CGIAR Platform on Excellence in Breeding

Tools and services that create synergies and accelerate genetic gains of breeding programs targeting the developing world

Executive Summary
CSIAR Platform on Excellence in Breeding

Tools and services that create synergies and accelerate genetic gains of breeding programs targeting the developing world

Executive Summary

The Proposal in its current form was developed with contributions by the following institutions.

Citation: Genetic Gains Working Group (2016) CGIAR Platform on Excellence in Breeding - Tools and services that create synergies and accelerate genetic gains of breeding programs targeting the developing world. Executive Summary of the Proposal submitted to the CGIAR Fund Council, 31 July 2016. International Maize and Wheat Improvement Center (CIMMYT), Mexico. D.F., MEXICO.
CGIAR Platform on Excellence in Breeding:
Tools and services that create synergies and accelerate genetic gains of breeding programs targeting the developing world

Executive Summary

Contents
1. Rationale .......................................................................................................................... 1
2. Vision .................................................................................................................................. 2
3. Agenda and funding mechanisms ......................................................................................... 4
4. Leadership and management ................................................................................................. 5
5. Module 1: Breeding program excellence ................................................................................ 6
6. Module 2: Trait discovery and breeding tools and services .................................................... 8
7. Module 3: Genotyping/sequencing tools and services ........................................................... 10
8. Module 4: Phenotyping tools and services ............................................................................ 12
9. Module 5: Bioinformatics and data management tools and services .................................... 14
10. Budget and associated outcomes ......................................................................................... 16
Rationale

**Importance of genetic improvement:** Over past decades, genetic improvement has increased agricultural productivity, reduced the pressure on forests, changed the nutritional profile of crops, and made agricultural production more resilient to diseases, pests and droughts. Without genetic improvement, food prices would be substantially higher and more forest land would have been lost to agriculture. It is estimated that between 30-60% of the yield increases in farmers’ fields can be traced back to breeders changing the genetic make-up of crops. Similarly, breeding has changed the resilience and productivity of fish and livestock production.

**The challenges ahead:** In spite of these successes, breeding programs targeting the developing world will need to deliver higher rates of genetic gain to meet the 21st century challenges of 50%-60% greater demand for food commodities, climate change and natural resource constraints. Climate change will reduce crop productivity by about 5% for every degree of warming above historical levels (Challinor et al. 2014). With rates of breeding gains at 1% or less, breeders must speed up their current efforts substantially.

**New opportunities:** The quest for accelerating genetic gains is one of the most rapidly evolving areas in science and technology, with the potential to generate a step change in the practice of breeding. Advances in genomics are just beginning to reveal the wealth of genetic diversity available within crops and animals, while new and more cost-effective genotyping and phenotyping technologies allow advancements to be attained with greater precision and speed, including through gene editing. The application of these technologies challenge breeding programs to capture, process and integrate much greater amounts of data within shorter time frames.

**Challenges to breeding programs that target the developing world:** The pace of technological modernization in breeding programs targeting the developing world is inadequate and varietal turnover in most countries is slow. Due to the diversity of environments, and the low return on investments, only 5% of private R&D investment in breeding goes toward low- and lower-middle income countries, even though they account for 45% of the global area sown to major staples, 48% of the world population and 84% of the world’s poor. Partial modernization of methods and approaches has taken place mostly in the larger CGIAR breeding programs, yet need to be urgently mainstreamed across commodities and also include smaller breeding programs.

**Why a CGIAR-led platform?** The CGIAR is the most prominent source of germplasm and breeding know-how to the developing world, providing valuable genetic diversity both in the form of parental lines for use in breeding and as finished cultivars for dissemination. Indeed, 94% of international germplasm exchange under the International Treaty on Plant Genetic Resources for Food and Agriculture originates from CGIAR centers. Increasing both the rate of genetic gain delivered directly by CGIAR breeding programs and improving their ability to support the modernization of national systems is the key scientific challenge facing the system, and the purpose for the proposed Platform.
Vision

To enable staple crop and animal breeding programs targeting the developing world to make step changes in increasing genetic gains of cultivars and breeds delivered to farmers, for impact on food and nutrition security, climate change adaptation and development at large.

How can the vision be achieved? - Achieving higher genetic gains in breeding programs targeting the developing world – CGIAR, national agricultural research systems (NARS), the private sector – will require the exploitation of economies of scale that can only be achieved by aggregating demand and capacity across programs, institutions, and commodities. Individually, even the largest CGIAR or NARS breeding programs are too small to support rapid modernization by adapting and mainstreaming state-of-the-art breeding technologies such as found in the multinational private sector. Together, coordinated and supported by a common platform, the programs serving smallholders in the developing world can raise the rate of genetic gain they deliver much more effectively.

Using synergies from the CGIAR portfolio: The Excellence in Breeding Platform is to be seen as part of the entire portfolio of research programs and platforms of the CGIAR. It will support the network of partners (CGIAR centers, NARS, local private breeding sector) that are developing new cultivars/breeds and conserving genetic resources within eight Agri-Food Systems CGIAR Research Programs (CRPs) and the Genebanks Platform. Through the Agri-Food System CRPs focused on key commodities1, 11 CGIAR centers maintain breeding pipelines and/or genebanks, access new technologies and know-how from a wide range of public and private sector partners, and help local public and private sector breeding programs improve their own programs.

