THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTUTUT AGRO MONTPELLER ET DE L'UNIVERSITE DE MONTPELLIER

En Génétique et Amélioration des Plantes
École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Portée par

Unité mixte de recherche AGAP Institut
Amélioration génétique et adaptation des plantes méditerranéennes et tropicales

Integration of genomic prediction in a recurrent selection scheme: the example of the CIAT-Cirad rainfed rice breeding program

Présentée par Cédric Bärtschi
Le 21 juin 2022

Sous la direction de Jean-Marc Bouvet
Co-encadrée par Cécile Grenier, Jérôme Bartholomé, Tuong-Vi Cao Hamadou

Devant le jury composé de

Florence Phocas, Directeur de recherche, INRAE Jouy-en-Josas	Rapporteure
Gilles Charmet, Directeur de recherche, INRAE Clermont-Ferrand	Rapporteur
Judith Burstin, Directrice de recherche, INRAE Dijon	Examinatrice
Jacques David, Professeur, SupAgro Montpellier	Examinateur
Cécile Grenier, Chercheuse, Cirad AGAP Institut Montpellier	Co-encadrante
Jean-Marc Bouvet, Directeur recherche, Cirad DGDRS Madagascar	Directeur de thèse

UNIVERSITÉ DE MONTPELLIER

L'INSTITUT agro Montpellier

Remerciements

Cette thèse aura été l'occasion de découvrir à quel point l'écriture est pour moi un exercice difficile. Je ne vais donc pas profiter de ces remerciements pour vous exposer ma prose sophistiquée. Néanmoins, je ne peux pas rendre cette thèse sans remercier toutes les personnes qui m'ont soutenu ces trois dernières années et permis d'achever le document que vous avez entre les mains.
Merci tout d'abord à mon équipe encadrante. Aucun mot ne saurait exprimer ma gratitude, ni leurs rendre toutes ces précieuses heures qu'ils ont investi pour moi. Merci à Cécile, qui aura toujours été là et qui m'aura encadrée avec enthousiasme et énergie malgré tous les soucis que j'ai pu lui donner. Merci à Jérôme, pour son calme, son pragmatisme et sa disponibilité peu importe le continent sur lequel il se trouvait. Merci à Vi et son perfectionnisme qui aura eu la lourde tâche de m'initier à la pratique de la génétique quantitative. Merci à Hugues, qui sans être dans mon encadrement m'aura donné de son temps et son énergie. Merci aussi à Yolima et toute son équipe en Colombie qui ont assuré le travail malgré la pandémie et les confinements. Finalement merci à Jean-Marc, qui malgré la distance et ses nombreuses responsabilités aura su se montrer disponible.
Merci également aux membres de mon comité de thèse Laurence Moreau, Mathias Lorieux et Patrice This qui auront fait de leur mieux pour aider et orienter le petit être borné que je suis.
J'aimerais aussi remercier tous les membres de mon jury, mes deux rapporteurs Florence Phocas et Gilles Charmet et à mes deux examinateurs Justin Burstin et Jacques David. La planification de ma soutenance a été un peu laborieuse et l'aurait été encore plus sans leur flexibilité et leur disponibilité.
Une thèse ce n'est pas que du travail, c'est aussi beaucoup de café. Merci à Fabien, Julien, Sergio et David pour tous ces bons moments et toutes ces pauses méritées ou pas. Merci aux autres doctorants, Charlotte, Aurélie, Benjamin et tant d'autres, mes compagnons d'infortune avec qui j'ai partagé rêves, ambitions et désillusions. Enfin, merci à tout mon entourage et leur soutient, à ma famille qui, bien que j'ais disparu durant trois ans n'ont pas oublié qu'ils ont un fils et un frère et à Lorna dont l'amour, la patience et la foi en mes capacités auront été précieux.

Table of content

List of figures

List of tables

Chapter 1 : General introduction

1.1 General context

Food demand is expected to increase between 45% and 51% by 2050 (van Dijk et al. 2021) as a consequence of the estimated world population growth, which should reach ~9 billion people in 2050 (UN 2019). Combined with an increased share of animal product in the diet resulting from greater incomes to a part of the world, the demand for crop and grass could even increase by up to 165% by 2100 (Bijl et al. 2017).

Two-thirds of caloric intake (Kromdijk and Long 2016) of the world population are covered by only four food crops: rice (*Oryza sativa* L.), wheat (*Triticum sp.*), maize (*Zea mays* L.) and soybean (*Glycine max* (L.) Merr.). So far, the combined production and stocks of these essential crops covered the world demand (FAO 2021), but future needs are not sure to be met (Ray et al. 2013). Increase in total production for almost all staple crops, with the exception soybean in developing countries, have been driven mainly by yield growth rather than surface increase (Fischer, Byerlee, and Edmeades 2009) (Figure 1-1). In the last 40 years, agricultural land has increased by 5% (Stehfest et al. 2019, data from FAOSTAT) and is expected to represent about 107.3% of the 2007/2009 reference surface by 2050 (Alexandratos and Bruinsma 2012). This value already includes the future loss of arable land for urbanization, soil degradation and conversion to forestry of protected area. If surface increase has participated in increase in production, it has never been and will not be the sole driver of production increase in the future.

From 1990 to 2010, the rate of yearly yield gain, relative to 2010 reference yields, was about 1% for wheat, rice and soybean and 1.5% for maize worldwide (Fischer, Byerlee, and Edmeades 2014). Part of this increase in yield was due to closing the yield gap which is the difference between the on-farm yield (FY) and the maximum potential yield (PY) for a cultivar in its target environment, with optimal

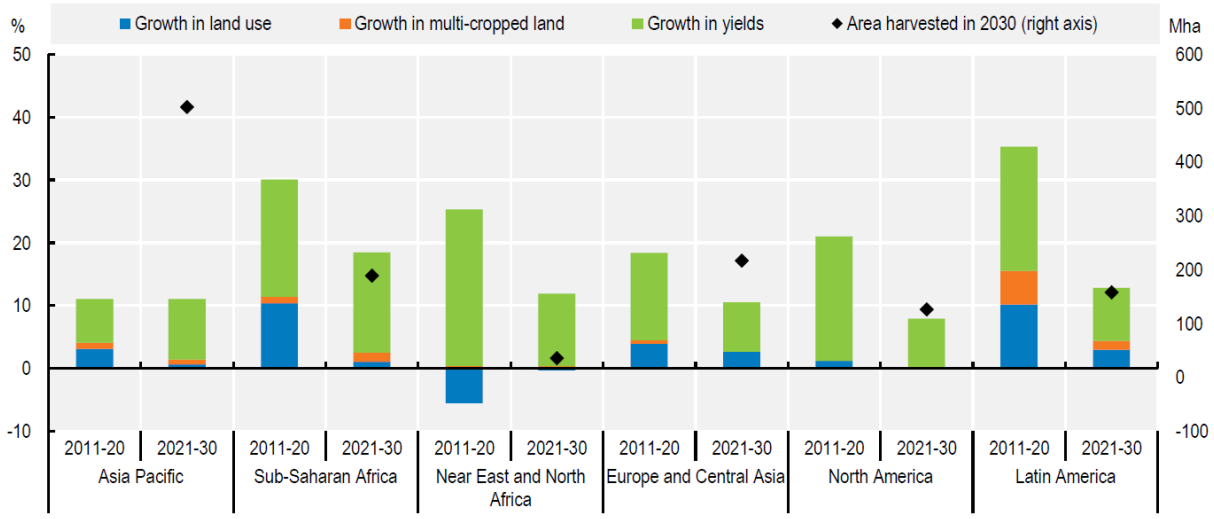


Figure 1-1: Decomposition of total production growth (2011-20 and 2021-30) into growth in land use, land intensification through growth in multi-cropped land, and growth in yields. It covers the following crops: cotton, maize, other coarse grains, other oilseeds, pulses, rice, roots and tubers, soybean, sugar beet, sugarcane, wheat and palm oil. (OECD and FAO 2021)

agronomic practice and no biotic stress. Since the 1990s, the yield gap has been reduced at a rate of 0.5% per year on average thanks to change in agricultural practices, switching to irrigation being among them (Fischer, Byerlee, and Edmeades 2014). However, the gap fluctuates across the years as it depends on the price the farmers are expecting for their crops. They will invest more in the inputs than allow them to close this gap only if they expect a return on investment. Closing the yield gap still has enormous potential to improve food security as it represents e.g. for rice in developing countries, between >50% to >120% of the currently achieved FY. Large yield gaps are mostly encountered in developing countries. This situation is caused by the socio-economic conditions and absence of infrastructure that might not evolve quickly (Fischer, Byerlee, and Edmeades 2014). History shows that farmers are keener to change cultivars than agriculture practices (Fischer, Byerlee, and Edmeades 2014). Hence, increasing the FY by increasing PY could mitigate food scarcity and give farmers time to adopt improved agriculture practices.

The PY of a variety is measured for its target environment. Changes in the environment will make old cultivars obsolete for new conditions and the PY could drop if new, adapted cultivars are not provided fast. Future yields are expected to be severely impacted by the increase in atmospheric CO_2 and the resulting climate change (IPCC 2022). The experts' report on climate change states that the higher temperature, change in rainfall patterns, and higher frequency of extreme events would put crop yield under even more pressure.

Plant breeders have the task to constantly develop cultivar with higher PY. The annual yield increases have been so far around 0.5% to 0.8% for rice wheat and soybean and 1.1% for maize but it is not expected to be sufficient to cover the increase in demand (Figure 1-2). Worst still, lower rates have been observed in the last 20 years (Fischer, Byerlee, and Edmeades 2014). Not only will crop breeding have to provide higher yielding cultivars but also products corresponding to the ideotypes that would

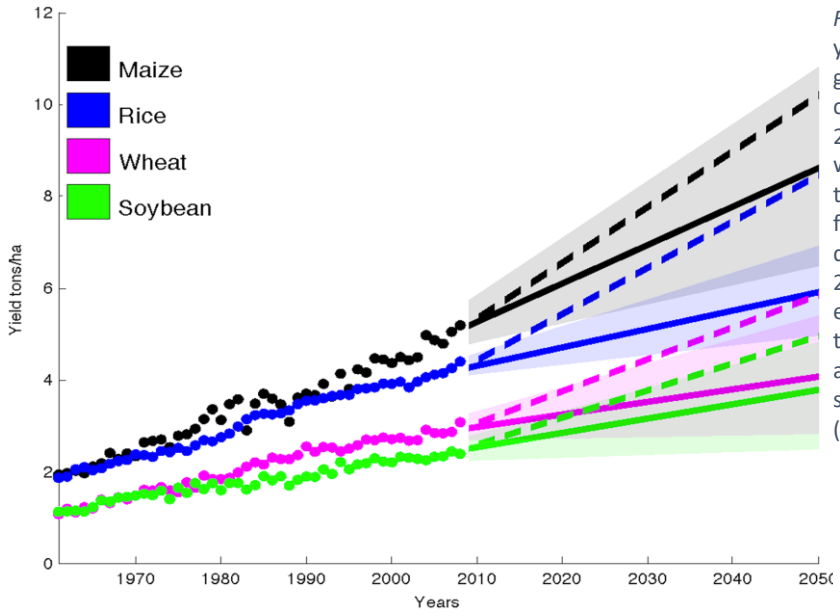


Figure 1-2: Global projections for yields. Observed area-weighted global yield 1961–2008 shown using closed circles and projections to 2050 using solid lines for maize, rice, wheat, and soybean. Shading shows the 90% confidence region derived from 99 bootstrapped samples. The dashed line shows the trend of the 2.4% yield improvement required each year to double production in these crops by 2050 without bringing additional land under cultivation starting in the base year of 2008. (Ray et al. 2013)

be adapted to new agricultural practices and new environmental conditions, not to mention new market demands. Breeding is, however, a lengthy endeavour. For annual crops it takes about 8 to 12 years between the cross and release of a new cultivar. To continue contributing to food security and to rapidly respond to the yet unknown challenges, plant breeding needs to improve its practice to enhance genetic gain and eventually accelerate the rate at which it provides adapted new cultivars.

1.2 Plant breeding

The role of a plant breeder is, as said by Bernardo (2008) (citing Dudley and Moll (1969)): *"(i) to create genetic variation mainly by crossing good by good, (ii) select the best progenies in the cross, and (iii) synthesize the best progenies into a new and improved cultivar*".

Plant breeding is the science of creating plants that fit human needs. It has been practiced intuitively since the dawn of agriculture 10,000 years ago, beginning with the unconscious domestication of species responsive to domestication and gradually evolved to an empirical science until the founding of the genetics as a research field. The discovery of Mendel's work at the beginning of the 20th century and the later foundation of quantitative genetics (Fisher 1919) have given the theoretical foundation to what has been successfully practiced for millennia already.

Besides the adaptative traits for cultivation, the required attributes that the selected crops needed to bear have evolved over the time following agricultural practices and societal requirements. Yield was and will stay central for every crop but the ideotype targeted to reach the highest yield did change over time. The ideotypes allowing the highest yields under animal- powered agriculture are not the same as the one adapted to mechanized agriculture. Similarly, different ideotypes are necessary for low input or high input agriculture. Finally, the environmental conditions in which the crop is grown are central.

Plant breeding can be seen from two angles: (i) the type of cultivar expected in farmers' field and (ii) the best methodology available to create the end product. The cultivar is the product that the farmers will acquire or purchase and grow in their fields (Table 1-1). To be released as a new variety, a cultivar must show uniformity, stability across the year, distinctiveness from existing cultivar as well as provide novelty. The methodology to develop the new cultivar is the choice of the breeder. It will be defined mostly by the type of end product required and the reproductive system of the species under selection. If we consider vascular plants, there are two sexual reproductive system: allogamy (open-pollination) and autogamy (self-pollination). While most wild species are allogamous (Zohary 2001) only a few important cereal crops such as maize (monoecious), pearl millet (*Pennisetum glaucum* (L.) R. Br.) (protogynous) or rye (*Secale cereal* L.) (self-incompatible) are spontaneous or obligate outcrossers. On another and while autogamy is relatively rare in the wild, most crops, such as wheat, rice, barley

(*Hordeum vulgare* L.), oat (*Avena sativa* L.) to only name a few, are autogamous species. In practice most plant will show both reproductive system but with a strong preference toward one.

Part of the breeders' work is two tweak those reproductive systems in one or another direction to get crosses from autogamous plants or to self the allogamous ones, to synthetize the expected improved cultivar type. Nucleic male sterility genes (ms-genes) are one of the genetic tools available to control autogamy (Rao, Devi, and Arundhati 1990). It can be used to facilitate large scale recombination and help make populations of intermating individuals and make an autogamous species behave like an allogamous one. The management of the breeding will depend on the degree of dominance of the male sterile allele. Essentially, one needs to ensure that the population under selection segregate for the ms-gene to have both male fertile and male sterile plants within a segregating progeny. As the male sterile plant does not produce fertile pollen, it will behave as allogamous plants. By harvesting their seeds, one can ensure offspring will be from a cross (F_1 or S_0). This type of gene exists in rice, sorghum and wheat. I will later discuss more in depth for in the case of an ms-gene in rice.

1.2.1 Breeding methods

Depending on the cultivar, different breeding approaches may be used. In the next sections, I will describe three common approaches that are relevant to this work.

1.2.1.1 Mass selection

Mass selection is the most straightforward selection method. First a population segregating for the traits of interest is either chosen (natural population, landraces, …) or created by crossing parents relevant for the breeding objectives. Then single plants are compared to each other in the population. If the goal is to find some good material to develop an improved population the breeder looks for above average plants. It can however also be used to purify an existing population or lines and here the goal will be to root out the out-of-type plants. Once the plants are selected, their seeds are bulked to produce a new population. For autogamous species, the population will progressively become more homozygous and mass selection can be a rapid and inexpensive method for increasing the frequency of desired genotypes in the population (Fehr, Fehr, and Jessen 1991). For the allogamous species mass selection can also be applied but in the context of recurrent selection (see below). This type of selection is mostly adapted for high heritability traits such as qualitative disease resistance (Fehr, Fehr, and Jessen 1991). It is also used in evolutionary plant breeding to discard unfit plants from various progenies families (Merrick et al., 2020).

Table 1-1: Non-exhaustive list of cultivar types and their main characteristics

Type	Reproductive method	Description	Advantage	Disadvantage	Examples
Synthetic or composite population	cross-pollination	Mix of closely related individuals spontaneously outcrossing	- Broad genetic base - Adaptability due to plasticity at the population level - Lower pressure on biotic stresses due to more diversity in the tolerance/resistance mechanisms	- Less homogenous (quality traits, architecture, …) - Risk of drift within the population if strong selection pressure occurs	rice, wheat, sorghum, maize, pearl millet, forage crops
Pure lines	self-pollination	Individuals in approximate homozygosity that share all the same genotypes. Genotype are said to be fixed when all positions have become homozygous because of inbreeding. The progeny will have the same genotypes.	- Easy to reproduce and to maintain as homogenous seed stock (purebred) - Homogeneous appearance and performance of the cultivar	- Sensitive to stress due to lack of adaptability resulting from the genetic uniformity of the crop - No evolution of the genetic make up of the cultivar	rice, wheat, sorghum, maize
Hybrids	cross-pollination	F_1 from two selected parent according to their combining ability. In self-pollinated species CMS can be used (three way system) GMS (two line hybrid (environmental-sensitive GMS))*	- Capacity to exploit the hybrid vigour (heterosis)	- No maintenance of hybrid the phenotypic value in later generations - Seeds have to be acquired yearly (often purchased at seed compagnies in charge of hybrids development)	maize, rice, cotton, wheat, barley
Clone	vegetative propagation/apomixis (asexual)	Plants genetically identical developed from a portion of the plant body / seeds without sexual reproduction	- Exact similarity with the parental plant - Easy to maintain a variety	- Extreme susceptibility to abiotic stresses susceptibly	potatoe, cassava, braccharia, many roots and tuber crop, Rice **

* CMS cytoplasmic male sterility; GMS genic male sterility in (Abbas et al. 2021)

** apomictic rice in development (Guiderdoni 2021)

1.2.1.2 Pedigree selection

Pedigree selection is a more sophisticated approach to selection. The starting point will also be a population with some diversity (normally a F_2 from a bi-parental cross but it can come from a population breeding scheme). The first step will be similar to mass selection as plants will be compared to each other in the base population. Then, at each following generation the progeny of each selected plant is planted and the genealogy of the families is recorded. Selection is then made among the families (offspring from the same plant) but also within families, always keeping track of which plant comes from which family. As the performance of the ancestor is known, the selection can be made on the actual performance observed in field but also relative to the family performance in the former generations. As the genetic relationship of lines is known this information can be used to maximize the genetic variability among the final set of material retained as best candidates for variety release (Fehr, Fehr, and Jessen 1991).

The advantage of the methods is that it discards inferior lines before complete fixation. However, for low heritability traits, selection has to be performed in more advanced generation when the lines become more homozygous (Collard and Mackill 2008).

1.2.1.3 Single seed descent

While pedigree breeding allows eventual accurate selection, it is labour intensive as well as time consuming. As each generation requires evaluation and selection with a thorough record of each selected plant, it must be conducted in conditions representing the target environment and the recurrency of the phenotyping is limited to the growth season. Additionally, selection in the early generations is done on segregating material. Hence much work might be invested in families that could be rejected in later generations. Single seed descent (SSD) is an attempt at tackling those issues in pedigree breeding. It rapidly advances large populations to a fixed homozygous state, where the evaluations are more accurate while keeping the workload low. To enable this, the generation advances are done through selfing and only a single seed is kept per family (sometimes increased to multiple seeds) from F_2 up to the generation where homozygosity is deemed sufficient (F_6 or F_7) (Fehr, Fehr, and Jessen 1991). This leads to rapid inbreeding with minimal work.

Using a single seed for each generation advance strongly reduces the workload. The populations of F_2 plants derived from a cross (F_1) must however be large enough that, even after four or five events of Mendelian sampling (the generation advances), all possible genotypes with the parental alleles (recombinant inbred lines or RILs) are still present in the population and selection can be realized on them.

1.2.1.4 Recurrent selection

When mass selection is practiced on autogamous species, the genotypes will increasingly become homozygous (if we set aside the small spontaneous outcrossing). When the same process is done on allogamous species, the frequency of favourable allele increases in the breeding population and so does its mean, but new genotypes will be produced at each generation. Recurrent selection (RS) as named by Hull (1946) is a cyclic method to improve population means for the traits under selection. Although any selection scheme can be seen as "recurrent" since new elite cultivars are eventually crossed together, RS specifically describes a breeding scheme that improves a population by short-term cycles of selection (Fehr, Fehr, and Jessen 1991). The increase in population mean is done by selecting the best individuals or families as they should carry more favourable alleles for a given trait. This will, over the cycles, increase the frequencies of the favourable allele and thus increase the probability of assembling the ideal genotype or at least superior genotypes (Gallais 2011).

Different methods are used to select the recombining unit. Mass selection is the simplest method and consists of comparing single plants within the whole population. Again, this is adapted to high heritability traits. If one wants to make repeated measures to evaluate the candidate materials (plants either selected on high heritability trait or randomly selected), as required for estimating quantitative traits, or to test the performance under multiple environments, advance generation might be necessary to increase the number of plants that can be observed. Different schemes exist for progeny testing, as the progenies can be generated by selfing or controlled crosses to obtain half-sibs or full-sibs families Considering the generation advance is done by selfing we will also get increasingly homogeneous material. A good review of the schemes and and their practical application on rice was done by Châtel and Guimarães (1997).

As mentioned earlier, many other plant breeding methods exist for cultivar development (hybrid development, backcross breeding …) but will not be presented in this short introduction.

1.2.2 Assessment of a breeding program

A plant breeding program can be assessed in several ways. Its commercial success is easily assessed by the number of successful varieties deployed and/or the royalties they generate. This is clearly important as it shows the adequacy of the product targeting and the adequacy of the available breeding populations to select for those targets. It says little, however, on the actual improvement of the breeding population and on the long-term trends. This success depends not only on the intrinsic quality of the variety but also on its competitiveness on the market. Another possibility is to measure the realized genetic gain, also named response to selection as progress in population mean may not be positive (e.g. plant height or cycle length) and is the change in population mean over the time. If records of population mean are often present in breeding program archive, a good measure of the

genetic gain remains challenging. Indeed, the evolution of the populations will be confounded with potential long-term climatic trend or changes in agronomic practices. Yet, corrections through statistical models are possible, if genetic connectivity between years exists (Rutkoski 2019). Given the suitable data are available several methods exist to measure the realized genetic gain but with limited accuracy (Rutkoski 2019).

1.2.3 The breeder's equation: a guide for program improvement

The measure of realized gain is important for the assessment of the past years but one might want to know how the current program will perform in the next years. The breeder's equation (Lush 1937) allows one to predict the genetic gain from one generation of selection based on the characteristics of the population under selection. The equation can take different forms depending on the estimators used and if the selection is applied on one or both of the parents, a common one is:

$$\Delta G_t = \frac{k r_{xg} \sigma_g}{L} \qquad \text{Eq. 1-1}$$

where ΔG_t, the genetic gain, is a function of the selection intensity which is the standardized selection differential Sdivided by the additive genetic variance of the estimated breeding values σ_x or $k = S/\sigma_x$, r_{xg} the correlation between the true breeding value and the estimator used and σ_g the standard deviation of the additive variance. It shows that the genetic progress depends on the selective pressure put on the population (k), the accuracy with which the breeding value is estimated (r_{xg}) and the available variability in the population (σ_g) and L the length of a breeding cycle. The estimation of the genetic gain over several cycles implies that the additive variance stays constant. The assumption at the base of the breeder's equation makes the actual values it returns potentially imprecise and have been experimentally shown as such (Rutkoski 2019). It is however still an important tool to understand the factors influencing genetic gain and to plan the improvement of a breeding cycle.

To accelerate genetic gain means to increase ΔG_t. This can be done by increasing the selection differential that will depend on the percentage of top (or bottom) selected in the evaluated population and the distribution of the additive effects of the evaluated population. By convenience it is assumed as standard normally distributed. To reduce this percentage the number of selected units can be decreased or the total size of the population can be increased. The population is generally at the maximum size the program permit if we consider classic phenotyping but reducing the top percentage carries the risk of genetic drift and fast loss of diversity. The r_{xg} stands for the correlation between the estimates used for selection and the true breeding value. Obviously, higher correlation leads to higher selection gain. It can be expressed in variance covariance as $r_{xg} = \sigma_{xg}/\sigma_x\sigma_g$. One way to increase the correlation is to increase the accuracy of the phenotyping which depends on good field practices. More replicates will increase the accuracy as well as more sophisticated field design, notably those allowing

the use of spatial modelling of the error. More replicates should lead to higher accuracy and higher genetic gain but, as for the phenotyping of a larger population, it requires resources the program might not have. The next parameter that can be increased is the additive genetic variance σ_g to introgress new material in the breeding populations but is tricky as new material might also reduce the population mean. The last term to influence the genetic gain is the length of the cycle. Here, less is better as shorter breeding cycles increase the genetic gain.

The rate of genetic gain still has a maximum because even with the optimal scheme, genetic gain is bound to the biology of the species. A plant requires a minimum space to grow which will limit the size of the population and the explored variability. The time necessary to grow them will also limit the reduction of the cycle length.

1.2.1 Simulation as an optimization tool

Experimentally testing new breeding schemes is in many cases not feasible as running various schemes simply to see which one performs the best requires too many resources, including a lot of time and a sufficiently large population to generate generalizable results. Real data are also affected by partly unknown processes, such as environment effects limiting the understanding of the causality. Finally, phenotypic data generated from field observations cannot represent the whole range of possible environmental conditions that might be relevant for the program.

Simulation has long been a tool for plant breeding and quantitative genetics. Its complexity and power increased with the power of available computers. From initial simple deterministic models, simulations are now using in many cases a stochastic approach. In deterministic simulation, the outputs are entirely defined by the input variables and the same output will be obtained for the same input. It is based on some mathematical equation that reflects reality. Typically, breeding scheme optimization this is the breeder's equation that lies at the core of the model as in (Atlin and Econopouly 2022).

Instead of using fixed input values, a stochastic simulation introduces uncertainty by sampling input values from probability distribution function that represent stochastic processes. The user can then fix the parameters of the distributions, but different values will be sample each time that the simulation is run. In the context of breeding scheme simulation, those stochastic processes are e.g. the probability of an allele to be inherited, the probability of a plant to be sampled or the distribution of the random error during phenotyping just to name a few. By choosing the distribution function parameters correctly, simulations can deliver valuable data on the long-term performance of specific breeding schemes under specific conditions and help decision making.

In the last decade, much stochastic simulation software has been developed specifically for breeding scheme optimization (e.g. ADAM-plant (Liu et al. 2019), AlphaSimR (Gaynor, Gorjanc, and Hickey 2021), BSL (Yabe, Iwata, and Jannink 2017), HaploSim (Coster and Bastiaansen 2010), MoBPS (Pook, Schlather,

and Simianer 2020), QU-Gene (Podlich and Cooper 1998), …). They all accommodate easily diploid organisms, additive and dominance effects and in most cases epistasis, however with varying levels of simplification. For integrating genomic information, they allow simulation of SNP but, to my knowledge, no other types of markers. These softwares differentiate themselves by their simplicity of usage and resources needed, some being very demanding in terms of RAM and CPU time.

For its simplicity, adequacy, on-the-team experience, and the large community of users, AlphaSimR was chosen for the simulation experiment of this thesis.

1.2.1.1 AlphaSimR

AlphaSimR (ASR) (Gaynor, Gorjanc, and Hickey 2021) is an R package specifically developed for stochastic simulation of breeding program. It consists of a collection of R functions that allows the users to script a simulation that can represent most breeding programs. It can handle diploids but also autopolyploid species. ASR combines the Markovian Coalescent Simulator (MaCS Chen et al 2009) for backward in time simulations when it generates founder haploids to meet users' defined population characteristics and the genedrop method is used for forward in time simulation when new haplotypes are generated.

1.3 Genomics assisted plant breeding

1.3.1 Molecular marker in plant breeding

The first technology to identify polymorphism on the nuclear DNA of an organism appeared in the 1970s (Grodzicker et al. 1974). With the increase of identified polymorphisms, reduction of the cost and simplification of the technology, molecular markers have taken an increasingly central place in geneticists' and breeders' work. The first record of marker-assisted breeding was in the early 1980s with isozyme makers (ref in (Yunbi Xu and Crouch 2008) to speed up the introgression of monogenic traits in tomato. Then the era of marker assisted selection (MAS) in plant breeding started.

The classical use of MAS is based on previous knowledge of which genomic regions control the trait of interest. Either biparental or multiparent populations are developed (for linkage mapping analysis) or a collection of diverse germplasm is assembled (for association mapping studies), genotyped and phenotyped for the trait of interest. Marker(s) statistically associated with a change in phenotype can then be identified as genetically closed to a quantitative trait locus (QTL). Following the first discovery, confirmation, validation, and sometimes a step of fine mapping or gene cloning are conducted before the marker can be used. Once a marker is proven to be tightly linked to QTL, it can be used to trace an allele on the genome responsible for the trait of interest (Collard and Mackill 2008). Once associated molecular markers are available (either the markers detected in the mapping or by designing new markers in the candidate region (e.g. KASP for Kompetitive Allele Specific PCR) individuals with the

favourable allele can be selected without the need of phenotyping. MAS proved to be especially useful for qualitative traits observed late in the life cycle or those that are expensive to measure.

