely consequences of combining different genomes. The result would be a purpose-specific "current value" for each accession, i.e. the extent to which that accession could enable a breeder or researcher to meet their needs. This would be used to select the most appropriate materials for specific current users. It could also be used more proactively, not just to select the most appropriate accessions, but to guide the creation and management of a large, dynamic set of user-oriented accessions, pre-bred by the genebanks (or others) and designed to meet current needs of researchers and breeders as effectively as possible.

New accessions would be added to the genebank's collection only where assessment of their genomes or heritage value demonstrates that they add significantly to the collection's overall value. The technology already exists to obtain a genome sequence in the field in real time[11]; this would be used to sequence a sample determine if the sample should be added to the collection or discarded, based on its complementarity to the existing collection.

In contrast, current accessions found to contribute little or no value to the collection could be archived or otherwise removed from the collection. Given the objective and relevant measures of value, such decisions could be made with a confidence that is not currently possible.

Conserving and delivering resources in the right way

FAO genebank management standards require that every accession be: (1) viable; (2) healthy; (3) genetically true-to-type; (4) protected against risks of loss; and (5) available to use at least internally for genebank management, and usually[12] also for distribution outside the genebank. Their status must be routinely monitored (which may include some destructive sampling, e.g. for viability and health testing); and there must be procedures in place for recovery of accessions that fall below thresholds for viability, quantity, health and genetic integrity.

These basic requirements will not change in the future genebank, but the assignment of a current value, future value and heritage value to each accession will enable management protocols to be tailored to the status of each accession, generating a much higher rate of return on investment in running the genebank. For example, the use-oriented part of the collection (high current value) should not need to be rigorously monitored for viability and should need only minimal protection against risks

---

[11] Parker J et al. Field-based species identification of closely-related plants using real-time nanopore sequencing. 2017. Sci Rep 7, 8345. https://doi.org/10.1038/s41598-017-08461-5

[12] Owners of sovereign rights or intellectual property rights over an accession may contractually restrict its use. Such restrictions limit the current use value of the accession, and may therefore influence decisions on how to manage it.

of loss: if an accession is indeed important for current use, it will be used, and will hence not lose viability or be at risk of extinction. Loss of viability, or extinction, or lack of immediate use (never needing to be regenerated due to low seed quantity), would indeed be good indicators that the accession does not in fact have a high current use value.

Moreover, the way the genebank conserves and delivers resources will be very different. The future genebank will incorporate deep innovations in protocols for process and materials management, to improve quality, increase reliability, enlarge capacity, and reduce costs. Highly repetitive tasks will be automated for consistently high quality, high throughput and low cost. Novel tests and procedures will be introduced where appropriate. Services will be outsourced where appropriate, to other genebanks or to other specialist service providers; and the genebank may in turn provide services to other genebanks. The ideal form of service provision depends on the context; in the next section we consider examples that are likely to be desirable and appropriate within the next 10 years.

6. The path to the future

The above vision of the ideal future genebank will take 20-30 years to achieve. In this section, we dissect out a number of specific areas where genebanks need the most urgent change over the next 10 years. As emphasised by Herrero *et al* 2020 (loc. cit.), this is not simply a technical path: "For systemic change and technological uptake to occur, there often needs to be an iterative process: private industries identify a business opportunity; governments identify the need for systemic change to achieve prosperity and well-being; a dialogue is initiated with citizens to enable attitudinal change; and finally innovations in policy, institutions and public investment encourage market shifts". Most of these elements are externalities beyond the genebanks' control, although genebanks will need to engage in appropriate fora, both as influencers of, and responders to, external change.

Conserving and delivering the right resources

The ability to conserve and deliver the right resources relies on being able to gather and use the information needed to identify, and where appropriate to acquire or create, the resources that satisfy specific, defined needs. Here we consider the various elements required to do this.

Evaluation

Phenotyping to assess agronomic value (evaluation) is often highlighted as the most severe bottleneck preventing more effective use of genetic resources. In response, some genebanks have drawn the questionable conclusion that *ad hoc*, opportunistic evaluation is better than nothing. Genebanks need a thorough review and revision of their approach to evaluation over the next ten years. This needs to include careful, dispassionate, strategic consideration of the purpose and limitations of phenotyping.

Part of the bottleneck is simply that phenotyping is very slow, laborious and expensive. Many agronomically important traits have low heritability, are highly variable across locations and across years, and show high GxE interactions[13]. Hence experiments to evaluate the potential agronomic value of materials have to be replicated, and repeated in multiple years and multiple locations, making them many times more costly than simple characterization trials. Locations must be representative of environments in potential target markets, which may be far distant from the researcher. For traits of response to the environment (such as diseases, soil quality, or temperature), specialist multi-treatment trials are needed for each response trait, with a further multiplication of costs. And some traits are intrinsically costly to measure, such as root traits and photosynthetic efficiency.

---

[13] A statistical interaction whereby the effects of a gene or genotype on a phenotype depends on the environment of growth.

Recent advances in high-throughput and image-based phenotyping will partly address this problem. However, investment will be needed to make these appropriate for genebanks. This must recognize that the scale of the problem is such that full phenotypic evaluation of every accession is impossible. It must also recognize that often the limitation is lack of knowledge about what to evaluate: research on developmental biology and the factors controlling yield and quality will progressively refine understanding of what biochemical, physiological, anatomical and morphological traits need to be assessed at different stages in a crop's life cycle and under different conditions. Phenotyping protocols must adapt accordingly. They will do so unendingly.