Figure 1. The Excellence in Breeding Platform develops and makes available tools that accelerate breeding gains across multiple Agri-Food System CRPs.

1 Crops: Banana, barley, bean, cassava, chickpea, cocoa, cowpea, forage crops, lentil, maize, millet, pigeonpea, plantain, potato, rice, sorghum, soybean, sweet potato, triticale, wheat and yam. Animals: buffalo, camel, cattle, chicken, goats, pigs, sheep.
The unique role of the Platform: The Platform agenda will foster best practices, with tools and processes that are effective and adaptable across a broad set of commodities. It will test, adapt and mainstream tools, provide expert advice as well as develop a knowledge base on a broad range of approaches for increasing breeding efficiency. Through the development of international public goods, lower budget and less advanced breeding programs will be able to capitalize on the bigger budgets of more advanced breeding programs and the private sector. The Platform will also broker access to cost-effective genotyping/sequencing, laboratory analysis services, and data analysis capacities, and support capacity strengthening and knowledge transfer beyond first users.

Benefits to small and large programs: Many countries and seed companies directly use the cultivars and breeds developed by the CGIAR while others both use them directly and adapt them through their own national or privately funded breeding programs. A CGIAR-led initiative will benefit both types of programs, either by NARS and local seed companies getting access to better cultivars and breeds, or by larger programs being enabled to prioritize, contribute to and adopt the tools and research management approaches adopted or developed within the CGIAR.

Voices from users:
What value do they wish to see generated by the Excellence in Breeding Platform?

**Major commodities:**
- Support to improve breeding programs and adopt new technologies, taking advantage of group negotiation, economies of scale and reduced opportunity costs
- Genotyping services, breeding information management and standards that support strive for excellence
- Greater capacity to analyze and manage molecular data
- Support for trait discovery in addition to breeding
- Synergistic sharing and learning of best breeding program management practices, knowledge and technologies

**Smaller commodities:**
- Access to otherwise unattainable modern tools and methodologies through group negotiation, economies of scale and reduced opportunity costs
- Ability to make modern tools and methodologies also available to partners
- Access to pipelined tools to integrate molecular and phenotypic data along with information on how best to integrate them into the breeding program
- Objective advice on breeding program improvement
- Access to consultancy services for mechanization, automation, biometrics and high-throughput data analysis
- Resource mobilization for high priority areas

**Fish and livestock**
- Connected data systems will enhance best practices
- Access to skills and visibility
- More efficient resource use and greater time efficiency
- Competitive sequencing prices through economies of scale
- Access to big data storage and computing capacity
- Identification of better germplasm and genetic improvement of indigenous breeds
Agenda and funding mechanisms

**Focus:** At its onset, the Platform’s agenda will contain five modules, jointly prioritized with the Agri-Food Systems CRPs and the Genebanks Platform:

1. Breeding program excellence
2. Trait discovery and breeding tools and services
3. Genotyping/sequencing tools and services
4. Phenotyping tools and services
5. Bioinformatics and data management tools and services

**Implementation principles:** The Platform will use a combination of communities of practice, consultancies, and contracted services to achieve its Theory of Change (Figure 2). It will capitalize on providers of innovation from the public and private sector within and outside the CGIAR and invest in the adaptation, documentation and mainstreaming of tools, methods and best practices. Being responsive to users’ needs, it will validate tools and services first with members before promoting them among the wider community of users. Focusing on the needs of breeding programs targeting the developing world implies that the Platform will place particular emphasis on the needs of resource-constrained breeding programs that often operate far away from service providers.

**Funding:** The Platform is designed to draw its input from communities of practice, which it actively facilitates, and provide its products as open-access tools and information which will become available through the internet. It will invest in training that supports virtual workshops, documentation and face-to-face training of personnel that are key for reaching a wider user community (“train-the-trainer approaches”). For this it will rely on pro-bono investors. Other, individual services will be provided on a pay-for-use basis such as genotyping and sequencing services, special requests for training, access to phenotyping platforms or computational infrastructure, or advice from consultants. As a solid user base is being established, the Platform will assess its ability to move to a membership or user-paid model over several years.

![Figure 2. Theory of Change for the Excellence in Breeding Platform in support of the Agri-Food Systems CRPs’ Theories of Change.](image-url)
Leadership and management

Leadership: The Platform will be led by a Platform Leader and Module Leaders, in collaboration with Expert Advisory Groups in each Module. A Platform Steering Committee will provide strategic oversight of the Platform and be accountable for the performance of the Platform and the systematic collection of user feedback. Expert Advisory Groups will provide specialized advice at the Module level and ensure ownership and feedback from the wider constituency.

Members: Value will drive use and orientation of Platform activities and this will mostly be determined by members. Members are active contributors to the Platform’s products. They may come from CGIAR centers, ARIs, NARS or the private sector. They provide tools and information developed as part of their own mandates and activities, and may receive platform funding to adapt or document such tools and information prioritized in view of meeting the Platform’s objectives. They contribute to testing the products before they are released to the wider public. They have an active interest in the Platform’s existence and performance due to the facilitated exchange of best practices, and access to better materials, sites and services. Members have the opportunity to be represented in Expert Advisory Groups that advise individual Modules and the Platform Steering Committee.