Molecular markers were successfully used for introgression in different crops (see review in (Bernardo 2016) and (Hasan et al. 2021)). However, despite the large amount of marker-trait association (~10,000 recorded in 2008, definitely more today), their use in breeding program seems to have remained limited (Bernardo 2008). There are several reasons for this, as the list of prerequisites for and constraints to applying MAS is long (Bernardo 2008; G.-L. Jiang 2013). Some of the important reason why MAS stayed anecdotic in plant breeding are:

1. QTL effects, especially when many minor QTLs are involved, are not consistent across the populations (genetic background effect) or environments (QTL x environment interaction)
2. Even for traits controlled by several major QTLs, a rapid pyramiding comes complicated as the probability to have favourable alleles for all QTLs targeted drops as the number of QTLs increase
3. QTL detection and association studies work well to identify major QTLs but most trait are controlled by many minor and the size of the population needed to detecte them are often prohibitive

Many economically important traits, foremost yield, are polygenic in nature and hence are not well suited for MAS limited to associated markers. Because of the rapid and cost-effective genotyping offering thousands of SNP markers, another method has been suggested to use allelic information not to identify and trace specific QTLs but to predict the expected phenotype considering genome wide makers. This method is known as genomic selection or genomic prediction (GP).

1.3.1 The concept of genomic prediction

Genomic prediction gained its fame with the article by Meuwissen et al. (2001) who. for the first time, predicted genomic estimated breeding value (GEBV) based on dense genome wide marker data. Older publications had however previously tested the concept (Bernardo 1994; Whittaker, Thompson, and Denham 2000).

The concept of GP is simple: train a statistical model with genotypic and phenotypic information from a reference population and then use the model to predict the performance of selection candidates based only on their genotypes. The rationale is that, assuming enough markers would be used to cover the genome, one or several will likely be in linkage disequilibrium with the QTLs of interest (Hayes, Visscher, and Goddard 2009). If we can estimate the effect that those markers have on the phenotype, we should be able, once those estimates are obtained, to predict phenotypes only based on genotype.

Studies on GP in the last decade have shown much potential and have been widely adopted by animal and plant breeding program alike to improve genetic gain.

1.3.2 Evaluation of prediction model

As with any other mathematical models, GP models need to be tested for the reliability of their inference and/or prediction. Predictive mathematical models are chosen based on their ability to minimize the error of prediction (Schrauf, de los Campos, and Munilla 2021). In general, a metric such as the means square error is used to select the best model. Any bias or scaling introduced by the model will increase the error. However, breeding bases its choice on ranking which is not affected by either of them. For this reason, the field of GP almost exclusively evaluates the models on their predictive accuracy (sometimes also predictive ability) which is the Pearson correlation between (GEBV) and observations (P) and which is less restrictive than a metric including bias and scaling. The correlation is used as such or is divided by the square root of the heritability as $\frac{\mathrm{cor}(GEBV,P)}{\sqrt{H^2}} = \mathrm{cor}(GEBV, TBV)$. The rationale behind it is well explained by Lorenz et al. (2011) but briefly the goal is to measure the model on its ability to predict the true breeding value (TBV) behind the observed P.

GP models or calibrations (combination of model and input data) are commonly tested by cross validation (CV). Considering one has a dataset including individuals with all the phenotypic data and all the genotypic data. This set will be partitioned into the training set (TS) and the validation set (VS). The effects of the model parameters; i.e., the marker effects, are estimated using the complete information of the TS. Then, those estimated effects are applied on the TS parameter values, the genotypes, to obtain a prediction. The correlation between those predictions and the observed "real" values are computed.

Two primary methods exist to partition the data in TS and VS. The first one is the k-fold cross validation. It starts by equally and randomly distributing the individuals in *k* folds (e.g. *k=5*). Then *k-1* folds are used as TS and the last fold as VS. This is repeated *k*-times so each individual is predicted once. The random sampling in *k* fold can be repeated as many times as necessary. The second approach is much simpler: x% of the dataset is used as TS and the remaining 100-x% are used as VS. Again, this is repeated as many times as necessary. Under the fully random approach it is unsure whether each individual is at least once in the VS. The two methods however do not show great difference in the final model selection (Cao T-V, personal communication).

When phenotypic data comes from a single environment the partitioning is straight forward and as described before. When the experimental design integrates more components such as different environments of phenotyping or in the case of hybrid breeding different tester used in which environment (Basnet et al. 2019) more complex CV schemes need to be developed to test the prediction scenarios we are interested in. The common nomenclature CV1 and CV2 was introduced by

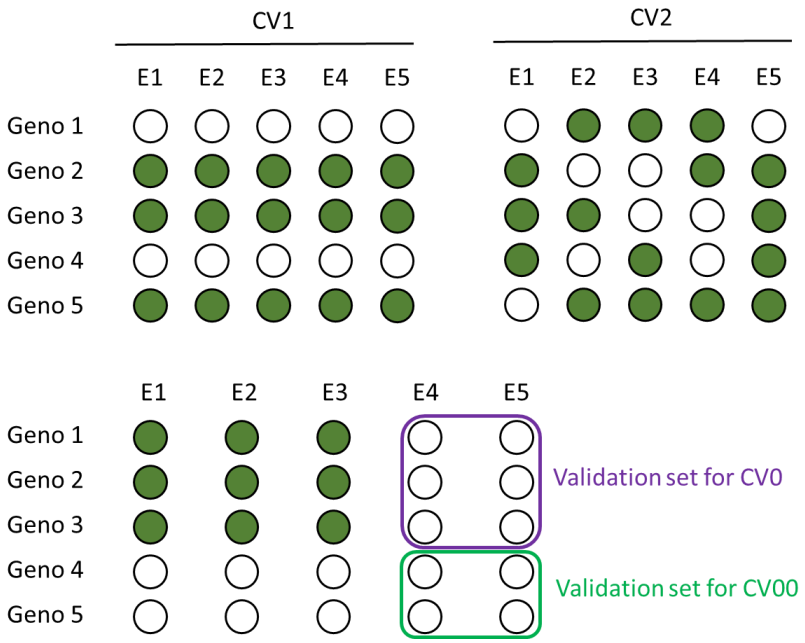


Figure 1-3: Schematic representation of four CV approaches CV1, CV2, CV0 and CV00 as presented in Burgueño (2012) and Jarquín (2017). Each line represents a genotype within the experimental design and each column within their respective CV approach an environment included in the design. A solid green circle means that the genotype x environment combination has a phenotype while, a white circle means that no phenotype is available and that the value will be predicted.

Burgueño (2012) and was later extended to CV0 and CV00 (Jarquín et al. 2017a) (Figure 1-3). CV1 represents the case where new individuals with only the genotype are predicted for already known environments. CV2 represents a case where some individuals are phenotyped only in a subset of the environments composing the experimental design and predicted in the remaining one. The proportion of genotypes that are phenotyped in several environments, sometimes called the overlap, can be played with, and can have an effect of the accuracy depending on the models used (Jarquín et al. 2020). CV0 represents a case where some individuals were phenotyped in several environment and their GEBV are predicted in an unknown environment. Finally, CV00 is the case where we predict new individuals in new environments based on calibration made on other individuals in other environments. Apart from CV, genomic prediction models can be evaluated on external population (external validation). This can be used to test a model explicitly calibrated to predict a defined population (validation population or breeding population). For example, if a model is calibrated with parental lines to predict the offspring or when a population at generation *t* is used in the calibration to predict the population at a later generation *t+n*.

1.3.3 Factors influencing prediction

The prediction accuracy of the GP models is influenced by different parameters. The predicted trait, by its number of QTLs involved in its expression, influences the potential accuracy (Daetwyler et al. 2010). The level of dominance and epistasis is also relevant for the accuracy of the precision as most of the models aim to capture only the additive component (Meuwissen, Hayes, and Goddard 2001; VanRaden 2008; de los Campos et al. 2013). Another parameter linked with the characteristic of the trait and strongly influencing accuracy is the heritability. It quantifies the amount of genetic variability

within the population and if heritability is low so is genetic variability. GP is expected to perform poorly (like any selection) when the genetic variability for the trait under selection is exhausted.

Heritability is also a metric for estimating the quality of the phenotyping (repeatability). Following the adage "*garbage in, garbage out*", the quality of the phenotyping has also an effect of the accuracy. The presence of outliers as well as the suitability of field experiment design have been explored and proven strongly influential (Ould Estaghvirou, Ogutu, and Piepho 2014; Bernal-Vasquez et al. 2014).

The technology used to molecularly characterize the population seems to have little impact on the precision (Elbasyoni et al. 2018) and, considering a widespread, validated technology is chosen (e.g. Genotyping-by-sequencing (Elshire et al. 2011), SNP-array), the price is probably the main argument for the choice here. One central parameter regarding the potential influence of the genotypic data in the prediction is the marker density. If the genotyping is sufficiently dense, the markers should be able to capture a large part of the genetic variance (Yang et al. 2010). Sufficient density will however be defined by the extent of the LD in the working population but also the distribution of the markers across the genome (Lorenz et al. 2011). While a minimum number of markers is required, gains in accuracy tend to stagnate above a certain threshold at which point denser genotyping will be unnecessary. The different marker treatments have also been tested for their impact on accuracy. The levels of missing data, the minimum minor allele frequency, the maximum LD tolerated between to marker or the imputation method (Grenier et al. 2015).

An essential factor to ensure accuracy of the GP model is an appropriate calibration set. It was simply put by Edwards et al. (2019) "*small numbers of close relatives and very large numbers of distant relatives are expected to enable predictions with higher accuracy*". In practice this means that adding a relatively small number of close relatives in the calibration set can improve the accuracy more than strongly increasing the size of the calibration set with distant genotypes (Edwards et al. 2019). GP works best when calibration and prediction sets are within a biparental family (Hickey et al. 2014). It still works with more distant individuals, but calibration set and prediction set must be sufficiently related to ensure useful levels of accuracy and a denser genotyping will be necessary (Lorenz and Smith 2015; Albrecht et al. 2011; Edwards et al. 2019). Finally, the calibration set also needs to be informative and capture as much as possible the genetic variability of the population on which GP is applied (Guo et al. 2014). The presence of structure within the population is also important for the accuracy of the prediction (Pszczola et al. 2012; Guo et al. 2014).

Finally, since the birth of GP, the type of mathematical model/statistical approach used to capture the markers effect have been heavily discussed (Onogi et al. 2015a). Some should be adapted to strongly polygenic traits while others should deal better with trait controlled by few major QTLs and many minor ones. It turns out that there is so far no single right answer and that the best performing model will depend on the trait predicted, the population, the signal to noise ratio and many other

uncontrolled parameters. The best approach still is to try several options but keep in mind that GBLUP is rarely a bad idea (personal observation). The literature abounds with reviews on the multiple model existing. Here are a few I found helpful: (de los Campos et al. 2013; Gianola 2013; Cuevas et al. 2018).

1.3.4 Prediction methods

The basic GP model can be formulated as follow:

$$Y = Z\beta + \varepsilon \quad \text{Eq. 1-2}$$

where Y is a vector of n phenotypes, Z a matrix of marker genotypes of n rows and p columns, β is the vector of p marker effects and ε the residuals of the model. This could be solved with an OLS (ordinary least square) method if the model did not suffer from the *large-p small-n* problem. This means that there are many more parameters p to predict, i.e. markers effects, than there are n data points. Indeed, the markers can easily be counted in tenths of thousands while the phenotypes are normally much lower. As clearly explained in Gianola (2013), the maximum-likelihood estimator of our model is $\hat{\beta} = (X'X)^{-1}X'Y$. However, $(X'X)$ is singular as soon as p>n and there is an infinity of solutions for our estimator. The dimensionality problem is normally combined with a strong expected multicollinearity within the markers used, as several adjacent markers can be in high LD hence highly correlated. To address that kind of ill-posed model, several statistical approaches are available.

1.3.4.1 Penalized methods

The first type of approach is based on variable selection, shrinkage of estimates or a combination of both and includes the following GP method.

The most common penalized method is the random regression (RR) BLUP, sometimes called SNP-BLUP, and is based on ridge regression (Whittaker, Thompson, and Denham 2000). Under this method the estimator becomes $\hat{\beta} = \left(Z'Z + \lambda I_p\right)^{-1} Z'y$, where the penalization λ is $\lambda = \frac{\sigma_e^2}{\sigma_{a0}^2}$ with σ_e^2 being the residual error variance and σ_{a0}^2 the marker variance and I_p is an identity matrix of order p.

LASSO for Least Absolute Shrinkage and Selection Operation (Tibshirani 1996) is another common penalized method. LASSO do not have a simple solution like RR-BLUP (or OLS) but instead it relies on iterative algorithm (Friedman, Hastie, and Tibshirani 2010). Compared to RR-BLUP, LASSO has the advantage of shrinking the estimates as well as performing variable selection by putting the effects of some $\hat{\beta}$ at zero. However, one limitation is that it will only select n markers within the p available.

Finally, the elastic net combines the penalization used in ridge-regression and in LASSO. It seems to perform better than LASSO at selecting variable for p>>n problem but still performs shrinkage as a ridge regression (Zou and Hastie 2005). Similarly to the LASSO, it also relies on an iterative algorithm to find the solution.

1.3.4.2 Bayesian methods

As previously mentioned, in case of p>>n the $Y = Z\beta + \varepsilon$ model cannot be solved by methods relying on maximum likelihood however we can still use methods relying on Bayesian inference. They belong to the first method used for GP and were already tested by Meuwissen et al. (2001).

The Bayesian approach estimates values for $\hat{\beta}$ as follows. First, let us give the standard Bayesian linear model used in GP:

$$p(\mu, \beta, \sigma^2|y, \omega) \propto p(y|\mu, \beta, \sigma^2)p(\mu, \beta, \sigma^2|\omega) \quad \text{Eq. 1-3}$$

$p(\mu, \beta, \sigma^2|y, \omega)$ is the posterior distribution, i.e. the solution, for μ the population mean, β the vector of marker effects and σ^2 the residual variance conditioned on y the vector of phenotypic data, and ω some hyperparameter(s) of the selected prior. The problem is solved to be proportional ($\propto$) to the product of the likelihood function $p(y|\mu, \beta, \sigma^2)$, μ, β, σ^2 conditioned on the data y and the prior distribution $p(\mu, \beta, \sigma^2|\omega)$ the same μ, β, σ^2 conditioned on the hyperparameter(s). By expending $p(\mu, \beta, \sigma^2|\omega) \propto \prod_{i=1}^{p} p(\beta_i|\omega)p(\sigma^2)$ one can see that the β_i marker effect is assigned some informative prior $p(\beta_i|\omega)$ conditionned on the hyperparameter, and the residual variance is assigned another prior (by convention a scale-inverse-chi-square distribution (de los Campos et al. 2013)). Considering the appropriate prior are used, the Bayesian method will spontaneously introduce regularization (Gianola 2013).

The choice of the prior $p(\beta_i|\omega)$ is defined by the *a priori* expected distribution for the marker effects to be sampled from, hence what kind of shrinkage is performed on the marker effects or if a combination of variable selection and shrinkage is performed. Various priors have been tested and composed the famous Bayesian alphabet: BayesA and BayesB (Meuwissen, Hayes, and Goddard 2001), BayesC (Habier et al. 2011), BayesCπ (Habier et al. 2011), Bayesian Lasso (Park and Casella 2008), … the list being not exhaustive and still growing. A review on the cited method as well as on some additional letters of the alphabet has been done by Gianola (2013).

1.3.4.3 Kernel method

A third common approach to solve the p>>n problem is based on the so-called “kernel trick”. In the context of quantitative genetics, the kernels are functions that allow to compute the distance between genotypes based on the genetic information. Instead of solving equation 1-2, the kernel method aims to solve the following equation:

$$Y = Wu + \varepsilon \quad \text{Eq. 1-4}$$

Y being again the vector of phenotypes, W being the design matrix for the genotypes and u the genotype effect with distribution $u \sim N(0, G\sigma_u^2)$. The marker information is used by the kernel to compute the variance-covariance matrix G.

The first one to use this approach was Bernardo (1994) with RFLP markers on maize. Later VanRaden (2008) introduced the linear kernel method GBLUP. He suggested several methods to compute the realized relationship matrix G the most used being:

$$G = \frac{(M - 2(p_j - 0.5))(M - 2(p_j - 0.5))'}{2\sum p_j(1 - p_j)} \quad \text{Eq. 1-5}$$

with M as the genotypic matrix with i individuals and j markers with element $m_{ij} \in \{0,1,2\}$ and p_j being the allele frequency of the *j*th marker for the complementary allele.

Another common kernel method is known as RKHS (Gianola and van Kaam 2008) and uses the following Gaussian kernel:

$$G = K(x_i, x_j) = exp\left[-\frac{\|x_i - x_j\|^2}{h}\right] \quad \text{Eq. 1-6}$$

$\|\cdot\|$ refers to the norm in the Euclidian space for the genotypes x_i and x_j and h is a bandwidth parameter. Several methods to compute the bandwidth are described in Cuevas et al. (2016). The advantage of the Gaussian kernel is that it can model every epistatic interaction of any order (Jiang and Reif 2015). The kernel method has also been used to integrate dominance effects in prediction model (Su et al. 2012; Vitezica, Varona, and Legarra 2013; Morais Júnior et al. 2017).

The kernel method shows its greatest potential in modelling multiple environments design. Once the matrix G is obtained it can be combined with other matrices to model genotype by environment (GxE) interactions. It was first used for GP by Burgueño (2012). The design matrix W (Eq. 1-4) would identify observation Y from multiple environments. An environment variance-covariance matrix Σ would be estimated and then the complete variance-covariance $M = \Sigma \otimes G$, $\otimes$ denoting the Kronecker product and u, would be distributed as $u \sim N(0, M)$. Here the covariance between environment would only be estimated on the data. Another approach was also suggested. Rather than using Eq. 1-4, the model was extended to contain multiple terms: a common genetic effect and an environment specific effect:

$$Y = W_g u_g + W_e u_e + \varepsilon \quad \text{Eq. 1-7}$$

Here W_g is a design matrix for the genotypes, u_g the vector of genotypic effect with a distribution $u_g \sim N\left(0, G\sigma_g^2\right)$, W_e the design matrix for the environment specific genotype and u_e the environment specific genotype effects with distribution $u_e \sim N\left(0, \begin{bmatrix} \sigma_{e1}^2 G & 0 & 0 \\ 0 & \sigma_{e2}^2 G & 0 \\ 0 & 0 & \sigma_{e3}^2 G \end{bmatrix}\right)$, as an example with three environment.

Those mixed models can all be solved either in a frequentist or Bayesian approach.

1.3.4.4 Machine learning

Some deep learning approaches have also been applied to predict phenotypes on genetic information (Montesinos-López et al. 2021). They are a research field in statistics by themselves and were not addressed during my thesis. Some reviews of different methods on their comparative performances with classical statistical methods has been made. The general conclusion, so far, is that despite their great potential the data set used in plant breeding are still too small and sometimes not of adequate quality to harness the full potential of those tools (Montesinos-López et al. 2021; Azodi et al. 2019).

1.3.5 Program improvement through GP

Depending on how it is used, GP can address all parameters of the breeder's equation (Hickey et al. 2017). In plant breeding the selection intensity (k) is often limited by the size of the population that the budget allows to phenotype. Considering genotyping is cheaper than phenotyping, this size limit can be move upward and a greater population screened will allow a higher selection intensity.

Genomic prediction can also be used for a more accurate evaluation of the tested material. The use of a genomic relationship matrix (GRM) can be used instead of the historically used relationship matrix based on the pedigree. This will help the adjustment of the model and return more accurate predictions of the true breeding values than the phenotype alone or the phenotype and a relationship matrix based on pedigree would have done.

The use of GP can accelerate and improve the introduction of new material in the breeding program helping maintain good additive variance in the breeding population (Gorjanc et al. 2016; Allier et al. 2020).

Finally, the breeding cycle length (L) can also be greatly shortened with the help of the GP. It allows to bypass the lengthy phenotypic evaluation step and directly select and cross superior material to start a new generation. It also permits the prediction of the value of the most recent material from field data and genotypes collected on previous generations (forward prediction). This use is especially visible in population improvement through recurrent selection where success is linked to the amount of recombination per time.

1.4 Rice

1.4.1 Origin and taxonomy

Rice (*Oryza sativa* L.) is a cereal crop belonging to the Poaceae family. The genus Oryza appeared around 15 million years ago (Stein et al., 2018). It contains 25 wild species and two cultivated ones: *Oryza glaberrima* and *O. sativa*. The *O. sativa* specie also known as Asian rice has been classified in five varietal group (Glaszmann 2008). This classification has since been extended and completed (Wang et al. 2018) confirming the grouping while expending it (and calling the groups "subpopulation"). The main subpopulations in terms of economic importance are the *indica*, *tropical japonica, subtropical* and *temperate japonica*. *Indica* is the subpopulation that

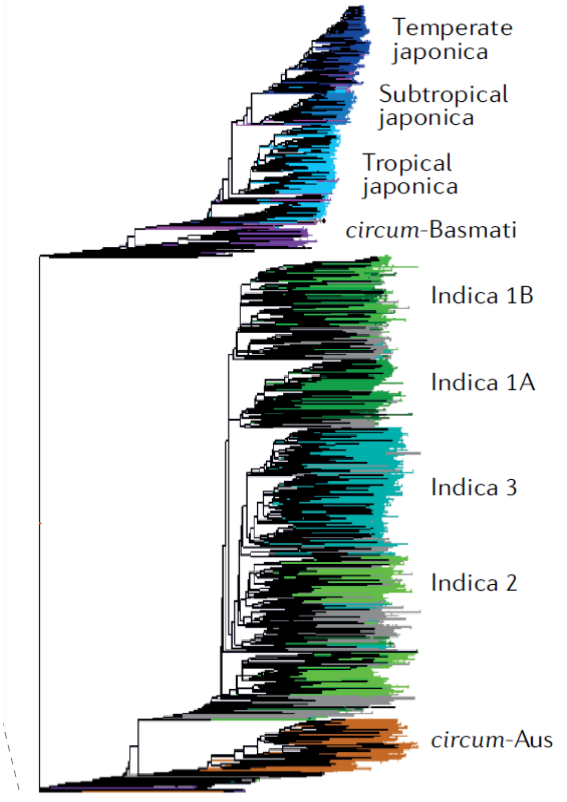


Figure 1-4: Subpopulations of *O. sativa* adapted from (Wing, Purugganan, and Zhang 2018)

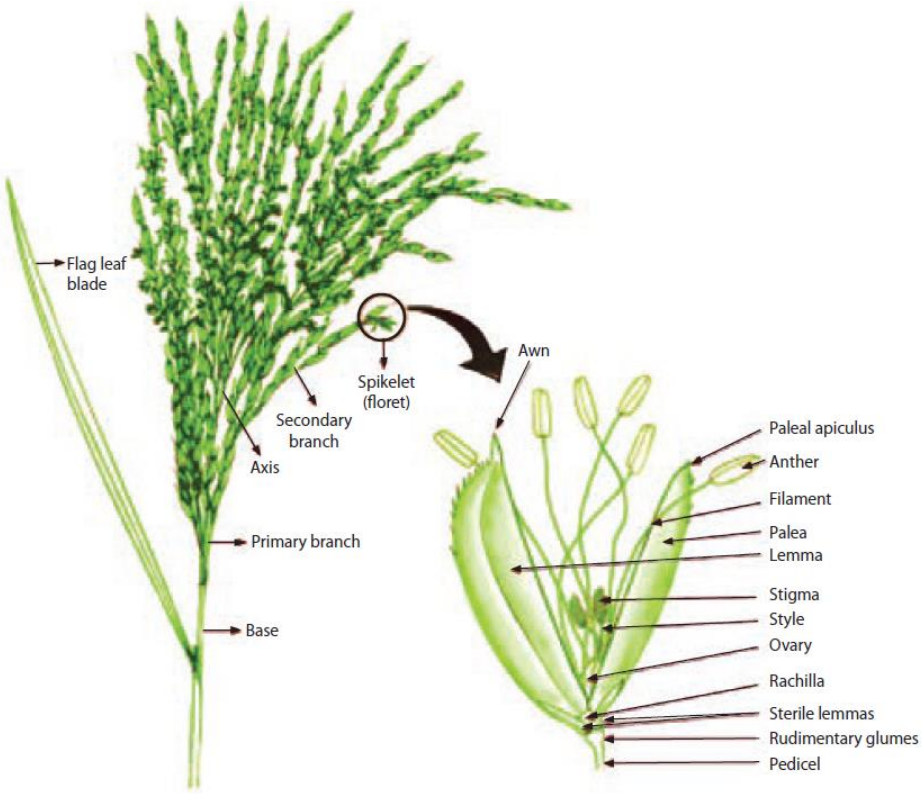


Figure 1-5: Morphology of the rice panicle and the spickelet (GRiSP 2013)

shows the most diversity which is related to its wider range of cultivation. It is grown essentially in optimal environmental condition, both in terms of water and sun. The different *japonica* subpopulations are closely related to each other and less diverse, restricted to specific cultivation areas. The *tropical japonicas* subpopulation s is more adapted to grow in area constrained by abiotic stress (water deficit, low soil fertility) while the *temperate japonica* one is more often encountered in subtropical and temperate zones under irrigated cropping systems. The *circum*-Basmati subpopulation is closely related to the *Japonica* one. It is essentially found in the Indian subcontinent and known particularly for their aromatic flavoured grain. The *circum*-Aus subpopulaitons are close relatives to *Indica*-type rice from the Ganges delta region. They are very diverse and are a known source of tolerance genes for various abiotic stresses (Casartelli et al. 2018).

1.4.2 Rice morphology

Rice is an annual herbaceous plant with plant height of <1 m to up to 5 m for some floating rices. Its life cycle can be summarized in three phases. The first phase between germination and tillering called the vegetative phase lasts between 50 and 100 days after sowing and is about half of the life cycle. The reproductive phase follows and starts with panicle initiation up to the end of flowering this will last 25 to 30 additional days. The panicle carries up to 400 spikelets each of them being only one floret giving one seed if fecundated (Figure 1-5). Rice is a mostly autogamous plant (0-6.8 % of allogamy) (Sahadevan and Namboodiri 1963) with selfing insured by cleistogamy. The last phase is ripening and

runs from fertilization to maturity of the grain. It lasts about 30 days. The length of the phase depends on the cultivar as it is a selection target. It depends also on the accumulation of degree days, a growth season being longer under temperate climate than under a tropical one. If rice is mainly cultivated as an annual crop, it can behave as a perennial as long as the meristem is preserved. This is called ratooning (GRiSP 2013).

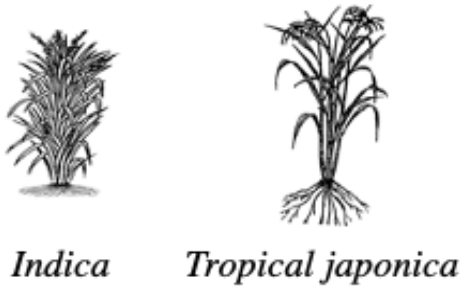


Figure 1-6: Representation of the morphological differences between improved *indica* and *tropical japonica* rice

The main parts of the rice grain are the endosperm, surrounded by the bran and the germ and encapsulated in the husk (or hull) composed of the palea and lemma (Figure 1-7). Rice is harvested with the husk as so-called "paddy-rice". A grain of paddy rice weighs between 10 and 45 milligrams from which about 20% is the inedible husk. The husk is removed post-harvest in fix mills to obtain the brown rice. It is usually further processed before consumption, going through a step of polishing removing the bran and the germ to get the ubiquitous white polished rice (GRiSP 2013). Under this form, the rice kernel resumes in 69% starchy endosperm which is the source of carbohydrates (90% of the milled grain), some micronutrient, a few proteins (6-7% of milled grain) and a very reduced fraction of mineral and vitamins, and traces of antioxidants (Luh 1991).

Although I described the main morphological characteristics of rice, large variability exists in the species, mainly reflecting the subpopulation division previously described. To only illustrate these differences in the two main subspecies (*indica* and *tropical japonica*), the Figure 1-6 highlight strong difference in tillering ability, plant height root depth and density. These characteristics will later be commented in terms of adaptability of these two groups to the ecosystems where rice is grown.