It also needs to be recognized that even if we could achieve the impossible and phenotype every accession, this would still have limited value because of GxG[14] interactions, because of the small effect of many genes, and because of epigenetic variation. The phenotype of an accession is only a weak indicator of its breeding value. This is fully revealed only by testing the progeny of crosses of the accession with elite varieties, not the accession itself (just like the stud value of a racing horse is determined by how many races its progeny have won more than by how many it has won).

The solution has to be better targeted, or "smart phenotyping", i.e. phenotyping only when the case has been made that the phenotyping will generate a sufficient return on investment. Breeders, for example, often undertake only observational trials to initially explore novel genetic resources that they know nothing about. They invest in progressively more, and more detailed and precise, phenotypic assessments as their breeding materials come closer to being commercially viable varieties, extending to dozens or even hundreds of locations in their target area before actual variety release. Genebanks need to recognize this principle: phenotyping must generate a reasonable return on investment (both in terms of financial investment, and in terms of scientific scrutiny – the confidence that users find value and assurance in what data that genebanks provide), or should not be undertaken.

Over the next ten years, genebanks will need to develop, justify and implement crop-specific smart phenotyping strategies. These will build on a number of existing methodologies and approaches[15,16]. The common feature is that they will evaluate, using a subset of methods, only subsets of accessions for subsets of possible traits. The accessions, the traits to be measured and the methods used will change over time. Subsets of accessions will be chosen to maximise the increase in knowledge of the diversity and genetic control of the traits in question. The knowledge gained will be used to direct future evaluation – including, where appropriate, testing of carefully generated progenies to determine their breeding value.

Characterization

Separate consideration must be given to the phenotyping of traits that characterise accessions – typically the size, shape and colour of various parts of the plant, and its phenology. These traits, also known as characterization descriptors, botanical descriptors, or simply descriptors[17], are highly heritable, easily visible and simple to record, consistently displayed regardless of environment, and

---

[14] A statistical interaction whereby the effects of one gene on a phenotype depend on what other genes are present in the genome.

[15] Crossa J et al. 2016. Genomic Prediction of Gene Bank Wheat Landraces. *Genes, Genomes, Genetics* **6** (7), 1819-1834; https://doi.org/10.1534/g3.116.029637

[16] Khazaei H et al. The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. 2013. PloS one **8**(5): e63107. doi:10.1371/journal.pone.0063107

[17] Appropriate terminology is not universally agreed. In some usages, the term "traits" refers only to traits that are important for determining agronomic value, and "descriptors" only to traits that are important for identifying the variety. In other usages, they are referred to as "evaluation descriptors" and "characterization descriptors"

vary among varieties. They are used by breeders in **DUS** testing for Intellectual Property claims through Plant Variety Protection, to prove that a candidate new variety is indeed **D**istinct from other varieties, **U**niform and **S**table from generation to generation. Characterisation in genebanks has a similar purpose, i.e. variety identification. In contrast to evaluation, it is necessary to characterize *all* accessions in a collection.

With advances in genotyping and sequencing, DNA-based data are becoming more precise than phenotypic characterization for variety identification, although they are still not effective for heterogeneous accessions. Genebank managers will need to plan a progressive (though not complete – see below) migration from phenotypic to appropriate DNA descriptors.

The migration from phenotypic to genotypic descriptors will not change the need to maintain a physical sample of the most original available material (often known as a seed file in seed genebanks). This will remain essential for quality control, in particular for purposes such as sorting good seed from bad seed, where hundreds of seeds from a new harvest may need to be screened every second – an impossible task for DNA descriptors. For this purpose, automated image-based phenotyping needs to be introduced more routinely, bringing the potential to record phenotype more precisely with a "digital seed file" and to use this as the basis for high-speed authentication. For example, with a controlled incident white light and a multi-spectral imager, colour can be recorded much more objectively than by eye, on the human-friendly CIE-Lab scale rather than RGB (red-green-blue). Similarly, size and shape can also be recorded more precisely and objectively by machine.

Special consideration must be given to traits at the intersection between evaluation and characterisation, i.e. traits that are good descriptors but also have agronomic significance, such as the orientation of leaves in the canopy. As it is feasible to record these economically for every accession, they may need their own phenotyping strategy.

Genotyping and sequencing

As described above, the long-term objective for genebanks is a digital catalogue of the functional genetic variants present in each accession, linked to corresponding information on genomes outside the genebank. This will enable DNA-based decisions on conservation and, in conjunction with knowledge of gene function and associated phenotypic data, on use. In terms of information management, this is readily achievable and tractable.

Over the next 10 years, the migration of genebanks towards DNA-based decisions will be fast, but must be crop-specific and responsive to emerging developments in sequencing and genomics. The optimal balance between whole-genome sequencing, partial sequencing, and genotyping depends on the complexity and diversity of the genomes, the breeding system of the crop and the amount of within-accession genetic heterogeneity, and is changing as the technology evolves and knowledge grows. The same applies to progress towards directly identifying the functional genetic variants, as opposed to markers that are genetically linked to those variants.