Contractors execute distinct activities prioritized by the Platform’s management processes. They may be sourced from members or non-members, based on “best-for-the-job” principles. They also include experts with private sector experience.

Users access the Platform’s products based on license agreements accepted when accessing the Platform’s materials and service-related information. They provide feedback through the Platform’s user review tool developed as part of Module 2. As part of its communication strategy, the Platform will make a distinct effort to increase interactions with far-from-source users, in particular NARS breeders, so as to provide value in proportion and appropriate to their needs.
Module 1: Breeding program excellence

Rationale
Breeding is a cyclical product development process that must be integrated from product profile design through to dissemination and cultivar replacement. Many breeding programs serving smallholders in the developing world do not have the capacity to manage this process end-to-end. Yet, there is surprisingly little standardized information available that would characterize germplasm development or genetic gains across the breeding programs of the CGIAR, NARS and the private sector, making it difficult to invest strategically or systematically improve performance.

With germplasm development being one of the most impactful activities carried out by the CGIAR, the CGIAR proposes, through this Platform, to establish and follow a common format to set demand-driven targets, drive and accelerate the transparent development of breeding germplasm and cultivars, drive dissemination, and measure program progress, both within the CGIAR system and with NARS partners. The advantage of common implementation of best practices will be the ability to compare and improve individual programs, which will be largely driven by breeders striving to get access to the information and services they need to deliver the best results possible. Donors will have the ability to more precisely follow the progress of breeding programs and deliver support where needed, or invest in “game changers”.

The goal of Module 1 is therefore to develop an excellence management and support process for breeding programs that will be broadly appreciated and accepted across the CGIAR and by its donors. While Module 1 will define metrics for the overall breeding program, Modules 2-5 will improve best practices within distinct areas.

Work packages

1. Developing a standard template and approaches to monitor breeding program performance, including a common set of metrics that will assess genetic gain.
   This work package will address three components of a standardized continuous improvement cycle: (i) a standard process to set demand-driven breeding targets and product profiles linked to needs assessments; (ii) components that can help assess breeding programs’ abilities to achieve their breeding targets (germplasm management and evaluation, phenotyping strategy, trait mobilization, molecular and data analysis, breeding methodology decisions); and (iii) a standard approach to define breeding program metrics.

2. Internal CGIAR breeding program assessment and development of strategic plans to set breeding program direction.
   Based on the approaches defined in work package 1, CGIAR breeding programs will execute an internal program assessment and collect the metrics defining breeding program scope and abilities to achieve their breeding targets. The insights will be incorporated into a commodity-and breeding program-specific strategic plan documenting current investments and prioritized improvements. The assessment, metrics and resulting plan will be peer-reviewed to achieve a similar standard across the system. Research managers and breeders can use the assessment and plans to prioritize improvements and monitor program success.

3. Support breeding programs for needed improvements through consultancies, investments, and/or simulations.
   Breeding program assessments and strategic plans will identify program gaps. Programs will be able to catch up on best practices through consultancies, workshops and the development of joint funding proposals targeted at highest priority needs. Module personnel will support the documentation of best practices as part of the Module 2 Toolbox, also for use as training material.
4. **Extend the assessment to NARS breeding programs.**
Similar characterizations of NARS breeding programs is highly desirable. Interested NARS will be involved in discussions as tools are being developed. The needs and constraints of CGIAR and NARS breeding programs differ. While CGIAR breeding programs need to address diverse needs, the mandates of NARS extend beyond where the CGIAR typically engages, and trade-off decisions may follow different incentives. NARS participation will be voluntary, yet can be promoted if investors decide to support improvement plans with NARS that complete the assessment of their programs.

5. **Standardized approaches to measure genetic gains in farmers’ fields.**
There has been much debate about how to reliably measure genetic gains in farmers’ fields and be able to link them to development goals. Through collaboration between breeders and socio-economists, current approaches will be reviewed and best practices established at various levels of investment or given different purposes, so that the CGIAR will emerge with a stronger approach for monitoring the reach of improved cultivars and their benefit to farmers, as well as more authoritative statements on germplasm-related impacts on achieving past and future development goals.

**Capacity building**

The Module improves skills in strategic planning, breeding program management and client focus, and publishes approaches to monitor breeding program performance, information that is so far lacking. A scientific workshop will be organized to conceptualize various impact assessment approaches and their purposes. The Module will manage a list of external experts, including form the private sector, that can assist in training and provide an outsider view of realistic improvements within individual programs.

**Benefits to users**

Benefits of this Module include: (i) isolated CGIAR and NARS breeding programs becoming part of a performance management framework that allows breeders and program leaders to improve their performance based on best practices used in similar circumstances including those from the private sector; (ii) systematic use of such information in reports and during external reviews; (iii) more authoritative support to joint and individual project proposals though improved prioritization of investment; (iv) more effective use of existing breeding resources and greater returns on investments in crop research.
Module 2: Trait discovery and breeding tools and services

Rationale

In the non-multinational breeding sector (CGIAR, ARIs, NARS, smaller companies), many of the newly emerging tools for trait discovery, mobilization and cultivar development originate in diverse commodity-specific pockets, driven by user demand. Tools developed in one commodity (specific crops, trees or animals) are therefore largely inaccessible to the broader breeding community. More widely available information is lacking even for established processes such as a lab information and field trial management.