1.4.3 Rice genetics

Rice is a diploid species with 12 chromosomes. With a genome size of ~ 390 Mb (Zhou et al., 2020) it is the smallest of all domesticated cereals. Its small genome, a great genomic synteny to other cereals as well as its cultural importance for the world has made it the model cereal. Rice was the first crop genome to be sequenced (International Rice Genome Sequencing Project and Sasaki 2005). Since then, a great part of the rice genetic resource has been sequenced and made available to all the scientific community (The 3,000 rice genomes project, 2014) and 12 new references genomes were built and made available to account for the large

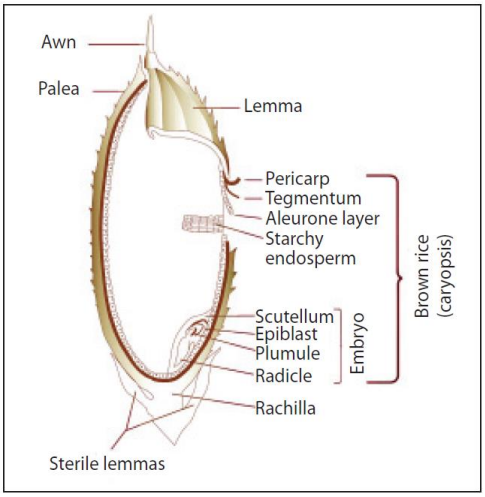


Figure 1-7: Morphology of the rice grain (GRiSP 2013)

diversity existing in the species (Zhou et al., 2020). As of today, about 56,000 loci were annotated and reported in database (http://rice.uga.edu, 2022.). Consequently, rice has benefited from a large community of scientists at the origin of great advances in the domain of genetic characterisation and exploitation.

1.4.4 Rice as a major crop

1.4.4.1 Culture

Oryza sativa is a very diverse species. It is the most important crop in terms of cultivated area with production zones spread from 53° North in the Amur River valley to 40° South in central Argentina. This range of latitude represents climate as diverse as wet-tropical and continental. Often rice cultivation is practiced over more than one season, up to three per year, under tropical climate. It is also cultivated under a variety of cropping systems which are often classified by the type of irrigation used (Figure 1-8).

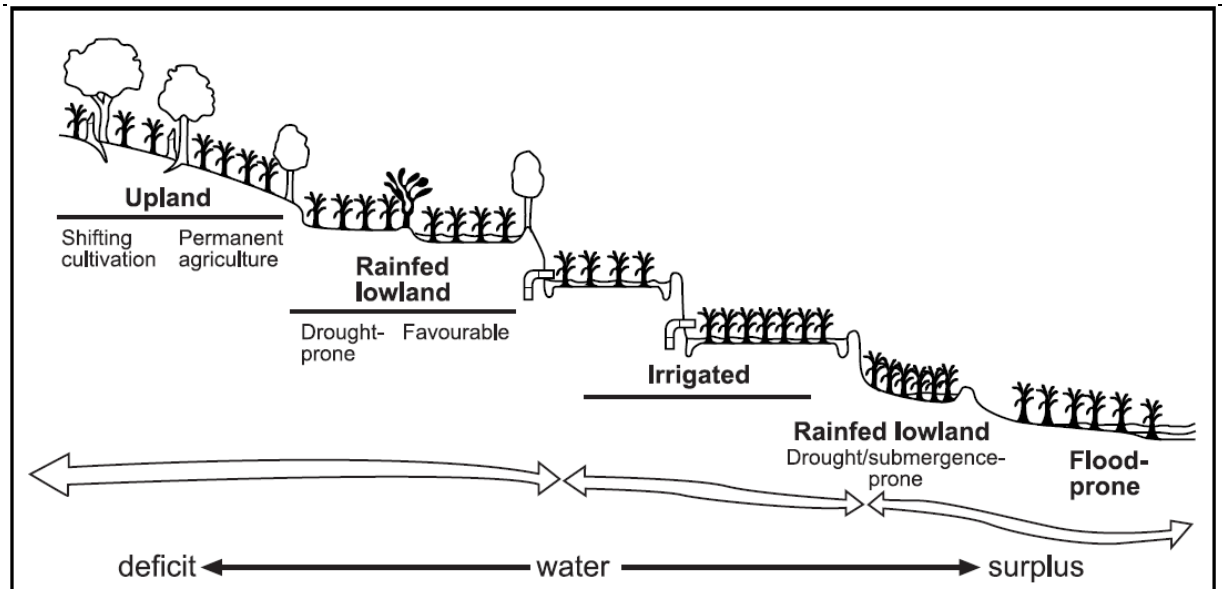
Most of the rice cultivated area is irrigated. As long as there is a source of water, this is a very reliable cropping system with stable yield. The second most common cropping system is rainfed lowland followed by rainfed upland. While rainfed lowland is expected to be submerged for a certain time when precipitation arrives, owing to bunded field preparation, upland rice is grown without submersion generally on levelled or sloping unbended fields. These cropping systems are more prone to water-stress as they rely on sufficient precipitation at the right time. As they rely on seasonal rainfalls, only one harvest per year is possible. The last cropping system is flood-prone or floating rice. This system is encountered in low-lying coastal area such as river deltas where the plants are expected to be submerge by at least 100 cm of water for more than 10 consecutive days. In lowland, irrigated and flood-prone condition rice can be either direct seeded on dry or wet soil or first grown in nurseries to by later transplanted in puddled field. Under upland conditions it is direct seeded or seeded in tilled land.

1.4.4.2 Economic importance

Rice is the 3rd most important crop in term of cultivated area with a total of 164 million ha in 2020 behind maize and wheat with 202 and 219 million ha, respectively (FAOSTAT, 2022). The crop is grown by more than 144 million farmers in more than 100 countries (CGIAR, 2013). In terms of consumption, it was the most important food crop in 2019 with 80.54kg/capita/year across the world against 64.94

Table 1-2: Rice production and yield by irrigation system *(GRiSP 2013)*

Type	Surface [million ha]	Average yield [t/ha]	Total production [%]
Upland	15	1	4
Rainfed lowland	52	2.3	19
Irrigated	93	3-9	75
Flood-prone	11	1.5	2



Upland	Rainfed lowland	Irrigated	Flood-prone
Level to steeply sloping fields; rarely flooded, aerobic soil; rice direct seeded on plowed dry soil or dibbled in wet, non-puddled soil	Level to slightly sloping, bunded fields; non-continuous flooding of variable depth and duration; submergence not exceeding 50 cm for more than 10 consecutive days; rice transplanted in puddle soil or direct seeded on puddle or plowed dry soil; alternating aerobic to anaerobic soil of variable frequency and duration	Leveled, bunded fields with water control; rice transplanted or direct seeded in puddle soil; shallow flooded with anaerobic soil during crop growth	Level to slightly sloping or depressed fields; more than 10 consecutive days of medium to very deep flooding (50 to more than 300 cm) during crop growth; rice transplanted in puddle soil or direct seeded on plowed dry soil; aerobic or anaerobic soil; soil salinity or toxicity in tidal areas

Figure 1-8: Rice irrigation system from (Halwart, Gupta, and WorldFish Center 2004)

for wheat (FAOSTAT, 2022) however with great disparities. In four countries (Bangladesh, Cambodia, Laos, and Vietnam) it is more than 200kg of rice that are consumed per capita per year. It covers 19% of the daily calories requirement and is considered a staple crop for billions of people especially in developing countries (OECD and FAO 2021). For this reason, it plays a central role in food security. As staple food, rice can also be effectively used to bring essential mineral micronutrients to people whose dietary diversity is low. Biofortification (the targeted increase in micronutrients within the consumed part of a crop) of rice has for this reason being actively researched (Bouis and Salzman 2017) in the last two decades to notably fight zinc deficiency (Kiran et al. 2022).

1.5 Optimization of a breeding scheme: the case of the CIAT-Cirad upland rice breeding program

1.5.1 Historic of the program

The involvement of Cirad in an international partnership for rice improvement in Latin America started in the 80ies, through a joint project with Embrapa Rice and Beans in Goiana (Brazil) in 1981 then with CIAT in 1992 (Châtel and Guimarães 1997). While the first phase aimed at enriching the Brazilian breeding program with genetic resources developed in Africa (through the IRAT, the former Cirad), the

second phase to foster and strengthen the breeding activities for the Latin America's upland ecosystems (Châtel et al. 1995). Since 1996, the CIAT-Cirad rice breeding program at CIAT has focused on broadening the genetic base of rice breeding populations. The strategy was to generate segregating material based on the development and improvement broad genetic base synthetic population using recurrent selection.

The first breeding population developed for upland conditions was synthesized with 26 elite donors of the *tropical japonica* group, specifically adapted to the tropical upland from Asia, Africa, Latin America and the Caribbean (LAC). Their recombination was facilitated with the use of an *indica* rice mutant bearing the nuclear male sterility gene (*ms*) (Singh and Ikehashi 1981). The first rice synthetic upland rice population was then generated with IR36-ms (Taillebois and Guimarães 1989). Since then, the population was improved through recurrent selection, enriched with additional elite breeding lines and cultivars and today various populations of upland rice are available and exploited in the program (Martinez et al. 2014).

1.5.2 Rice in LAC

Rice arrived in the New World shortly after the arrival of the first Europeans. It was only in the 20th century that rice consumption per capita gained in importance in Latin America and the Caribbean , increasing from 10 to 30kg (Calvert et al. 2006). The trend persisted in countries like Panama, Guyana and Surinam where annual consumption in 2019 was above 100kg per capita per year (FAOSTAT). While most countries are consuming their internally produced rice, imports represent an important fraction of the rice consumed in many countries, raising the issue of self-sufficiency in countries where rice accounts for part of people's diets.

In the LAC region, rice production is defined according to the availability of irrigation systems and environmental conditions. While some countries only have a very reduced portion of the rice cultivated area under irrigation (8% in Bolivia), others are almost exclusively growing rice under irrigated systems (Argentina, Chile, Paraguay and Uruguay) (Rice Atlas in (Andrade et al. 2021)). Of the rice growing areas in LAC, rainfed systems represent less than half the total area, with 1.1 million ha (Rice Atlas for the 16 countries where data was available, CIAT access). As seen in Figure 1-8, the rainfed systems that only depend on the rainfall pattern can be separated in two ecosystems; the lowland where rice has enough water during its entire cycle, and the upland where hydric deficits are often present during the crop cycle. The uplands are often associated with low soil fertility or challenging soil properties such as aluminium toxicity and acidity.

The savannah ecosystems are those on which the CIAT-Cirad breeding program has focused since its inception. This ecosystem is characterized by high aluminium toxicity (> 70% saturation), soil acidity (pH~4) and frequent drought spells during the crop cycles. The uplands and particularly the savannah

of Colombia or cerrados in Brazil, are ecosystems that offer great potential to grow rice under rainfed conditions and in crop rotation systems with other species such as soybeans or forage grass. Such areas are often said to have potential for providing rice to the rest of the world (A. Castro, personal communication).

1.5.3 The breeding objectives

The main objective of the CIAT-Cirad rice breeding program is to develop germplasm highly adapted to the upland ecosystems. *Tropical japonica* rice, which is the best subspecies for this ecosystem due to its tolerance to drought stress, was the germplasm of choice for developing such cultivars.

The specific objectives of the program are to develop highly productive rice adapted to rainfed conditions, with earliness to escape potential drought and enable rotation, resistance to the main biotic stress prevailing in these ecosystems, notably blast fungus (*Magnaporthe oryzae*), and with nutrient dense grain, in particular zinc, to improved nutrition for the millions suffering from malnutrition.

1.5.4 Developing a RS population using ms-gene

For rice population, two types of populations can be developed using the ms-gene: monocytoplasmic population and polycytoplasmic population. The development of the two types differs at the "sterilization" step, the introduction of the ms-gene in the genetical background of interest (Châtel and Guimarães 1997).

For a monocytoplasmic population, elites are crossed with an ms-carrying genotype. To ensure that the offspring carry the ms-allele, the sterilizing genotype must be homozygous and thus can only be the mother in the cross. The seeds will all be produced by the sterilizing genotype that will be the only source of cytoplasm, hence monocytoplasmic. In this case, 50% of the genetic variability in the population is from the ms-source.

Polycytoplasmic populations are made to reduce the genetic contribution of the ms-source as well as to diversify the cytoplasm origin. The offspring of the first ms-source x Elite are backcrossed with the Elite parent. With one backcross the genetic contribution of ms-source is reduced to 25% of the variability, ensuring that the cytoplasm of the parent can be mixed.

After the first cross (monocytoplasmic) or the backcross (polycytoplasmic), the plants are selfed and the next generation are recombined. After recombination, only sterile plants are harvested and the seeds, that are either of genotypes [Ms:ms] or [ms:ms], are the starting population of the recurrent selection.

Some other, more sophisticated methods have been applied to synthetize populations. For example, to synthetize the CNA-IRAT 5 population, first IR36 [ms:ms] was crossed to Palawan and the F1 selfed

to an F_2. The F_2 was then used as an ms-donor and crossed to 26 varieties (Taillebois and Guimarães 1989).

1.5.5 The breeding scheme

The program uses a breeding scheme composed of two parts: a population improvement based on RS and a product development part using the diversity from the population and selecting progenies by pedigree breeding (Figure 1-9).

As seen in chapter 1.2.1, RS is the cyclic (i) evaluation of progeny families from a breeding population, (ii) selection of its best performing families and (iii) recombination of the selected entries to generate an improved version of the population. The RS relies on large numbers of crosses among a large number of parental lines and can thus be labour intensive for species requiring manual castration like rice. To facilitate outcrossing in the field, populations segregating for the IR36 male sterility gene can be used. The individuals homozygous for the recessive allele [ms:ms] produce sterile pollen and any seeds carried by those plants are the results of an outcross. The remaining genotypes [Ms:-] will produce fertile pollen and carry seeds that are the results of a selfing. By selecting for or against male sterility, outcross or selfed progeny can be targeted.

In the first season, a population of about 3000 $S_{0:1}$ individuals, with 25% sterile [ms:ms] plants, are sown together. Under those conditions, the fertile genotypes play the role of males and the sterile genotypes of females which will be crossed with neighbouring males. Those sterile plants will carry S_0 seeds and will be harvested at the end of the season. The *S* nomenclature describes the generation of selfing of the observed individual/lines, e.g. a S_0 plant represents an individual plant that went through 0 generations of selfing (i.e., a plant coming from a cross), while $S_{0:1}$ represents a family derived from the original cross (S_0) but after one generation of selfing and bulking.

During the second season, ~3000 S_0 plants are grown again in a single plot. At the end of the crop cycle, a manageable number of fertile [Ms:ms] plants (~200) carrying $S_{0:1}$ seeds are harvested. Those seeds are the result of selfing. For each of these families, some $S_{0:1}$ seeds are stored for later crossing (if selected upon progeny testing) while the rest is used for generation advance and progeny testing. Those steps occur preferably under a controlled environment to maximize success. For progeny testing, fertile $S_{0:1}$ plants are harvested in bulk to get $S_{0:2}$ seeds. Then, the same procedure is used to get $S_{0:3}$ seeds from $S_{0:2}$ plants. The evaluation of the S_0 is done on either their $S_{0:2}$ or on their $S_{0:2}$ and $S_{0:3}$ derived progenies following the same experimental design in two different environments. The phenotypes measured at the $S_{0:2}$ (or $S_{0:2}$ and $S_{0:3}$) are thus the progeny mean representative of the individual S_0 plants extracted from the population. The best families are selected and their $S_{0:1}$ seeds that were set aside in storage are used to recombine and synthesize the improved population. The whole process of

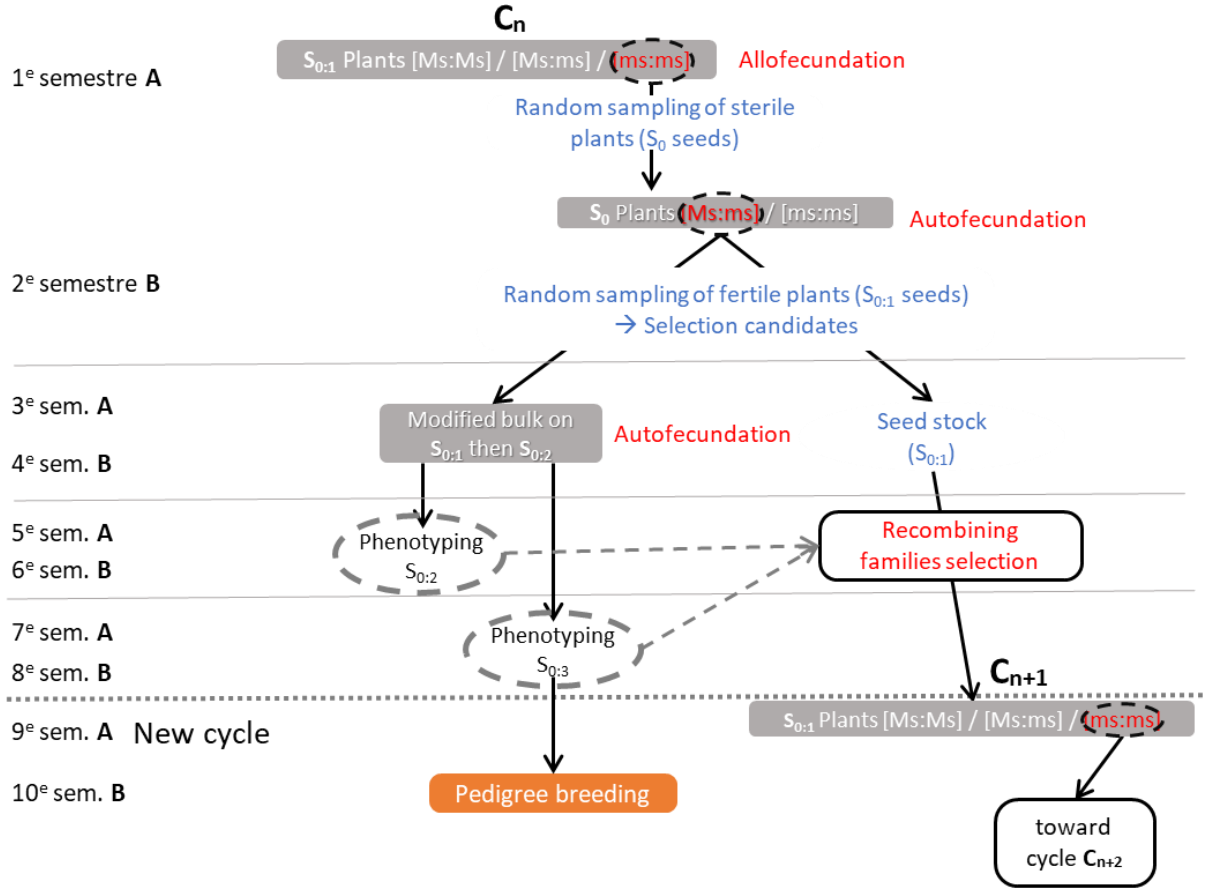


Figure 1-9: Schematic description of the CIAT-Cirad upland rice breeding scheme

generation advance and phenotyping at two generation in two environments takes three years, which brings the length of one breeding cycle to four years.

The best candidates selected for the population improvement scheme are also the base material for the pedigree breeding. Further genetic fixation is required to develop an inbred cultivar. Selfing and bulk harvest are performed for two or more generations to fix the genotypes. Then a few cycles of pedigree breeding are conducted to select the best individuals within the best families.

After fixed elite lines are developed, a seed multiplication step is applied to produce the seed quantity necessary for the several rounds of yield testing at the end of which candidate varieties are selected.

The generation advancement and phenotyping work is time consuming and slows down the population improvement part of the program. It takes up to two to three grow-out seasons to select the families for recombination while the recombination is performed with the seed collected on fertile S_0 plants. To increase the number of recombinations per unit time, efforts have been made to replace the progeny testing with genomic prediction on S_0 genotypes and decouple the recombination from the phenotyping.

1.6 Objectives: improvement of the breeding scheme

We have seen that the current breeding scheme has some potential for improvement. The RS part of the scheme lasts four years (including the establishment of the population with S_0 plants) under the current progeny testing scheme but the material is ready for recycling at the end of the first year. This is suboptimal as we have seen that RS shows a greater potential when the number of crossing events is increased in time (Gorjanc, Gaynor, and Hickey 2018). Presently, the three additional years are for evaluation and are not necessary to prepare the material for the crossing step. With a tool like GP available, there is now the opportunity to dissociate timewise phenotyping and the actual selection.

The first objective of my thesis was to test the potential of early GP on the material generated by the CIAT-Cirad RS-scheme. This is addressed in chapter 2. Different CV-schemes were used to test the predictive ability of the data from the two available generations, $S_{0:2}$ and $S_{0:3}$. Multi-environment calibrations were also tested through CV to evaluate the benefit of including the two sites currently used by the program. This experiment gave us the opportunity to see that the material available for GP in the CIAT-Cirad breeding program is adapted for performing GP based RS. While the CV were realised within generations, our goal is to use GP to predict the line ability of S_0 plants (Gallais 1979) and not S_0 at a specifically the generations $S_{0:2}$ or $S_{0:3}$. For this reason, we designed another experiment relying on $S_{0:4}$ phenotypes as references.

The second objective, addressed in chapter 3, was to predict a more advanced generation, $S_{0:4}$, with calibration realized on earlier phenotyping of different genotypes. This objective aimed to test the calibration/validation performed with multiple generations of progenies. Here again, we tested if calibration could utilize data from the two available sites. The complexity was however increased by testing calibration that would confound generation and site effects but could reduce the phenotyping to one year. We also extended our understanding on the GxE dynamic by testing different predictive models with different assumptions on the GxE variance structure. Finally, we tested methods to select the calibration set in an attempt to maximize the utility of the phenotyping in $S_{0:3}$.

Once we validated that progeny testing at early generation is appropriate to predict breeding values of S_0, to implement the change and apply genomic selection on S_0 plants, we needed to also check the predictive ability of our approach when calibration and predicted material would not come from the same cycle of RS. The third and last objective was then to evaluate the long-term effect of the forward GP for the selection of S_0 families in RS. This will be presented in chapter 4. As no realistic approach in field existed to validate prediction across generations, this objective was addressed through stochastic simulations. This was also the opportunity to not limit our reference set to $S_{0:4}$ as before but to effectively compare the predictions to S_0 derived double haploid. We also took advantage of the

simulations to compare two calibration schemes based on different approaches to the phenotyping. Finally, the selection process was realized on an index using the phenotypic recorded obtained from different traits.

I will finish this work by summarizing the results, discussing the limitations and making some recommendations on the evolution of the CIAT-Cirad breeding program and on future direction for the research.

1.7 Literature cited

Abbas, Adil, Ping Yu, Lianping Sun, Zhengfu Yang, Daibo Chen, Shihua Cheng, and Liyong Cao. 2021. Exploiting Genic Male Sterility in Rice: From Molecular Dissection to Breeding Applications. *Frontiers in Plant Science* 12 (March): 629314. https://doi.org/10.3389/fpls.2021.629314.

Albrecht, Theresa, Valentin Wimmer, Hans-Jürgen Auinger, Malena Erbe, Carsten Knaak, Milena Ouzunova, Henner Simianer, and Chris-Carolin Schön. 2011. Genome-Based Prediction of Testcross Values in Maize. *Theoretical and Applied Genetics* 123 (2): 339–50. https://doi.org/10.1007/s00122-011-1587-7.

Alexandratos, Nikos, and Jelle Bruinsma. 2012. World Agriculture towards 2030/2050: The 2012 Revision, 154.

Allier, Antoine, Simon Teyssèdre, Christina Lehermeier, Laurence Moreau, and Alain Charcosset. 2020. Optimized Breeding Strategies to Harness Genetic Resources with Different Performance Levels. *BMC Genomics* 21 (1). https://doi.org/10.1186/s12864-020-6756-0.

Andrade, Robert, Sergio Urioste, Tatiana Rivera, Benjamin Schiek, Fridah Nyakundi, Jose Vergara, Leroy Mwanzia, Katherine Loaiza, and Carolina Gonzalez. 2021. Where Is My Crop? Data-Driven Initiatives to Support Integrated Multi-Stakeholder Agricultural Decisions. *Frontiers in Sustainable Food Systems* 5 (December): 737528. https://doi.org/10.3389/fsufs.2021.737528.

Atlin, Gary N., and Bethany Fallon Econopouly. 2022. Simple Deterministic Modeling Can Guide the Design of Breeding Pipelines for Self-pollinated Crops. *Crop Science*, csc2.20684. https://doi.org/10.1002/csc2.20684.

Azodi, Christina B., Emily Bolger, Andrew McCarren, Mark Roantree, Gustavo de los Campos, and Shin-Han Shiu. 2019. Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits. *G3; Genes|Genomes|Genetics* 9 (11): 3691–3702. https://doi.org/10.1534/g3.119.400498.

Basnet, Bhoja Raj, Jose Crossa, Susanne Dreisigacker, Paulino Pérez-Rodríguez, Yann Manes, Ravi P. Singh, Umesh R. Rosyara, Fatima Camarillo-Castillo, and Mercedes Murua. 2019. Hybrid Wheat Prediction Using Genomic, Pedigree, and Environmental Covariables Interaction Models. *The Plant Genome* 12 (1): 0. https://doi.org/10.3835/plantgenome2018.07.0051.

Bernal-Vasquez, Angela-Maria, Jens Möhring, Malthe Schmidt, Manfred Schönleben, Chris-Carolin Schön, and Hans-Peter Piepho. 2014. The Importance of Phenotypic Data Analysis for Genomic Prediction - a Case Study Comparing Different Spatial Models in Rye. *BMC Genomics* 15 (1): 646. https://doi.org/10.1186/1471-2164-15-646.

Bernardo, Rex. 1994. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. *Crop Science* 34 (1): 20–25. https://doi.org/10.2135/cropsci1994.0011183X003400010003x.

Bernardo, Rex. 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science* 48 (5): 1649. https://doi.org/10.2135/cropsci2008.03.0131.

Bernardo, Rex. 2016. Bandwagons I, Too, Have Known. *Theoretical and Applied Genetics* 129 (12): 2323–32. https://doi.org/10.1007/s00122-016-2772-5.

Bijl, David L., Patrick W. Bogaart, Stefan C. Dekker, Elke Stehfest, Bert J.M. de Vries, and Detlef P. van Vuuren. 2017. A Physically-Based Model of Long-Term Food Demand. *Global Environmental Change* 45 (July): 47–62. https://doi.org/10.1016/j.gloenvcha.2017.04.003.

Bouis, Howarth E., and Saltzman Amy. 2017 Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. *Global Food Security* 12. https://doi.org/10.1016/j.gfs.2017.01.009

Burgueño, Juan, Gustavo de los Campos, Kent Weigel, and José Crossa. 2012. Genomic Prediction of Breeding Values When Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers. *Crop Science* 52 (2): 707. https://doi.org/10.2135/cropsci2011.06.0299.

Calvert, L, L Sanint, M Châtel, and J Izquierdo. 2006. Rice Production in Latin America at Critical Crossroads, 26.

Campos, Gustavo de los, John M Hickey, Ricardo Pong-Wong, Hans D Daetwyler, and Mario P L Calus. 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193 (2): 327–45. https://doi.org/10.1534/genetics.112.143313.

Casartelli, Alberto, David Riewe, Hans Michael Hubberten, Thomas Altmann, Rainer Hoefgen, and Sigrid Heuer. 2018. Exploring Traditional Aus-Type Rice for Metabolites Conferring Drought Tolerance. *Rice* 11 (1): 9. https://doi.org/10.1186/s12284-017-0189-7.

Châtel, Marc, Elcio Perpétuo Guimaraes, Yolima Ospina, and Jaime Borrero. 1995. Upland Rice Improvement : Using Gene Pools and Populations with Recessive Male-Sterile Gene. *CIRAD-CA* 29.

Châtel, Marc Henri, and Elcio P Guimarães. 1997. *Recurrent Selection in Rice, Using a Male-Sterile Gene*. CIAT.

Collard, Bertrand C.Y, and David J Mackill. 2008. Marker-Assisted Selection: An Approach for Precision Plant Breeding in the Twenty-First Century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1491): 557–72. https://doi.org/10.1098/rstb.2007.2170.

Coster, Albart, and John Bastiaansen. 2010. Package "HaploSim" v1.8.4. https://cran.r-project.org/web/packages/HaploSim/HaploSim.pdf.

Cuevas, Jaime, José Crossa, Víctor Soberanis, Sergio Pérez-Elizalde, Paulino Pérez-Rodríguez, Gustavo de Los Campos, O. A. Montesinos-López, and Juan Burgueño. 2016. Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. *The Plant Genome* 9 (3): 0. https://doi.org/10.3835/plantgenome2016.03.0024.

Cuevas, Jaime, Italo Granato, Roberto Fritsche-Neto, Osval A. Montesinos-Lopez, Juan Burgueño, Massaine Bandeira e Sousa, and José Crossa. 2018. Genomic-Enabled Prediction Kernel Models

with Random Intercepts for Multi-Environment Trials. *G3; Genes|Genomes|Genetics* 8 (4): 1347–65. https://doi.org/10.1534/g3.117.300454.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185 (3): 1021–31. https://doi.org/10.1534/genetics.110.116855.

Dijk, Michiel van, Tom Morley, Marie Luise Rau, and Yashar Saghai. 2021. A Meta-Analysis of Projected Global Food Demand and Population at Risk of Hunger for the Period 2010–2050. *Nature Food* 2 (7): 494–501. https://doi.org/10.1038/s43016-021-00322-9.

Edwards, Stefan McKinnon, Jaap B. Buntjer, Robert Jackson, Alison R. Bentley, Jacob Lage, Ed Byrne, Chris Burt, et al. 2019. The Effects of Training Population Design on Genomic Prediction Accuracy in Wheat. *Theoretical and Applied Genetics*, March. .

