Archetype for implementing Rapid Cycle Genomic Selection (RCGS) in CGIAR-NARES-SME breeding network pipelines

INITIATIVE ON Accelerated Breeding

Examples for pure line crops; RTB crops and hybrid crops

May 2024

Attribution

Output of the Consortium on Rapid Cycle Genomic Selection, convened by BMGF, 4th – 5th September 2023, with participants from Accelerated Breeding and Breeding Resources Initiatives and Quantitative Genetics Specialists from Genetic Innovation, Advanced Research Universities, and Industry.

Editors:

Dorcus Gemenet – Accelerated Breeding Initiative (ABI)

Christian Werner – Accelerated Breeding Initiative (ABI)

Lorena Guimaraes – Accelerated Breeding Initiative (ABI)

Michael Quinn – Accelerated Breeding Initiative (ABI)

Bhoja Raj Basnet – Accelerated Breeding Initiative (ABI)

Biswanath Das – Accelerated Breeding Initiative (ABI)

Giovanny Covarrubias-Pazaran – ABI-IRRI

Keith Gardner – ABI-CIMMYT

Chris Gaynor – Bayer

Marle Rose Labroo – Bayer

Gary Atlin – Bill and Melinda Gates Foundation

Jim Lorenzen – Bill and Melinda Gates Foundation

Young Wha Lee – Bill and Melinda Gates Foundation

Jeff Ehlers – Bill and Melinda Gates Foundation

Acknowledged for participating in the convening and contributing to follow-up discussions and consultations:

CGIAR-NARES Breeding teams; Clay Sneller – Ohio State University; Aaron Lorenz – University of Minnesota; Jessica Rutkoski – University of Illinois; Kelly Robbins – Cornell University Jean-Luc Jannink – Cornell University; Jeff Endelman – University of Wisconsin-Madison; John Hickey – Bayer; Brian Gardunia – Bayer; Valentine Wimmer – KWS.

Contents

Glossary

CGIAR – Consultative Group on International Agricultural Research

NARES – National Agricultural Research and Extension Services

BMGF – Bill and Melinda Gates Foundation

SME – Small and medium-sized enterprises

ABI – Accelerated Breeding Initiative

BRS – Breeding and Research Services

RCGS – Rapid cycle genomic selection

RTB – Roots, tubers, and bananas

Early Testing 1 – The first year of field testing

RGA – Rapid generation advance

DH – Doubled haploid

TPE – Target population of environments

OFVT – On-Farm Verification Trials

OCS – Optimal cross-selection

QA/QC – Quality Assurance and Quality Control

MAS – Marker assisted selection

GRM – Genomic relationship matrix

GEBV – Genomic Estimated Breeding Value

GEGCA – Genomic estimated general combining ability

GPCP – Genomic prediction of cross performance

SAH – Semi-autotrophic hydroponics

Summary

Accelerated Breeding Initiative's aim is to develop a continuous stream of climate-resilient, preferred, inclusive, high-yielding, and nutritious varieties, against the present-day global challenges such as climate change and environmental degradation.

We are committed to increasing genetic gains in farmer's fields by developing and availing to seed systems new varieties that perform well under farmer conditions and have clear advantages in essential traits that can be recognized by farmers and end-users in order to incentivise the farmers to adopt the new varieties in lieu of their familiar varieties, and to do so at a pace that matches the increasingly erratic growing conditions.

The rapid cycle genomic selection (RCGS) archetype seeks to enable selection and recycling of new parents as fast as possible – preferably two-to-three-year cycles, using data that is accurately representing the target population of environments (TPE) in which the future potential varieties will grow. This will ensure that breeding populations are continuously improved to provide high-value alleles and haplotypes for adaptation to biotic and abiotic stresses, and high-value parents from which superior products are identified and released as new varieties. To achieve this we need to: i) Take advantage of the BRS-procured genotyping services with mid-density marker panels that can estimate genomic relationships and marker effects; ii) Deploy new testing strategies that sample the TPE better, using relatively similar resources, e.g. by use of sparse testing where alleles are replicated across locations in the TPE without physical replications or with partial replications, and using the family structure and genomic relationships to provide connectivity, as well as the additional use of managed stress trials to mimic different environments; iii) Build and leverage on efficient CGIAR-NARES-SME breeding networks to scale up testing sites at Early Testing 1 and develop innovative ways for resource reallocation between Early Testing and later stages of testing; iv) Recycle new parents for population improvement faster using the improved accuracy from better testing strategies.

In this document we:

- Articulate a value proposition for working as a breeding network to CGIAR-NARES-SME partners.
- Provide clear steps on how to redesign existing pipelines to optimally implement a rapid cycle genomic selection scheme, with example breeding schemes in pure line crops, roots, tubers and bananas (RTB), and hybrid crops.
- We identify the support and services needed by the CGIAR-NARES-SME breeding networks to implement this archetype.

1.0 Introduction

1.1 Objective

Increase genetic gain in farmers' fields from improved breeding populations developed through accurate rapid selection of new crossing parents from Early Testing 1 genotypes (the first stage of field testing) broadly tested across the target population of environments (TPE).

1.2 Rationale

Recurrent selection and crossing of 'elite-by-elite' parents to develop the next generation leads to stacking of positive alleles of essential traits for a given Target Product Profile (TPP). Rapid recycling of crossing parents from Early Testing 1 trials can substantially reduce the generation interval and enable higher rates of genetic gain. In a conventional breeding pipeline, selection of both parents and commercial products is based on genetic merit estimates based on phenotypic data from the selection candidates themselves.

These candidates are replicated within and across environments to estimate their merit, but the phenotypic information in Early Testing 1 is often not enough to calculate accurate breeding values (value as a parent). This is because Early Testing 1 mainly occurs at a very few research stations managed by breeders. Since the environmental conditions and agronomical practices in farmers' fields, especially in sub-Saharan Africa, are highly heterogeneous, conditions on research stations for Early Testing 1 may not be representative of conditions on famers' fields. As most selection pressure is applied at this stage (up to 95% depending on the crop), a low genetic correlation between performance on-station and on-farm will result in selection of new crossing parents and advancement of candidate products (varieties) that may not be suitable to meet farmers' requirements. To circumvent this, Early Testing 1 can be designed to better represent the farmer conditions by testing in more locations in the TPE than is currently done and to combine this with genomic selection (GS) where genotypic information from relatives substantially enhances the amount of information available to provide a more accurate estimate of breeding value.

This information is obtained using genetic markers modelled in the form of the genomic relationship matrix (GRM) or directly modelling the markers, which allows the additive value of DNA markers (representing haplotypes) carried by a selection candidate to be estimated across all its relatives (mainly full and half sibs). In pipelines that use GS to select parents, genotypes can be sparsely tested across the environments with relatives (full and half sibs) replacing conventional replicates. Early Testing 1 genotypes can be distributed and tested among CGIAR-NARES-SME testing sites to achieve a more robust testing that captures the target population of environments (TPE) in more countries. Leveraging mid-density marker genotyping provided by Breeding and Research Services (BRS) Genotyping Lab Shared Services, a genotype's performance can be predicted across the entire CGIAR-NARES-SME testing network, resulting in improved prediction accuracy and a better translation of breeding advancements into genetic gains in farmers' fields.