As a result, many CGIAR, ARI, NARS, smaller company researchers and breeders invest in developing their own commodity-specific solutions rather than adapting and adopting others. The most recent tools are not adopted as users do not have information on how their application would benefit them, nor the ability to quickly embed the tools in their workflows; they also have no access to training or interpretation expertise, and lack understanding of how to restructure their workflows to optimize the return on investment.

Module 2 aims to generate a platform where: (i) tools to accelerate and/or strengthen trait characterization, mobilization and cultivar development are identified from the public and (accessible) private sector; (ii) information on how to restructure workflows and incorporate tools to maximize their benefit is exchanged; (iii) tools are more widely tested, needs and investment/development priorities are set; (iv) accelerated awareness and uptake of the best tools are supported through consultancies, and virtual or face-to-face training.

Existing tools and know-how from the Agri-Food Systems CRPs will form the initial basis of the “catalogue” of tools and methods made available in this module (Figure 3). This will be complemented by solutions from Modules 3-5 and those available from ARIs and the private sector. It is estimated that the module will initially reach 60 key users from the CGIAR, NARS and ARIs associated with the Agri-Food Systems CRPs, reaching several thousands of users by 2022.

---

**Figure 3.** Interrelations of the Trait Discovery and Breeding Module with Agri-food Systems-CRPs and the Genebanks Platform, the Bioinformatics Module and external providers of software, know-how and computational capacity.
Work packages

1. **Development of a toolbox to support trait discovery, mobilization and breeding.**
   This work package focuses on developing, updating and maintaining a web-based platform to facilitate the description, review and use of tools and workflows relevant to modern and state-of-the-art discovery and breeding applications, resulting in a practical toolbox for researchers and breeders that can be structured by use cases and types of users. The web platform will use an “Amazon” review type approach (such as AnswerHub) to facilitate reviews and issue reports. The Toolbox will be updated by members adding and reviewing new tools/approaches/suppliers, and unused or poorly reviewed entities will be removed, to work towards a knowledge base of best practices.

2. **Fostering toolbox adoption and its dynamic use.**
   This work package documents and uploads new tools and approaches for trait discovery and breeding, with a strong emphasis on user participation, priority setting and capacity development. The Module Leader will work with a Community of Practice, made up of diversity analysts, trait discovery specialists and breeders from the Agri-Food Systems CRPs and the Genebanks Platform and external members, and the leaders of other Modules to document "missing" features, develop relevant use cases and prioritize needs for approach and tool development. It includes approaches for pipelined data analysis. Input from Module 1 will be used to assess how ready various CGIAR centers and NARS may be to use new interventions.

3. **Source and ground-truth innovative ideas through an incubator.**
   This work package focuses on brainstorming and discussion of blue sky ideas through forums held to discuss physical and virtual cross-disciplinary blue sky ideas. The objective of these forums is to enable researchers within and outside Agri-Food Systems CRPs to propose and discuss the application of high-payoff novel approaches to plant and animal trait discovery and breeding targeting the developing world. Participation of key donors is required to enable potential incubation of project ideas. It is anticipated that these incubator forums would be held once to twice a year, each with an attendance of 30 to 50 people.

Capacity building

The web platform structure, associated information and training modules developed through this Module’s toolbox follow use cases. The Platform will invest in members documenting tools, and in Platform personnel and consultants adapting those tools for a wider range of users, as part of a web-based “Toolbox”. Capacity building will be enabled through virtual meetings, face-to-face training associated with other meetings, and expert advice. It is anticipated that 40-125 users will be trained annually. Primary recipients of capacity building in the first phase are CGIAR, advanced research institution (ARI) and NARS scientists involved in upstream breeding implementation of Agri-Food Systems CRPs. In a second phase, capacity building activities will be streamlined based on user needs for low- to high-tech applications.

Benefits to users

The Toolbox will support users in trait mobilization and cultivar development, through cutting-edge approaches that are being tested in view of value addition to breeding programs. Users will have (i) lower tool self-development and maintenance investment; (ii) be able to access a wider range of solutions and innovations; (iii) access practical recommendations on use; and (iv) benefit from more informed selection and implementation of tools and approaches. Overall, a greater rate of successful upstream breeding projects and their faster implementation will result in greater use of genetic diversity and faster breeding progress by CGIAR and NARS programs.
Module 3: Genotyping/sequencing tools and services

Advances in genomics and molecular breeding technologies provide new opportunities to accelerate gene and trait discovery, and enhance breeding outputs. If applied effectively, the correct scale and form of genotyping accelerates the rate of genetic gain and reduces the cost per unit genetic gain. This is evident through the widespread adoption and routine implementation of genotyping by multinational seed and animal breeding companies. Applications by the CGIAR and NARS have so far focused mostly on trait discovery, while routine implementation as a forward breeding tool is not as advanced, both due to the cost of genotyping and lack of supporting systems and advice.

The cost of genotyping is an issue of economy of scale. Single projects or breeding programs do not have enough purchasing power to negotiate prices with the private sector. Through aggregation of demand, informed forecasting and streamlining of processes, better terms and conditions can be negotiated. On the other hand, if genotyping cannot be conducted in a time effective manner, it will not be adopted in breeding. Negotiating defined and guaranteed data turnaround times with service suppliers is therefore crucial.