Elbasyoni, Ibrahim S., A.J. Lorenz, M. Guttieri, K. Frels, P.S. Baenziger, J. Poland, and E. Akhunov. 2018. A Comparison between Genotyping-by-Sequencing and Array-Based Scoring of SNPs for Genomic Prediction Accuracy in Winter Wheat. *Plant Science* 270 (May): 123–30. https://doi.org/10.1016/j.plantsci.2018.02.019.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Edited by Laszlo Orban. *PLoS ONE* 6 (5): e19379. https://doi.org/10.1371/journal.pone.0019379.

FAO. 2021. *Food Outlook – Biannual Report on Global Food Markets*. FAO. https://doi.org/10.4060/cb7491en.

Fehr, W. R, Elinor L Fehr, and Holly J Jessen. 1991. *Principles of Cultivar Development*. Ames, Iowa: W.R. Fehr.

Fischer, R A, Derek Byerlee, and G O Edmeades. 2009. Can Technology Deliver on the Yield Challenge to 2050?, 46.

Fischer, Tony, Derek Byerlee, and Greg Edmeades. 2014. *Crop Yields and Global Food Security: Will Yield Increase Continue to Feed the World?* ACIAR Monograph Series 158. Canberra: ACIAR.

Fisher, R. A. 1919. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* 52 (2): 399–433. https://doi.org/10.1017/S0080456800012163.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33 (1). https://doi.org/10.18637/jss.v033.i01.

Frouin, Julien, Denis Filloux, James Taillebois, Cécile Grenier, Fabienne Montes, Frédéric de Lamotte, Jean-Luc Verdeil, Brigitte Courtois, and Nourollah Ahmadi. 2014. Positional Cloning of the Rice Male Sterility Gene Ms-IR36, Widely Used in the Inter-Crossing Phase of Recurrent Selection Schemes. *Molecular Breeding* 33 (3): 555–67. https://doi.org/10.1007/s11032-013-9972-3.

Gallais, André. 1979. The Concept of Varietal Ability in Plant Breeding. *Euphytica* 28 (3): 811–23. https://doi.org/10.1007/BF00038955.

Gallais, André. 2011. *Méthodes de création de variétés en amélioration des plantes*. 1ère édition. Savoir Faire. Versailles, France: Edition Quae.

Gaynor, R Chris, Gregor Gorjanc, and John M Hickey. 2021. AlphaSimR: An R Package for Breeding Program Simulations. Edited by D-J de Koning. *G3 Genes*|*Genomes*|*Genetics* 11 (2): jkaa017. https://doi.org/10.1093/g3journal/jkaa017.

Gianola, Daniel. 2013. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. *Genetics* 194 (3): 573–96. https://doi.org/10.1534/genetics.113.151753.

Gianola, Daniel, and Johannes B. C. H. M. van Kaam. 2008. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* 178 (4): 2289–2303. https://doi.org/10.1534/genetics.107.084285.

Glaszmann, J. C. 2008. A Varietal Classification of Asian Cultivated Rice (*Oryza Sativa* L.) Based on Isozyme Polymorphism. In , 83–90. https://doi.org/10.1142/9789812814265_0008.

Gorjanc, Gregor, R. Chris Gaynor, and John M. Hickey. 2018. Optimal Cross Selection for Long-Term Genetic Gain in Two-Part Programs with Rapid Recurrent Genomic Selection. *Theoretical and Applied Genetics* 131 (9): 1953–66. https://doi.org/10.1007/s00122-018-3125-3.

Gorjanc, Gregor, Janez Jenko, Sarah J. Hearne, and John M. Hickey. 2016. Initiating Maize Pre-Breeding Programs Using Genomic Selection to Harness Polygenic Variation from Landrace Populations. *BMC Genomics* 17 (1): 30. https://doi.org/10.1186/s12864-015-2345-z.

Grenier, Cécile, Tuong-Vi Cao, Yolima Ospina, Constanza Quintero, Marc Henri Châtel, Joe Tohme, Brigitte Courtois, and Nourollah Ahmadi. 2015. Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. *PLOS ONE* 10 (8): e0136594. https://doi.org/10.1371/journal.pone.0136594.

GRiSP, (Global Rice Science Partnership). 2013. *Rice Almanac: Source Book for the Most Important Economic Activities on Earth*. Fourth Edition. Los Baños, Philippines: IRRI.

Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical Mapping of Temperature-Sensitive Mutations of Adenoviruses. *Cold Spring Harbor Symposia on Quantitative Biology* 39 (0): 439–46. https://doi.org/10.1101/SQB.1974.039.01.056.

Guiderdoni, Emmanuel. 2021. Synthetic Apomixis in Rice : Close to the Grail. ISFRG.

Guo, Zhigang, Dominic M. Tucker, Christopher J. Basten, Harish Gandhi, Elhan Ersoz, Baohong Guo, Zhanyou Xu, Daolong Wang, and Gilles Gay. 2014. The Impact of Population Structure on Genomic Prediction in Stratified Populations. *Theoretical and Applied Genetics* 127 (3): 749–62. https://doi.org/10.1007/s00122-013-2255-x.

Habier, David, Rohan L Fernando, Kadir Kizilkaya, and Dorian J Garrick. 2011. Extension of the Bayesian Alphabet for Genomic Selection. *BMC Bioinformatics* 12 (1): 186. https://doi.org/10.1186/1471-2105-12-186.

Hasan, Nazarul, Sana Choudhary, Neha Naaz, Nidhi Sharma, and Rafiul Amin Laskar. 2021. Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes. *Journal of Genetic Engineering & Biotechnology* 19 (August): 128. https://doi.org/10.1186/s43141-021-00231-1.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased Accuracy of Artificial Selection by Using the Realized Relationship Matrix. *Genetics Research* 91 (01): 47. https://doi.org/10.1017/S0016672308009981.

Hickey, John M, Tinashe Chiurugwi, Ian Mackay, Wayne Powell, and Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants. 2017. Genomic Prediction Unifies Animal

and Plant Breeding Programs to Form Platforms for Biological Discovery. *Nature Genetics* 49 (August): 1297.

Hickey, John M., Susanne Dreisigacker, Jose Crossa, Sarah Hearne, Raman Babu, Boddupalli M. Prasanna, Martin Grondona, et al. 2014. Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. *Crop Science* 54 (4): 1476–88. https://doi.org/10.2135/cropsci2013.03.0195.

http://rice.uga.edu. n.d. Http://Rice.Uga.Edu.

Hull, F. G. 1946. Recurrent Selection for Specific Combining Ability in Corn. *Journal of American Society of Agronomy* 37: 134–46. https://doi.org/10.2134/agronj1945.00021962003700020006x.

International Rice Genome Sequencing Project, and Takuji Sasaki. 2005. The Map-Based Sequence of the Rice Genome. *Nature* 436 (7052): 793–800. https://doi.org/10.1038/nature03895.

IPCC. 2022. Chapter 5: Food, Fibre, and Other Ecosystem Products. In *IPCC WGII Sixth Assessment Report*. https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_FinalDraft_Chapter05.pdf.

Jarquín, Diego, Reka Howard, Jose Crossa, Yoseph Beyene, Manje Gowda, Johannes W. R. Martini, Giovanny Covarrubias Pazaran, et al. 2020. Genomic Prediction Enhanced Sparse Testing for Multi-Environment Trials. *G3; Genes|Genomes|Genetics* 10 (8): 2725–39. https://doi.org/10.1534/g3.120.401349.

Jarquín, Diego, Cristiano Lemes da Silva, R. Chris Gaynor, Jesse Poland, Allan Fritz, Reka Howard, Sarah Battenfield, and Jose Crossa. 2017. Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. *The Plant Genome* 10 (2): plantgenome2016.12.0130. https://doi.org/10.3835/plantgenome2016.12.0130.

Jiang, Guo-Liang. 2013. Molecular Markers and Marker-Assisted Breeding in Plants. In *Plant Breeding from Laboratories to Fields*, edited by Sven Bode Andersen. InTech. https://doi.org/10.5772/52583.

Jiang, Yong, and Jochen C. Reif. 2015. Modeling Epistasis in Genomic Selection. *Genetics* 201 (2): 759–68. https://doi.org/10.1534/genetics.115.177907.

Kiran, Aysha, Abdul Wakeel, Khalid Mahmood, Rafia Mubaraka, Hafsa, and Stephan M. Haefele. 2022. Biofortification of Staple Crops to Alleviate Human Malnutrition: Contributions and Potential in Developing Countries. *Agronomy* 12 (2): 452. https://doi.org/10.3390/agronomy12020452.

Kromdijk, Johannes, and Stephen P. Long. 2016. One Crop Breeding Cycle from Starvation? How Engineering Crop Photosynthesis for Rising CO_2 and Temperature Could Be One Important Route to Alleviation. *Proceedings of the Royal Society B: Biological Sciences* 283 (1826): 20152578. https://doi.org/10.1098/rspb.2015.2578.

Liu, Huiming, Biructawit Bekele Tessema, Just Jensen, Fabio Cericola, Jeppe Reitan Andersen, and Anders Christian Sørensen. 2019. ADAM-Plant: A Software for Stochastic Simulations of Plant Breeding From Molecular to Phenotypic Level and From Simple Selection to Complex Speed Breeding Programs. *Frontiers in Plant Science* 9. https://doi.org/10.3389/fpls.2018.01926.

Lorenz, Aaron J., Shiaoman Chao, Franco G. Asoro, Elliot L. Heffner, Takeshi Hayashi, Hiroyoshi Iwata, Kevin P. Smith, Mark E. Sorrells, and Jean-Luc Jannink. 2011. Genomic Selection in Plant

Breeding. In *Advances in Agronomy*, 110:77–123. Elsevier. https://doi.org/10.1016/B978-0-12-385531-2.00002-5.

Lorenz, Aaron J., and Kevin P. Smith. 2015. Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. *Crop Science* 55 (6): 2657–67. https://doi.org/10.2135/cropsci2014.12.0827.

Luh, Bor S., ed. 1991. *Rice*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-3754-4.

Lush, Jay Laurence. 1937. *Animal Breeding Plans*. Iowa State College Press. Ames, Iowa.

Martinez, César P., Edgar A. Torres, Marc Châtel, Gloria Mosquera, Jorge Duitama, Manabu Ishitani, M. Selvaraj, et al. 2014. Rice Breeding in Latin America. Book_section. Plant Breeding Reviews: Volume 38. Wiley-Blackwell. Amérique latine. 2014. https://agritrop.cirad.fr/575285/.

Meuwissen, T H, B J Hayes, and M E Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157 (4): 1819–29.

Montesinos-López, Osval Antonio, Abelardo Montesinos-López, Paulino Pérez-Rodríguez, José Alberto Barrón-López, Johannes W. R. Martini, Silvia Berenice Fajardo-Flores, Laura S. Gaytan-Lugo, Pedro C. Santana-Mancilla, and José Crossa. 2021. A Review of Deep Learning Applications for Genomic Selection. *BMC Genomics* 22 (1): 19. https://doi.org/10.1186/s12864-020-07319-x.

Morais Júnior, Odilon P., JoAAo Batista Duarte, FlAA¡vio Breseghello, Alexandre S G Coelho, Tereza C O Borba, Jordene T Aguiar, PAAricles C F Neves, and Orlando P Morais. 2017. Relevance of Additive and Nonadditive Genetic Relatedness for Genomic Prediction in Rice Population under Recurrent Selection Breeding. *Genetics and Molecular Research* 16 (4). https://doi.org/10.4238/gmr16039849.

OECD and FAO. 2021. *OECD-FAO Agricultural Outlook 2021-2030*. OECD-FAO Agricultural Outlook. OECD. https://doi.org/10.1787/19428846-en.

Onogi, Akio, Osamu Ideta, Yuto Inoshita, Kaworu Ebana, Takuma Yoshioka, Masanori Yamasaki, and Hiroyoshi Iwata. 2015. Exploring the Areas of Applicability of Whole-Genome Prediction Methods for Asian Rice (Oryza Sativa L.). *Theoretical and Applied Genetics* 128 (1): 41–53. https://doi.org/10.1007/s00122-014-2411-y.

Ould Estaghvirou, Sidi Boubacar Ould, Joseph O. Ogutu, and Hans-Peter Piepho. 2014. Influence of Outliers on Accuracy Estimation in Genomic Prediction in Plant Breeding. *G3; Genes|Genomes|Genetics* 4 (12): 2317–28. https://doi.org/10.1534/g3.114.011957.

Park, Trevor, and George Casella. 2008. The Bayesian Lasso. *Journal of the American Statistical Association* 103 (482): 681–86. https://doi.org/10.1198/016214508000000337.

Podlich, D. W., and M. Cooper. 1998. QU-GENE: A Simulation Platform for Quantitative Analysis of Genetic Models. *Bioinformatics* 14 (7): 632–53. https://doi.org/10.1093/bioinformatics/14.7.632.

Pook, Torsten, Martin Schlather, and Henner Simianer. 2020. MoBPS - Modular Breeding Program Simulator. *G3 Genes|Genomes|Genetics* 10 (6): 1915–18. https://doi.org/10.1534/g3.120.401193.

Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of Direct Genomic Values for Animals with Different Relationships within and to the Reference Population. *Journal of Dairy Science* 95 (1): 389–400. https://doi.org/10.3168/jds.2011-4338.

Rao, M. Krishna, K. Uma Devi, and A. Arundhati. 1990. Applications of Genic Male Sterility in Plant Breeding. *Plant Breeding* 105 (1): 1–25. https://doi.org/10.1111/j.1439-0523.1990.tb00447.x.

Ray, Deepak K., Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. Edited by John P. Hart. *PLoS ONE* 8 (6): e66428. https://doi.org/10.1371/journal.pone.0066428.

Rutkoski, J. E. 2019. Estimation of Realized Rates of Genetic Gain and Indicators for Breeding Program Assessment. *Crop Science* 59 (3): 981–93. https://doi.org/10.2135/cropsci2018.09.0537.

Rutkoski, Jessica E. 2019. A Practical Guide to Genetic Gain. In *Advances in Agronomy*, 157:217–49. Elsevier. https://doi.org/10.1016/bs.agron.2019.05.001.

Sahadevan, P. C., and K. M. Narayanan Namboodiri. 1963. Natural Crossing in Rice. *Proceedings / Indian Academy of Sciences* 58 (3): 176–85. https://doi.org/10.1007/BF03051950.

Schrauf, Matías F., Gustavo de los Campos, and Sebastián Munilla. 2021. Comparing Genomic Prediction Models by Means of Cross Validation. *Frontiers in Plant Science* 12 (November): 734512. https://doi.org/10.3389/fpls.2021.734512.

Singh, R. J., and H. Ikehashi. 1981. Monogenic Male-sterility in Rice: Induction, Identification and Inheritance [1]. *Crop Science* 21 (2): 286–89. https://doi.org/10.2135/cropsci1981.0011183X002100020020x.

Stehfest, Elke, Willem-Jan van Zeist, Hugo Valin, Petr Havlik, Alexander Popp, Page Kyle, Andrzej Tabeau, et al. 2019. Key Determinants of Global Land-Use Projections. *Nature Communications* 10 (1): 2166. https://doi.org/10.1038/s41467-019-09945-w.

Su, Guosheng, Ole F. Christensen, Tage Ostersen, Mark Henryon, and Mogens S. Lund. 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. Edited by Abraham A. Palmer. *PLoS One* 7 (9): e45293. https://doi.org/10.1371/journal.pone.0045293.

Taillebois, James, and Elcio P Guimarães. 1989. CNA-IRAT 5 Upland Rice Population. https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1085653/1/IRRN19890001.pdf.

Tibshirani, Robert. 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

UN. 2019. World Population Prospects. 2019. https://population.un.org/wpp/.

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91 (11): 4414–23. https://doi.org/10.3168/jds.2007-0980.

Vitezica, Zulma G, Luis Varona, and Andres Legarra. 2013. On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. *Genetics* 195 (4): 1223–30. https://doi.org/10.1534/genetics.113.155176.

Wang, Wensheng, Ramil Mauleon, Zhiqiang Hu, Dmytro Chebotarov, Shuaishuai Tai, Zhichao Wu, Min Li, et al. 2018. Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice. *Nature* 557 (7703): 43–49. https://doi.org/10.1038/s41586-018-0063-9.

Whittaker, John C., Robin Thompson, and Mike C. Denham. 2000. Marker-Assisted Selection Using Ridge Regression. *Genetical Research* 75 (2): 249–52. https://doi.org/10.1017/S0016672399004462.

Xu, Yunbi, and Jonathan H. Crouch. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science* 48 (2): 391–407. https://doi.org/10.2135/cropsci2007.04.0191.

Yabe, Shiori, Hiroyoshi Iwata, and Jean-Luc Jannink. 2017. A Simple Package to Script and Simulate Breeding Schemes: The Breeding Scheme Language. *Crop Science* 57 (3): 1347. https://doi.org/10.2135/cropsci2016.06.0538.

Yang, Jian, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, et al. 2010. Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nature Genetics* 42 (7): 565–69. https://doi.org/10.1038/ng.608.

Zohary, Daniel. 2001. Domestication of Crop Plants. In *Encyclopedia of Biodiversity*, 217–27. https://doi.org/10.1016/B0-12-226865-2/00079-1.

Zou, Hui, and Trevor Hastie. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Chapter 2 : Impact of early genomic prediction for recurrent selection in an upland rice synthetic population

Avant-propos

Ce chapitre a fait l'objet d'une publication dans la revue G3. Il aura été le premier article abordé durant la thèse. Ça a été l'occasion de découvrir les données phénotypiques, de faire les premières statistiques exploratives ainsi que de traiter les données aberrantes. Grâce à cet article j'ai également pour la première fois manipulé des données génotypiques et réalisé les différentes étapes de contrôle qualité afin que les génotypes soient prêts pour de la prédiction.

Une fois tout ce travail préparatif terminé, j'ai finalement réalisé mes premières prédictions génomiques. Dans un premier temps le package R de prédiction génomique BGLR (Perez and de Los Campos 2014) a été utilisé puis, après une réorientation stratégique, c'est le package ASReml-R (Butler et al. 2007) qui a été utilisé. J'aurais eu ainsi l'opportunité de tester les deux outils, malheureusement sans avoir le temps d'en comparer les résultats.

Vous trouvez ci-dessous la liste de mes précieux co-auteurs ainsi que le DOI. J'ai préféré garder une identité graphique unique tout au long du document de thèse plutôt que d'y coller l'article tel qu'il a été publié en ligne.

Cédric Baertschi[1,2], Tuong-Vi Cao[1,2], Jérôme Bartholomé [1,2,3], Yolima Ospina[4], Constanza Quintero[4], Julien Frouin[1,2], Jean-Marc Bouvet[1,2,5], and Cécile Grenier[1,2,4]

[1]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France,

[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France,

[3]Rice Breeding Platform, International Rice Research Institute, Metro Manila, Philippines,

[4]Alliance Bioversity-CIAT, Recta Palmira Cali, Colombia, and

[5]CIRAD, Dispositif de Recherche et d'Enseignement en Partenariat "Forêts et Biodiversité à Madagascar", Antananarivo, Madagascar

DOI: 10.1093/g3journal/jkab320

2.1 Abstract

Population breeding through recurrent selection is based on the repetition of evaluation and recombination among best-selected individuals. In this type of breeding strategy, early evaluation of selection candidates combined with genomic prediction could substantially shorten the breeding cycle length, thus increasing the rate of genetic gain. The objective of this study was to optimize early genomic prediction in an upland rice (Oryza sativa L.) synthetic population improved through recurrent selection via shuttle breeding in two sites. To this end, we used genomic prediction on 334 S_0 genotypes evaluated with early generation progeny testing ($S_{0:2}$ and $S_{0:3}$) across two sites. Four traits were measured (plant height, days to flowering, grain yield, and grain zinc concentration) and the predictive ability was assessed for the target site. For days to flowering and plant height, which correlate well among sites (0.51–0.62), an increase of up to 0.4 in predictive ability was observed when the model was trained using the two sites. For grain zinc concentration, adding the phenotype of the predicted lines in the nontarget site to the model improved the predictive ability (0.51 with two-site and 0.31 with single-site model), whereas for grain yield the gain was less (0.42 with two-site and 0.35 with single-site calibration). Through these results, we found a good opportunity to optimize the genomic recurrent selection scheme and maximize the use of resources by performing early progeny testing in two sites for traits with best expression and/or relevance in each specific environment.

Keywords: rice; recurrent selection; genomic prediction; GxE; grain zinc concentration

2.2 Introduction

Population improvement strategies are recognized as methods to exploit the genetic diversity of a crop and enrich the genetic basis of breeding programs. In rice, population breeding through recurrent selection (RS) was suggested as a valuable option in countering the decline in genetic diversity among the improved rice germplasm from Latin America and the Caribbean (LAC) (Cuevas-Pérez *et al*. 1992; Guimarães *et al.* 1996). RS in rice started in South America in 1985 (Taillebois and Guimarães 1989) and later spread to most of the continent through a Food and Agriculture Organization funded initiative (Châtel *et al*. 2005; Martínez *et al.* 2015). In the region, RS was applied to rice synthetic populations, each composed of several elite materials, carefully chosen as founders, which had intercrossed for various generations (Guimarães 2005). Following recurrent cycles of selection and recombination, several thousand S_0 plants (S being used here to define the number of selfing cycles) are available for use in the breeding program either as new parents for population improvement or as S_0 progenies for variety development. The particularity of RS breeding in rice as performed in various countries in LAC is that it uses a recessive nuclear male sterility (*ms*) gene to facilitate outcrossing (reviewed in Frouin *et al*. 2014). This gene allows random recombination among a large number of parental plants at each cycle. Different ways are used to improve populations carrying the *ms* gene (Châtel and Guimarães 1997). The most common practice is to evaluate a moderate number (200-300) of candidates randomly drawn from the synthetic population. The evaluation of the candidates is then performed through progeny testing, with more or less fixed families ($S_{0:2}$, $S_{0:3}$ or $S_{0:4}$ depending on the trait and required experimental design, obtained through several cycles of inbreeding and bulk harvest). Subsequently, parental lines are selected to be used for the next recombination cycle. Among others, two compromises have to be made that have a direct impact on the genetic gain achieved by the RS breeding scheme: (i) the number of candidate units evaluated through progeny testing with direct impact on the selection intensity and; (ii) the required degree of fixation of those progenies prior to phenotyping, which would affect the breeding cycle length and influence the precision of genetic variance estimates.

Since its introduction by Meuwissen *et al*. (2001), genomic prediction (GP) has been widely adopted by animal and plant breeders alike. By allowing rapid selection of superior genotypes and accelerating the breeding cycle, GP has shown great potential since the advent of this new breeding paradigm in crop species in 2007 (Bernardo and Yu 2007). The value of GP in the context of RS is fairly evident as the selection based on genomic estimated breeding value (GEBV) can be applied to a very large population of genotyped entries through the calibration of a prediction model performed on a reduced set of training units. Furthermore, the average progeny phenotypic values associated with the genomic matrix of the respective S_0 individuals could allow a more precise estimate of the genetic variation in the case of early generation segregating candidate units. GP was simulated on multiparent populations

(Guo *et al*. 2011; Heffner *et al*. 2011; Bian and Holland 2017; Allier *et al*. 2019) directly related or not to a breeding program to assess the potential use of GP in genetic improvement through RS. However, few simulation studies have assessed the potential value of GP for crop synthetic populations (Müller *et al*. 2017; Schopp *et al*. 2017; Müller *et al*. 2018). Theoretically and through simulation approaches, recurrent genomic selection has the particular advantage of managing both the genetic gain and the maintenance of genetic diversity in the breeding program (Gorjanc *et al.* 2018; Allier *et al*. 2020). In a simulated wheat breeding program, the inclusion of a step of population improvement with rapid recycling of early material proved to be greatly superior in terms of genetic gain compared to a program relying solely on biparental crosses between elite material to generate diversity (Gaynor *et al.* 2017). Similarly, recurrent genomic selection in soybean (Ramasubramanian and Beavis, 2020) and maize (Zhang *et al.* 2017) showed the long-term potential of RS combined with GP. GP has already been applied to material from RS for rice in single (Grenier *et al.* 2015; Morais Júnior *et al*. 2017) and multi-environment contexts (Morais Júnior *et al*. 2018). In these studies, the results showed relatively good predictive ability (PA) for various simple and complex traits such as plant height, flowering date and grain yield. In both cases, however, the calibrations were based on material that underwent some degrees of fixation through plant selection and a few cycles of selfing. A significant jump in efficiency in these schemes is expected by calibrating on early generation candidates from S_0 progenies to save time in building the models and to accelerate the recycling of the selected germplasm.

GP integrating genotype by environment interaction (GxE) has proven successful, showing greater PA than the single environment prediction, provided environments are positively correlated. An approach to multi-environment GP was proposed and applied by Burgueño *et al.* (2012) where the authors modelled the environment and genotype covariance structure and used it within a mixed model framework. Later, GxE was incorporated in a GP model by separately capturing the main marker effect, common to all environments, and an environment-specific marker effect (Lopez-Cruz *et al.* 2015). This method is easy to implement and showed good results for wheat breeding under multiple environmental conditions. Additionally, it has the advantage that it enables working with different genotype covariance structures. Genotype covariances based on either a linear kernel (GxE GBLUP) or a Gaussian kernel (GxE RKHS) have been tested, and the Gaussian kernel allows a more flexible structure than the linear kernel and potentially better prediction (Cuevas *et al.* 2016). To optimize calibration with the multi-environment data, various strategies of genome-based models including GxE were proposed (Jarquin *et al.* 2020). The authors compared different partitioning of the calibration sets among the multiple sites where the population was tested, with different degree of overlapping of the genotypes between environments. Sparse testing designs in which subset of the genotypes are tested in each location was presented as a method to reduce the experimental effort and optimize the use of breeding program resources.

The current study was conducted in the context of a collaborative rice breeding program between CIAT (International Center for Tropical Agriculture, member of the CGIAR centers) and Cirad (French Agricultural Research Centre for International Development). The CIAT-Cirad rice breeding program has historically conducted population development and improvement through RS. Its current RS program based on progeny testing is conducted in two locations; at CIAT-HQ in Palmira, where rice is cultivated all year round under irrigated conditions, and in Santa Rosa, an experimental site where rice is grown under rainfed conditions during the main cropping season. While aiming to implement early GP in our RS scheme, we were also interested in making optimal use of all the data gathered in both locations (target and not target) for the breeding program. The main objective was to evaluate the PA of the GP model including the GxE interaction to obtain reliable estimates of the breeding value of selection candidates in the target site.

2.3 Material and methods

2.3.1 Development of PCT27 population

The genetic material used in this study belongs to the tropical japonica group of rice (*Oryza sativa* L.). Several synthetic populations developed in the CIAT-Cirad rice breeding program were improved for adaptation to upland ecosystems and acid soils. In 2015, Grenier *et al.* (2015) used a training set defined with 348 $S_{2:4}$ lines derived from four populations to study the potential of GP in an RS scheme. Of the 348 families at the S_2 generation, marker-assisted-selection for the *ms* gene (Frouin *et al.* 2014) helped to select [ms:Ms] male fertile plants in 35 randomly sampled families. One single plant per family was selfed, and the seeds of each of 35 plants were mixed in equal proportion to generate a candidate population hereafter referred to as PCT27 (Figure 2-1). Two recombination cycles were performed at CIAT-HQ in Palmira under irrigated conditions in a bundled field isolated from other rice experimental plots by at least 50 m to avoid pollen contamination and without any selection pressure. At each cycle, a population of about 3,000 plants was established with male sterile and male fertile plants randomly distributed within the plot. The recombining units were then collected by harvesting male sterile plants pollinated by any male fertile plants in the vicinity. At the third cycle of recombination, 334 S_0 fertile plants were randomly extracted from the population to constitute our reference population. All entries were advanced to the $S_{0:2}$ and then $S_{0:3}$ generation by bulk harvesting seeds from 15 to 20 male fertile plants per line per generation. Additionally, 50 temporal checks from the same population were also advanced by bulk method to the generations $S_{0:2}$ and $S_{0:3}$ and were used to test the generation effect and the year effect within the site. The terms line and genotype were used indifferently in this work to refer to the S_0 plants and their bulked offspring at $S_{0:2}$ or $S_{0:3}$ if specified. either generation.