At present, the genome of most crops is too complex for routine whole-genome sequencing of more than just a few accessions. Large-scale, whole-genome sequencing has been possible for a few crops, for example rice, with its relatively small genome, but even here it remains unclear how many more genomes must be sequenced to reveal all the genes and all the functional genetic variants that can be found in the rice genepool. It also remains unclear whether or when whole-genome sequencing will become so routine that the standard DNA fingerprint could become a whole-genome sequence.

Hence, each crop will need its own strategy for progress in genotyping and sequencing. The key common theme, however, must be a route towards cataloguing functional genetic variants, and not simply building up vast quantities of data on DNA sequences with unknown function. In particular, many or most SNPs appear to have no clear functional significance; management decisions based on these may be highly suspect, whether they are decisions on conservation or on use. For example, rationalising a collection based on a threshold level of similarity from all known SNPs is not ideal, as it may miss a single variant of high importance.

A second common element of the strategy will be how to manage material that has been sequenced. For most crops, each accession is a population of genomes with different sequences and different genetic variants. For the purposes of gene discovery, it is important to phenotype the exact same genome that has been sequenced or genotyped. This requires the sequenced genome to be managed, conserved and distributed separately from the accession from which the genome was taken. However, if all accessions are genotyped or sequenced, and a pure line is conserved and distributed separately from the original, the size of the collection would be (at least) doubled – an expense to be avoided.

Implementation of the "smart phenotyping" strategy described above will be important here: the purified line should only be conserved in the genebank if it is to be prioritised for phenotyping in the near future. If the sequence information is to be used for decision-making without the prospect of immediate phenotyping, it would be economically more viable not to maintain the pure line, and instead to repeat the purification and sequencing at some future date, following a decision to phenotype. Ideally, decisions on which purified lines should be maintained for phenotyping will be supported by "smart algorithms" that objectively quantify the importance and urgency of phenotyping the lines.

Breeder- and researcher-oriented materials

Distinguishing between the current and future use value of accessions brings the opportunity to optimise management of different parts of the collection specifically to meet the current needs of breeders and researchers. This in turn requires closer interaction with breeders and researchers, to achieve a deeper understanding of their needs. Traditionally, genebanks have viewed their role as supplying material that the breeders do not already have. However, the lack of suitable diversity to breed with is only one of several constraints that breeders may have. More commonly, their main problem is an inability to measure their target traits with sufficiently high precision and throughput for effective selection.

A range of options needs to be built up over the next 10 years to segment collections. At its simplest, this can be further enhancement of efforts to define relevant subsets of accessions. Depending on the need, these may be traditional core or mini-core subsets intended simply to make the task of phenotyping more achievable. Alternatively, accessions may be selected based on specific user-defined criteria - combinations of passport, phenotypic and genetic data - to create subsets that are more likely to contain what the user wants.

As has already been demonstrated for rice, demand for subsets can be radically increased by sequencing all the accessions in the subset and making that information available to users. The reason is that this enables users to conduct their own genome-wide association studies, which is becoming an increasingly important first step in understanding the genetic control of a trait. In this case, a single sequenced subset can be used to support gene discovery for multiple traits. Hence a short-term objective for genebanks should be to replicate this for all crops, by sequencing the genomes of diverse core collections of all crops.

In addition, the genebank can and should become more proactive in designing and creating novel genetic resources in support of breeders and researchers. Importantly, it must complement rather than duplicate breeders' own pre-breeding work, and hence must undertake such efforts in consultation, and collaboration, with breeders. Breeders' pre-breeding initiatives are typically trait-based, focusing on introgressing high-value traits from undesirable genomes into elite breeding lines. Depending on the scope of the breeders' initiatives, genebanks may play a role in the trait-based approach by "pyramiding" multiple known high-value traits into easily useable materials. Alternatively, they could be more exploratory, combining divergent genomes that have never previously been crossed, with the aim of exposing large amounts of novel phenotypic diversity by creating radically different genomes, supporting rapid response to change. A range of possible crossing designs already exist, such as MAGIC (Multi-parent Advanced Generation Inter-Cross) and NAM (Nested Association Mapping); their exploratory value can be maximised by using genomic information to select the parents.

The resulting materials could become a large proportion of the genebank's collection, but would need to be highly dynamic, relinquishing combinations that are unwanted or that become unwanted.

Digital genebanks

As effective conservation and use of genetic resources becomes increasingly dependent on precise genetic understanding of the potential contribution of any accession to specified breeding and research targets, genebanks will need to provide online access to detailed genetic and phenotypic information about all the accessions they hold. This will be the primary resource guiding management and use of the collections.

The challenge of developing an appropriate information system is vast. Much of the genetic and phenotypic information needed is outside the scope of information for which genebank standards have been developed. Most of the information is collected by disparate communities of users, who follow their own standards, and would not follow genebank standards even if they existed. Moreover, the information is needed for genetic resources that are not in genebanks just as much as for those that are, so that users can determine what novel options can be provided by the genebank. Hence the "digital genebank" needs to be part of a global information system on plant genetic resources, in the form of distributed systems based on shared standards and best practices, with participants including everyone who creates information on genetic resources.