In addition, the choice of optimal technology is not always obvious and made more complex given the rapid evolution of genotyping and associated data analysis approaches. The development and sharing of use cases and implementation guides are critical enabling devices empowering users to be more aware of the suites of technologies and suppliers available.

The Genotyping Module is intended to enable access to appropriate genotyping technologies across breeding programs associated with the Agri-Food Systems CRPs. It will do so by lowering the costs of genotyping, ensuring appropriate data turnaround time, ensuring that the most appropriate technologies can be chosen for each application, and creating an environment of change that drives adoption of more efficient and cost-effective technologies and enables the move from low-throughput to high-throughput systems.

Work packages

1. Capacity enhancement.
This work package will offer experience and guidance on the use of genotyping in different research and breeding applications. Through the development of use cases and implementation guidelines, it will help groups optimize and apply genotyping effectively in their work. Knowledge dissemination will be conducted through the online Toolbox developed in Module 2 and through face-to-face or online training. The Platform will maintain a list of internal and external experts to advise individual members on the use of genotyping in different research and breeding applications.

2. Broker access to genotyping supplies and services.
This Module will negotiate with groups and companies for cost, turn-around time and quality of genotyping supplies and services. The objective will be to fund the verification of 10 assays for existing validated trait markers per Agri-food Systems CRP, when a new supplier is suggested by an Agri-Food Systems CRP or external user, or approaches the Platform directly. This will enable Agri-Food Systems CRPs to judge the service supplied. Members of the Platform will have access to the best service providers through a common contract.

3. Prospect new approaches and customization of new tools.
This work package will work in close collaboration with ARIs and the private sector to assess the latest technology improvements and developments for both DNA extraction and genotyping. ARIs are often the earliest adopters of technologies for discovery, whereas the private sector adopts in an industrialized production-oriented manner for breeding. Use cases of collaboration with ARIs from individual Agri-Food Systems CRPs will be examined to provide information on the applicability of the latest technologies.
The Genotyping module is aligned with three currently-funded projects:

1. The High-ThroughPut Genotyping facility (HTPG), led by ICRISAT, will provide low-cost and fast-turnaround genotyping facilities to CGIAR and partners, reducing costs to US $1 per sample.

2. The Integrated Genotyping Support and Service (IGSS), a public-private partnership between Diversity Arrays Technology (DArT Pty. Ltd.) and the BecA-ILRI hub, supports the development of a commercial high-density, high-throughput genotyping and information management service in Africa.

3. Seeds of Discovery (SeeD), led by CIMMYT, provides support to breeders and other scientists in integrating DNA marker technology and genomic tools in the development and release of new cultivars.

**Capacity building**

Capacity building in the first two years of the project will focus on those scientists directly involved in genotyping. Training will focus on building capacity in forecasting, and in defining and implementing quality sampling and tracking procedures for effective service implementation. In addition, specific expertise and training will be provided to convert currently used gel-based markers to more efficient SNP-based markers. In parallel, the Community of Practice associated with the Genotyping Module will work on specific use case implementation guides, for upgrading the skills of CGIAR and NARS breeders through virtual and face-to-face meetings and aligned with training done in Module 2. Strategies to incorporate new genotyping or sequencing technologies in breeding programs and associated costs must also be taken into account.

**Benefits to users**

The Genotyping Module will provide users with access to lower cost, timely, quality guaranteed, state-of-the-art genotyping services. More effective and cost efficient services will enable breeders, seed producers and germplasm bank curators to adopt genotyping technologies more widely within their programs, to the benefit of (i) proactive understanding and use of diversity, (ii) applying greater precision and intensities during selection, (iii) rapidly recombining the best alleles, (iv) saving on phenotyping and nursery costs, or (v) ensuring better quality control and fewer errors in the breeding or seed production process.
Module 4: Phenotyping tools and services

Mechanization, automation, precision and high-throughput phenotyping have attracted huge investments in high-income countries, offering high leverage investments and quick wins that could be implemented in CGIAR and NARS breeding programs. These advances increase genetic gains by minimizing environmental variability in field trials, increasing the relevance of selections, increasing the speed and precision with which traits can be measured, allowing greater population sizes by reducing the cost of non-destructive data collection, combination with genomic selection training populations and reducing human errors.

Although some phenotyping challenges are specific to a single commodity, many phenotyping traits and methods are generic. Expertise in this area, however, is scarce, often requires interdisciplinary insights, and should be shared for more rapid progress. For example, precision and high-throughput remote or ground sensing technologies generate massive amounts of phenotype data, yet a bottleneck is formed by the need to convert this data into real-time “breeding values” and link it with genotype information for routine use in breeding programs. Joining efforts among CGIAR programs and with public and private sector capacities in high-income countries will greatly accelerate the identification and extension of best practices among CGIAR breeding programs and NARS partners.

Module 4 develops a CGIAR-wide Community of Practice and a hub for phenotyping support services to facilitate mechanization and automation, reduce the cost of routine phenotyping for selection, and increase adoption of high-throughput phenotyping tools and genotype, environment and management (GxExM) analysis tools across the Agri-Food Systems CRPs and external users. The module will focus on vetting, promoting and understanding existing technologies across the CGIAR, NARS, ARI and the private sector rather than the development of novel phenotyping tools. The focus will be on “quick wins”: low-cost/state-of-the-art phenotyping equipment, inexpensive service providers, and data analysis approaches that can be shared across commodities.