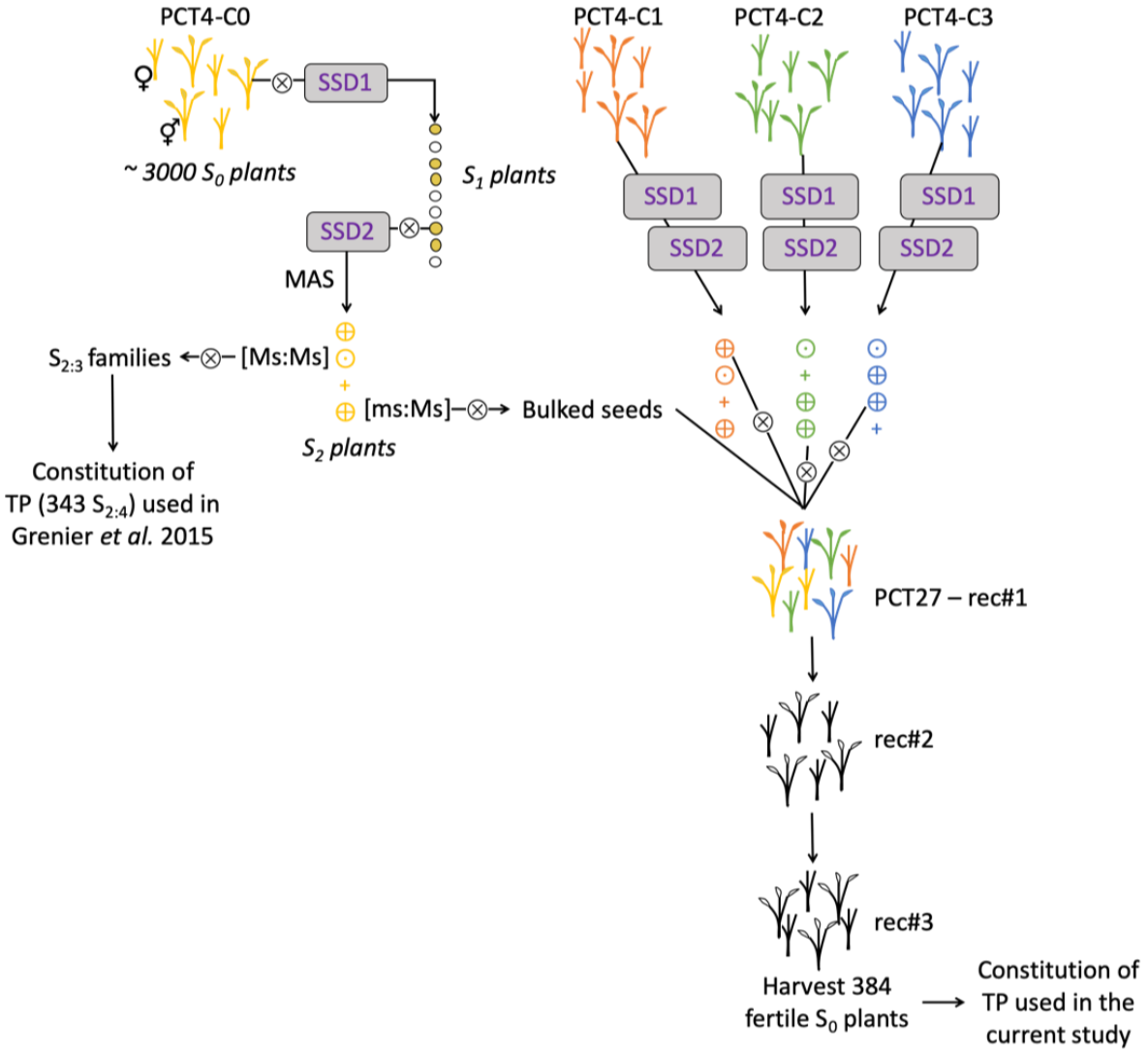


Figure 2-1: Process followed for the development of the PCT27 population. Populations PCT4-C0, PCT4-C1, PCT4-C2 and PCT4-C3 were described in Grenier et al. (2015). Each population contains about 3,000 plants with half male fertile plants (♂) that can be selfed and half male sterile plants (♀). "SSD" is the single descend method of generation advance applied to 100 male fertile plants per population.⊗ indicates the selfing process. The "MAS" (marker-assisted selection) process was performed for the selection of S2 plants based on genotypic profile at the ms gene. Genotyped plants are symbolized as + for plants with the [ms:ms] genotype, ⊕ for the [ms:Ms] genotype and ⊙ for the [Ms:Ms] genotype. "rec" are recombination cycles performed by harvesting all male sterile plants from the population without any selection pressure. For PCT27—&rec#1 this first recombination cycle was done among the progenies of 35 families randomly extracted among the four populations

2.3.2 Genotyping

Leaf tissues were sampled on the 334 S_0 plants and DNA extraction was performed as in Grenier *et al.* 2015. Genotyping was done by genotyping-by-sequencing (GBS) approach (Elshire *et al.* 2011). The detailed method is described in Appendix 1 and the genetic characterization of the population can be seen in supplementary Tables and Figures. As a result of the genotyping and subsequent genetic analysis, the population was characterized by 9,928 SNP markers fairly well distributed among the 12 rice chromosomes (STable 2-2, SFig 2-1). The MAF distribution among the 334 S_0 reflects a population

where rare alleles were not depleted, which fits well with long-term objectives of a population breeding program (SFig 2-2). The degree of allelic fixation varied greatly between the genotypes but remained relatively low for individuals at the S_0 generation (STable 2-2). Considering the rather large average LD (STable 2-4) and the slow LD decay observed (SFig 2-3), the average marker density (1 SNP every 40 kb) was considered good enough to allow the capture of all linked QTLs with the SNP matrix in hand. The whole population was characterized, with a total absence of structure, which provides a good base for setting up a GP scheme through CV (SFig 2-4).

2.3.3 Field trial and phenotyping

Field phenotyping was performed at two locations in Colombia. One site was an experimental field at CIAT-HQ in Palmira (PAL) located in the Valle del Cauca, Colombia (3.50° N - 76.35° W, 1000 masl). At this location, rice evaluation trials are conducted under irrigated conditions and can be performed all year round due to favourable environmental conditions and good water availability that enable the irrigation scheme throughout the crop cycle. As it is not a rice prone area, no severe disease pressure is naturally present, and the rice crop usually expresses its full potential. On the other hand, Santa Rosa (SRO) is an experimental site, owned by the Colombian National Federation of rice growers (Fedearroz) located in the Oriental plains of Colombia, in the department of Meta, Colombia (4.03° N - 73.48° W, 300 masl). At this site, the rice crop is established through direct seeding and the trials are conducted under rainfed conditions during the main cropping season, between May and September. The predominance of rice cultivation in this area, the climatic conditions of hot and humid summers during the main growing season and the natural occurrence of various strains of pathogens (bacterial, fungal or viral) make this site a hot spot for disease screening.

Four trials were conducted during two consecutive years, 2017 and 2018, using different semesters for each location. Field trials were established in PAL on 04/12/2017 and 10/12/2018 and in SRO on 12/05/2017 and 30/05/2018. At each site, the experimental design followed a lattice with 16 blocks and three repetitions. The 50 temporal check lines (only $S_{0:2}$ in the 2017 trials and $S_{0:2}$ and $S_{0:3}$ lines in the 2018 trials) were randomly distributed across the design within each repetition of the two sites and two-year trials. In PAL, the trials were established after transplanting 3-week-old seedlings in a bundled field. The plot size was two rows of 17 plants with 25 cm between plants and between rows. Fertilizer application followed a split application with NPK nutrients added at 25 and 35 days after transplanting. Irrigation was maintained continuously in order to ensure a 25 cm layer of water in the field until a week prior to the crop maturation period. In SRO, the trials were established by direct sowing of two 4 m-long rows, spaced by 26 cm at a density of 1 gram of seed per linear meter. Split fertilizer application was performed according to the recommended application for growing tropical

japonica rice in upland soil conditions. Phytosanitary treatment was applied in SRO to prevent blast outbreaks. For all four trials, a similar design was applied, but with a different randomization.

Four traits were measured following the IRRI Standard Evaluation System (IRRI 2013) on the whole training population including the 50 temporal checks. Flowering date (FL) was expressed as the number of days after crop establishment – being either the date after either transplantation (PAL) or sowing (SRO) – when 50% of the plants within a plot reached anthesis. Plant height (PH) was calculated as the average height measured in centimetres of five plants with their panicle extended. Grain yield (YLD) was obtained by weighing the grains collected within each plot after discarding the plants at the start and end of each plot. For each harvested plot, percent humidity was measured and used to correct the weight of collected grains, expressed in grams per plot, for a relative humidity of 14%. The YLD value was neither adjusted for the plot size nor for the count of fertile plants. The grain zinc concentration (ZN), expressed in parts per million (ppm), was measured on a subsample of collected grains polished in Teflon equipment, using energy dispersive X-ray fluorescence spectrometry (X-supreme 8000, Oxford Instrument, Shanghai, CN) available at the CIAT-HQ Nutritional Laboratory. The exact same procedure was used for generation $S_{0:2}$ in 2017 and generation $S_{0:3}$ in 2018.

The 50 temporal checks were phenotyped as $S_{0:2}$ in 2017 and as $S_{0:2}$ and $S_{0:3}$ in 2018. This allowed measurement of the non-confounded year within site effect on the $S_{0:2}$ and the generation effect in 2018 by analysing the data from the $S_{0:2}$ and $S_{0:3}$ lines as presented in Appendix 2.

2.3.4 Statistical models for genomic prediction

Raw data were visually explored for outliers as described in Appendix 3. Based on clean data, Pearson's correlation between phenotypic BLUPs obtained in PAL and SRO was computed for generations $S_{0:2}$ and $S_{0:3}$ using the 334 S_0 families phenotyped in both generations.

All the models were estimated using ASReml-R v3.0 (Butler *et al.* 2007). GP was done independently in each generation. For single-site calibration, the following model was used

$$Y_{ijk} = \mu + r_i + b(r)_{ij} + g_k + \varepsilon_{ijk} \qquad \text{(Model 1)}$$

The fixed effects were the intercept μ and the replicate effect r_i. The random part was composed of the block effect b_{ij} nested in replicate with distribution $b \sim \mathrm{N}(0, I\sigma_b^2)$, the genotype effect g_k that represents the progeny means with distribution $g \sim \mathrm{N}(0, M\sigma_g^2)$ and the residual ε_{ijk} with $\varepsilon \sim \mathrm{N}(0, I\sigma_\varepsilon^2)$. The variance σ_b^2 is associated with the blocks, while σ_g^2 and σ_ε^2 are the genotypic and error variances, respectively. The two variance-covariance matrices used are I for the identity matrix and M representing the genotype variance-covariance computed according to either of the two prediction methods described below.

For the two-site approach the following model was used

$$Y_{ijkl} = \mu + s_i + r(s)_{ij} + b(r(s))_{ijk} + g_l + gs_{il} + \varepsilon_{ijkl} \quad \text{(Model 2)}$$

The fixed effects were the same as for Model 1, with an additional fixed site effect s_i. The random part of Model 1 was completed with the genotype (progeny means) by site interaction gs_{il} with distribution $gs \sim \mathrm{N}\left(0, \begin{bmatrix} M_{PAL}\,\sigma^2_{gsPAL} & 0 \\ 0 & M_{SRO}\,\sigma^2_{gsSRO} \end{bmatrix}\right)$ and the residual ε_{ijkl} with distribution $\varepsilon \sim \mathrm{N}\left(0, \mathrm{I} \otimes \begin{bmatrix} \sigma^2_{\varepsilon PAL} & 0 \\ 0 & \sigma^2_{\varepsilon SRO} \end{bmatrix}\right)$.

In addition to the three variances described in Model 1, Model 2 includes two site-specific genotype by site interaction variances σ^2_{gsPAL} and σ^2_{gsSRO} as well as two site-specific error variances $\sigma^2_{\varepsilon PAL}$ and $\sigma^2_{\varepsilon SRO}$. The error variance-covariance is modeled by the Kronecker product of the identity matrix and the variances matrix.

To compute the variance structure (*M*) for the genotype effect and genotype by site interaction (M_{PAL}, M_{SRO}), two different kernels were used. In the first approach, GBLUP, $M = M_{PAL} = M_{SRO}$, where M was based on the linear kernel $\mathrm{M} = \frac{XX`}{N}$ (Lopez-Cruz *et al.* 2015), a proportional of the matrix proposed by VanRaden (2008) was used, with *X* being the genomic data with genotypes coded as *-1, 0, 1* and N the number of markers. The second approach, RKHS, was based on the reproducing kernel Hilbert space approach by Gianola and van Kaam (2008). Three different variance-covariance structures were computed: one for the complete data (M_0) and one for each site independently ($\mathrm{M}_{\mathrm{PAL}}, \mathrm{M}_{\mathrm{SRO}}$), all based on the Gaussian kernel $K_e(x_{me}, x_{ne}) = \exp(-h_i \parallel x_{me} - x_{ne} \parallel^2)$, for x_{me}, x_{ne} being two marker genotype vectors and $(m, n) \in \{1, \ldots, N\}^2$. The bandwidth *h* controls the decay rate of the correlation between the lines, smaller *h* giving a sharper correlogram. We computed *h* with the method proposed by Pérez-Elizalde *et al.* (2015) and the provided R function *marg.fun*. A gamma prior distribution for *h* was used, the shape parameter was set at 3 and the scale parameter set at 1.5. Three different bandwidth parameters were computed as the method relies partially on phenotypes, and hence yields different kernels depending on the site. New bandwidth parameters were estimated at each cross-validation (CV) cycle based on the BLUP adjusted phenotypes of the sampled training set, as in Pérez-Elizalde *et al.* (2015). For both methods, the genotypic information was based on 9,928.

Models 1 and 2 with identity matrix as variance-covariance matrices were used to compute broad sense heritability. H^2 at trial level (generation within site) was used as a measure for repeatability and computed using the formula:

$$H^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_\varepsilon}{NR}} \quad \text{(Eq. 1),}$$

2.3.5 Cross-validation schemes for evaluating predictive ability

Several CV schemes were used with different partitioning of the population among the two sites (Figure 2-2, STable 2-3).

In the first instance, only phenotypic data from the target site of selection SRO was considered (Figure 2-2). In that scenario, predictions were obtained based on Model 1 with a calibration based on a single site (SIN_{SRO}). Various calibration set sizes (s) were tested, $s \in \{25, 50, 100, 200\}$.

For the two-site CV procedures, Model 2 was used. Calibrations were constructed with either a balanced (BAL) or imbalanced (IMB) representation of both sites. BAL1 represents a calibration method where both sites were represented by an equal number of phenotyped S_0 families. Sets of "s" S_0 were selected and their phenotypes in both sites were used for the training (Figure 2-2). This corresponds to a CV1 in Burgueño (2012). For BAL2, "s" refers to the number of S_0 families observed in SRO and in PAL, however, only a fraction of the families was observed in both sites (i.e., the overlap), the remaining families being observed in only one of them. An overlap of 50% of the total number families included in the calibration was targeted. For "s" S_0 families observed in both sites, the total number of S_0 families was then $\frac{3}{2}$s. For the imbalanced (IMB) scenario, the whole population was phenotyped in PAL and only a fraction of size "s" was phenotyped in SRO (Figure 2-2).

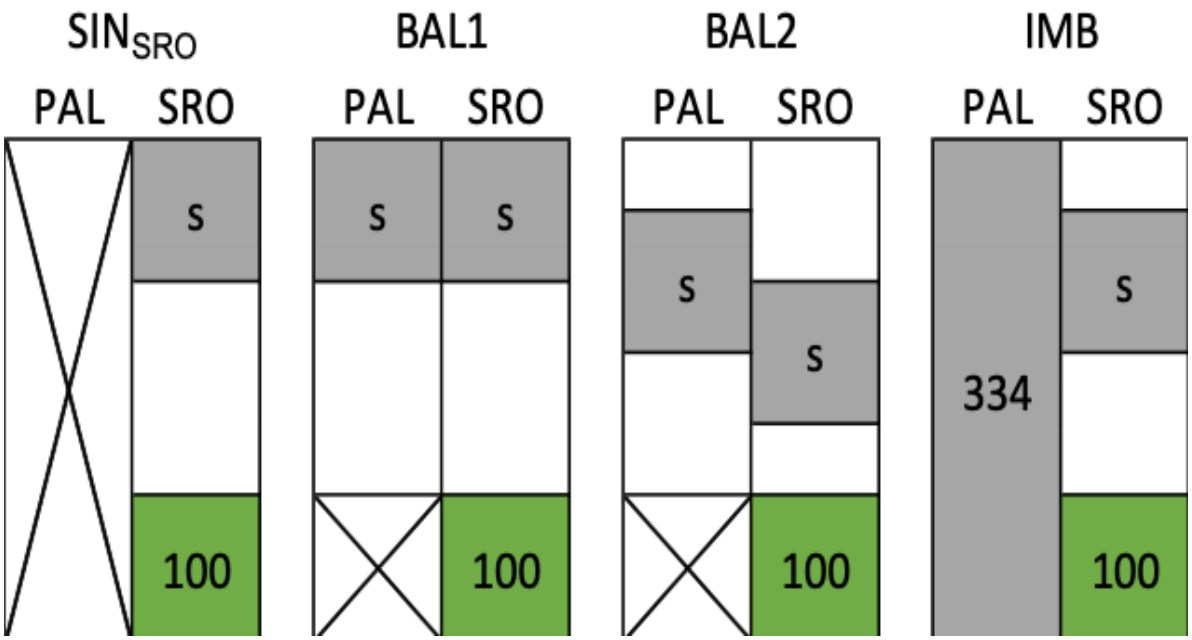


Figure 2-2: The four scenarios of cross-validations to evaluate the prediction accuracy in Santa Rosa (SRO). The first scenario (SINSRO) uses phenotypic information from a single site, while the three others include Palmira (PAL) phenotypes in two-site models. In the latter case, the level of information between locations is either balanced (BAL) or imbalanced (IMB). The grey area represents the genotypes included in the training set with a varying size "s" to calibrate the model and the green area represents the validation set fixed to 100 genotypes.

The same CV procedures were applied to each generation and with both GP models (GBLUP and RKHS). The GEBVs in SRO were obtained for the S_0 included in the validation set, defined as the set for which no phenotype at SRO was recorded. In each scenario, 100 alternative samplings were performed for which the PA was measured as $\text{cor}(\hat{Y}, GEBV)$. The reference $\hat{Y}$ was obtained with the complete SRO phenotypes using Model 1 and $M = I$, I being an identity matrix and computed as $\hat{Y}_k = \mu + g_k$. GEBVs were obtained with the models including molecular information as $\hat{Y}_k = \mu + g_k$ for SIN_{SRO} or $\hat{Y}_{SRO,l} = \mu + s_{SRO} + g_l$ for the other predictions (BAL1, BAL2, and IMB). For each scenario, the mean and the standard deviation of PA were computer on the 100 iterations.

To ensure that the variation in accuracy between the CV procedures was only due to the size differences in the training set, the correlations were always computed on the predictions for 100 genotypes randomly selected from the validation sets. However, for BAL2 with a training set size of 200, the validation set was reduced to 34 genotypes as those were the only genotypes with no phenotypic records that could be used for the validation with this strategy (STable 2-3). The PA was still computed, but as the correlation was computed on only 34 points, the results must be considered with caution.

2.3.6 Effects if the calibration parameters on the predictive abilities

To investigate the response of the PA to the calibration parameters, linear models were fitted to the PA obtained from the 100 iterations with each scenario. Depending on the scenario, the independent variables were year, GP method, CV scenario, training set size and all their combinations. Proportion of variance associated with one or more main effect, errors or interactions were estimated through the Eta^2, as $Eta^2 = SSq_{effect}/SSq_{total}$, where SSq_{effect} is the sum of squares for the effect under consideration and SSq_{total} is the total sum of squares of all effects, errors and interactions in the ANOVA study. Throughout the text, this ratio is expressed as a percentage.

2.4 Results

2.4.1 Effect of sites and generations on the phenotypic performance

The phenotypic data were collected in two sites and on the same S_0 progeny at two generations. In each site, the phenotyping was done in 2017 for 334 families at the $S_{0:2}$ generation and in 2018 for the same 334 families at the $S_{0:3}$ generation.

For most traits recorded in the two locations, the mean phenotypic values differed between sites (Table 2-1, Figure 2-3). While the differences between sites were moderate for FL and PH, they were large for YLD and ZN, with more than 60% change in the 2017 trials. The $S_{0:2}$ families evaluated in 2017 had later flowering, shorter plant height, lower yield and higher zinc concentration in SRO than in PAL. However, this tendency did not hold for the 2018 trials. The differences between sites in PH were greater at the 2018 trials, with taller $S_{0:3}$ plants in SRO. For each trait, the spread of the data was

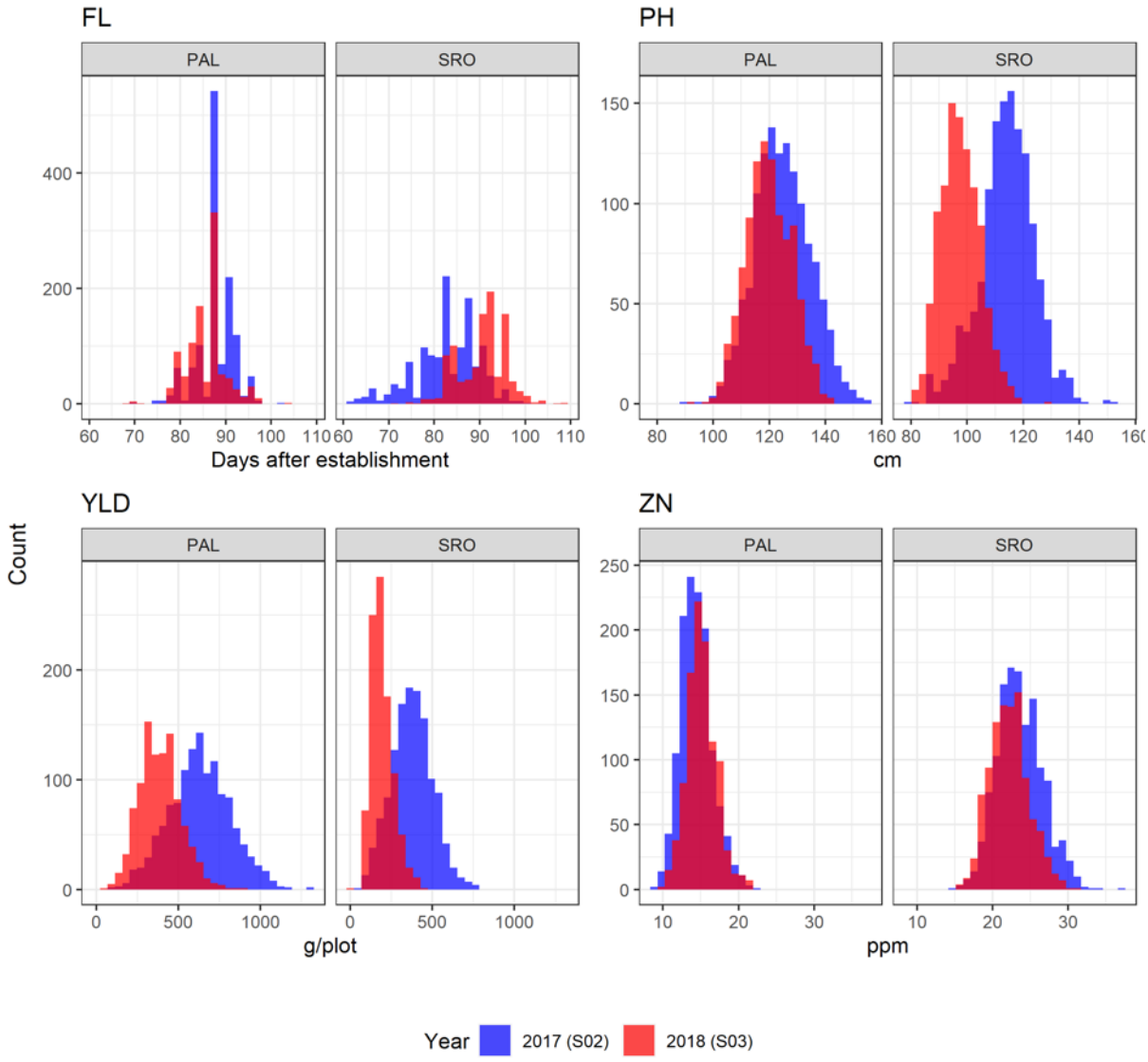


Figure 2-3: Histograms of the raw phenotypic values of the four traits: flowering day (FL), plant height (PH), grain yield per plot (YLD) and grain Zn concentration (ZN). The two environments: Palmira (PAL, irrigated) and Santa Rosa (SRO, rainfed) are represented. Outliers were discarded as presented in Appendix 2.

consistent across site and year with 0.4 to 4 points of difference in the coefficient of variation. The highest coefficient of variation was observed for YLD in 2018 (34%), and was higher than in the 2017 trial (27%). The trait broad sense heritability (H^2) at trial level showed large differences between traits and across sites and years. This measure of trial repeatability ranged from 0.52 for YLD in the PAL_2017 trial to 0.96 for FL in SRO_2017. Heritability was systematically higher in SRO than in PAL, and similar or slightly increased in the 2018 trials for all traits in all locations, but for FL and YLD measured in SRO. As the year and the generation effect were confounded, 50 temporal checks were used to untangle the potential effects of generation and year. The significance of the fixed effect and variance decomposition among the 50 temporal checks showed that differences were exclusively due to year effect and neither a significant generation effect nor a significant genotype by generation interaction could be observed (STable 2-5).

Table 2-1: Descriptive values of the experiments in all trials (site x generation combinations) with mean, standard error (SE), coefficient of variation (Cvar) and broad sense heritability (H2) from Model 1.

				$S_{0:2}$ generation in 2017			
Trait [a]	Site	Mean	SE	min	max	Cvar	H^2 (SE)
FL	PAL	88.24	0.24	75	102	3.88	0.69 (0.03)
	SRO	82.17	0.37	61	96	7.93	0.96 (<0.01)
PH	PAL	125.62	0.62	88.4	155.4	7.76	0.61 (0.04)
	SRO	116.65	0.59	94.2	151.8	6.68	0.79 (0.02)
YLD	PAL	673.85	10.33	237.5	1311.5	24.07	0.52 (0.05)
	SRO	398.54	9.75	54.3	755.1	27.6	0.75 (0.02)
ZN	PAL	14.3	0.18	8.8	22	14.39	0.71 (0.03)
	SRO	23.8	0.21	15.9	37.1	12.64	0.81 (0.02)
				$S_{0:3}$ generation in 2018			
Trait [a]	Site	Mean	SE	min	max	Cvar	H^2 (SE)
FL	PAL	85.7	0.33	68	103	5.04	0.74 (0.02)
	SRO	90.54	0.36	72	108	5.76	0.78 (0.02)
PH	PAL	119.84	0.55	92.5	142.67	6.71	0.76 (0.02)
	SRO	97.63	0.53	80.8	128	7.09	0.80 (0.02)
YLD	PAL	387.54	8.3	54.6	901.1	32.23	0.56 (0.04)
	SRO	191.4	7.37	10.7	461.6	33.91	0.58 (0.04)
ZN	PAL	15.14	0.16	10.05	21.9	12.82	0.75 (0.02)
	SRO	22.21	0.18	15.3	30.8	11.51	0.81 (0.02)

[a] Traits: days to flowering (FL), plant height (PH), grain yield per plot (YLD), grain Zn concentration (ZN)

For each trait scored in each year, an analysis of the variance components was performed on the combined data from both sites using Model 2 (Table 2-2). The proportion of variance explained by the genotype effect was greater that of the combined genotype by site interaction effects from both sites (GxS_{PAL} and GxS_{SRO}) only for FL in 2018, and PH recoded in both years. As a result, greater heritability was observed for these traits/years combination with H^2 = 0.57, 0.50 and 0.62 for FL_2018, PH_2017 and PH_2018, respectively. The lowest genotype contribution to the explanation of variance was encountered for YLD, with large interaction effects and error effects associated with a particular site for each year, resulting in low H^2 in both years (H^2 = 0.19 and 0.11 in 2017 and 2018, respectively). For ZN, the genotype effect represented a third of the combined GxS interaction variances in both year trials, leading to similar and moderate H^2 for both years (H^2 = 0.38 and 0.40 for 2017 and 2018, respectively). The variance decomposition for each trait was coherent with the site correlation observed within years (Table 2-3). The highest correlations between SRO and PAL were observed for PH (r^2 = 0.62) and FL (r^2 = 0.62) in 2018. For the same traits in 2017, the correlations were lower (r^2 = 0.55 and 0.51 for FL and PH, respectively). The site correlation was the lowest for YLD in both years (r^2 between 0.13 and 0.20) and intermediate for ZN with comparable values in both years (r^2 = 0.41 and 0.42 in 2017 and 2018, respectively).

Table 2-2: Variance decomposition and broad sense heritability (H2) from Model 2 by trait and generation. GxSPAL and GxSSRO are the genotype by site interaction variances associated with PAL and SRO, respectively. Bloc stands for the variance associated with bloc within replicate within site. ResidualPAL and ResidualSRO are the residual variances associated with PAL and SRO, respectively.