2.0 Process for successful implementation of rapid cycle genomic selection schemes

i) Develop efficient CGIAR-NARES-SME partnerships with joint priority setting and decision making on regional Target Product Profiles that represent the needs of all network partners and share roles and responsibilities.
ii) Redesign the existing breeding pipelines and schemes of the priority regional Target Product Profiles.

2.1 Efficient CGIAR-NARES-SME partnerships

The rapid cycle genomic selection archetype requires effective CGIAR-NARES-SME breeding networks with joint priority setting, joint decision making and shared benefits, in a way that all partners benefit from the partnership. Working with Accelerated Breeding's (ABI) TRANSFORM work package, many breeding teams across crops and Centres have been working on building these networks. The archetype focuses on population improvement of essential traits in each Target Product Profile (TPP) that represents the needs of all partners in a given market segment (MS). The stream of improved parents and high-value haplotypes coming from the population improvement process is then used to develop and deliver climate-resilient, market-preferred varieties by the CGIAR-NARES-SME breeding network partners.

For breeding pipelines jointly run by the CGIAR-NARES-SME networks, the population improvement hub will be coordinated by the CGIAR partner while testing is done by all network partners in the CGIAR-NARES-SME network according to the relative testing capability of partners and the ease of international germplasm transfer within the TPE region. However, this applies only to those pipelines jointly run by the CGIAR-NARES-SME breeding networks and does not exclude the NARES-SME partners from running other pipelines according to national priorities.

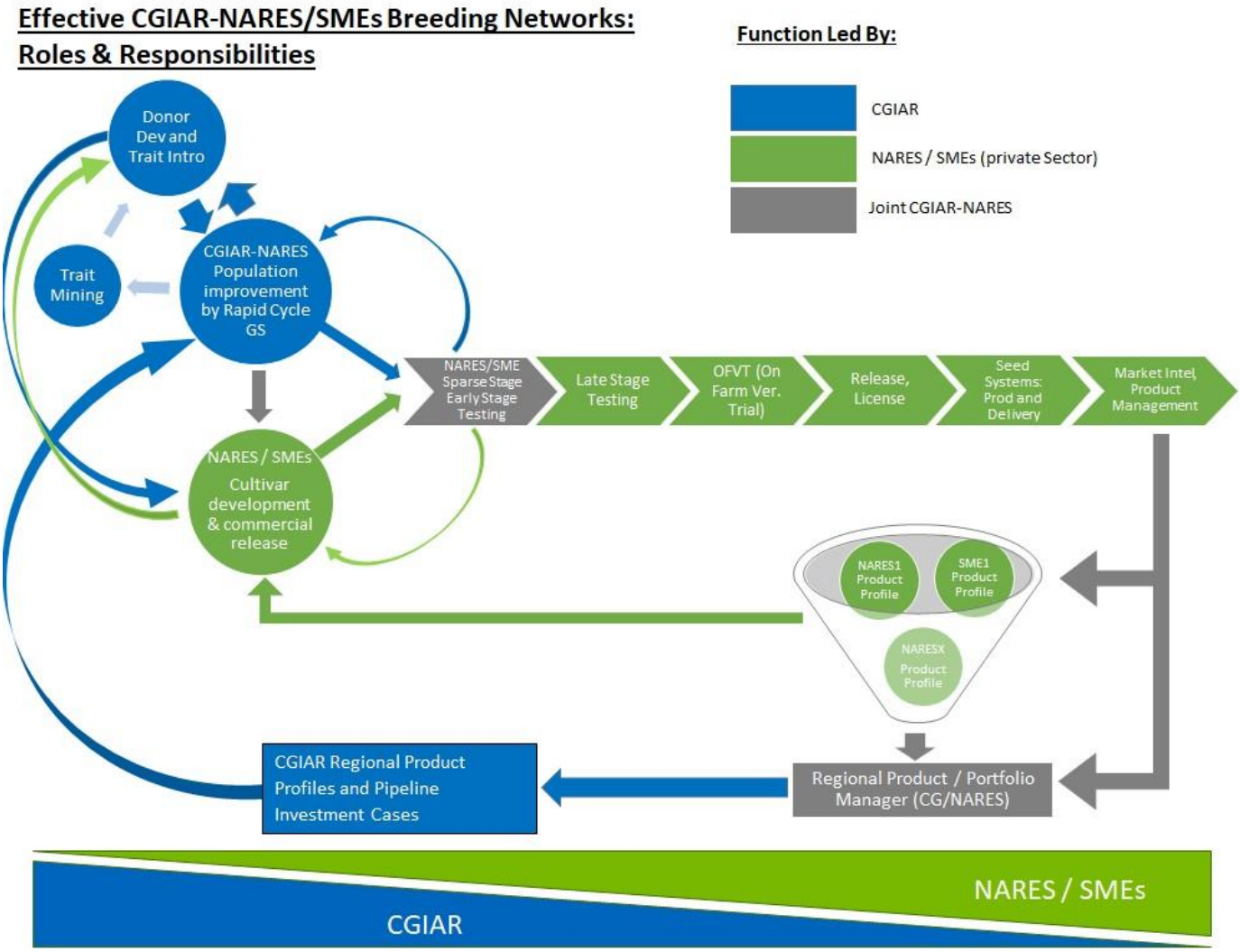
For scaling purposes, once the strategy is well optimized and is operational across networks, some NARES partners might take up more responsibility by running pipelines, backstopped by the CGIAR partner, on behalf of the breeding network. However, to achieve this, there is need for well-defined data and germplasm exchange protocols and data governance agreements within the networks.

This archetype does not imply that population improvement is entirely separated from product identification, rather, that the continuous stream of high-value parents and haplotypes are continuously taken up and tested in later stages of the product

development process (Late Testing and On-farm Verification trials) towards variety release and delivery to seed systems.

To achieve this, there is need for clearly shared roles and responsibilities among the CGIAR-NARES-SME partners with some partners taking more responsibility in earlier stages and some in later stages of the population improvement and product development process.

Fig. 1 Effective partnerships in CGAIR-NARES-SME breeding networks with role-sharing according to testing capability



2.1.1 Potential role sharing in the rapid cycle genomic selection archetype

CGIAR partners incorporating joint decisions with NARES-SME partners

- Develop methods for designing gender intentional TPPs based on market intelligence.
- Identify and deploy high-value alleles and haplotypes that address regional priorities.
- Coordinate the rapid cycle genomic selection population improvement network, manage the population improvement hub, extract high-value parents for product development by CGIAR-NARES-SME partners.
- Backstop NARES-SME partners in product identification in later stages of the pipelines that leverages on the phenotyping and dissemination capacity of NARES-SME partners.

NARES-SME partners with backstopping from CGIAR

- Population improvement for market segments not served by the CGIAR-NARES-SME network.
- Managing high-quality testing sites for the regional network for both Early Testing 1 trials and late-stage trials for product identification.
- Identify, extensively test, and nationally release finished products.
- Validate the performance of new products in extensive on-farm testing and make the investment case for varietal replacement to national authorities.
- Provide feedback into developing the gender intentional Target Product Profile development process.