Work packages

1. Establishment of best practices for phenotyping and environmental analysis.
   This work package assesses high-throughput phenotyping tools and methods, GxExM analyses approaches and gene-to-phenotype models, and barriers to their adoption by breeding programs targeting the developing world in order to accelerate the development and implementation of best practices among a wider range of users. It will develop online learning tools and videos in collaboration with members, organize joint capacity building workshops and facilitate access to existing phenotyping platforms for training or use.

2. Support services from ARIs for A. experimental design, data management, and analysis of precision and high-throughput phenotype data, and B. GxExM analysis and genotype-to-phenotype predictions.
   This work package is to ensure that CGIAR and NARS breeding programs receive the support and capacity needed to design well-conceived high-throughput phenotyping, and protocols for GxExM analysis and gene-to-phenotype predictions. The core idea would be to provide access to existing expertise, technology, and know-how available, particularly from ARIs. It will establish the necessary linkages with environmental data managed and standardized by the CGIAR Platform on Big-Data, Information and Knowledge.

3. Coordination and procurement of phenotyping services for routine analyses of physico-chemical composition and functional properties in plant and animal materials in support of breeding.
   Breeding programs can increase their operational efficiency by identifying and using dedicated service laboratories within and outside the CGIAR system to generate routine phenotype data (e.g. physico-chemical composition and functional properties). This work
package will identify reliable laboratories, making sure such information is available to users and negotiating group rates. Specific targets include a survey of laboratories, capacities and costs, quality audits of proposed service laboratories within and outside the CGIAR, and a list of vetted service labs with negotiated costs for generic physico-chemical and plant functionality phenotyping services which can be updated annually.

4. Needs assessment and consultancy services to improve mechanization and automation across CGIAR and NARS breeding programs, as well as infrastructure and management of research stations.

The work package assesses the current state of mechanization and automation across the CGIAR and NARS, suitable equipment and mechanization tools available for each commodity, and current state of research station infrastructure. Based on this information, we will recommend “quick wins” to participating breeding programs and generate an inventory of robust and easy-to-service mechanization and automation approaches suitable for developing world conditions. Given that quite a range of CGIAR and NARS breeding programs are co-located, the information will also be used to develop recommendations for high-priority investment needs and implement training courses for experiment station managers.

Capacity building

Human capacity for phenotyping will be developed by making information available through the web platform, virtual meetings, exchange visits and biannual meetings among members, linked to relevant conferences. As part of the Toolbox in Module 2, a phenotyping website with updated resources, new literature, training materials and phenotyping protocols will be maintained as a resource base that can be used in training courses or workshops, particularly with NARS. Coordinated assessments of the current state of needs, infrastructure, mechanization and automation will be used to develop recommendations for high-priority investments with greatest systems pay-off.

Benefits to users

Users of the Phenotyping Module will benefit from (i) up-to-date information on relevant and tested phenotyping tools and approaches; (ii) accelerated learning and implementation of best phenotyping practices in support of faster turn-around of data, more precise selection and greater selection intensities; (iii) investments in safe-bet equipment (versus those that may soon become obsolete or do not warrant the investment); (iv) access to a pool of consultants to help set up phenotyping platforms and resolve technical problems; (v) access to labs that provide high quality physico-chemical composition and functional properties assessments at best prices and with appropriate turn-around times; (vi) freeing up of resources through mechanization/automation and for expensive in-house lab analyses; (vii) linkages to existing phenotyping entities in high-income countries (such as the International Plant Phenotyping Network; European Plant Phenotyping Network), the private sector or specific CGIAR crops (such as the Expert Working Group on Wheat Phenotyping). System-wide benefits include: (i) greater coordination of research protocols resulting in greater translatability of information across labs, environments, and crops; (ii) increased medium- and long-term returns on investment in crop research; and (v) increased scope and better standardization of data provide more reliable inputs for crop simulation models that can be used to predict big picture scenarios, helping the CGIAR and other organizations to set science agendas.
Module 5: Bioinformatics and data management tools and services

Modern breeding programs rely on information infrastructure that transforms increasingly large and diverse data into valuable information and place it in the hands of researchers and breeders at critical decision points. Advances in phenotyping, remote sensing, and genomic technologies enable the collection of vast quantities of valuable breeding data. In their raw form, and without standardized protocols and nomenclature, these data are difficult to store, manage and utilize.

However, when coupled with effective data collection, data management, analysis pipelines and decision support tools, these data can be transformed into coherent information that leads to new insights, more effective breeding decisions and greater rates of genetic gain. Software tools, computer processing and storage capacity, as well as practical experience with integrated breeding workflows, have become more accessible. The development and implementation of application programming interfaces (APIs) enable coherent, interoperable, adaptable data management and analysis workflows (“pipelines”) from available, tested and individually evolving bioinformatics components developed by ARIs, CGIAR centers and the private sector.

The grand challenge for the CGIAR and this Platform is to enable access by CGIAR, NARS and SME breeding programs to up-to-date, robust bioinformatics and biometrics tools and services that support trait discovery and selection decisions. The Bioinformatics Module will coordinate the technical implementation of software and pre-packaged software pipelines with multiple partners, developing practical solutions existing breeding and research workflows. Existing tools will be identified and adapted for integration via the nascent Breeding Application Programming Interface (BrAPI), where they are backed by reputable institutions as part of their core mandate. This Module will allow users to save time and resources when acquiring bioinformatics tools and reduce the considerable risk involved in tool acquisition, development and implementation.