		$S_{0:2}$ generation in 2017			$S_{0:3}$ generation in 2018		
Trait [a]	Variance component	Variance	proportion	H^2 (SE)	Variance	proportion	H^2 (SE)
FL	Genotype	4.92	0.11	0.25 (0.03)	7.86	0.22	0.57 (0.03)
	GxS_{PAL}	<0.001	<0.001		<0.001	<0.001	
	GxS_{SRO}	26.44	0.62		4.49	0.13	
	Bloc	0.93	0.02		1.89	0.05	
	$Residual_{PAL}$	5.59	0.13		9.23	0.26	
	$Residual_{SRO}$	4.93	0.12		12.4	0.35	
PH	Genotype	21.87	0.17	0.50 (0.04)	22.25	0.26	0.62 (0.03)
	GxS_{PAL}	7.93	0.06		6.8	0.08	
	GxS_{SRO}	7.95	0.06		3.14	0.04	
	Bloc	5.67	0.05		4.35	0.05	
	$Residual_{PAL}$	57.48	0.46		29.9	0.34	
	$Residual_{SRO}$	24.16	0.19		20.36	0.23	
YLD	Genotype	1796.61	0.05	0.19 (0.05)	498.32	0.03	0.11 (0.05)
	GxS_{PAL}	4148.64	0.12		3220.8	0.19	
	GxS_{SRO}	3919.93	0.12		540.88	0.03	
	Bloc	1732.23	0.05		1160.68	0.07	
	$Residual_{PAL}$	16676.45	0.49		9301.29	0.53	
	$Residual_{SRO}$	5768.75	0.17		2674.47	0.15	
ZN	Genotype	1.49	0.14	0.38 (0.04)	1.31	0.16	0.40 (0.04)
	GxS_{PAL}	0.16	0.02		0.27	0.03	
	GxS_{SRO}	3.05	0.29		2.28	0.27	
	Bloc	0.61	0.06		0.44	0.05	
	$Residual_{PAL}$	2.02	0.19		1.62	0.19	
	$Residual_{SRO}$	3.11	0.30		2.53	0.30	

[a] Traits: days to flowering (FL), plant height (PH), grain yield per plot (YLD), grain Zn concentration (ZN)

2.4.2 Predictive abilities with calibration using single environment data

The effects of different parameters used for the calibration of the model were first investigated for the PA from the single environment CV in SRO; SIN_{SRO} (Figure 2-4). Similar global average PA were achieved for all traits combining set sizes, years and GP methods (PA = 0.30, 0.33, 0.27 and 0.24 for FL, PH, YLD and ZN, respectively) (STable 2-6). The linear model including all the factors taken individually, their first-order interaction and one second-order interaction explained 33 to 59% of the observed variation of PA (Table 2-4), indicating that a large proportion of the variability was due to the sampling of the CV method.

Table 2-3: Pearson's phenotypic correlations and p-value for each phenotypic trait (BLUPs obtained from Model 1) recorded in the two sites PAL and SRO within each year of field trial.

Trait [a]	$S_{0:2}$ generation in 2017	$S_{0:3}$ generation in 2018
FL	0.554 (<0.001)	0.624 (<0.001)
PH	0.509 (<0.001)	0.620 (<0.001)
YLD	0.206 (<0.001)	0.134 (0.014)
ZN	0.408 (<0.001)	0.424 (<0.001)

[a] Traits: days to flowering (FL), plant height (PH), grain yield per plot (YLD), grain Zn concentration (ZN)

Table 2-4: Analysis by trait of the factors influencing the variability of the predictive ability. The results are for the CV SIN_{SRO} scenario. Eta^2 is the proportion of variance associated with each effect and R^2 is the coefficient of determination obtained from a linear model applied to the data from the 100 iterations (n=1600).

		SIN_{SRO}	
Trait [a]	Factor [b]	Eta^2	R^2
FL	Year	0.105	0.333
	GP method	0.009	
	Set size	0.215	
	Year:GP method	0.000	
	Year:Set size	0.003	
	GP method:Set size	0.001	
	Year:GP method:Set size	0.001	
PH	Year	0.043	0.592
	GP method	0.000	
	Set size	0.529	
	Year:GP method	0.002	
	Year:Set size	0.017	
	GP method:Set size	0.001	
	Year:GP method:Set size	0.001	
YLD	Year	0.004	0.395
	GP method	0.001	
	Set size	0.386	
	Year:GP method	0.001	
	Year:Set size	0.003	
	GP method:Set size	0.000	
	Year:GP method:Set size	0.001	
ZN	Year	0.027	0.358
	GP method	0.001	
	Set size	0.327	
	Year:GP method	0.000	
	Year:Set size	0.001	
	GP method:Set size	0.001	
	Year:GP method:Set size	0.001	

[a] Traits: days to flowering (FL), plant height (PH), grain yield per plot (YLD), grain Zn concentration (ZN)
[b] Factors: Year: 2017 ($S_{0:2}$), 2018 ($S_{0:3}$); GP method: GBLUP, RKHS; Set size: 25, 50, 100, 200

The training set size accounted for most of the PA variance explained by the model for all the traits. The largest training set size greatly improved the PA for all the traits (Eta^2 = 22%, 53%, 39% and 33% for FL, PH, YLD and ZN, respectively). The year factor described a lower proportion of the total explained variance, with a maximum of 11% of the explained variance in PA for FL. For all traits GP model explained only a very limited proportion (<1%) of the variance. For most traits (PH, YLD and ZN), the average PA was greater when predictions were performed with the GBLUP model. For this reason, the rest of the paper will focus on the results achieved with the GBLUP model. However, the results for RKHS can be found in supplementary data (STable 2-7).

2.4.3 Predictive abilities with calibration using single and two-environment data

Two CV scenarios including SRO and PAL (BAL1, BAL2) were compared to SIN_{SRO} including only SRO data to investigate the combined effect of the training set composition and its size (Figure 2-5). The calibrations were tested in the two different years for their ability to predict line performance in SRO. When the two sites were included in the training set, the main source of variation was the number of phenotypes from SRO and PAL included in the training set. Comparing the PA associated with the set

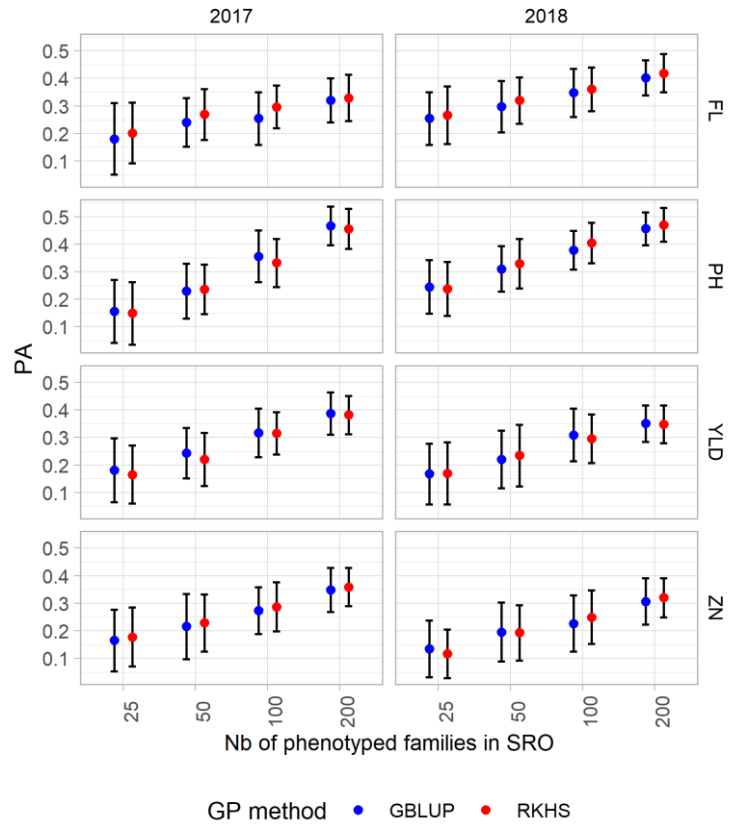


Figure 2-4: Mean predictive ability (PA) for the single-site model in Santa Rosa (SRO) for the four traits: flowering day (FL), plant height (PH), grain yield per plot (YLD) and grain Zn concentration (ZN), scored in two years (2017 and 2018). Four training set sizes (25, 50, 100 and 200) and two GP methods (GBLUP and RKHS) are considered. The bars represent the standard deviation.

size "s" in the case of BAL1 and BAL2 with the PA obtained with the same "s" in SIN_{SRO} allowed us to assess the effect due to the addition of phenotypes from PAL to the training set. Globally, across all set sizes, PA in the BAL2 scenario was greater for all the traits considered (STable 2-8), with average PA ranging from 0.23 for ZN to 0.38 for PH.

While training set size was the factor explaining most of PA variation (>22%) for all traits, year effect had some importance (11%) but only for FL. CV methods on the other hand accounted only for a small fraction of the PA variation. The highest gains in PA provided by any two-site CV scenarios compared to the single-site model were obtained for the training set size of 50 to predict PH_2017 (PA increase of +0.07) using the BAL2 model.

2.4.4 Two-site calibration as a sparse testing approach

So far, we have compared single-site prediction with two-site prediction methods to predict the phenotype of families that were never observed, based solely on between-family information exchange. Another possible approach is to take advantage of the population information by phenotyping all the families in one environment other than the one targeted for the prediction. As PAL is easier to manage, being free of main rice pathogens and closer to the research institute, we tested a scenario with unbalanced representation of the sites in the training sets (IMB), where all 334 families phenotyped in PAL and only a subset of a varying set size "s" phenotyped in SRO were considered.

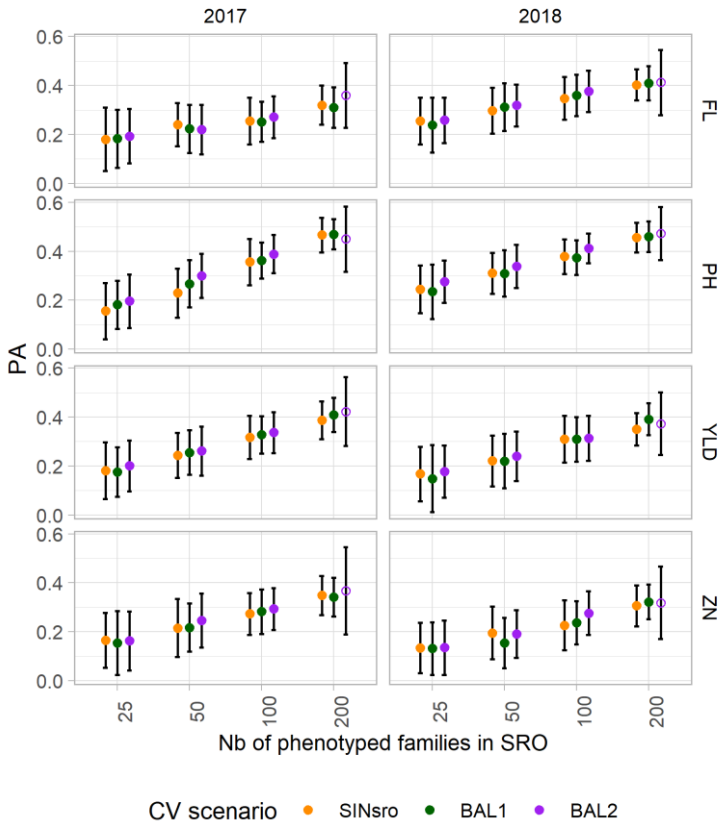


Figure 2-5: Mean predictive ability (PA) of the GBLUP model to predict phenotypes at Santa Rosa (SRO) for the three CV scenarios: single-site data in SRO (SINSRO) and two-site data with balanced information from the two sites (BAL1 and BAL2). The results for both years (2017 and 2018) and the four traits are presented. The bars represent the standard deviation. The open dot for is for the cross-validation obtained from only 34 genotypes.

The PA were improved by including the phenotypes of the whole population in PAL in the training set, and this was consistently observed for all traits, although to a different extent (Figure 2-6, STable 2-9). The largest differences in average PA were observed for FL (SIN_{SRO} = 0.29, IMB = 0.56) and PH (SIN_{SRO} = 0.33, IMB = 0.62). However, for both traits the increase of "s" did not yield a much higher PA with the IMB method. Average ZN predictions also benefited from PAL information, but less so (SIN_{SRO} = 0.24, IMB = 0.45). For those three traits, the average PA with the IMB method was rather close to the phenotypic correlation between the two sites (dotted line in Figure 2-6). Conversely, for YLD the average PA was similar between SIN_{SRO} (0.27) and IMB (0.34), with values above the indirect phenotypic prediction as represented by the site correlation. The partition of factor effects in the linear model revealed that the proportion of variance explained by the CV method depended on the traits (7% for YLD compared to ≥50% for all other traits) (Table 2-5). Only for YLD did the set size account for a large fraction (32%) of the explained PA variance. The contribution of the year effect to the total PA variance was low (≤1%) for YLD and ZN while still contributing to a small portion of the variance for FL and PH (10% and 6.5%, respectively). For both traits, average PA was higher in the $S_{0:3}$ 2018 trials.

Table 2-5: Analysis by trait of the factors influencing the variability of PA. The data are the PA for the CV scenarios comparing SIN_{SRO}, BAL1 and BAL2 or SIN_{SRO} and IMB. Eta^2 is the proportion of variance associated with each effect and R^2 is the coefficient of determination obtained from a linear model applied to the data from the 100 iterations (n=2400 for the model including SIN_{SRO}, BAL1 and BAL2 scenarios and n=1600 for the model including SIN_{SRO} and IMB scenarios)

		SIN_{SRO}/BAL1/BAL2		SIN_{SRO}/IMB	
Trait [a]	Factor [b]	Eta^2	R^2	Eta^2	R^2
FL	CV	0.003	0.342	0.619	0.792
	Year	0.116		0.104	
	Set size	0.215		0.020	
	CV:Year	0.000		0.010	
	CV:Set size	0.002		0.026	
	Year:Set size	0.003		0.000	
	CV:Year:Set size	0.003		0.000	
PH	CV	0.009	0.539	0.620	0.853
	Year	0.019		0.065	
	Set size	0.492		0.094	
	CV:Year	0.001		0.018	
	CV:Set size	0.004		0.050	
	Year:Set size	0.012		0.004	
	CV:Year:Set size	0.002		0.002	
YLD	CV	0.004	0.407	0.072	0.404
	Year	0.009		0.001	
	Set size	0.390		0.319	
	CV:Year	0.000		0.003	
	CV:Set size	0.003		0.006	
	Year:Set size	0.001		0.001	
	CV:Year:Set size	0.001		0.002	
ZN	CV	0.004	0.322	0.499	0.630
	Year	0.020		0.001	
	Set size	0.291		0.096	
	CV:Year	0.000		0.019	
	CV:Set size	0.003		0.015	
	Year:Set size	0.001		0.001	
	CV:Year:Set size	0.003		0.000	

[a] Traits: days to flowering (FL), plant height (PH), grain yield per plot (YLD), grain Zn concentration (ZN)

[b] Factors: CV: SIN_{SRO}, BAL1, BAL2 or IMB; Year: 2017 ($S_{0:2}$), 2018 ($S_{0:3}$); Set size: 25, 50, 100, 200

2.5 Discussion

2.5.1 Evaluation of early generation progenies

The training population with which we tested the various CV scenarios had the expected characteristics for applying GP, both in terms of marker density relative to the specific population LD and total absence of structure among the 334 S_0 genotypes (Appendix 1 and supplementary tables and figures).

Our progeny phenotyping method could not capture the within-line variations, as we recorded traits as the mean of the evaluated plot (FL, PH) or from the bulked harvested plot (YLD, ZN). For most combinations of traits and sites the difference in H^2 between the $S_{0:2}$ and $S_{0:3}$ progeny testing was limited and fell within the confidence interval of each other. However, the H^2 of the $S_{0:2}$ progenies was significantly higher for FL and YLD in SRO and was significantly lower for PH in PAL. This lack of consistency suggested that changes were driven more by environmental causes than by the degree of allelic fixation within the genetic material. This was supported by the temporal checks for which a significant year effect could be observed for all traits and sites, while no effect of the generation was observed. We concluded that the changes in mean between $S_{0:2}$ and $S_{0:3}$ within sites were essentially driven by the environment effect. As the phenotypic variance due to generation was minor compared to the variance associated with the year, generation advance did not seem to influence the PA. For

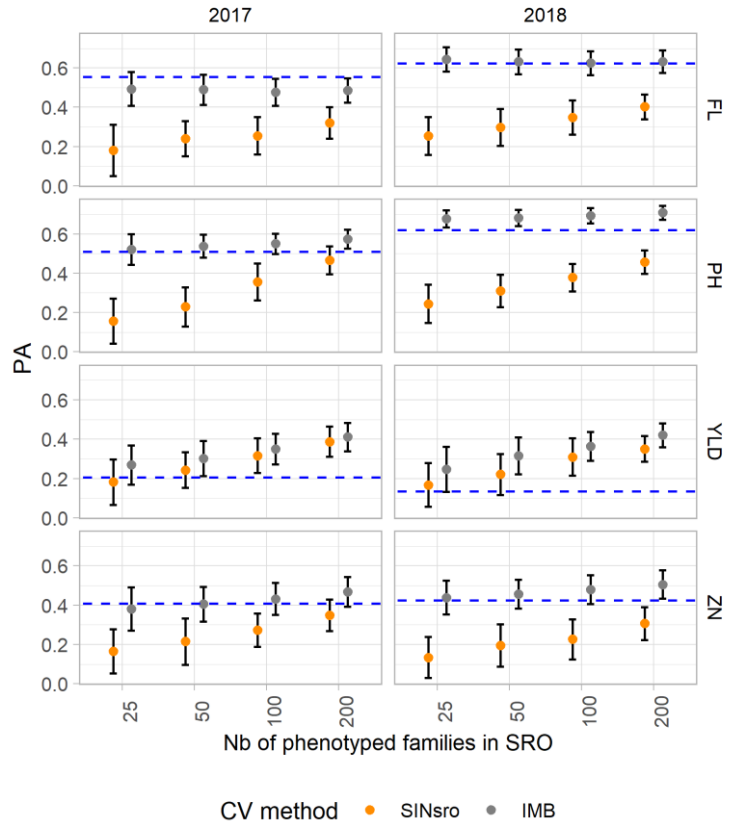


Figure 2-6: Mean predictive ability (PA) of the GBLUP model to predict phenotypes at Santa Rosa (SRO) for two CV scenarios: single-site data in SRO (SIN_{SRO}) and two-site data with complete information in Palmira and incomplete in target site SRO (IMB). The results for both years (2017 and 2018) and the four traits are presented. The bars represent the standard deviation. Dotted blue lines indicate the phenotypic correlation between sites.

time and economic reasons, calibration on $S_{0:2}$ phenotypes could thus be preferred as it allows a reduction of the breeding cycle length and cost.

2.5.2 Potential of early genomic prediction

We first tested GP models on the early generation phenotypes collected in a single environment. As expected, regardless of the generation, the four traits showed differences in mean PA. FL and PH were overall the best predicted traits, followed by ZN and YLD. This was fairly consistent with what is reported in the literature where FL and PH generally show high PA in absolute terms and relative to yield parameters (Combs and Bernardo 2013; Spindel *et al.* 2015; Ben Hassen *et al.* 2017; Ben Hassen *et al.* 2018). However, when comparing with another GP study performed on families derived from rice synthetic populations much higher PA for FL was achieved than in Grenier *et al.* (2015), where average PA for FL reached only a maximum of 0.29 for the population of 343 $S_{2:4}$ lines. Conversely, maximum PA for PH (0.46) was comparable to the PA obtained for the 343 $S_{2:4}$ (0.50) (Grenier *et al.* 2015), but lower than the PA obtained for the 174 $S_{1:3}$ (0.52) (Morais Júnior *et al.* 2018). The maximum PA for YLD (0.39) was slightly higher than the maximum reported for the rice diversity panel of 369 elite breeding lines evaluated in replicated yield trials (0.30) (Spindel *et al.* 2015), but lower than that reported for the 174 $S_{1:3}$ lines (0.44) (Morais Júnior *et al.* 2018), despite an H^2 for YLD that was higher in our study (H^2 = 0.58) than in the two others aforementioned (0.44 in $S_{1:3}$ lines and 0.32 in the diversity panel). Overall, for these commonly reported traits, the PA obtained in our study did not greatly differ

from what was reported for GP in rice diversity panels or synthetic populations (as reviewed in Ahmadi *et al.* 2020).

Although various studies on maize and spring wheat have proven the effectiveness of the GP-based approach for kernel zinc concentration, to our knowledge no study applying GP to rice for grain zinc concentration has yet been reported. Grain zinc concentration is a complex trait greatly influenced by soil and other associated factors (Jin *et al.* 2013; Hindu *et al.* 2018; Velu *et al.* 2018; Naik *et al.* 2020), so there are great hopes that GP will simplify the process of breeding rice for nutritional quality. On average, the PA for ZN in a single environment were low (0.26 and 0.24, for 2017 and 2018, respectively). However, the maximum PA in SIN_{SRO} reached 0.36 with 200 $S_{0:2}$ progenies (2017 data and RKHS model), which is comparable to the average estimated PA obtained with the 5-fold CV1 model applied to the HarvestPlus association mapping panel of 330 wheat lines (PA = 0.36) (Velu *et al.* 2018).

2.5.3 Effect of the GP methods on predictive ability

In the context of single-site analysis, we found that the two prediction methods, GBLUP and RHKS, induced some differences in PA only for FL. While GBLUP uses a linear kernel that models only the additive effects, RKHS uses a Gaussian kernel that carries the additive effects and the additive-additive epistatic effects at every possible order (Jiang and Reif 2015). RKHS has been reported to perform better than the linear model in the presence of epistasis (González-Camacho *et al.* 2012; Jiang and Reif 2015; Onogi *et al.* 2015). Epistasis has been reported in FL (Hori *et al.* 2016), PH (Yu *et al.* 2002; Shen *et al.* 2014), YLD (Luo *et al.* 2001; Xing *et al.* 2002) and ZN (Lu *et al.* 2008; Norton *et al.* 2010), however, both GP methods performed similarly for the traits we looked at in our population. The phenotypes we considered were all progeny means, which represent the breeding value or additive effect of our tested S_0 (Falconer 1960). Different and opposed epistatic effects can appear in the same family and have probably impeded RKHS from capturing them accurately. Limited differences between the two GP methods have also been reported in previous studies testing predictions for rice collections of fixed accessions (reviewed by Ahmadi *et al.* 2020) or $S_{1:3}$ lines extracted from synthetic populations (Morais Júnior *et al.* 2018). Given our phenotypes and considering the PA, GBLUP appeared as the most appropriate method in our context of population breeding considering single or two-site phenotyping data in our calibration models.

2.5.4 Prediction of the target environment using the two-site calibration model

Most of the contrasts in phenotypic records observed between the two sites were due in large part to the differences in crop establishment, soil conditions, climatic and biotic constraints as well as field management. Between irrigated and rainfed conditions, not only yield performance was expected to be affected by the environmental conditions, but also the grain zinc concentration, these two traits

showing lower correlation between sites. Under flooded conditions, the soil oxygen and redox potential will drop and trigger the formation of non-available zinc or its adsorption onto different compounds, depending on the soil type (reviewed in Rehman *et al.* 2012). As PAL is subject to continuous flooding, low zinc availability was expected and, consequently, observed ZN was much lower than in SRO.

Knowing the environment effect on the trait expressions and the phenotypic correlation between the sites, we tested the potential of GP including two sites with various CV schemes involving several factors. Of all the factors tested in the scenarios, the training set size had the most influence on PA. Training set size explained most of the differences observed for all the traits. The year of phenotyping was best in explaining the PA variations only for FL, which could be related to climatic differences and/or small changes in crop establishment date, which both are known to affect crop phenology. The CV methods SIN_{SRO}, BAL1 and BAL 2, accounted for a small portion of the variance explained by the models. In general, the two BAL scenarios showed a limited advantage over SIN_{SRO} for all traits. The prediction of unobserved genotypes for a specific environment using a two-site model was as precise as that obtained with a single-site model. Indeed, the prediction of gs_{iSRO} is based on the same amount of information as the g_k from a SIN_{SRO} calibration. For this reason, the two-site calibration could perform better only if g_k is more precise and relatively larger (larger associated variance) than gs_{iSRO}, but this is expected only for well-correlated environments. BAL1 and BAL2 differed in the number of genotypes repeated over the two sites. In BAL1, 100% of the genotypes included in the calibration had phenotypes in both sites (the overlapping proportion), whereas only 50% of the included genotypes had phenotypes of both sites in BAL2. The effect of the overlap proportion was tested by Jarquín *et al.* (2020) in a study that assessed the effects of data allocation on the PA of genomic-enabled prediction models. With their GxE model (M3 as presented in their article), the use of overlapping sets of genotypes improved the precision. In our case, the tendency was the reverse. Maintaining similar efforts in phenotyping in both sites while reducing the overlap (BAL1 and BAL2 with the same "s" progenies) resulted in higher precision in the predictions, but only for a specific case of PH_2017 with small training set size of 50 genotypes. For PH, exploring more of the population genetic variability within relatively small training set sizes might have had a greater impact thanks to a higher phenotypic correlation between sites.

While neither BAL1 nor BAL2 could greatly improve PA compared to SIN_{SRO}, calibrating with the whole population phenotyped in PAL and only a subset "s" of the population in SRO for predicting in SRO (IMB) generated substantial improvement of PA for all traits. The interest of this sparse testing method lies in borrowing information within lines across environments (Lopez-Cruz *et al.* 2015). However, if the phenotypes are not correlated between sites, benefits from the inclusion of both environments are expected to be low, as we found with YLD, where less improvement of PA was achieved through

IMB than for the other traits. Generally, sparse testing is in most cases more precise than the prediction of unobserved genotypes in known environments, regardless of the calibration method used (Burgueño *et al.* 2012; Jarquín *et al.* 2014; Lopez-Cruz *et al.* 2015; Ben Hassen *et al.* 2018; Millet *et al.* 2019). However, as the predicted lines must be observed in at least one environment, the burden on the phenotyping still remains, but the effort can lead to an increase in PA for traits with strong to moderate environment correlation, as was the case for FL, PH and ZN. For ZN, which has only a moderate site correlation, the IMB yielded a large gain in PA even with a drastic reduction of the phenotyping effort in SRO (from SIN_{SRO}_s25 = 0.14 for to IMB_s25 = 0.44 with the 2018 data). Overall, the sparse testing provided an improvement in the prediction of ZN in the rice synthetic population, with average PA (IMB_s200 = 0.51 with the 2018 data) in the range of those reported for spring wheat (Velu *et al.* 2016) and maize (Mageto *et al.* 2020).

2.5.5 Optimization of calibration procedure for GP

In our study, we tested the calibration of a GP model using phenotypic records gathered from early progeny testing in two sites. The potential of using two-site data and sparse testing for the model calibration, was considered as a satisfactory measure to predict most traits, even for YLD, despite a slightly reduced advantage compared to what was reported for the other traits.

We have demonstrated that the calibration using phenotypic data collected on progeny testing at two successive early generations could deliver relatively good and comparable PA. This opens up possibilities for rapid cycling RS, with recycling of parental lines from the genotyping of S_0 plants, based on the breeding value of the S_0. Yet, there is still a need to confirm that the models do predict well the performance of more advanced generations for inbred line development. Indeed, the units to derive in the pedigree breeding scheme should be selected on the basis of “varietal ability” (Gallais 1979), which is the expected value of all lines within a family at fixation. This will be explored in our next study, with an external validation of the GP models using a different set of S_0 progenies extracted from the PCT27 and brought to near fixation.

We are aware that the optimized scheme we suggest, based on random sampling of the training set, genome-wide markers considered as random effects, and random allocation of genotypes to sparse testing could be improved further still by considering other criteria known to increase the performance of GP. It remains to be seen whether PA can be improved by optimized assembly of the training set as performed in various studies (Rincent *et al.* 2012; Bustos-Korts *et al.* 2016; Rincent *et al.* 2017; Akdemir and Isidro-Sánchez 2019; Mangin *et al.* 2019), by inclusion of particular weights for some specific loci (Spindel *et al.* 2016; Bhandari *et al.* 2019; Frouin *et al.* 2019) or by use of an efficient method to proceed to sparse testing in the context of GxE models (Ahmadi *et al.* 2020).

Notwithstanding optimization of the calibration to develop efficient prediction models to fit our scheme, we ought also to consider the gain of applying GP-aided RS in our rice breeding program. So far, only the PA within generations has been tested, starting with the extraction of S_0 fertile plants of the C_n cycle. Prediction of S_0 in C_{n+1} would be done with calibration based on data from the previous cycle C_n. This has been tested through simulation (Müller *et al.* 2017; Ramasubramanian *et al.* 2020) and showed that the persistency of PA across cycles could be achieved with the accumulation of data from several past cycles. Simulation studies will be performed on our population to optimize the long-term use of GP-aided RS and define how and when it is best to upgrade the calibration model. The simulation will also offer the opportunity to improve the prediction and apply genomic selection while maintaining enough genetic diversity for further use of the population.

2.6 Data Availability

All supplementary tables, figures and the data used in this study are available at Figshare: https://figshare.com/s/4544ab2020c736dc9bb3.