The depth, complexity and amount of this information requires infrastructure and expertise that will be far beyond the resources of any single player. An effective global information system will likely take more than ten years to develop, and is perhaps the most speculative, wishful component of the new vision. Technically, it need not take so long, if investment and planning were mobilized, but these will need to be considerable, bringing together communities that currently work largely independently on various aspects of plant biology, and making their disparate data management systems jointly searchable.

Genebank managers will need to focus on realistic targets, developing linkages with other players who specialise in big data, and contributing the elements that can only come from genebanks. They will need to add the information they create directly, and they will need to work with their users to help ensure that user-created information is also added to the global information system. The genebank manager will need to ensure that the information remains connected to the physical material that they maintain in the genebank.

The task of building the global information system, managing the huge amounts of genetic and phenotypic data, and providing a searchable interface, would have to be undertaken by (and in collaboration with) specialists outside the genebank community. The beginnings of a global information system are in place (http://www.fao.org/plant-treaty/areas-of-work/global-information-system/en/: GLIS). In addition, a large number of initiatives also exist to develop different aspects of the required infrastructure, such as GODAN (Global Open Data for Agriculture and Nutrition:

https://www.godan.info/), Genesys for genebanks (https://www.genesys-pgr.org/), GGBN (Global Genome Biodiversity Network: http://www.ggbn.org/ggbn_portal/), DivSeek International Network (https://divseekintl.org/), crop-specific consortia such as IRIC (International Rice Informatics Consortium: http://iric.irri.org/), and many others. These all need to be linked and searchable through GLIS to build a truly effective global information system in which genebanks can operate digitally.

The first step towards making all these systems searchable through GLIS is to identify model use cases (specific examples of cross-database searches that would most help users) and to explore options to address concerns about fair and equitable sharing arising from the use of digital data. So far, even the specification of use cases has been limited to individual initiatives such as those named above.

Conserving and delivering resources in the right way

Genebank standards are well developed[18], and CGIAR genebanks have a good record of adherence to these standards. However, there remain a few problem areas in urgent need of improvement, which in turn requires investment in conservation research to determine how to improve.

Beyond these problem areas, the ability to conserve and deliver resources better is primarily about organizational optimisation. Procedures for managing materials, information and processes must be streamlined to maximize efficiency, maintaining consistently high and demonstrable quality control while reducing costs in a system that matches throughput capacity to demand.

Seed longevity in storage

The basic framework for understanding, improving and predicting seed longevity in storage was established in the 1970s[19], based on moisture content and temperature. However, recent research has demonstrated that this is unsatisfactory. Unexplained variation in longevity in storage is very high, genotypic specific responses have been demonstrated, many samples do not live as long as predicted by the seed viability equations, and current viability testing standards do not adequately predict the longevity of seeds in storage. This in turn necessitates high expenditure on testing viability and on rejuvenating seeds more frequently than should otherwise be necessary.

Critical research is ongoing (e.g.[20,21]), but building on this research effort needs to remain a vital element of improving how genebanks work, such as through improving understanding of dormancy and aging processes in seeds, and improved methodologies for testing viability and packing seeds.

Cryopreservation

Cryopreservation is still challenging for many crops. After more than 20 years of research and protocol development, CGIAR has managed to secure more than 50% of the banana and potato collections it manages into liquid nitrogen at -196°C. This is a game-changing technology, that for the first time secures vegetative propagated and recalcitrant seed species in long-term storage, precluding the need for annually replanting and regenerating live materials in the field and tissue culture laboratory. CGIAR is one of the very few leaders in the discipline and an opportunity exists not only to ensure that other relevant crops in CGIAR-managed collections – including seed of recalcitrant species - are cryopreserved, but also for CGIAR to provide a specialist service for the cryopreservation of collections

---

[18] FAO. 2014. Genebank Standards for Plant Genetic Resources for Food and Agriculture. E-ISBN 978-92-5-108262-1. Available online at http://www.fao.org/3/a-i3704e.pdf

[19] Ellis RH & Roberts EH. 1980. Improved equations for the prediction of seed longevity. Annals of Botany, 45, 13-30.

[20] Hay FR et al. Seed longevity phenotyping: recommendations on research methodology. 2019. *Journal of Experimental Botany* **70** (2), 425–434, https://doi.org/10.1093/jxb/ery358

[21] Fleming MB et al. Exploring the fate of mRNA in aging seeds: protection, destruction, or slow decay? *Journal of Experimental Botany* **69** (18) 4309–4321, https://doi.org/10.1093/jxb/ery215

worldwide, including potentially of crops such as coconut, coffee and cacao that are currently not included in the mandates of CGIAR genebanks. The application of cryopreservation technology requires highly skilled staff, long-term planning and sustained execution, which is impossible to achieve without reliable long-term investment. This is an area where CGIAR's mandate plus its related strengths in phytosanitary research and control puts it in a unique position to play a global role to great effect.

Automated process and materials management

Large organizations have achieved major improvements through automating repetitive tasks and also the progression through multi-task processes. Although requiring substantial initial investment, automation has been essential for survival, simultaneously reducing costs, increasing throughput and enhancing quality, albeit sometimes requiring a change in expectations and standards. The human eye and brain will remain the most powerful "super computer", but automated approaches, while performing at a different standard, may be the most efficient. Inter alia, they enable 24/7 working and facilitate a complete redesign of workflows, for the better.