Work packages

1. **Analyze pipelines and tools to clarify user and technical requirements related to selected trait discovery and breeding case studies.**
   This work package will define and document prioritized case studies with distinct user groups as part of Module 2 to ensure practical relevance and ownership by users, whose involvement in Agile Software Development, Iteration Planning and User Acceptance Testing is critical to the success of the outputs of the module.

2. **Develop modular and adaptable pipelines to support breeding workflows for the selected case studies and implement a common BrAPI.**
   Configurable data management, analysis and decision support pipelines will be defined and implemented for the case studies. The current landscape of available tools will be mapped and grouped logically into pipelines, with common BrAPI, developed based on an analysis of existing APIs, and building on established standards and mechanisms to ensure compatibility and interoperability across tools at the database, analysis and user interface levels.

3. **Support the development of databases and tools to complement and expand the usefulness of existing bioinformatics initiatives.**
   Several CGIAR and external projects develop fundamental bioinformatics resources, systems and tools that are beneficial to all crop and livestock breeders and researchers. Case study breeding workflows will reveal gaps in terms of tools, user interfaces and desirable links between tools and with external databases. This work package will develop and adapt new databases and/or tools to address such gaps, with investment conditional on compliance with the BrAPI and associated data standards. In cases where a commercial software option would offer the best value, the Platform provides an avenue to collectively negotiate preferential prices.
4. Define and implement metadata and data standards to facilitate data flows between tools and data sharing.
Defining data standards and defining and capturing a suitable amount of metadata in ontologies is a “must have” to ensure linkage across and within workflows and data sets. Metadata and data standards will be operationalized through the BrAPI. They will be aligned with standards developed by the Global Information System of the International Treaty on Plant Genetic Resources, and build on the work of the existing Crop and Agronomy Ontology Community of Practice and other initiatives.

5. Establish a Community of Practice and provide access to bioinformatics and biometrics advice, services and resources.
A Community of Practice will be formed from the bioinformatics and biometrics teams to: develop relevant component tools; facilitate collaboration for large-scale work; establish core operational guidelines; generate and share knowledge on tools and best practices; assess and plan for the sustainability of supported tools; and address capacity development needs. Emerging technologies and methodologies will be tested, compared and validated with users. Investments will be made in training biometricians at universities in CGIAR target countries and in brokering access to computational infrastructure.

6. Advance sustainable deployment and maintenance of tools and applications.
The tools used in the modular pipeline to support selected breeding and research workflows will need to continually evolve; therefore, sustainability of software development and maintenance will be an important selection criterion for choosing a tool. It is therefore imperative to plan ahead for resources to support post-deployment maintenance and improvement of tools supporting the developed pipelines.

Capacity building
The Bioinformatics Module will augment the knowledge and skills of crop improvement scientists with a view to improving their access to and use of information, analytical tools, and associated methodologies and technologies. While Module 2 will facilitate joint training opportunities, both virtually and face-to-face, the Community of Practice of bioinformatics and biometrics specialists will provide trainers, mentors and technical backstopping to partners of Agri-Food Systems CRPs. Train-the-trainer approaches will be used for face-to-face sharing of knowledge with the large number of NARS and SME partners and collaborators. Capacity building will be crucial to create a critical mass of appropriately skilled NARS scientists who will hasten the adoption and application of modern breeding approaches in their respective regions and countries.

Benefits to users
The Bioinformatics Module will deliver tools and technologies to support the modernization of breeding and trait discovery, and increase the speed at which better varieties can be developed and attendant economic benefits generated. It will provide a framework for more formalized collaboration among the diverse initiatives in this area and thereby reduce redundant investments, thus allowing each member to progress further. This implies joint priority setting and standards that often require larger users with their own bioinformatics and biometrics investments to make compromises, adjust workflows and redirect their own bioinformatics support. To smaller breeding programs that cannot afford to make significant investments in their own bioinformatics and biometrics, it will give access to otherwise inaccessible tools and approaches, and to bioinformatics and biometrics capacities within the system. By mapping the software landscape and common user interest, it will bring coherence to the various efforts, which will result in improved interoperability and compatibility. The Platform will for the first time establish an overview of bioinformatics system use, user feedback and future priorities, which can be used by investors.
Budget and associated outcomes

Tables 1 and 2 provide the base and uplift budgets for the Excellence in Breeding Platform aligned with the Outcome targets in Table 3.

Table 1. Base budget for the Excellence in Breeding Platform.