2.1.2 Value proposition for NARES-SME partners

- ***More say in how populations are developed representing partners' needs.*** Partners jointly decide based on network data, which parents are recycled at the population improvement hub, and which products get advanced into later stages of testing for product identification. This is achieved through joint advancement meetings.
- ***A rapidly improving stream of high-value lines or clones from Early and Late Testing that can be used either as parents or released directly by NARES-SME***. Network partners co-own the populations, which serve as the primary source of climate-change adaptation in the regions they serve. The partners will also have easy access to unselected lines or clones drawn from the improved populations as needed to develop adapted products.

- ***Fast access to new high-value haplotypes for biotic and abiotic stress tolerance***. The networks will ensure that donors combining new disease and stress-tolerance alleles with good agronomic performance and the network-approved quality profile are quickly developed and deployed.
- ***Support to improve phenotyping and information management.*** Gains from RCGS are highly dependent on the accuracy of Early Testing 1 phenotyping. The network needs to strongly support NARES phenotyping, information management and data analysis, both technically and financially. This technical support will also benefit NARES own breeding pipelines for national priorities.
- ***Technical and financial support for implementing RCGS in their own pipelines***. For those programs with breeding pipelines for national priorities, technical support for their own in-house RCGS efforts can be fairly inexpensively provided in the form of genotyping slots, support for national breeding informatics systems, and implementing analytics.

2.2 Redesign of the existing breeding pipelines and schemes: Best practices for successful implementation of RCGS

Rapid cycle genomic selection is a method for accelerating and increasing the accuracy with which selection of new parents in the TPE is done in existing breeding pipelines, not separate diversity panels or 'special populations for GS'. The re-design necessitates doing several things differently.

New parents are recycled faster with 2-3-year cycle length targets using Early Testing trial data

This needs to be coupled with methods for reducing time to stage 1 testing by: i) Using improved technologies and methods such as rapid generation advancement (RGA), speed breeding and doubled haploid (DH) technology; ii) Improved efficiency in multiplying planting material and minimizing cross-border exchange transactional costs. Breeding programs should work with BRS-Trialling and Nursery to improve efficiency. Breeding teams should work with quantitative genetics specialists to determine the best recycling strategy that balances speed and accuracy of selecting new parents.

Accuracy of selecting parents from Early Testing 1 is improved by sampling the TPE earlier and using a genomic relationship matrix that leverages information from relatives to estimate breeding value

CGIAR-NARES-SME breeding networks should aim to evaluate Early Testing 1 trials in a minimum of 10 and up to 30 testing sites as available to the network partners. The Early Testing 1 trial is split among the crop breeding network partners in a sparse-testing manner, where full-sibs and half-sibs are split and tested across partner sites with connectivity achieved by using genomic information. Replacing replicates with relatives in

sparse testing removes the need to evaluate every candidate at multiple sites across the TPE at Early Testing 1. Instead, the aim is to ensure that the population of alleles experiences a representative sample of TPE conditions, such that breeding values derived from that process are equally representative of the TPE. Sparse testing greatly reduces the number of seed needed to evaluate each candidate in Early Testing 1. For instance, instead of increasing enough seed to evaluate a candidate in two replicates at each of three-to-five locations, only enough planting material for a single plot (plus remnant) is required. This could eliminate a season of seed increase/multiplication, hence allowing the breeding cycle to be shortened. CGIAR-NARE-SME breeding networks should work with quantitative genetics specialists and biometricians to define the best way to design the sparse testing.

A selection index for essential TPP traits and inbreeding control methods are routinely used to select parents

Selection of traits is aligned with the Target Product Profiles by using the Desired Gains index that leverages quantitative genetic parameters and the desired gains towards the TPP traits to assign weights to different TPP essential traits, in a manner that ensures the population is moving towards the right direction with respect to all essential traits. It is expected that with improved Early Testing 1 strategies, it will be possible to measure all essential TPP traits at this stage. Rapid cycling may also lead to faster conversion of variance into gain, hence leading to inbreeding. To balance the trade-off between genetic gain and genetic variance, specific crosses for population improvement are made using an inbreeding control algorithm such as optimal cross selection (OCS), that leverages on the contribution theory. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to implement the selection indices and OCS.

More streamlined germplasm management, exchange, data sharing and data analytics processes are implemented

To achieve robust Early Testing 1 across network partners, seed/planting material preparation and shipment protocols need to be improved to match with the crop calendars in the partner countries. Pedigree correctness need to be assured amidst these exchanges using low density markers for quality assurance and control (QA/QC) of parents and F_1. Breeding network data needs to get back in time to drive joint recycling and advancement decisions, hence the need to efficiently use the data management system (databases: Enterprise Breeding System (EBS), Breeding Management System (BMS), BreedBase) with timely input of data by all partners. Due to increased volume of data from improved Early Testing 1 strategies, state-of-the-art analytical pipelines need to be adopted to ensure faster turn-around of performance reports to inform decisions during product advancement meetings (PAM) and parent recycling decisions. To achieve these, CGIAR-NARES-SME breeding networks should work with BRS-Trialling and Nursery, BRS-Lab services, BRS-Breeding IT, and the Breeding Analytics team.

More targeted use of the genotyping shared service procured by BRS-Lab Services for low-density (diagnostic/marker-assisted selection [MAS]) and mid-density genotyping (GS)

All selection candidates included in Early Testing 1 need to be genotyped at a density permitting calculation of the GRM (about 1000-5000 SNPs) using the GI-contracted mid-density service for the crop. Because this genotyping is relatively expensive (approx. $10), it may be most cost-effective to restrict it to lines/clones that have survived both the visual and marker-assisted selection steps. As genotyping costs change, other plans may become more efficient. At current diagnostic genotyping costs (ca. $2/sample) it is likely to be most efficient to screen with diagnostic markers only those lines/clones surviving the visual selection step. However, the visual selection and marker-assisted selection can be done concurrently if doing them sequentially adds a full season to the breeding scheme. For efficient support from the BRS-Lab services, there is need to correctly forecast genotyping needs for QA/QC, MAS and GS and to share clear crop calendars to ensure fast data turn-around to drive decisions. Streamlined sampling and shipment protocols should be mainstreamed into breeding operations to avoid time wasted on poor sample quality and repeat sampling which leads to data delays. CGIAR-NARES-SME breeding networks should work with BRS-Lab Services and ABI-ACCELERATE to achieve the desired level of efficiency.