An example is the process of retrieving and preparing seeds for use. Currently, technicians enter the cold store with the list of packets to be retrieved. Each packet is typically on a different shelf, so servicing a typical request for a few hundred samples may require the technicians to remain in the cold store (at 0°C) for 1-2 hours to visit all the relevant shelves and retrieve the packets. They then return to the preparation room, and leave the packets for a day to equilibrate without risk of condensation. The next day, they open the packets, separate the samples into seeds for use and remnant seeds to be returned to the cold store, and re-dry, package, seal and label. They then take the remnant seeds back to the appropriate place in storage. Genebanks processing a large number of samples per year typically prepare some samples for use in advance, to minimize the number of times they have to open up and repack bulk samples in storage. If the whole process were automated on a mobile platform, it could be completed inside the cold store, thus halving the number of visits to the shelves and eliminating the potentially damaging effect of fluctuating seed temperatures and moisture content. It would also address potential Occupational Health and Safety hazards, which have already driven some national genebanks to automate storage and retrieval in order to eliminate the need for staff to move frequently between cold store and room temperature.

Another example is the maintenance of clonal accessions in tissue culture. Unlike seed accessions, which can be preserved alive for decades at very low cost, cultures have to be regenerated much more frequently, requiring labour-intensive re-culturing. The cost of maintaining such cultures is orders of magnitude greater than the cost of preserving seeds; indeed, clonal collections are far more expensive than seed collections despite being far smaller. The repetitive nature of re-culturing makes it ideal for automation, and the benefits of automation will be correspondingly great.

Automated process and materials management also introduces the possibility of tighter control through remote management, allowing managers to control processes no matter where they are.

Some preliminary attempts have been made to automate various operations in large genebanks, for example for planting, phenotyping, harvesting, seed sorting, viability testing, packing, labelling, and the storage and retrieval of materials from a seed store. This includes, as an example, remote management of seed sorting. The most widely adopted and successful advance has been the introduction of bar-coded or QR-coded labels for inventory management and for tracking samples through workflows.

With the demonstrated benefits of automation in several areas, it is timely and important to now undertake a much deeper revision of process and materials management, especially in the larger genebanks. Many small, independent initiatives have been undertaken already by various genebanks supported by various robotics companies; the time is ripe to create a global genebank consortium to invest in automation R&D.

Crop wild relatives, forages, trees

This group of species represents a special case. They span a wide range of biologies, with highly diverse, specific requirements for growth, harvesting, storage, dormancy breaking etc. Hence, they require specialised and highly tailored conservation protocols. Moreover, the costs of conservation and use are relatively high, and therefore collections are generally small, and throughput is low.

For all these reasons, these difficult species are likely to benefit the most from conservation research and specialisation. With high diversity and low throughput, they are unlikely to benefit significantly from automation in the same way; many activities will probably continue to be managed manually by specialist staff.

Phytosanitary controls

CGIAR germplasm health units are pursuing international recognition for their phytosanitary controls under the so-called "GreenPass" system, which would streamline movement of material from genebank to users and vice versa (https://www.genebanks.org/news-activities/news/greenpass/). This is an important area of consideration for all genebanks, but too large to explore here.

Service provision

Over the next 10 years, genebanks must also move away from the concept that every genebank must do everything in-house. As appropriate, genebanks may outsource services to other genebanks or to other service providers, and they may indeed usefully offer services to other genebanks.

Having said this, some core tasks must always be undertaken in-house by all genebanks:

- The basic quality management and data management (but not software development) required to manage the genebank
- Maintenance of active collection for distribution to at least local and national users
- Participation in relevant networks
- Interaction & collaboration with users & stakeholders within the country
- Providing training, raising awareness, and lobbying for support

The concept of service provision needs to be addressed not only among CGIAR genebanks, but among all genebanks and potential service providers. CGIAR genebanks could offer to undertake key services on behalf of national genebanks if they need it; or CGIAR genebanks could outsource some services to NARS and national genebanks in developing countries that have the requisite capacities. Beyond promoting efficiencies from a technical perspective, this could be a key way for developing countries to buy into and participate in the global system, and thus contribute to benefit-sharing as envisioned by the Plant Treaty. Hence, this has political as well as financial implications.

No specific recommendations for outsourcing are made in this section, since they would depend on higher-level decisions on how CGIAR genebanks will be organized in the future (see the next section). This section merely presents a number of options. These must be reviewed at system level for each genebank to determine the optimal solution for each service in the context of the system as a whole.

For example, to ensure secure conservation, every genebank must have an active collection available for use by a defined user community; but long-term storage could be provided by another genebank, and safety backup storage *must* be in another genebank (ideally on a different continent). Genebanks with excess capacity for long-term conservation may provide it as a service to genebanks that lack that capacity. Similarly, effective international distribution requires specialist technical and legal knowledge of phytosanitary procedures and transboundary movement of genetic resources.

Optimising service provision needs to take into account the cost (time and money) of movement of material to a service provider, especially international movements that are subject to additional regulations and phytosanitary testing. Countries in the same geographic or political region may have agreements minimizing such regulations and testing, so special care must be taken if considering outsourcing to a service provider in a different geopolitical region.