<table>
<thead>
<tr>
<th>Base Budget</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Platform Management</td>
<td>$456,794</td>
<td>$479,634</td>
<td>$503,615</td>
<td>$528,796</td>
<td>$555,236</td>
<td>$582,998</td>
</tr>
<tr>
<td>1. Breeding Excellence</td>
<td>$1,824,627</td>
<td>$1,915,858</td>
<td>$2,011,651</td>
<td>$2,112,234</td>
<td>$2,217,846</td>
<td>$2,328,738</td>
</tr>
<tr>
<td>2. Trait Discovery and Breeding</td>
<td>$2,146,559</td>
<td>$2,253,887</td>
<td>$2,366,582</td>
<td>$2,484,911</td>
<td>$2,609,156</td>
<td>$2,739,614</td>
</tr>
<tr>
<td>3. Genotyping Services</td>
<td>$936,116</td>
<td>$982,922</td>
<td>$1,032,068</td>
<td>$1,083,672</td>
<td>$1,137,855</td>
<td>$1,194,748</td>
</tr>
<tr>
<td>4. Phenotyping</td>
<td>$1,534,011</td>
<td>$1,610,712</td>
<td>$1,691,248</td>
<td>$1,775,810</td>
<td>$1,864,600</td>
<td>$1,957,830</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$10,000,000</td>
<td>$10,500,000</td>
<td>$11,025,000</td>
<td>$11,576,250</td>
<td>$12,155,062</td>
<td>$12,762,816</td>
</tr>
</tbody>
</table>

Table 2. Uplift budget for the Excellence in Breeding Platform.

<table>
<thead>
<tr>
<th>Uplift Budget</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Platform Management</td>
<td>$552,097</td>
<td>$579,702</td>
<td>$608,687</td>
<td>$639,122</td>
<td>$671,078</td>
<td>$704,631</td>
</tr>
<tr>
<td>1. Breeding Excellence</td>
<td>$2,588,140</td>
<td>$2,717,547</td>
<td>$2,853,424</td>
<td>$2,996,095</td>
<td>$3,145,900</td>
<td>$3,303,195</td>
</tr>
<tr>
<td>2. Trait Discovery and Breeding</td>
<td>$3,219,785</td>
<td>$3,380,774</td>
<td>$3,549,813</td>
<td>$3,727,304</td>
<td>$3,913,669</td>
<td>$4,109,352</td>
</tr>
<tr>
<td>3. Genotyping Services</td>
<td>$1,528,296</td>
<td>$1,604,711</td>
<td>$1,684,947</td>
<td>$1,769,194</td>
<td>$1,857,654</td>
<td>$1,950,536</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$15,000,000</td>
<td>$15,750,000</td>
<td>$16,537,500</td>
<td>$17,364,375</td>
<td>$18,232,594</td>
<td>$19,144,224</td>
</tr>
<tr>
<td>Module</td>
<td>2022 outcomes</td>
<td>Scope - Base</td>
<td>Scope - Uplift</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Module 1: Breeding Excellence</strong></td>
<td>Breeding excellence assessment process</td>
<td>Main CGIAR breeding programs, 3-5 NARS</td>
<td>Including sub-units in CGIAR breeding programs, &gt; 10 NARS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standardized metrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genetic gains assessments</td>
<td>4 studies (AFS funding)</td>
<td>&gt; 6 studies (AFS funding)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best practices documentation in ToolBox</td>
<td>4 modules/use cases per year</td>
<td>&gt; 10 modules/use cases per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert consultations</td>
<td>7 per year</td>
<td>&gt; 12 per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Investment and ROI increases</td>
<td>20%</td>
<td>&gt; 30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Module 2: Trait discovery and breeding tools and services</strong></td>
<td>ToolBox (all Modules)</td>
<td>&gt; 5,000 users</td>
<td>&gt; 10,000 users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best practices documentation for trait discovery and breeding in ToolBox</td>
<td>18 modules/use cases per year</td>
<td>25 modules/use cases per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pipelined analyses approaches tested, adapted and promoted</td>
<td>5</td>
<td>&gt;10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>40 participants per year</td>
<td>125 participants per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert consultations</td>
<td>&gt; 4 per year</td>
<td>&gt; 6 per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Module 3: Genotyping/sequencing tools and services</strong></td>
<td>Common genotyping services</td>
<td>5-10 users</td>
<td>&gt; 15 users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marker conversions to SNP-based platforms or best practices documentation for genotyping/sequencing in ToolBox</td>
<td>8 use cases per year</td>
<td>&gt;12 use cases per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>16 participants per year</td>
<td>55 participants per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert consultations or external marker conversions to SNP-based platforms</td>
<td>5 per year</td>
<td>&gt; 7 per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Module 4: Phenotyping tools</strong></td>
<td>HTP phenotyping</td>
<td>3 successful institutional users reducing phenotyping cost by &gt;25%</td>
<td>&gt; 5 successful institutional users reducing phenotyping cost by &gt;25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GxExM analyses</td>
<td>&gt; 5 routine users</td>
<td>&gt; 10 routine users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mechanization and automation</td>
<td>&gt; 3 institutional beneficiaries</td>
<td>&gt; 8 institutional beneficiaries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best practices documentation for phenotyping/mechanization/automation in ToolBox</td>
<td>7 modules/use cases per year</td>
<td>&gt;15 modules/use cases per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>13 participants per year</td>
<td>50 participants per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert consultations</td>
<td>&gt; 4 per year</td>
<td>&gt; 4 per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Module 5: Bioinformatics and data management tools and services</strong></td>
<td>Software tools</td>
<td>&gt; 1,000 users</td>
<td>&gt; 2,500 users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contributors of BrAPI compatible components</td>
<td>&gt; 5 institutions</td>
<td>&gt; 10 institutions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of computational infrastructure in other institutions</td>
<td>3 users</td>
<td>&gt; 6 users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>75 participants per year</td>
<td>185 participants per year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>Members</td>
<td>10</td>
<td>&gt; 30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>