More streamlined product identification protocols at Late Testing and On-Farm Verification Trial (OFVT) for product advancement

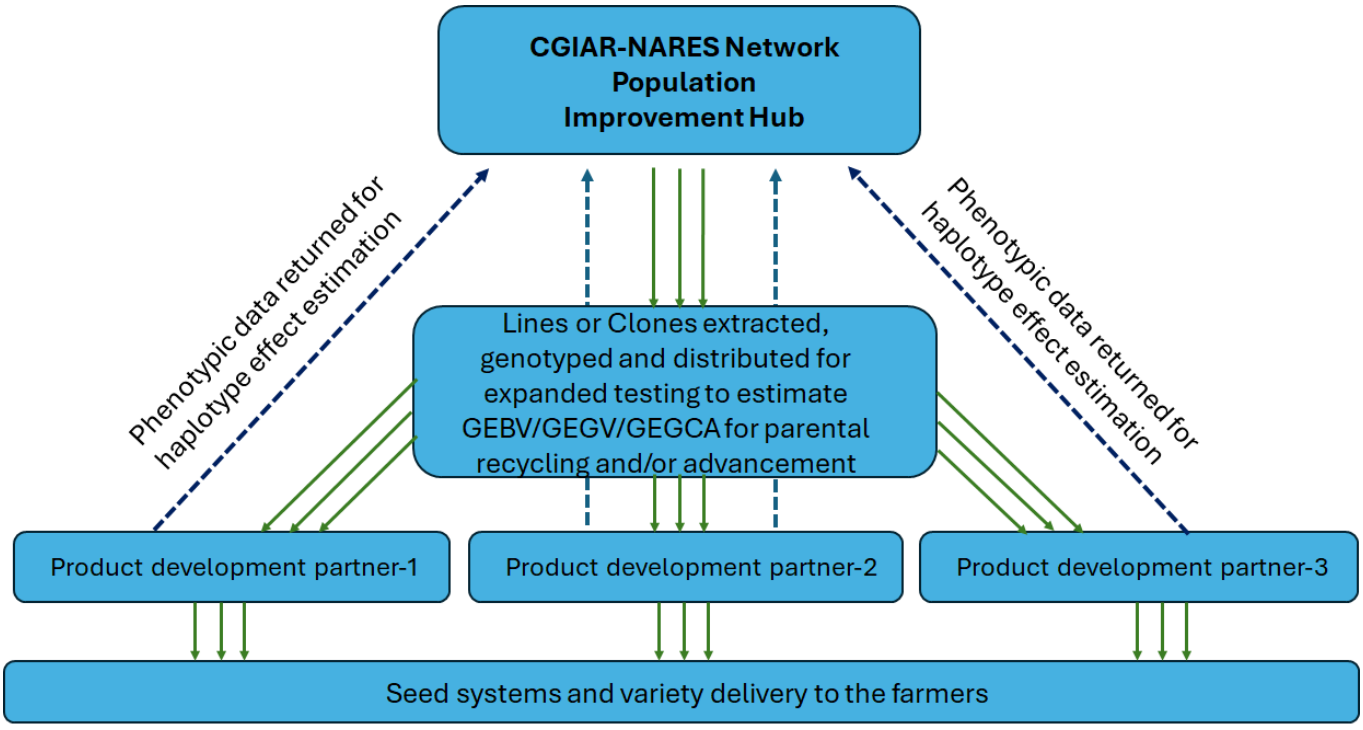
The singular most important reason for an explicit and streamlined population improvement effort within breeding pipelines is to ensure that climate-resilient, market-preferred products can be identified and advanced within those pipelines. A continued stream of potential products will be identified from the high value haplotypes and parents that come from the population improvement effort. CGIAR-NARES-SME need to identify potential products during the sparse testing at Early Testing 1, as well as from additional testing for high heritability traits and managed stress trials as applicable. Those genotypes that meet thresholds for Target Product Profile essential traits will be advanced to later stages of testing towards candidate product identification. For accurate selection of candidate varieties for advancement, Late Testing should be robust, with each genotype being tested in up to 30 environments (Locations) within the TPE (market segment). The testing should be focused on all essential TPP traits, with each TPP trait being measured against the best commercial (local) check that is used as a benchmark variety. Late Testing should also adopt a connectivity check strategy to allow estimation of annual genetic gain trends using historical data from the programs. After Late Testing, the candidate varieties need to be verified for performance under farmers' conditions and how testing environments represent target environments. On-farm verification trials should be done in > 30 sites. The purpose of on-farm verification trials is to provide a breeding program with i) information on the performance of the candidate varieties under farmer's conditions; ii) information on what is preferred by farmers, end-users and why the preference; iii) information on gender-disaggregated preferences of candidate varieties to inform gender

intentional breeding strategies. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists and Biometricians to implement better Late-Testing and OFVT trials.

Mainstreamed resource re-allocation between Early Testing and Late Testing

Given that Early Testing 1 becomes more robust and more accurately representing the TPE, resources can be re-allocated from later stages of testing to accommodate the more extensive Early Testing 1. For example, an RTB crop whose breeding scheme normally has the following stages: seedling nursery, clonal evaluation stage, preliminary yield trial, advanced yield trials, uniform yield trial and on-farm verification stages, with Early Testing 1 happening at clonal evaluation stage, might no longer need a preliminary yield trial, hence reallocating those resources. Similarly, for a crop such as Maize, currently having a breeding scheme with stage 1, stage 2, stage 3, stage 4, stage 5 and on-farm verification, given that testcrossing with all testers now happens at Early Testing 1 (stage 1), and phenotyping at Early Testing 1 becomes more extensive, the new breeding scheme might skip stage 2 and stage 3 and go directly to stage 4, 5, and on-farm. Breeding teams should work with quantitative genetics specialists to explore the most optimal ways to re-allocate the resources between Early and Late Testing.

Fig 2. Graphical representation of the RCGS archetype in a breeding pipeline



3.0 Example breeding schemes in different crops

Here we give example breeding schemes in pure line crops, RTB and hybrids with inbred parents based on the RCGS archetype. It should be noted that the example breeding schemes provided do not cover all the breeding schemes used for crops bred by the CGIAR-NARES-SME networks, e.g. open pollinated variety (OPV) breeding schemes, breeding schemes for synthetics and for those hybrid crops that use cytoplasmic male sterility (CMS). For the schemes not represented by examples given here, breeding teams will work with quantitative genetics specialists to develop appropriate breeding schemes for their respective pipelines and crops during implementation.

3.1 Example RCGS archetype breeding scheme in pure line crops

Line crops bred by CGIAR-NARES-SME breeding networks include: i) cereals like Rice, Wheat, Barley, and ii) legumes like Beans, Groundnut, Cowpea, Chickpea, Soybean, Lentils, Grasspea, among others. Breeding methods in pure line crops are generally the same except in some operational adjustments e.g. seed set is not a bottleneck in cereals while it may be a bottleneck in legumes and might need additional operational interventions to go to Early Testing 1 faster.

Population improvement

Select crosses and conduct quality control genotyping:

To balance the trade-off between genetic gain and genetic variance, we leverage on the contribution theory to select crosses rather than use truncation selection to select parents. Between 50-200 crosses that constitute the next generation are selected from Early Testing 1 data using optimal cross selection (OCS). The crosses are selected based on a GEBV-based desired gains index which allows multiple traits, captured in the TPP, to be integrated in a selection decision. F_1s need to be confirmed using quality-assurance and quality control genotyping (QA/QC). CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to define the optimal number of crosses depending on their pipeline size and available resources.

Create F3 or F4-derived lines per cross via single-seed descent:

One main purpose of RCGS is to shorten the breeding cycle as far as possible. To do this, inbred lines need to be generated as quickly as possible via single-seed descent (SSD). Modelling indicates that there is no benefit in terms of gain per cycle to inbreeding beyond

F_4. In some cases, F_3 -derived lines are optimal if using them allows the breeding cycle to be shortened by a year (Atlin and Econopouly 2022). If the SSD system can advance three generations per year, it is possible to generate and increase F_4-derived lines within two years of the initial cross. If the program can only advance two generations annually, it may be best to use F_3-derived lines as selection candidates. Between 50 – 500 lines per cross would be adequate. However, this number needs to be determined depending on pipeline size and program resources. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to determine the optimal numbers for their respective pipelines. The number of lines that the SSD pipeline needs to deliver depends on the selection pressure expected at the visual and marker-marker selection step as discussed below.