Examples of services that can be outsourced by some genebanks, to other genebanks or beyond, include:

Service	Comments
Safety backup	Offered by only a very few genebanks, including the Svalbard Global Seed Vault
Long-term base conservation	Offered by major genebanks
Active collection for global distribution	Offered by international genebanks
Active collection for regional distribution	Offered by international genebanks
Regeneration	Must be in a suitable environment for the material
Re-culturing tissue cultures	Offered by specialist tissue culture laboratories, which may be based in a genebank conserving clonal material
Phenotyping	Must be in a suitable environment
Health & viability testing	Only feasible to outsource in the host country *
Genetic integrity testing	Offered by specialist DNA labs (for DNA-based testing, both generally for conformity to the accession standard and specifically for unintentional presence of transgenes or other contaminants)
Research in seed longevity	Offered by specialist genebanks/universities
Conservation of CWR, forages, trees	Offered by specialist genebanks
Cryopreservation research	Offered by specialist genebanks/universities
Large-scale implementation of cryopreservation	Offered by specialist genebanks
Phytosanitary control	Required to different levels depending on whether germplasm distribution occurs at a national, regional or international scale. Intercontinental distribution could be offered by specialist germplasm health units.
Research in seed longevity	Offered by specialist genebanks/universities
Training & capacity building in services to be undertaken in-house	Offered by capacity building specialists
Georeferencing & spatial analysis	Offered by GIS specialists
Genomic services	Offered by specialists – usually not genebanks
Big data management	Offered by specialists – usually not genebanks
Providing public access to data	Offered by specialists – usually not genebanks
Engaging in the policy discussion	Offered by policy specialists

* Normally, health tests must be in-country because they are prerequisite to export, and viability tests should be in-country to avoid the cost of associated health tests for out-of-country tests. There may be a few exceptions where neighbouring countries have specialised agreements enabling quick and easy transfers of material for testing.

CGIAR system

How the CGIAR system is restructured under One CGIAR will have a major influence on how the CGIAR genebanks should, and could, be organized. Currently, the disposition of CGIAR genebanks is largely the result of historical legacies, and to a degree is commodity-based, in line with the largely commodity-based organization of CGIAR Centres. Changing CGIAR to frame everything around transforming water and agriculture systems in the context of the climate crisis, rather than around commodities, presents a fundamentally new need and opportunity to restructure the genebanks accordingly.

In particular, such reorganization could potentially present the opportunity to merge the eleven CGIAR genebanks into a smaller number. This could enable considerable cost-savings through economies of scale. In particular, noting that the benefits of automation increase with increasing size of organization, consolidating into a small number of large genebanks will increase the importance of automating all critical tasks that are amenable to it. Equally importantly, it provides an opportunity to reconfigure collections at both Centre and crop levels, dispersing different subsets of a collection among more than one of the geographical locations, to ensure that germplasm in demand is located where it is most needed. The political challenges of such a reorganization may be insurmountable but the concept is still worth exploring.

In an extreme form, it could even take the form of a single central genebank responsible for long-term conservation of *all* crops, combined with a set of "hubs" in strategic locations responsible for regeneration and distribution to users. There are effective, internationally-renowned, national-scale precedents for this structure, for example in the USA and India, which have central facilities respectively in Fort Collins and New Delhi. In practice, for both political and phytosanitary reasons a single, central global facility is unlikely to be workable, but it may be possible to envisage three CGIAR facilities focussing on long-term conservation, in Asia, Africa and Latin America, while a larger number of hubs carry out distribution and regeneration functions.

Such a separation of function would bring the potential to achieve major efficiencies of scale on conservation, together with the benefits of specialising different activities in different locations, while ensuring that regeneration and distribution are undertaken in the most effective, fit-for-purpose locations. Focussing only on seed crops, a possible division of responsibilities is shown in the following table:

Facility		**Central facility**		**Hub ***
Storage conditions **		**LTS**	**MTS**	**MTS**
Activity	**Viability test**	Yes	Yes	No
	Health test	Yes	Yes	Phytosanitary compliance
	Genetic integrity test	DNA fingerprint	DNA fingerprint	Digital seed file
	Distribution to users	No	No	Yes
	Regeneration	No	No	Yes

* Regeneration hub is located in a good environment for production of high-quality seed; distribution hub is located in a convenient location for international distribution. Where appropriate, a regeneration hub and distribution hub may be in the same location, and the central facility could also be a hub
** LTS = Long-term storage; MTS = medium-term storage

An equivalent scenario would apply to clonal crops, with the following differences:

- Cryopreservation in the central facility would be the equivalent of long-term storage for seed crops;
- Viability tests would be for viability or vigour of plantlets or tissue cultures rather than of seeds;
- Digital images of reference plant materials instead of digital seed files.
- Regeneration and rejuvenation of accessions would occur centrally to minimize international movement of clonal material.

Such a scenario would introduce the possibility of an entirely new type of high-volume, high-tech, robotic genebank that can be operated by a smaller number of staff than a typical manually-operated genebank. This in turn implies a different composition of managerial and technical staff: managerial staff would have much greater responsibility; a new breed of technical staff would be required for machine and process maintenance; and there would be a shift in staff balance, with a higher proportion dedicated to information processing and management.

Rather than merging by geopolitical region, an alternative basis for merging genebanks could be by agricultural system, in line with the revised CGIAR structure. However, considering the increasingly global distribution of most crops, this would probably not be appropriate.