2.7 Acknowledgments

The authors would like to thank all the scientists, field workers, lab assistants from Alliance Bioversity-CIAT who have contributed to the data collection. Special thanks go to Joe Tohme and Maria Fernanda Alvarez for their support. Additional thanks are due to the FLAR Grain Quality Laboratory, the HP-CIAT Nutritional Laboratory for grain quality evaluation and to Fedearroz for the access to field facilities at their research station in Santa Rosa. This work was supported by the CIRAD - UMR AGAP HPC Data Center of the South Green Bioinformatics platform (http://www.southgreen.fr/).

2.8 Fundings

This work was part of C.B.'s PhD study. The authors acknowledge the support from HarvestPlus, part of the CGIAR Research Program Agriculture for Nutrition and Health (A4NH), for co-funding the PhD scholarship and for providing the funds to carry out the field trial experiments, and the CGIAR Research Program RICE, for additional support in genotyping and other field-related activities.

2.9 Literature cited

Ahmadi, N., J. Bartholomé, C. Tuong-Vi, C. Grenier, 2020 Genomic selection in rice: empirical results and implications for breeding. In. *Quantitative Genetics, Genomics and Plant Breeding*, 2nd Edition. M. S. Kang editor. CABI, Wallingford, , pp 243-258. doi:10.1079/9781789240214.0243.

Akdemir, D., J. Isidro-Sánchez, 2019 Design of training populations for selective phenotyping in genomic prediction. Scientific Reports. 9(1). doi:10.1038/s41598-018-38081-6.

Allier, A., S. Teyssèdre, C. Lehermeier, A. Charcosset, L. Moreau, 2020a Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. Theor Appl Genet. doi:10.1007/s00122-019-03451-9.

Allier, A., S. Teyssèdre, C. Lehermeier, L. Moreau, A. Charcosset, 2020b Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics. 21(1). doi:10.1186/s12864-020-6756-0.

Ben Hassen, M., J. Bartholomé, G. Valè, T.-V. Cao, N. Ahmadi, 2018a Genomic Prediction Accounting for Genotype by Environment Interaction Offers an Effective Framework for Breeding Simultaneously for Adaptation to an Abiotic Stress and Performance Under Normal Cropping Conditions in Rice. G3. 8(7): 2319–2332. doi:10.1534/g3.118.200098.

Ben Hassen, M., T.-V. Cao, J. Bartholomé, G. Orasen, C. Colombi *et al.*, 2018b Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theor Appl Genet. 131(2): 417–435. doi:10.1007/s00122-017-3011-4.

Bernardo, R., J. Yu, 2007 Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science. 47(3): 1082–1090. doi:10.2135/cropsci2006.11.0690.

Bhandari, A., J. Bartholomé, T.-V. Cao, N. Kumari, J. Frouin *et al.*, 2019 Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLoS One. 14(5): e0208871. doi:10.1371/journal.pone.0208871.

Bian, Y., J. B. Holland, 2017 Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity. 118(6): 585–593. doi:10.1038/hdy.2017.4.

Burgueño, J., G. de los Campos, K. Weigel, J. Crossa, 2012 Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Science. 52(2): 707. doi:10.2135/cropsci2011.06.0299.

Bustos-Korts, D., M. Malosetti, S. Chapman, B. Biddulph, F. van Eeuwijk, 2016 Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. G3. 6(11): 3733–3747. doi:10.1534/g3.116.035410.

Butler, D. G., Cullis, B.R., A. R. Gilmour, Gogel, B.G. and Thompson, R. 2017. ASReml-R Reference Manual Version 3. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Châtel, M. H., Y. Ospina, F. Rodriguez, V. H. Lozano, 2005 CIRAD/CIAT Rice Project: Population Improvement and Obtaining Rice Lines for the Savannah Ecosystem, pp 237-254 in *Population improvement: A way of exploiting the rice genetic resources of Latin America.* FAO. Rome.

Châtel, M. H., E. P. Guimarães, 1997 Recurrent selection in rice, using a male-sterile gene. CIAT, Colombia.

Combs, E., R. Bernardo, 2013 Accuracy of Genomewide Selection for Different Traits with Constant Population Size, Heritability, and Number of Markers. The Plant Genome. 6(1). doi:10.3835/plantgenome2012.11.0030.

Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez *et al.*, 2016 Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. The Plant Genome. 9(3). doi:10.3835/plantgenome2016.03.0024.

Cuevas-Pérez, F. E., E. P. Guimarães, L. E. Berrio Orozco, D. I. González, 1992 Genetic base of irrigated rice in Latin America and the Caribbean 1971 to 1989. Crop Science. 32(4): 1054–1059.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett *et al.,* 2014 NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size from genetic data. Molecular Ecology Resources. 14(1): 209–214. doi:10.1111/1755-0998.12157.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6(5), e19379. doi: 10.1371/journal.pone.0019379

Falconer DS. 1960 Introduction to quantitative genetics. Oliver and Boyd. Edinburgh/London.

Frichot, E., O. François, 2015 LEA: An R-package for landscape and ecological association studies. Methods Ecol Evol 6, 925–929.

Frouin, J., D. Filloux, J. Taillebois, C. Grenier, F. Montes *et al.*, 2014 Positional cloning of the rice male sterility gene ms-IR36, widely used in the inter-crossing phase of recurrent selection schemes. Molecular Breeding. 33(3): 555–567. doi:10.1007/s11032-013-9972-3.

Frouin, J., A. Labeyrie, A. Boisnard, G. A. Sacchi, G. A. and N. Ahmadi, 2019 Genomic prediction offers the most effective marker assisted breeding approach for ability to prevent arsenic accumulation in rice grains. PLoS One 14, e0217516–22.

Gallais A. 1979 The concept of varietal ability in plant breeding. Euphytica. 28(3): 811–823. doi:10.1007/BF00038955.

Gaynor, R. C., G. Gorjanc, A. R. Bentley, E. S. Ober, P. Howell *et al.*, 2017 A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. Crop Science. 57(5): 2372–2386. doi:10.2135/cropsci2016.09.0742.

Gianola, D., J. B. C. H. M. van Kaam, 2008 Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics. 178(4): 2289–2303. doi:10.1534/genetics.107.084285.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One. 9(2): e90346. doi:10.1371/journal.pone.0090346.

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. E. Cairns *et al.*, 2012 Genome-enabled prediction of genetic values using radial basis function neural networks. Theor Appl Genet. 125(4): 759–771. doi:10.1007/s00122-012-1868-9.

Gorjanc, G., R. C. Gaynor, J. M. Hickey, 2018 Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor Appl Genet. 131(9): 1953–1966. doi:10.1007/s00122-018-3125-3.

Grenier, C., T.-V. Cao, Y. Ospina, C. Quintero, M. H. Châtel *et al.*, 2015 Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. PLoS One. 10(8): e0136594. doi:10.1371/journal.pone.0136594.

Guimarães, E. P., J. Borrero, Y. Ospina, 1996 Genetic diversity of upland rice germplasm distributed in Latin America. Pesquisa Agropecuária Brasileira. 31(3): 187–194.

Guimarães, E. P., 2005 Population improvement: A way of exploiting the rice genetic resources of Latin America. Guimarães EP (Ed.) Rome.

Guo, Z., D. M. Tucker, J. Lu, V. Kishore, G. Gay, 2011 Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet. 124:261–275. doi: 10.1007/s00122-011-1702-9.

Heffner, E. L., J.-L. Jannink, M. E. Sorrells, 2011 Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. The Plant Genome. 4(1): 65–75. doi:10.3835/plantgenome2010.12.0029.

Hill, W. G., 1981 Estimation of effective population size from data on linkage disequilibrium. Genet Res. 38(3): 209–216. doi:10.1017/S0016672300020553.

Hindu, V., N. Palacios-Rojas, R. Babu, W. B. Suwarno, Z. Rashid *et al*,. 2018 Identification and validation of genomic regions influencing kernel zinc and iron in maize. Theor Appl Genet. 131(7): 1443–1457. doi:10.1007/s00122-018-3089-3.

Holland, J. B., W. E. Nyquist, C. T. Cervantes-Martinez, 2003. Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:9-111.

Hori, K., K. Matsubara, M. Yano, 2016 Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. Theor Appl Genet. 129(12): 2241–2252. doi:10.1007/s00122-016-2773-4.

IRRI. 2013 Standard Evaluation System (SES) for Rice. 5th Edition.

Isik F., J. B. Holland, C. Maltecca, 2017 Genetic Data Analysis for Plant and Animal Breeding. Cham (Switzerland): Springer International Publishing. doi:10.1007/978-3-319-55177-7.

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt *et al*., 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet. 127(3): 595–607. doi:10.1007/s00122-013-2243-1.

Jarquín, D., R. Howard, J. Crossa, Y. Beyene, M. Gowda *et al*., 2020 Genomic Prediction Enhanced Sparse Testing for Multi-environment Trials. G3. 10(8): 2725–2739. doi:10.1534/g3.120.401349.

Jiang, Y,. J. C. Reif, 2015 Modeling Epistasis in Genomic Selection. Genetics. 201(2): 759–768. doi:10.1534/genetics.115.177907.

Jin, T., J. Zhou, J. Chen, L. Zhu, Y. Zhao and Y. Huang, 2013 The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. Breeding Science. 63(3): 317–324. doi:10.1270/jsbbs.63.317.

Kawahara, Y., M. de la Bastide, J. P. Hamilton, H. Kanamori, W. R. McCombie *et al*., 2013 Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice. 6(1). doi:10.1186/1939-8433-6-4.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland *et al*., 2015 Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3. 5(4): 569–582. doi:10.1534/g3.114.016097.

Lu, K., L. Li, X. Zheng, Z. Zhang, T. Mou and Z. Hu, 2008 Quantitative trait loci controlling Cu, Ca, Zn, Mn and Fe content in rice grains. Journal of Genetics. 87(3): 305–310. doi:10.1007/s12041-008-0049-8.

Luo, L. J., J. W. Stansel, G. S. Khush, A. H. Paterson, 2001 Overdominant Epistatic Loci Are the Primary Genetic Basis of Inbreeding Depression and Heterosis in Rice. II. Grain Yield Components. Genetics. 158: 1755–1771.

Mageto, E. K., J. Crossa, P. Pérez-Rodríguez, T. Dhliwayo, N. Palacios-Rojas *et al*., 2020 Genomic Prediction with Genotype by Environment Interaction Analysis for Kernel Zinc Concentration in Tropical Maize Germplasm. G3. 10(8): 2629-2639. doi: 10.1534/g3.120.401172.

Mangin, B., R. Rincent, C.-E. Rabier, L. Moreau, E. Goudemand-Duguem, 2019 Training set optimization of genomic prediction by means of EthAcc. PLoS ONE 14, e0205629–21.

Martínez, C. P., E. A. Torres, M. H. Châtel, G. Mosquera, J. Duitama *et al*., 2014 Rice Breeding in Latin America pp 187-278 in *Plant Breeding Reviews: Volume 38,* edited by Jannink JL. John Wiley & Sons, Hoboken New Jersey. doi: 10.1002/9781118916865.ch05.

Meuwissen, T. H., B. J. Hayes, M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4): 1819–1829.

Millet, E. J., W. Kruijer, A. Coupel-Ledru, S. Alvarez Prado, L. Cabrera-Bosquet *et al*., 2019 Genomic prediction of maize yield across European environmental conditions. Nature Genetics. 51(6): 952–956. doi:10.1038/s41588-019-0414-y.

Morais Júnior, O. P., F. Breseghello, J. B. Duarte, O. P. Morais, P. H. N. Rangel *et al.*, 2017 Effectiveness of Recurrent Selection in Irrigated Rice Breeding. Crop Science 57, 3043–3058.

Morais Júnior, O. P., F. Breseghello, J. B. Duarte, A. S. G. Coelho, T. C. O. Borba *et al.*, 2018 Assessing Prediction Models for Different Traits in a Rice Population Derived from a Recurrent Selection Program. Crop Science. 58(6): 2347. doi:10.2135/cropsci2018.02.0087.

Müller, D., P. Schopp, A. E. Melchinger, 2017 Persistency of Prediction Accuracy and Genetic Gain in Synthetic Populations Under Recurrent Genomic Selection. G3. 7(3): 801–811. doi:10.1534/g3.116.036582.

Müller, D., P. Schopp, A. E. Melchinger, 2018 Selection on Expected Maximum Haploid Breeding Values Can Increase Genetic Gain in Recurrent Genomic Selection. G3. 8(4): 1173–1181. doi:10.1534/g3.118.200091.

Naik, S. M., A. K. Raman, M. Nagamallika, C. Venkateshwarlu, S. P. Singh *et al*., 2020 Genotype × environment interactions for grain iron and zinc content in rice. J Sci Food Agric. 100(11): 4150–4164. doi:10.1002/jsfa.10454.

Norton, G. J., C. M. Deacon, L. Xiong, S. Huang, A. A. Meharg and A. H. Price, 2010 Genetic mapping of the rice ionome in leaves and grain: identification of QTLs for 17 elements including arsenic, cadmium, iron and selenium. Plant and Soil. 329(1–2): 139–153. doi:10.1007/s11104-009-0141-8.

Onogi, A., O. Ideta, Y. Inoshita, K. Ebana, T. Yoshioka *et al*., 2015 Exploring the areas of applicability of whole-genome prediction methods for Asian rice (Oryza sativa L.). Theor Appl Genet. 128(1): 41–53. doi:10.1007/s00122-014-2411-y.

Pérez-Elizalde, S., J. Cuevas, P. Pérez-Rodríguez, J. Crossa, 2015 Selection of the Bandwidth Parameter in a Bayesian Kernel Regression Model for Genomic-Enabled Prediction. JABES. 20(4): 512–532. doi:10.1007/s13253-015-0229-y.

Perrier, X., J.-P. Jacquemoud-Collet, 2006 DARwin software. http://darwin.cirad.fr.

Purcell, S. B Neale, K. Todd-Brown, L Thomas, M. A. R. Ferreira *et al.*, 2007 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559–575.

R Core Team. 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramasubramanian, V., W. D. Beavis, 2020 Factors affecting Response to Recurrent Genomic Selection in Soybeans. bioRxiv. doi: https://doi.org/10.1101/2020.02.14.949008 (preprint posted February 14, 2020).

Rehman, H., T. Aziz, M. Farooq, A. Wakeel, Z. Rengel, 2012 Zinc nutrition in rice production systems: a review. Plant Soil. 361(1–2): 203–226. doi:10.1007/s11104-012-1346-9.

Rincent, R., A. Charcosset, L. Moreau, 2017 Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. Theor Appl Genet. 130(11): 2231–2247. doi:10.1007/s00122-017-2956-7.

Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel *et al*., 2012 Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). Genetics. 192(2): 715–728. doi:10.1534/genetics.112.141473.

Risterucci, A. M., L. Grivet, J. A. K. N'Goran, I. Pieretti, M. H. Flament and C. Lanaud, 2000 A high-density linkage map of Theobroma cacao L. Theor Appl Genet. 101(5): 948–955. doi:10.1007/s001220051566.

Schopp, P., D. Müller, F. Technow, A. E. Melchinger, 2017 Accuracy of Genomic Prediction in Synthetic Populations Depending on the Number of Parents, Relatedness, and Ancestral Linkage Disequilibrium. Genetics. 205(1): 441–454. doi:10.1534/genetics.116.193243.

Shen, G., W. Zhan, H. Chen, Y. Xing, 2014 Dominance and epistasis are the main contributors to heterosis for plant height in rice. Plant Science. 215–216: 11–18. doi:10.1016/j.plantsci.2013.10.004.

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard *et al*., 2016 Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity. 116: 395-408. doi:10.1038/hdy.2015.113;

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard *et al*., 2015 Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. Mauricio R, editor. PLoS Genet. 11(2): e1004982. doi:10.1371/journal.pgen.1004982.

Swarts, K., H. Li, J. A. Romero Navarro, D. An, M. C. Romay *et al* 2014 Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome. 7(3). doi: 10.3835/plantgenome2014.05.0023

Taillebois, J., E. P. Guimarães, 1989 CNA-IRAT 5 upland rice population. https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1085653/1/IRRN19890001.pdf.

VanRaden, P. M., 2008 Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science. 91(11): 4414–4423. doi:10.3168/jds.2007-0980.

Velu, G., J. Crossa, R. P. Singh, Y. Hao, S. Dreisigacker *et al.* 2016 Genomic prediction for grain zinc and iron concentrations in spring wheat. Theor Appl Genet. 129:1595–605. doi:10.1007/s00122-016-2726-y.

Velu, G., R. P. Singh, L. Crespo-Herrera, P. Juliana, S. Dreisigacker *et al*., 2018 Genetic dissection of grain zinc concentration in spring wheat for mainstreaming biofortification in CIMMYT wheat breeding. Scientific Reports. 8(1). doi:10.1038/s41598-018-31951-z

Waples, R. S., 2006 A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv Genet. 7(2): 167–184. doi:10.1007/s10592-005-9100-y.

Wimmer, V., T. Albrecht, H.-J. Auinger, C.-C. Schoen, 2012 synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics. 28(15): 2086–2087.

Xing, Y., Y. Tan, J. Hua, X. Sun, C. Xu and Q. Zhang, 2002 Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet. 105(2): 248–257. doi:10.1007/s00122-002-0952-y.

Yu, S. B., J. X Li, C. G. Xu, Y. F. Tan, X. H. Li and Q. Zhang, 2002 Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet. 104(4): 619–625. doi:10.1007/s00122-001-0772-5.

Zhang, A., H. Wang, Y. Beyene, K. Semagn, Y. Liu *et al*., 2017 Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. Front Plant Sci. 8. doi:10.3389/fpls.2017.01916.

2.10 Appendix

2.10.1 Appendix 1

2.10.1.1 Genotyping-by-sequencing ("GBS") and data treatment

DNA libraries were prepared at the Regional Genotyping Technology Platform (http://www.gptr-lr-genotypage.com) hosted at Cirad, Montpellier, France. For 949 S_0 plants extracted from the PCT27, including the 334 considered in the training set, genomic DNA was extracted from the leaf tissues of a single S_0 plant grown in PAL, using a MATAB lysis buffer (Risterucci *et al.* 2000) and purified using the NucleoMag C-Beads protocol from Macherey-Nagel. Each DNA sample was diluted to 20 ng/µL and 150 ng was digested separately with two restriction enzymes PstI and MseI. DNA libraries were then single-end sequenced in a single-flow cell channel (i.e. 96-plex sequencing) using an Illumina HiSeq2000 (Illumina, Inc.) at the Regional Genotyping Platform (http://get.genotoul.fr/) hosted at INRA, Toulouse, France.

The fastq sequences were aligned to the rice reference genome, Os-Nipponbare-Reference-IRGSP-1.0 (Kawahara *et al.* 2013) using Bowtie2 with the default parameters (option very sensitive). Non-aligning sequences and sequences with multiple positions were discarded. Single-nucleotide polymorphism (SNP) calling was performed using the Tassel GBS pipeline v5.2.29 (Glaubitz *et al.* 2014). The filters applied to loci are the missing data (<20%), the depth for each data point (>10), the minor allele frequency (>2.5%) and the bi-allelic status of SNPs. To limit the probability of under-calling a heterozygous site, the read depth for SNP calling was set to a minimum of 10, so that the probability of under-calling a heterozygous site was limited to a theoretical maximum of 0.2% (Swarts *et al.* 2014). Missing data were imputed using Beagle 4.1 embedded in the R package Synbreed v0.11-22 (Wimmer *et al.* 2012).

After quality control, 9,928 SNPs remained for the genetic characterization of the training set and the genomic prediction step. All following analyses were thus performed on the 334 S_0 plants, the latter used in the GP models. Graphical representation of the SNP distribution across the 12 chromosomes was performed using the Synbreed package (Wimmer *et al.* 2012). LD was calculated by computing the

pairwise LD measure r^2 as in (Hill and Robertson 1968) with PLINK1.09 using every pair of variants within a 50 variants window (Purcell *et al.* 2007). Non-linear regression modeling was performed using the *nls* function in the statistical package R v3.3.0 (R-Core Team 2017) to represent the LD on each chromosome. The effective population size was computed using the linkage disequilibrium method (Hill 1981; Waples 2006) with the software NeEstimator V2.01 (Do *et al.* 2014). Inference of population structure was performed using the *snmf* function from the R package LEA (Frichot *et al.* 2015). Population structure was graphically investigated by first computing Euclidean distances between the genotypes and then building a neighbor joining tree. The computation and graphical representation were done with DARwin V6.0.021 (Perrier and Jacquemoud-Collet 2006).

2.10.1.2 Genetic characterization of the population

The 9,928 SNP markers were fairly well distributed among the 12 rice chromosomes (Figure S1), with an average marker density of one SNP every 40 kb ranging from 27.3 kb to 64 kb (Table S1). For half the markers, the average distance between the nearest neighbours was 9.9 kb, ranging from 3.1 to 15.5 kb according to the chromosomes. The distribution of MAF in the population (Table S1, Figure S2) followed a beta distribution with beta = 5.45 and alpha = 1.37 showing a great proportion of less frequent alleles. Half the loci had an MAF below 15.7%. Across the whole genome, the average heterozygosity per locus was 30%, with loci having a minimum of 2.7 to a maximum of 100% heterozygous genotypes (Table S1). The 334 S_0 genotypes were either relatively fixed (0.08% of heterozygous loci) or fairly heterozygous (41% of heterozygous loci), and half the population was heterozygous for at least 29% of the loci (Table S1). The effective population size of the PCT27 measured based on the LD among the 334 S_0 was N_e = 40. Pairwise LD in the population across the 12 chromosomes was rather large with an average r^2 of 0.59 for distance between 0 and 25kb (Table S2). The LD decreased to 50% of its initial value at a slow rate (300 to 400 kb) (Figure S3). No structure was found in the population, as illustrated by the neighbour joining grouping based on similarity distances (Figure S4).

2.10.2 Appendix 2

2.10.2.1 Analysis of the year and generation effect

For the 334 S_0 families of the PCT27 used in this work, the generation $S_{0:2}$ was phenotyped in 2017 and the generation $S_{0:3}$ in 2018. To measure the year effect disconnected from the generation effect, 50 families (temporal checks) from the same PCT27 were observed in 2017 and 2018 as generation $S_{0:2}$. Similarly, to evaluate the extent of the generation effect that would result from inbreeding, the same 50 temporal checks were observed as $S_{0:2}$ and $S_{0:3}$ in 2018. This was done for all traits. An alpha-lattice design with eight unbalanced blocks and three replicates was used for each trial.

To reduce the block effect resulting from the sampling of temporal checks, each block was enhanced with two spatial checks (SC), one plot of IR64 (*indica* mega variety) and one plot of L23 (tropical japonica inbred line from the CIAT-Cirad upland breeding program) and so centred the block value on the SC mean value.

To assess the year effect, the following mixed model was applied by site to the data of the two years for the 50 temporal checks at generation $S_{0:2}$ and the two spatial checks.

$Y_{\{ijklm\}} = \mu + y_{\{i\}} + r(y)_{\{ij\}} + SC(y)_{\{ik\}} + b\big(r(y)\big)_{\{ijl\}} + g_{\{m\}} + yg_{\{im\}} + \varepsilon_{\{ijklm\}}$, the fixed part of the model was composed of the intercept μ, the year effect y, the replicate effect r and the SC variable, which discriminates the two spatial checks from each other and from the PCT27 lines $k = \{PCT27, IR64, L23\}$. The random part was composed of the line (genotype) effect g with distribution $g \sim \big(0, I\sigma_g^2\big)$, the line by year interaction gy with distribution $yg \sim \big(0, I\sigma_{yg}^2\big)$ and the error ε with distribution $\varepsilon \sim (0, I\sigma_\varepsilon^2)$.To assess the generation effect, the following model was applied to the data of the 50 temporal checks at generation $S_{0:2}$ and $S_{0:3}$ and the two spatial checks in 2018.

$Y_{\{ijklm\}} = \mu + r_{\{i\}} + b(r)_{\{ij\}} + SC(G)_{\{kl\}} + g_{\{m\}} + Gg_{\{jm\}} + \varepsilon_{\{ijklm\}}$

The parameter annotation was the same as for the analysis of the year effect, with additionally G as the fixed effect of the generation and Gg as the random interaction between the line and the generation with distribution $Gg \sim \big(0, I\sigma_{Gg}^2\big)$.

The results can be seen in supplementary Table S4.

2.10.3 Appendix 3

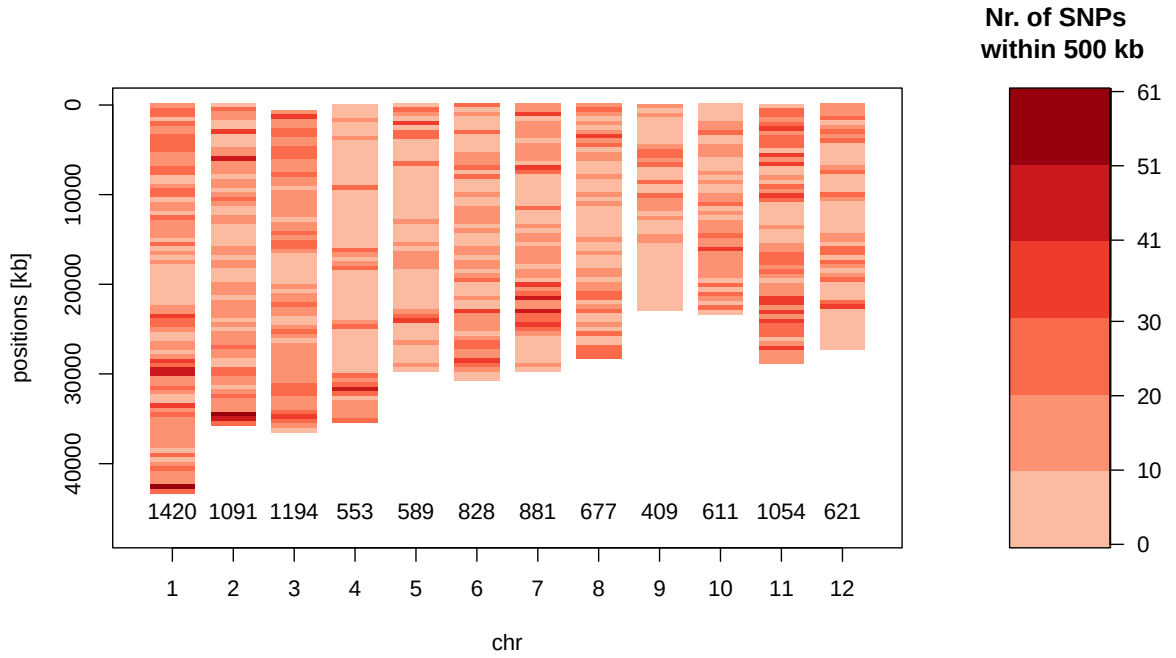
2.10.3.1 Phenotypic data preparation

In PAL, where the number of plants was known, any single plot with less than 14 established plants was identified and the data was removed if it strongly differed from the other two replicates. After this first round of cleaning, a mixed model was applied to each trial separately with the intention to discard the plot phenotypic scores of progenies with inconsistent records among the repetitions within a site, either as a result of poor crop establishment or unexpected problem on the specific plot. Within each trial, the plots with a residual in absolute value at more than four standard residuals were labelled as outliers and removed. The model used to remove outliers was formulated as follows and was the same for each trial:

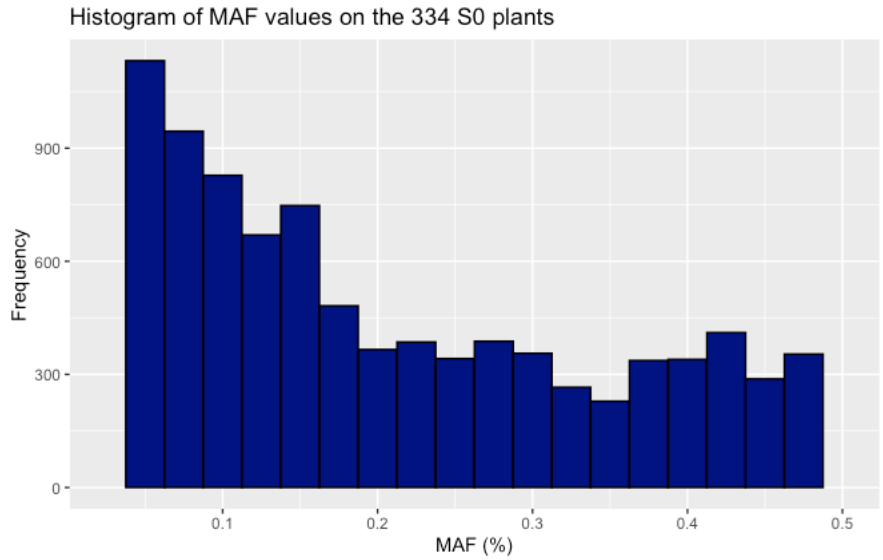
$Y_{ijk} = \mu + r_i + b(r)_{ij} + g_k + \varepsilon_{ijk}$

where μ is the general intercept, r is the fixed effect for the replicates, b is a random block effect nested in r with distribution $b \sim \mathrm{N}\big(0, I\sigma_b^2\big)$, g is the random genotype effect with distribution $g \sim N\big(0, I\sigma_g^2\big)$, and ε is the random error term with distribution $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$. This model, as well as all the following mixed models was fitted using ASReml-R v3.0 (Butler *et al.* 2007).