Increase and screen the selection candidates in a visual observation and increase nursery:

The lines created by SSD should be grown out as progeny rows or plots, preferably in the main season, providing an opportunity to select for highly heritable traits such as plant type, seed or grain shape and colour, maturity, height, etc. It is expected that a visual selection intensity of 10-20% is likely to be applied by most programs.

Lines that survive the visual selection step can be screened using marker-assisted selection. At current diagnostic genotyping costs (ca. $2/sample) it is likely to be most efficient to screen with diagnostic markers only those lines surviving the visual selection step. However, visual selection and marker-assisted selection can be done concurrently if doing them sequentially increases the breeding process by a season. The total initial segregating population size will depend on the expected selection pressure during visual selection and diagnostic marker selection steps, to achieve a final target population size needed for Early Testing 1 evaluation. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to define the optimal segregating population size for their respective pipeline.

In addition to selection, the purpose of this nursery is to generate enough seed to support Early Testing 1 sparse testing as well as any disease or managed stress screens to be applied at Early. Sufficient remnant seed of each selection candidate also needs to be retained to allow those lines selected out of Early Testing 1 to be increased for later stages of testing and shared with all partners towards product identification.

Manage Early Testing 1 trial design, distribution to partners and mid-density genotyping for genomic selection:

The population improvement hub is responsible for designing the sparse Early Testing 1 trial and ensuring that it is distributed to CGIAR-NARES-SME partners safely and in good time for planting. All selection candidates included in Early Testing 1 need to be genotyped at a density permitting calculation of the GRM (about 1000-5000 SNPs) using the GI-contracted mid-density service for the crop. Because this genotyping is relatively expensive (approx. $10), it may be most cost-effective to restrict it to lines that have

survived both the visual and marker-assisted selection steps. As genotyping costs change, other plans may become more efficient.

Managed stress screening:

Early Testing 1 agronomic trials should be primarily conducted to sample the TPE in a sparse testing manner, with linkage among trials provided by full sibs and the GRM rather than full replication of selection candidates over locations. Additional Early Testing 1 trials at key research stations will be conducted to generate specific phenotypes (e.g. drought response, optimal management response, performance under managed flooding, disease screening platforms, cooking time, among others). These will need to include the full set of the selection candidates because there are usually only one or two locations per managed stress screening. The size of populations should be able to achieve this as well as representing the population structure across the sparse testing sites. Breeding teams should work with quantitative genetics specialists in establishing the best population sizes for their respective pipelines.

Support data management and analysis for parent selection and advancement:

Genotypic and phenotypic data need to be managed to permit accurate selection. Data must be centrally managed using a GI-supported breeding database system (EBS, BMS, or BreedBase) linked to a genomics database (GIGWA for most programs). NARES partners must be supported to electronically receive experimental designs, collect and upload data, and access results using these systems. The CGIAR Breeding Analytics team have developed a full analytical pipeline for GS, integrating the desired gains index and optimal cross selection as well as environmental co-variates. Breeding networks will need support from a quantitative genetics specialist to implement this pipeline.

Table 1: Example two-year RCGS breeding scheme timetable using SSD, assuming advancement of three generations per year

Year	Season	Activity 1	Activity 2	Notes
0	**Main**	Sow potential parents so that they flower ~2 weeks after yield data is captured		
0/1	**Main / Off 1**	Make F_1s	Quality control genotyping of parents	Re-cycle only the very best, highly elite lines. E.g. ~500 crosses selected using OCS
1	**Off 2**	Make F_2s	Quality control genotyping of F1s	SSD with > 50 plants per cross
1	**Main1**	Make $F_{2.3}$s		SSD with > 50 plants per cross
1	**Main2**	Make $F_{3.4}$s	Visual observation and selection of individuals within families	Sow F3 bulk as spaced plants.
2	**Off 1**	Visual selection nursery of $F_{3:5}$multiplication / observation rows.	Visual observation and selection of individuals within very best families. Only select from best 150 families (2^{nd} year yield data on parents is now in). Genotype to support selection (optional) if resources permit.	Plant out at least 1 row per line, > 30 lines per cross.
2	**Off 2**		Genotyping of selected lines (optional if there is time and if not genotyped in previous step)	
	Main	Early Testing. Capture the TPE + year effect as best as possible. Managed stress trials likely required. Sparse testing can be used	Mid-density genotyping of each line mandatory if not done in off-season.	Trial e.g. 500 best lines from 150 best families.

3.1.1 Example archetype for RTB crops

Vegetatively propagated crops bred by CGIAR and its partners include Cassava, Yam, Potato, Sweetpotato and Banana. They are commonly referred to as roots, tubers and bananas (RTB) crops.

In general, the RTB crops present some unique challenges:

1. ***Low multiplication ratio and long multiplication time:*** crossing parents are always heterogenous, and each seed represents a potential variety. Planting material for multi-environment evaluation requires clonal (asexual) propagation following germination of a single plant. This usually results in a low multiplication ratio and long multiplication time. A rapid way to develop clones for prompt multi-

location testing would be by in-vitro propagation. However, this is a relatively expensive exercise given the large population sizes at Early Testing 1 and the current breeding program operating budgets. The planting material is normally bulky which also presents a challenge for multi-environment testing.

2. ***Restricted cross-border dissemination of clones:*** cross-border sharing of clones presents increased risk of transfer of plant pathogens across countries. This necessitates costly indexing and treatment of clones for transfer (between $6-11 per sample). The current breeding program operating budgets may not support these treatments for a large population at Early Testing 1.
3. ***Issues with cross-incompatibility and flowering synchronization***: Several of the RTB crops have problems achieving all the targeted crosses due to cross-incompatibility and lack of synchronization in flowering of parents of the targeted crosses. Some crops may suffer from this more than others. This necessitates more research also dedicated to solving problems with flowering and crossing compatibility.

A RCGS archetype for RTB crops must consider these constraints for efficient implementation of the breeding strategy. Here, we provide breeding scheme options relevant to RTB as examples.

Population improvement

Most of the RTB crops are amenable to breeding within one pool, except Banana (Plantain and Matooke), which due to biological reasons can only be improved across pools and ploidies. Recent studies have indicated that diploid RTB crops such as Cassava might also benefit from reciprocal recurrent selection schemes. Here we give examples of RCGS schemes both in one pool and two pools, however, CGIAR-NARES-SME breeding networks will need to work with quantitative genetics specialists to develop RCGS schemes for respective crop specificities and available resources.

Selecting parents or crosses: One-pool breeding strategy

Depending on the program size, 50-200 crosses with 50-200 individuals per cross are recommended. Crossing parents are selected using genomic prediction of cross performance (GPCP) combined with optimal cross selection, based on the desired gains index including essential traits captured in the Target Product Profile (TPP). The number of parents is derived from the crosses identified by GPCP and OCS and not restricted to a recommended value. All parents for the crosses should come from Early Testing 1 data. CGIAR-NARES-SME Breeders will work with quantitative genetics specialists to apply GPCP + OCS and optimize the targeted number of crosses and individuals per cross, depending on the pipeline size, available resources, as well as logistical and biological limitations.