Another, more rational basis would be to merge genebanks by crops that require similar types of operations for conservation and use. This would maximise the benefits arising from any economies of scale. At a minimum, we can envision one seed genebank and one clonal genebank. There is very little operational overlap between these two crop groups, and a high degree of operational overlap among crops within each of the two groups. There are thus major benefits to be gained from such a merger, with little additional benefit from further merging. Species maintained partly as clones and partly as seeds would be maintained in both genebanks.

Among the seed crops, conservation protocols vary with breeding system (self-pollinated, cross-pollinated by insects, cross-pollinated by wind), plant habit (annual, perennial herb, tree), and seed storage behaviour and size. Where protocols differ significantly, clearly centralisation brings little benefit from economies of scale. Hence, it may be appropriate to consider more than one seed genebank, perhaps based on breeding system, or perhaps creating a specialist centre for wild species and trees, these being the most expensive collections to maintain, with the most to gain from specialist management.

The logistics of distributing material to users is a major factor in centralization. International shipment incurs high costs and significant delays, resulting from the need to comply with onerous but essential phytosanitary, biosafety, policy and import/export regulations. The issue is especially problematic for clonal crops, where risk that material dies before it reaches the intended recipient is significantly higher than for seeds. The scale of the difficulty may decrease if the ability to select the right accessions results in a reduction in the distribution of material, but even so remains a key consideration.

In a number of cases, countries that rely on close collaboration have sought to minimise such problems through specialised bilateral or regional agreements that streamline transboundary movement between member countries. Genebanks can and should benefit from the strategic placing of distribution hubs within key regions or geopolitical blocks which share phytosanitary risks. Hence, the consolidation to fewer genebanks would go hand-in-hand with a number of distribution hubs. Furthermore, operating multi-crop distribution hubs will be significantly more efficient than the current system where each Centre determines the need for such hubs in the areas of their focus regardless of where other Centres are operating.

Options for centralisation will be limited to some extent by the wide differences in adaptation among crops. Genebanks already have to accommodate wide differences in adaptation among varieties in their collections, in some cases outsourcing regeneration to more appropriately located partners. Considering whether and where regeneration hubs should be located should take account of various environmental criteria, such as:

- It is possible but expensive to grow crops in a climate with unsuitable rainfall, such as dryland crops in Philippines or rice in Syria.
- Some species and varieties are killed by the same low or high temperatures that are essential for the growth of others. This necessitates the use of multiple growing locations. For this purpose, the most economic location for the genebank headquarters is in a country that has both hot and cold areas, most likely a tropical country with hot lowlands and cool highlands (a few illustrative examples include Peru, Colombia, Ecuador, Kenya, Tanzania, India, Nepal). This avoids the overheads of international travel to reach all necessary growing environments.
- Some species and varieties require specific sets of daylength conditions. Accommodating all possible needs requires growing locations at a wide range of latitudes, which are likely to span multiple countries.

Global integration

Globally, genebanks are often viewed as isolationist, not just in wanting to do everything in-house, but also in prioritising the safe conservation of crop diversity over its sustainable use. This view, real or imaginary, is limiting the contribution of genebanks to sustainable use of PGRFA and needs to be changed.

It is important to recognise that genebanks are only one category among the various stakeholder groups that manage crop diversity, and conserve only a small proportion of the total diversity of crops and their wild relatives. Relationships among and within stakeholder groups, at least for a single crop, may be visualised as a triangle:

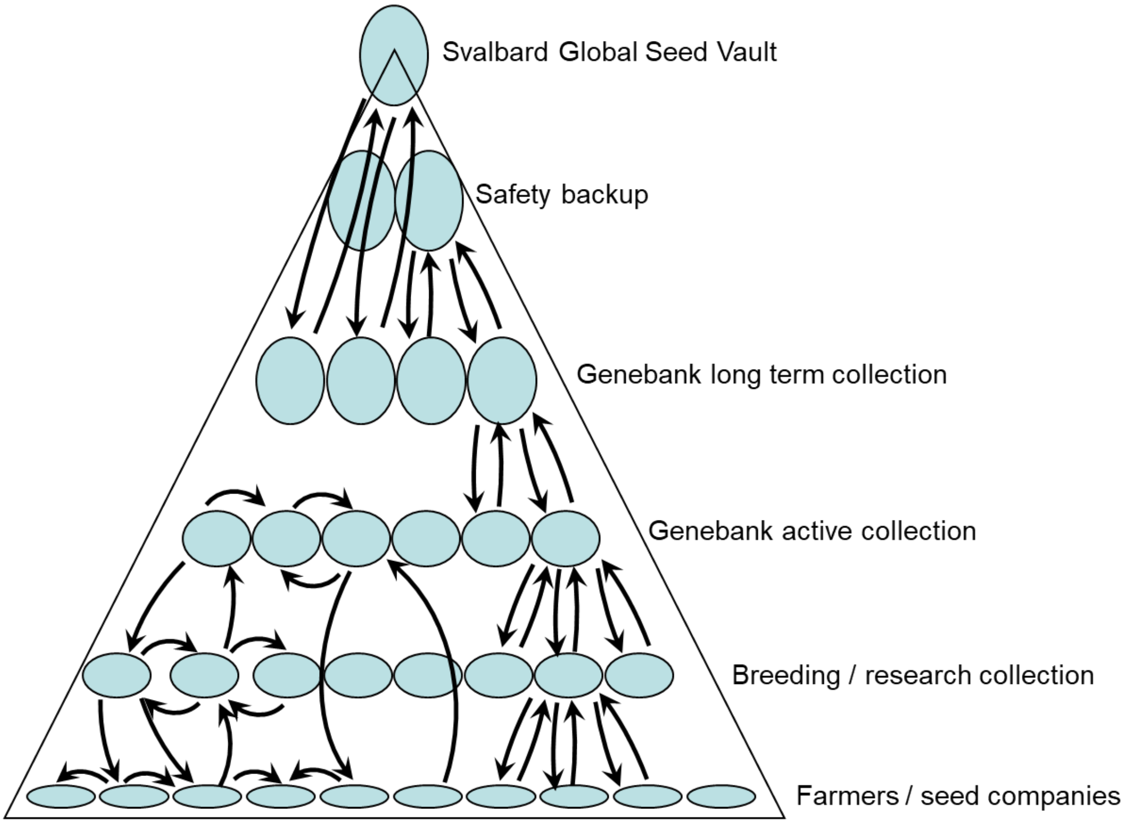

Figure x. A simplified depiction of relationships among stakeholder groups that manage plant genetic resources for food and agriculture (adapted from Global Rice Strategy 2010).

At the base of the triangle, an estimated 570 million largely smallholder farmers manage *in situ* the vast majority of the plant genetic resources on earth. Traditionally, farmers save some of each harvest for replanting the next season or for exchange or sale to other farmers for replanting, while using or selling the remainder for consumption. In a complementary, linked, model, there is a separation of function between farmers (often contracted by seed companies) that produce seed for replanting, and farmers that produce commodities for consumption. Some farmers also play a role as “citizen scientists”, interacting with genebanks, researchers and breeders in their search for better livelihoods from farming. Thus, farmers may source their seeds from their own harvests, from each other (perhaps via “community seedbanks”), from seed companies, or directly from breeders, research groups or genebanks. Every season brings a new decision on what to plant, what to discard, and what to retain, based on current market conditions and opportunities. In future, such decisions will be supported by novel tools currently under development, enabling variety selection, long-range weather forecasts and epidemic predictions.

A level up from farmers and seed companies are the breeders and researchers who focus on improving crop production, through a wide range of activities across the full spectrum of upstream-downstream research for development, such as breeding, pre-breeding, genetics, genomics, agronomy, developmental biology and socio-economics. They include public, private, international, multinational, regional, national, institutional, non-governmental entities. Ideally, there should be breeding and research activities for each crop in every country where that crop is a high national priority, maintaining so-called “working collections.” They focus on improved varieties and production systems for their countries; thus, one breeder supports many farmers. International breeding programmes target a broader range of farmers from multiple countries, and typically achieve this by working in conjunction with national programmes. Breeders and researchers source their seeds from other breeding and research programmes and from genebanks or even farmers, as appropriate. Every season again brings a new decision on what new materials to create, what to discard and what to retain, based on current assessment of market needs, opportunities and bottlenecks in relation to the longer-term strategy.

The next level up from breeders and researchers are genebanks that maintain active collections. Each genebank should service the needs of a specific group of users, most often breeders and researchers but also (especially for under-developed and under-used crops, forages and trees where the difference between traditional and modern varieties is not large) farmers themselves. With annual management decisions based on a longer-term view of change as well as on immediate market needs and opportunities, their collections are deliberately and necessarily less dynamic than those of breeders, researchers and farmers. When they do add new materials into their collections, they may do so from farmers, breeders, researchers or even other genebanks.

Higher levels denote the progressively more specialist elements of safe conservation of accessions in genebanks, with the Svalbard Global Seed Vault providing the ultimate in long-term security of the diversity of crops that can be conserved in seed banks. As depicted in the diagram, long-term conservation is a specialist investment in the future with no short-term return on investment, and not every genebank needs to do it. CGIAR, however, by its mandate and binding agreements with the international community, must have a facility for long-term conservation; and indeed it could and does provide this as a service to other genebanks, especially to smaller genebanks and those in low-income countries.

Linkages must be strengthened within and among these levels, and by extension between *ex situ* and *in situ*/on farm conservation, guided by appropriate external stakeholder groups specialising in economic, social, marketing and national and international policy aspects. Information flows will increasingly be the glue holding the system together. But optimising this is the most sensitive to externalities. It is difficult to set specific targets for the role of CGIAR genebanks within this system of information flows, beyond the need to continue engaging in national and international policy fora. In particular, CGIAR needs to be more proactive in promoting significant dialogue in the context of the Plant Treaty.

7. Conclusion

The innovations proposed in this document would modernise and transform the CGIAR genebanks to be efficient and effective in a way never previously considered achievable, playing an even more indispensable global role in the conservation and sustainable use of plant genetic resources for food and agriculture. Not all of them will be possible. But harmonizing those that are with the ongoing “One CGIAR” change process, the CGIAR genebanks will participate fully in CGIAR’s battle to address the climate crisis. Looking beyond the 2030 Sustainable Development Goals, these innovations will prepare CGIAR to respond rapidly and effectively to future challenges, whatever they may be.