Selecting parents or crosses: Two-pool recurrent selection breeding strategy

All parents or crosses need to be selected based on Early Testing 1 data to leverage on speed. To select parents for recycling within pool, genomic estimated general combining ability (GEGCA) must be used with an inbreeding control method like OCS. To select parents and crosses for hybrids across pools, genomic prediction of cross performance

(GPCP) should be used, which is the predicted mean GEGCA+GESCA (genomic estimated specific combining ability) of a cross. The GEGCA and GESCA are predicted using hybrid genotypes. In RTB crops the parents are non-inbred, meaning that many loci are heterozygous and will be segregating in the progeny of any cross. Therefore, it is impossible to infer hybrid genotypes from parents' genotypes and, consequently, both parents and hybrids would have to be genotyped. Between 50-200 crosses, with 50-200 individuals per cross across pools, (and 25-100 within pool crosses) are recommended. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to decide on the best methods given crop-specific challenges e.g. multiple ploidies in Banana, and available resources and on the population sizes and structure for respective pipelines.

Managing Early Testing 1 trial design, distribution among CGIAR_NARES-SME Partners and mid-density genotyping

Distribution of true seed, visual selection, marker-assisted selection and mid-density genotyping

Given the across-border exchange challenges with clones, distribution of true seed might be the easier/safer option. The population improvement hub is responsible for designing Early Testing 1 trials and distributing the Early Testing 1 planting material among the partners via true seed. The sparse testing design should allow the connectivity of the family structure among network partners, by splitting the full sibs adequately among partners. A full set of the families should remain at the population improvement hub to be used in recycling when data comes back from CGIAR-NARES-SME partners. Partners, with backstopping from the CGIAR population improvement hub, are responsible for germinating the true seed, carrying out visual selection, and marker-assisted selection as needed at seedling nursery stage before Early Testing 1, according to set out protocols and standard operating procedures developed by the breeding network partners. Only material surviving visual and marker-assisted selection will move into Early Testing 1. The total initial segregating population size will depend on the expected selection pressure during visual selection and marker-assisted selection steps (normally between 80-95% in RTB), to achieve a final target population size needed for Early Testing 1. CGIAR-NARES-SME breeding networks should work with quantitative genetics specialists to define the optimal segregating population sizes for their respective pipelines. The CGIAR-NARES-SME network partners are responsible for sampling all genotypes going into Early Testing 1 for genotyping using the mid-density panel for genomic selection. The population improvement hub is responsible for coordinating this sampling and mid-density genotyping.

Distribution of clones multiplied via semi-autotrophic hydroponics (SAH)

Semi-autotrophic hydroponics technology is a potentially cost-efficient way for rapid clonal propagation. It is amenable to large population sizes such as those expected at Early Testing 1. However, if not germinated directly from seed, an additional cost of pathogen indexing and cleaning is needed (~$6-11 per sample). In the case of exchange of Early Testing 1 trials among CGIAR-NARES-SME partners in the form of clones, the population improvement hub will be responsible for crossing, visual selection and marker-assisted

selection in the seedling nursery, and distribution of clones that advance to Early Testing 1 (that survive visual and marker-assisted selection steps). A set of clean clones at Early Testing 1 is maintained at the population improvement hub to be used as new parents when data is returned from the network partners.

Managed stress screening

Early Testing 1 agronomic trials should be primarily conducted to sample the TPE, with linkage among trials provided by full sibs and the GRM rather than full replication of selection candidates over locations. Additional Early Testing 1 trials at key research stations may be conducted to generate specific phenotypes (e.g. drought response, optimal management response, performance under managed flooding, disease screening platforms, among others). These will need to include the full set of the selection candidates because there are usually only one or two locations per managed stress screening. The size of populations should be able to achieve this as well as representing the population structure across the sparse testing sites. Breeding teams should work with quantitative genetics specialists in establishing the best population size for their respective pipelines.

Support data management and analysis for cross selection, recycling of parents and advancement of products

CGIAR-NARES-SME partners are responsible for Early Testing 1 and returning data for joint data-driven decisions towards either selecting crosses and recycling new parents or advancing potential products. Genotypic and phenotypic data must be centrally managed using a GI-supported breeding database system (EBS, BMS, or BreedBase) linked to a genomics database (GIGWA or any other platform). NARES partners must be supported to electronically receive experimental designs, collect and upload data, and access results using these systems. The CGIAR Breeding Analytics team have developed a full analytical pipeline for GS, integrating the desired gains index and optimal cross selection, as well as environmental co-variates. Breeding networks will need support from a quantitative genetics specialist to implement this pipeline.

Role of regional germplasm exchange hubs

Recently, Crops to End Hunger (CtEH) initiative funded the development of the regional Roots, Tubers, and Bananas East Africa Germplasm Exchange Laboratory (RTB-EAGEL) based at the Kenya Plant Health Inspectorate Service (KEPHIS). Rapid-cycle genomic selection can leverage such a platform to transfer advanced clones among network partners. Other purposes of such a platform include maintenance and transfer of released varieties across regions.

Fig 3: Example of an RTB RCGS breeding scheme recycling using data from CET-clonal evaluation trial in a one-pool breeding strategy.

Numbers will be depended on the respective pipeline/crop and breeding teams should work with quantitative genetics specialists to optimally define these. Furthermore, given that Early Testing 1 (CET) is more extensive, resources can be reallocated from other testing stages e.g. Early Testing 2 (PYT) and Late Testing 2 (UYT) might no longer be necessary.

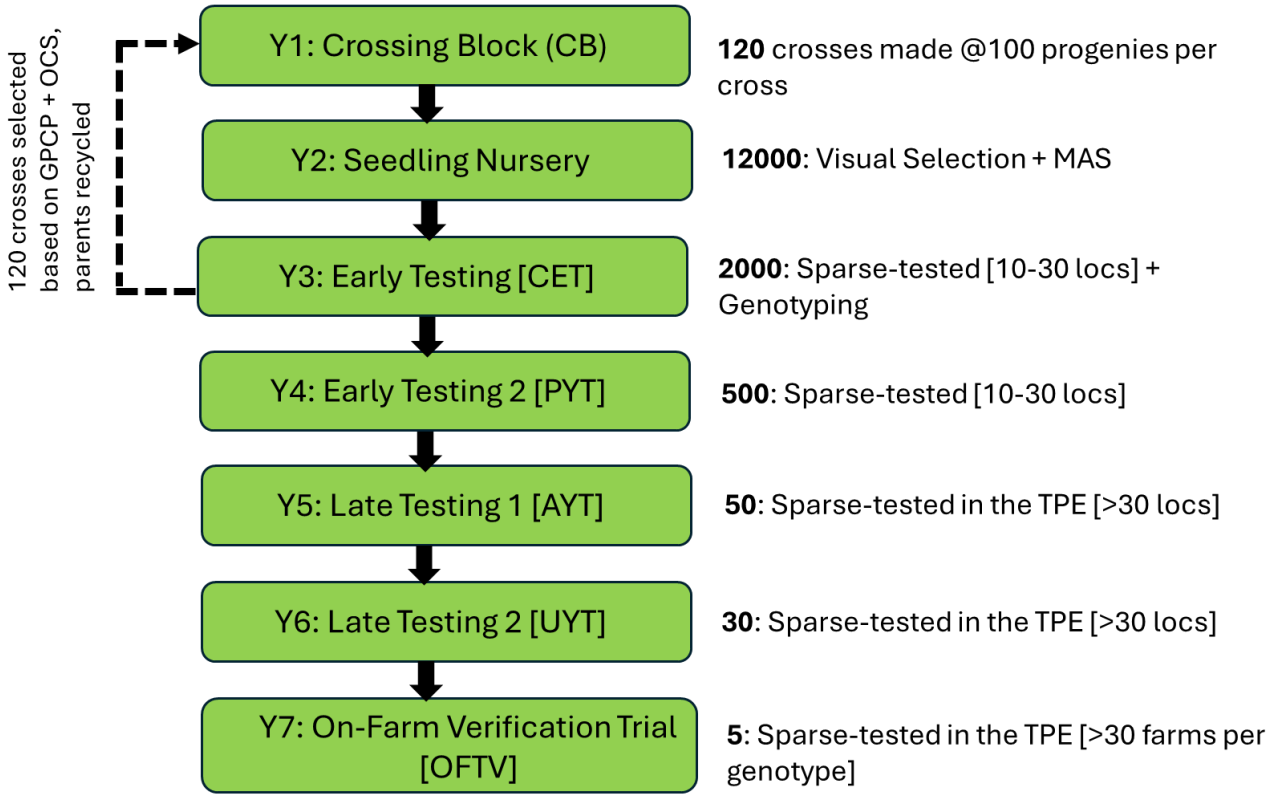
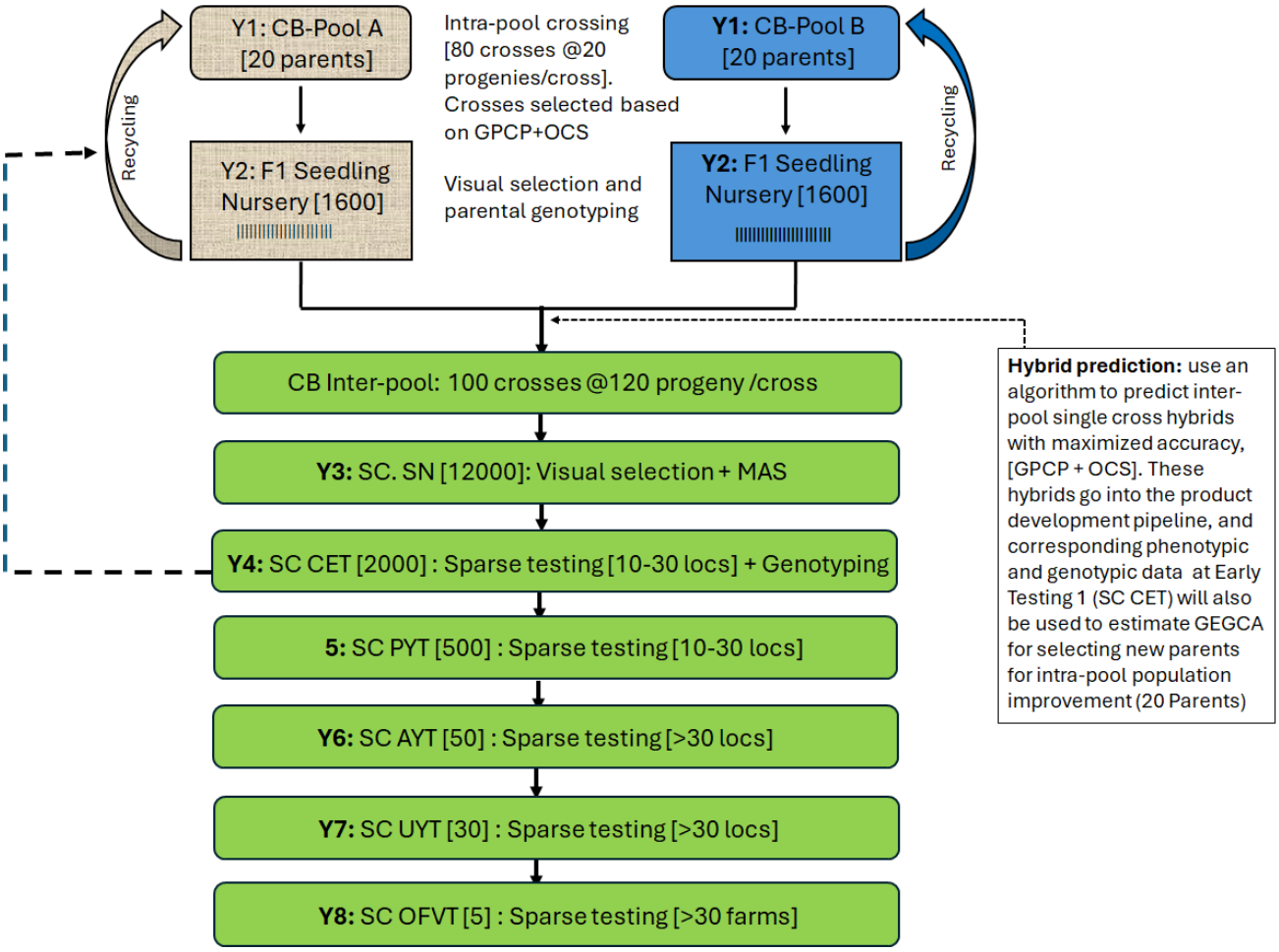


Fig 4: An example of an RTB RCGS breeding scheme recycling using data from CET-clonal evaluation trial in a two-pool breeding strategy.

Numbers will change depending on the actual pipeline/crop and breeding teams should work with quantitative genetics specialists to establish what is optimal for their respective pipelines. Furthermore, given that Early Testing 1 is more extensive, resources can be reallocated from other testing stages e.g. Early Testing 2 and Late Testing 2 might no longer be necessary.



3.1.2 Example RCGS archetype for hybrid crops with inbred parents

Hybrid crops with inbred parents bred by the CGIAR and its partners include Maize, Pearl Millet, Sorghum, among others. Population improvement involves improving the inbred lines using information from their test hybrids. Here, we give an example of a hybrid reciprocal recurrent selection breeding scheme that does not use cytoplasmic male sterility.

Example RCGS archetype for hybrid crops with inbred parents

Population improvement

Select parents, make crosses and derive lines: 20-50 parents of the next generation will be selected based on genomic-estimated general combining ability (GEGCA) across testers from the opposite pool as estimated from testcross evaluation. The GEGCA for each trait will be combined in a desired gains selection index to select the parents. Closed, complementary populations will be established for each side of the heterotic pattern. Within-pool crosses need to be done based on optimal cross selection (OCS) among the high GEGCA parents, to balance genetic gain and genetic variance. Between 25-100 within-pool crosses and between 50-200 lines per cross (population) are recommended. However, breeding teams should work with quantitative genetics specialists to determine the optimal numbers for their respective pipelines. Selection candidates are DH lines or lines generated through inbreeding. If through inbreeding, line selection as parents for the next cycle of crosses will generally be at the F3 (S1.2) stage to reduce cycle time. In future, lines can also be selected from F2 (S0.1) if this can further remove a full year from the cycle, with measures to generate commercial selection candidates with adequate uniformity put in place. Breeding teams should work with quantitative genetics specialists to optimally implement this crossing and line derivation step.

Visual selection and marker-assisted selection of lines:

The lines from each cross should be grown out as progeny rows, preferably in the main season, providing an opportunity to select for highly heritable traits. Lines that survive the visual selection step can be screened using marker-assisted selection. At current diagnostic genotyping costs (ca. $2/sample) it is likely to be most efficient to screen with diagnostic markers only those lines surviving the visual selection step. However, if doing visual selection and marker-assisted selection sequentially adds a full season to the breeding process, the two can be done concurrently. In addition to selection, the purpose of this nursery is to generate enough seed to make testcrossing and line maintenance for the following stages.

Managing testers and sparse testcrossing of lines:

All testers available to a given pipeline across different stages will be used to create testcrosses at Early Testing 1. Testers should be the latest commercially relevant female single cross hybrids of the opposite pool. The testers should be replaced every cycle, with the process of identifying new testers being documented as a standard operating procedure for the crop across pipelines and partners. In sparse-testcrossing, lines will be nested within testers. For example, if there are 5 testers, and 1000 lines, 200 of the lines are crossed to each tester, with about 50 lines being crossed to all testers. As the implementation of the strategy continues and logistics are efficiently in place more aggressive testcrossing strategies like where no line is crossed to all testers can be implemented. The family structure (populations) of the lines should be split among the testers.

Manage Early Testing 1 trial design, distribution to partners and mid-density genotyping for GS:

Sparse testing will be used to phenotype testcrosses at Early Testing 1. Because in sparse testing design, the family structure provides replication of haplotypes across the TPE, the full- and half-sib populations need to be distributed among testing sites in a manner that provides the replication needed to achieve adequate precision. For example, if there are 20 testing sites, there should be at least one full sib tested in each testing site. A 5-10% partial replication across sites can be applied. Breeding teams should work with quantitative genetics specialists to design the sparse testing trials and ensure that the numbers are adequate to provide allele replications across testing sites as well as any managed stress trials that may need full sets of testcrosses. The population improvement hub is responsible for designing the sparse Early Testing 1 trial and ensuring that it is distributed to CGIAR-NARES-SME partners safely and in good time for planting. All selection candidates included in Early Testing 1 (including lines, testers and testcrosses) need to be genotyped at a density permitting calculation of the GRM (about 1000-5000 SNPs) using the GI-contracted mid-density service for the crop. Because this genotyping is relatively expensive (approx. $10), it may be most cost-effective to restrict it to lines that have survived both the visual and marker-assisted selection steps. As genotyping costs change, other plans may become more efficient.

Managed stress screening:

Early Testing 1 agronomic trials should be primarily conducted to sample the TPE, with linkage among trials provided by full sibs and the GRM rather than full replication of selection candidates over testing sites. Additional Early Testing 1 trials at key research stations will be conducted to generate specific phenotypes (e.g. drought response, optimal management response, performance under managed flooding, disease screening platforms, etc). These will need to include the full set of the selection candidates because there are usually only one or two locations per managed stress screening. The size of populations should be able to achieve this as well as representing the population structure across the sparse testing sites. Breeding teams should work with quantitative genetics specialists in establishing the best population size for their respective pipelines.

Support data management and analysis for parent selection and advancement:

Genotypic and phenotypic data need to be managed to permit accurate selection. Data must be centrally managed using a GI-supported breeding database system (EBS, BMS, or BreedBase) linked to a genomics database (GIGWA for most programs). NARES partners must be supported to electronically receive experimental designs, collect, and upload data, and access results using these systems. The CGIAR Breeding Analytics team have developed a full analytical pipeline for GS, integrating the desired gains index and optimal cross selection as well as environmental co-variates. Breeding networks will need support from quantitative genetics specialists to implement this pipeline.

Selection for product advancement:

Given that all testers used within a pipeline are used at Early Testing 1, additional testcrosses (full sibs of the testcrosses in Early Testing 1) could be planted at research stations to visually select potential genotypes that meet all minimum TPP thresholds for

the commercial pipeline. This is with an aim of increasing the selection intensity for the products advancing to later stages of testing. Hybrids meeting the minimum TPP thresholds for qualitative traits and have high GEBV for the quantitative traits as estimated from the full-sibs in Early Testing 1 trial, will be advanced to later stages of product development, in addition to the selections made for advancement to later stages of product development by CGIAR-NARES-SME from the Early Testing 1 sparse trials.

Table 2: Example of a 3-year breeding scheme for RCGS in hybrid crops with inbred parents. The lines are derived at F3. For a breeding scheme using DH technology, breeding teams should work with quantitative genetics specialists.

Year	Season	Activity 1	Activity 2	Notes
1	**Off-season**	Grow parents, Make intra-pool F_1: 30 parents, 100 crosses		Only elite lines selected based GEGCA desired gain index are used as parents. Crosses among the selected parents are made based on OCS
1	**Main**	Grow F_1, Make F_2: 100 populations with 100 individuals per population	QA/QC of the F_1	
2	**Off-season**	Grow F_2, Make $F_{2:3}$: 80 populations with 80 individuals per population		
2	**Main**	Grow $F_{2:3}$ lines, select F_3 derived lines for Early Testing 1: 20 populations of 50 individuals per population	Visual selection + MAS.	Seed increase for Early Testing 1, testcrossing, remnant seed maintained
3	**Off-season**	Sparse-testcross 1000 best F_3 derived lines + 5 testers: 200 lines each per tester, (10 individuals per population per tester for 20 populations)	Genotype selected lines + Testers with mid-density panel for GEGCA estimation	
3	**Main**	Distribute the 1000 testcrosses Testing 1 sparse trials: 10-30 sites	Managed stress screening on-station of full set	More full sibs per testcross than needed for the core population improvement Early Testing 1 can be developed and evaluated for product development

4.0 Accountability to the crop breeding network success and key performance indicators

To ensure successful implementation, each CGIAR-NARES-SME partner needs to be accountable to other partners regarding responsibilities under their implementation. The breeding network needs to evaluate their collective performance and health, hence needing to put in place core key performance indicators. Some suggested indicators include:

1. Number of half- and full-sib families developed by the population improvement hub
2. QA/QC results for parental and F_1 purity
3. Seed distribution rate and adherence to crop calendars
4. Number of locations used in testing (across the network)
5. Heritability across locations from individual partners
6. Heritability across locations (joint across all partners)
7. Genetic variance
8. Selection intensity (for recycling)
9. Average age of parents
10. ….

Dorcus Gemenet, Breeding Optimization Lead & ACCELERATE Lead, d.gemenet@cgiar.org

Michael Quinn, Accelerated Breeding Lead, m.quinn@cgiar.org

CGIAR is a global research partnership for a food-secure future. CGIAR science is dedicated to transforming food, land, and water systems in a climate crisis. Its research is carried out by 13 CGIAR Centers/Alliances in close collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations and the private sector. www.cgiar.org

We would like to thank all funders who support this research through their contributions to the CGIAR Trust Fund: www.cgiar.org/funders.

To learn more about this Initiative, please visit this webpage.

To learn more about this and other Initiatives in the CGIAR Research Portfolio, please visit www.cgiar.org/cgiar-portfolio

INITIATIVE ON
Accelerated Breeding