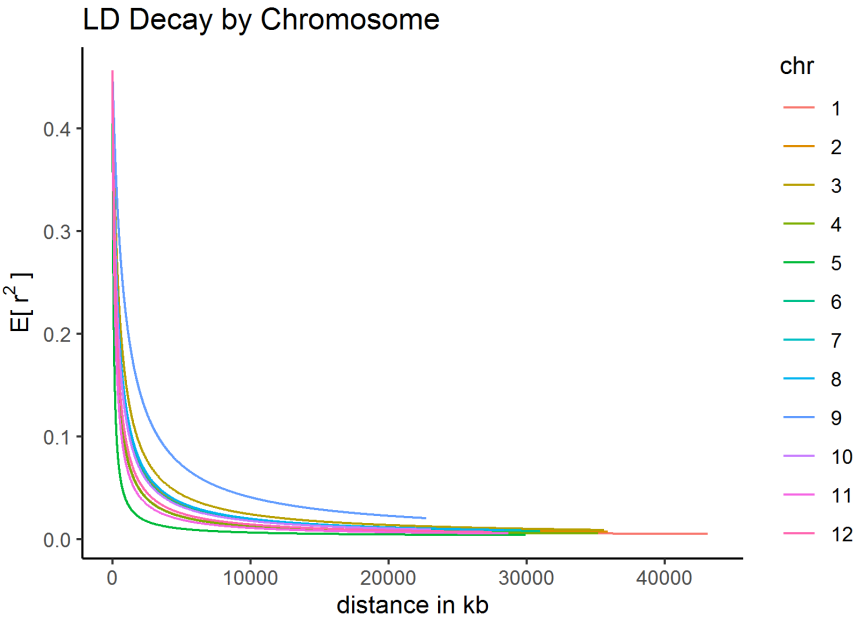
2.11 Supplementary Figures



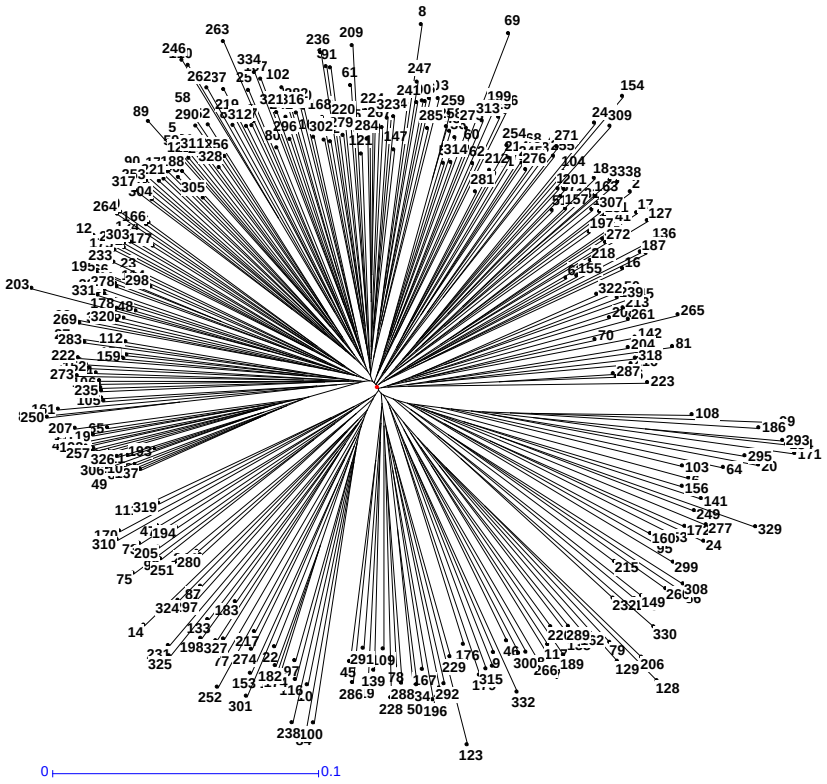
SFig 2-1: Density of SNP markers in the calibration set (334 S_0 plants) in the 12-chromosome R package Synbreed (Wimmer et al., 2012)



SFig 2-2: MAF distribution among the population of 334 S_0 plants.



SFig 2-3: Linkage disequilibrium, measured as r^2 between all pairs of markers considered in a 50 variants window (PLINK1.09). R^2 values plotted against distances between markers in kb as a nonlinear regression model based on Hill and Weir's (1988) equation.



SFig 2-4: Neighbor joining tree of Euclidean distance between the 334 S_0 plants (Darwin)

2.12 Supplementary Tables

STable 2-1: Genetic characterization of the population. Summary information on the distribution, MAF and heterozygosity of the 9 928 SNP loci

A		The rice 12 chromosomes												
	Chr	Os01	Os02	Os03	Os04	Os05	Os06	Os07	Os08	Os09	Os10	Os11	Os12	All
	Size (bp)	43 179 539	35 891 034	36 403 807	35 391 250	29 954 718	31 000 883	29 653 386	28 382 303	22 879 956	23 160 871	28 806 417	27 498 173	372 202 337
	Number	1 420	1 091	1 194	553	589	828	881	677	409	611	1 054	621	9 928
	Density	30 408	32 897	30 489	63 999	50 857	37 441	33 659	41 924	55 941	37 906	27 331	44 280	40 594
Distribution of distances (bp) between two adjacent SNP loci	Minimum	1	1	1	1	1	1	1	1	1	1	1	1	1
	1st Quartile	32	34	40	19	16	17	18	32	17	27	12	21	23
	Median	10 143	11 494	10 538	8 953	6 676	7 492	7 123	5 476	9 666	12 314	3 094	10 400	9 082
	Average	30 408	32 897	30 489	63 999	50 857	37 441	33 659	41 924	55 941	37 906	27 331	44 280	37 490
	3rd Quartile	35 317	38 327	39 304	48 947	39 108	38 980	36 728	49 010	64 973	48 416	33 443	53 539	41 156
	Maximum	2 197 892	711 533	794 696	1 951 759	2 209 631	1 196 222	1 499 537	773 794	1 188 875	677 358	528 305	792 342	2 209 631
Distribution of the minor allele frequency (MAF)	Minimum	0.021	0.022	0.017	0.02	0.021	0.018	0.022	0.022	0.022	0.021	0.02	0.021	0.017
	1st Quartile	0.066	0.069	0.069	0.086	0.057	0.104	0.051	0.132	0.088	0.08	0.074	0.087	0.075
	Median	0.131	0.162	0.162	0.193	0.113	0.143	0.153	0.235	0.16	0.124	0.183	0.148	0.157
	Average	0.183	0.203	0.198	0.225	0.168	0.202	0.206	0.247	0.185	0.171	0.219	0.201	0.201
	3rd Quartile	0.268	0.317	0.283	0.392	0.263	0.288	0.343	0.36	0.243	0.231	0.35	0.296	0.307
	Maximum	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.497	0.499	0.5	0.5	0.5	0.5
Distribution of heterozygosity	Minimum	4.20%	3.70%	3.40%	2.90%	4.20%	3.40%	4.20%	3.70%	4.20%	4.20%	4.00%	2.90%	2.90%
	1st Quartile	12.10%	13.20%	12.50%	15.30%	10.80%	17.90%	9.50%	22.40%	17.20%	15.00%	13.50%	16.40%	14.00%
	Median	22.20%	27.20%	26.90%	29.80%	21.90%	25.10%	25.30%	36.10%	26.60%	21.90%	29.00%	25.30%	25.90%
	Average	26.80%	30.70%	28.10%	34.00%	28.10%	29.20%	28.90%	34.30%	30.80%	26.10%	30.50%	29.60%	29.50%
	3rd Quartile	38.80%	42.30%	40.40%	46.40%	40.10%	41.70%	45.60%	45.40%	33.80%	36.50%	43.80%	40.10%	42.00%
	Maximum	98.70%	100.00%	100.00%	100.00%	98.90%	99.50%	99.50%	97.60%	98.90%	99.70%	100.00%	95.30%	100.00%

STable 2-2: Genetic characterization of the population: Observed heterozygosity (Ho) among the 334 genotypes

	Ho	Bin of Ho	Frequency	Percentage	Cumulative percentage
Min.	0.08743	[0.08 - 0.118]	3	0.9	0.9
1st Qu.	0.27065	[0.118 - 0.156]	4	1.2	2.1
Median	0.29296	[0.156 - 0.194]	4	1.2	3.3
Mean	0.29466	[0.194 - 0.232]	7	2.1	5.4
3rd Qu.	0.32386	[0.232 - 0.27]	61	18.3	23.7
Max.	0.41015	[0.27 - 0.308]	126	37.7	61.4
		[0.308 - 0.346]	90	26.9	88.3
		[0.346 - 0.384]	32	9.6	97.9
		[0.384 - 0.422]	7	2.1	100

STable 2-3: Summary of the cross-validation (CV) procedures used in the study. Calibration set represents the number of lines in the respective sites (PAL and SRO) and is not the total number of lines representing either PAL or SRO or both sites. Validation set represents the number of lines with no phenotype in SRO considered for the validation

CV method	Set size (s)	Calibration set			Validation Set	
		PAL	SRO	ntot		
SINsr	25	0	25	25	100	
	50	0	50	50	100	
	100	0	100	100	100	
	200	0	200	200	100	
BAL1	25	25	25	25	100	
	50	50	50	50	100	validation on lines with no PAL
	100	100	100	100	100	observation
	200	200	200	200	100	
BAL2	25	25	25	37.5	100	
	50	50	50	75	100	
	100	100	100	150	100	
	200	200	200	300	34	
IMB	25	334	25	334	100	
	50	334	50	334	100	validation on lines with PAL
	100	334	100	334	100	observations
	200	334	200	334	100	

STable 2-4: Average linkage disequilibrium (r2) between marker pairs according to chromosomes and the distance between markers, considering loci with MAF >2.5%. In italics are r2 with values less than initial r2/2

	The rice 12 chromosomes													
Distance range (kb) between markers	**Os01**	**Os02**	**Os03**	**Os04**	**Os05**	**Os06**	**Os07**	**Os08**	**Os09**	**Os10**	**Os11**	**Os12**	**Average**	**std**
]0:25]	0.610	0.620	0.674	0.581	0.338	0.608	0.632	0.645	0.744	0.541	0.510	0.580	0.590	0.100
]25:50]	0.448	0.526	0.583	0.530	0.273	0.391	0.485	0.529	0.586	0.422	0.349	0.513	0.470	0.096
]50:75]	0.458	0.515	0.567	0.366	0.378	0.470	0.476	0.451	0.450	0.440	0.333	0.426	0.444	0.064
]75:100]	0.406	0.464	0.591	0.351	0.265	0.433	0.471	0.454	0.446	0.428	0.332	0.466	0.426	0.082
]100:150]	0.387	0.451	0.505	0.333	0.394	0.433	0.405	0.450	0.433	0.429	0.301	0.428	0.412	0.055
]150:200]	0.380	0.435	0.478	0.310	0.480	0.457	0.421	0.448	0.450	0.392	0.294	0.357	0.409	0.063
]200:250]	0.319	0.386	0.451	0.338	0.363	0.400	0.358	0.429	0.394	0.393	0.273	0.382	0.374	0.048
]250:300]	*0.303*	0.390	0.437	0.331	0.335	0.372	0.384	0.360	*0.360*	0.385	*0.249*	0.332	0.353	0.048
]300:400]	*0.269*	0.347	0.392	*0.270*	0.326	0.334	0.337	0.338	0.379	0.299	*0.218*	0.304	0.318	0.049
]400:500]	*0.231*	*0.305*	0.380	*0.258*	0.302	0.314	0.331	0.323	*0.332*	0.290	*0.206*	*0.224*	*0.291*	0.052
]500:750]	*0.198*	*0.265*	*0.327*	*0.256*	0.239	*0.263*	*0.279*	*0.294*	*0.365*	*0.245*	*0.186*	*0.230*	*0.262*	0.050
]750:1000]	*0.165*	*0.245*	*0.248*	*0.197*	*0.111*	*0.197*	*0.182*	*0.253*	*0.355*	*0.194*	*0.168*	*0.154*	*0.206*	0.063

STable 2-5: Fixed effect and variance decomposition for 50 Temporal Checks randomly distributed across the design within each repetition, considering A) 50 $S_{0:2}$ lines in the two sites in 2017 and 2018 trials, following the model y = mu + year + rep:year + bloc:rep:year + genotype + genotype:year + error and B) 50 $S_{0:2}$ and 50 $S_{0:3}$ lines in the two sites in the 2018 trials, following the model y = mu + rep + Gen:rep + genotype + genotype:generation + error. The p-values for the fixed year effect are obtained by the Wald test and the p-values for the random effect by the likelihood ratio test

A

		Year effect[a]		Variance decomposition[b]				
Trait	Site	2018-2017	p-value	G	p-value	GxY	p-value	(GxY)/G
FL	PAL	-2.545	**	9.54	***	1.64	ns	0.17
	SRO	2.908	***	28.38	***	9.70	ns	0.34
PH	PAL	-4.27	***	30.67	***	8.76	ns	0.29
	SRO	-7.222	***	14.87	***	2.55	ns	0.17
YLD	PAL	457.519	***	2050.98	***	0.00	ns	0.00
	SRO	-75.879	***	2295.96	***	196.30	ns	0.09
ZN	PAL	0.145	**	1.68	***	0.00	ns	0.00
	SRO	2.231	***	3.82	***	1.36	ns	0.36

[a] 2018-2017: difference between year 2018 and 2017

[b] G: genetic variance; GxY: genotype by year interaction variance; (GxY)/G ratio of the genetic variance and the genotype by year interaction variance

B

		Generation effect[a]		Variance decomposition[b]				
Trait	Site	$S_{0:3}$-$S_{0:2}$	p-value	G	p-value	GxGen	p-value	(GxGen)/G
FL	PAL	0.624	ns	16.95	***	0.00	ns	0.000
	SRO	0.5	ns	14.70	***	0.29	ns	0.020
PH	PAL	0.245	ns	55.82	***	0.00	ns	0.000
	SRO	-0.381	ns	27.40	***	0.00	ns	0.000
YLD	PAL	29.167	ns	1626.61	***	0.00	ns	0.000
	SRO	-2.166	ns	1555.21	***	211.78	ns	0.136
ZN	PAL	-0.221	ns	2.64	***	0.13	ns	0.049
	SRO	0.297	ns	4.25	***	0.25	ns	0.058

[a] $S_{0:3}$-$S_{0:2}$: difference between generation $S_{0:3}$ and $S_{0:2}$

[b] G: genetic variance; GxGen: genotype by generation interaction variance; (GxGen)/G ratio of the genetic variance and the genotype by generation interaction variance

STable 2-6: Average predictive ability for the SINSRO scenarios across all traits, years, GP methods and calibration set sizes

		2017					2018					
Trait		25	50	100	200	Mean	25	50	100	200	Mean	General Mean
FL	Mean	0.194	0.257	0.278	0.327	0.264	0.263	0.311	0.357	0.413	0.336	0.300
	GBLUP	0.184	0.242	0.257	0.322	0.251	0.257	0.300	0.351	0.404	0.328	0.290
	RKHS	0.205	0.272	0.298	0.332	0.277	0.270	0.322	0.363	0.422	0.344	0.310
PH	Mean	0.155	0.235	0.347	0.464	0.300	0.244	0.322	0.394	0.466	0.357	0.328
	GBLUP	0.158	0.232	0.360	0.469	0.305	0.247	0.313	0.380	0.458	0.350	0.327
	RKHS	0.151	0.238	0.335	0.459	0.296	0.241	0.332	0.408	0.473	0.363	0.330
YLD	Mean	0.176	0.234	0.318	0.387	0.279	0.171	0.231	0.306	0.351	0.265	0.272
	GBLUP	0.184	0.246	0.320	0.390	0.285	0.170	0.224	0.312	0.353	0.265	0.275
	RKHS	0.167	0.223	0.317	0.384	0.273	0.173	0.238	0.299	0.350	0.265	0.269
ZN	Mean	0.173	0.225	0.283	0.356	0.259	0.127	0.197	0.240	0.316	0.220	0.240
	GBLUP	0.167	0.219	0.276	0.351	0.253	0.136	0.198	0.229	0.309	0.218	0.236
	RKHS	0.179	0.231	0.290	0.361	0.265	0.118	0.195	0.252	0.323	0.222	0.244

STable 2-7: Average predictive ability for the BAL1, BAL2, IMB and IMB scenarios across all traits, years and calibration set sizes using RKHS

	2017					2018					General
Trait	25	50	100	200	Mean	25	50	100	200	Mean	mean

FL	Mean	0.252	0.282	0.298	0.322	0.288	0.364	0.403	0.439	0.469	0.419	0.354
	BAL1	0.117	0.147	0.171	0.186	0.155	0.244	0.291	0.349	0.380	0.316	0.236
	BAL2	0.136	0.156	0.168	0.215	0.169	0.254	0.311	0.357	0.384	0.326	0.248
	IMB	0.553	0.552	0.553	0.556	0.553	0.687	0.688	0.688	0.690	0.688	0.621
	SIN_{SRO}	0.205	0.272	0.298	0.332	0.277	0.270	0.322	0.363	0.422	0.344	0.310
PH	Mean	0.276	0.339	0.413	0.480	0.377	0.363	0.422	0.480	0.542	0.452	0.414
	BAL3	0.191	0.267	0.357	0.443	0.315	0.231	0.314	0.388	0.478	0.353	0.334
	BAL4	0.217	0.299	0.400	0.447	0.341	0.291	0.350	0.422	0.503	0.391	0.366
	IMB	0.545	0.551	0.560	0.572	0.557	0.690	0.693	0.701	0.713	0.699	0.628
	SIN_{SRO}	0.151	0.238	0.335	0.459	0.296	0.241	0.332	0.408	0.473	0.363	0.330
YLD	Mean	0.192	0.242	0.318	0.381	0.283	0.183	0.244	0.304	0.376	0.277	0.280
	BAL5	0.145	0.205	0.310	0.364	0.256	0.141	0.223	0.283	0.395	0.260	0.258
	BAL6	0.180	0.241	0.318	0.411	0.288	0.167	0.231	0.309	0.388	0.274	0.281
	IMB	0.275	0.298	0.325	0.362	0.315	0.253	0.286	0.326	0.369	0.308	0.312
	SIN_{SRO}	0.167	0.223	0.317	0.384	0.273	0.173	0.238	0.299	0.350	0.265	0.269
ZN	Mean	0.220	0.258	0.289	0.328	0.274	0.196	0.238	0.272	0.293	0.250	0.262
	BAL7	0.124	0.176	0.215	0.248	0.191	0.106	0.139	0.176	0.196	0.154	0.172
	BAL8	0.164	0.206	0.230	0.276	0.219	0.102	0.156	0.196	0.189	0.161	0.190
	IMB	0.414	0.419	0.423	0.425	0.420	0.460	0.463	0.465	0.464	0.463	0.442
	SIN_{SRO}	0.179	0.231	0.290	0.361	0.265	0.118	0.195	0.252	0.323	0.222	0.244

STable 2-8: Average predictive ability for the SINSRO, BAL1 and BAL2 scenarios across all traits, years and calibration set sizes using GBLUP

		2017					2018					General
Trait		25	50	100	200	Mean	25	50	100	200	Mean	mean
FL	Mean	0.188	0.230	0.261	0.335	0.254	0.253	0.312	0.364	0.413	0.335	0.295
	BAL1	0.185	0.225	0.254	0.313	0.244	0.242	0.316	0.362	0.412	0.333	0.289
	BAL2	0.196	0.223	0.273	0.369	0.265	0.260	0.321	0.379	0.422	0.346	0.305
	SIN_{SRO}	0.184	0.242	0.257	0.322	0.251	0.257	0.300	0.351	0.404	0.328	0.290
PH	Mean	0.181	0.268	0.372	0.467	0.322	0.254	0.322	0.390	0.467	0.358	0.340
	BAL1	0.184	0.270	0.364	0.472	0.323	0.238	0.313	0.376	0.461	0.347	0.335
	BAL2	0.199	0.303	0.392	0.460	0.339	0.278	0.341	0.414	0.480	0.378	0.359
	SIN_{SRO}	0.158	0.232	0.360	0.469	0.305	0.247	0.313	0.380	0.458	0.350	0.327
YLD	Mean	0.188	0.256	0.330	0.411	0.296	0.167	0.230	0.313	0.375	0.272	0.284
	BAL1	0.177	0.258	0.330	0.412	0.294	0.152	0.223	0.312	0.393	0.270	0.282
	BAL2	0.203	0.264	0.339	0.433	0.310	0.180	0.243	0.316	0.381	0.280	0.295
	SIN_{SRO}	0.184	0.246	0.320	0.390	0.285	0.170	0.224	0.312	0.353	0.265	0.275
ZN	Mean	0.163	0.229	0.285	0.359	0.259	0.136	0.182	0.249	0.320	0.222	0.241
	BAL1	0.157	0.220	0.285	0.344	0.252	0.133	0.156	0.239	0.324	0.213	0.232
	BAL2	0.165	0.250	0.296	0.382	0.273	0.138	0.193	0.279	0.327	0.234	0.254
	SIN_{SRO}	0.167	0.219	0.276	0.351	0.253	0.136	0.198	0.229	0.309	0.218	0.236

STable 2-9: Average predictive ability for the SIN_SRO and IMB scenarios across all traits, years and calibration set sizes using GBLUP

		2017					2018					General
Trait		25	50	100	200	Mean	25	50	100	200	Mean	mean
FL	Mean	0.341	0.368	0.368	0.406	0.371	0.453	0.468	0.489	0.520	0.483	0.427
	IMB	0.498	0.493	0.479	0.489	0.490	0.649	0.636	0.628	0.636	0.637	0.563
	SIN_{SRO}	0.184	0.242	0.257	0.322	0.251	0.257	0.300	0.351	0.404	0.328	0.290
PH	Mean	0.341	0.386	0.456	0.522	0.427	0.464	0.498	0.538	0.584	0.521	0.474
	IMB	0.524	0.541	0.553	0.575	0.548	0.680	0.684	0.695	0.710	0.692	0.620
	SIN_{SRO}	0.158	0.232	0.360	0.469	0.305	0.247	0.313	0.380	0.458	0.350	0.327
YLD	Mean	0.228	0.275	0.336	0.402	0.310	0.211	0.271	0.340	0.387	0.302	0.306
	IMB	0.272	0.304	0.353	0.414	0.336	0.252	0.319	0.367	0.422	0.340	0.338
	SIN_{SRO}	0.184	0.246	0.320	0.390	0.285	0.170	0.224	0.312	0.353	0.265	0.275
ZN	Mean	0.276	0.314	0.356	0.411	0.339	0.289	0.329	0.356	0.409	0.346	0.343
	IMB	0.385	0.409	0.436	0.471	0.425	0.443	0.460	0.483	0.509	0.474	0.450
	SIN_{SRO}	0.167	0.219	0.276	0.351	0.253	0.136	0.198	0.229	0.309	0.218	0.236

Chapter 3 : An optimized multigeneration multisite genomic prediction model for recurrent genomic selection in an upland rice population

Avant propos

Le chapitre qui suit constitue la suite du chapitre 2. Une fois la prédiction génomique testée par validation croisée à l'intérieur des génération $S_{0:2}$ et $S_{0:3}$, nous l'avons appliquée entre deux sets de notre population. Le premier set, servant à la calibration, était constitué du matériel utilisé au chapitre précédent alors que le second, développé spécialement pour la validation, était constitué de nouveaux génotypes phénotypés en $S_{0:4}$. Plutôt que de prédire à l'intérieur d'une même génération, l'utilisation des phénotypes $S_{0:4}$ comme références nous a permis de nous approcher au plus de lignées fixées. Les différents caractères aillant montrer différentes réponses à l'utilisation des données des deux sites au chapitre précédent, nous avons testé différentes structures de variance-covariance avec le double objectif de comprendre mieux les phénomènes derrières ces différences entre caractères et idéalement de trouver l'approche la plus adaptée pour chacun d'entre eux. Il aura été écrit à six mains par Hugues de Verdal, Cécile Grenier et moi-même et fera l'objet d'une publication dans la revue Rice.

Hugues de Verdal[12*], Cédric Baertschi[12], Julien Frouin[12], Yolima Ospina[3], Maria Fernanda Alvarez[3], Jérôme Bartholomé[123], Cécile Grenier[123*]

[1]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France.

[3]Alliance Bioversity-CIAT, A.A.6713, Km 17 Recta Palmira Cali, Colombia

3.1 Abstract

Genomic selection (GS) is a good option to improve the genetic gain of recurrent selection in rice breeding programs. The present study assessed the impact of the addition of multigeneration multisite genomic prediction models that could significantly increase the predictive ability of GS and therefore, the genetic gain of the CIAT-Cirad rice breeding program.

Of a synthetic population PCT27, a fraction was used for calibrating models (PCT27A), while another set (PCT27B) was considered for validating them. All S_0 plants from PCT27 were advanced by selfing to the $S_{0:2}$, $S_{0:3}$ and $S_{0:4}$ generation by bulk harvesting seeds. Progenies were phenotyped at $S_{0:2}$ and $S_{0:3}$ generations for PCT27A and at $S_{0:4}$ generation for PCT27B in two distinct environments: Santa Rosa as the target site and Palmira as a surrogate site with distinct characteristics but with important potential contributions to accelerate breeding cycles.

Predictive ability (PA) of genomic predictions were estimated using several scenarios and models, according to the presence of one or two growing environments, one or several phenotyping generations, the presence of genetic by environment interaction and the size and composition of the training set.

The results indicated that selection intensity can be increased by GS models calibrated on a fraction of the population. Breeding cycles can be accelerated with models calibrated with early generation families ($S_{0:2}$). Despite relatively low PA achieved when including two locations in the training set, the gain in time realized by phenotyping in the surrogate site during the off-season lead to genetic and economic gains.

3.2 Introduction

In the literature, several studies have demonstrated empirically or by simulation the interests of genomic selection (GS) models for crops breeding, for example wheat (Crossa et al. 2010; Heffner, Jannink, and Sorrells 2011; Rutkoski et al. 2012), maize (Bernardo and Yu 2007; Zhao et al. 2012; Crossa et al. 2013) and barley (Lorenz, Smith, and Jannink 2012; Endelman et al. 2014; Sorrells 2015) among others. Regardless of the trait and species considered, predictive ability (PA), i.e. the estimated correlation between phenotypic and predicted values, was always higher with GS than with classical selection based on phenotypes and pedigree. In the case of rice, the potential of GS to accelerate genetic gain has been highlighted previously (Onogi et al. 2015b; Isidro et al. 2015; Spindel et al. 2015; Grenier et al. 2015; Wang et al. 2017; Ben Hassen, Cao, et al. 2018; Bhandari et al. 2019; Nour Ahmadi et al. 2020). The main observations extracted from a review of GS applied to rice (Ahmadi et al. 2020), resume that markers set size does not have to be large (Spindel et al. 2015; Bhandari et al. 2019), the population structure needs to be accounted for (Isidro et al. 2015; Grenier et al. 2015; Ben Hassen, Cao, et al. 2018), and the relatedness between the training set and the breeding population remains essential to insure high effective PA. GS models in rice breeding have been used to select among genebank accessions (Tanaka and Iwata 2018; Wissuwa et al., in press), to predict among biparental crosses derived from parental breeding lines (Ben Hassen, Bartholome, et al. 2018), progenies or synthetic populations (Grenier et al. 2015; Morais Júnior et al. 2018; Baertschi et al. 2021). The integration of genomic prediction (GP) into the rice breeding program is expected to increase genetic improvement for polygenic traits such as yield and adaptation to climate change.

The ways in which GS can increase genetic gain over a conventional pedigree breeding program are multiple (Spindel and Iwata 2018; Bartholomé, Prakash, and Cobb 2021). Considering the breeder's equation, almost all parameters could be improved using GS. A greater precision in prediction has a direct impact on genetic gain (Falconer and MacKay 1996), therefore, even a small improvement in PA can have a consequent impact in terms of genetic gain (Onogi et al. 2015b; Yang Xu et al. 2021). In addition, predicting the genomic estimated breeding value (GEBV) of non-phenotyped genotypes included in the candidate population would significantly increase the intensity of selection. As genotyping is getting more affordable relative to the phenotyping cost, selection could be performed on a larger number of individuals. Yet, the inclusion of a larger set or entries will have to be balanced with the necessity to limit population structuration and to maintain relatedness between the training set and candidate population. While applying high selection intensity on a large population could theoretically maximize a benefit to the breeding program, such value would essentially depend on population genetic diversity and the prevalence of superior genotypes.

Furthermore, it has been shown previously that PA could be improved with multi-environment models rather than using single-environment models (Burgueño, Campos, et al. 2012; Lopez-Cruz et al. 2015; Crossa et al. 2016; Cuevas et al. 2016; Cuevas et al. 2017; Ben Hassen et al. 2018; Jarquin et al. 2020; Yang Xu et al. 2021). Multi-environment trials are commonly performed in plant breeding, with trials in environments more or less close to the production environment, making it possible to evaluate the genotype and its phenotypic stability for traits under selection, or genotypes with high adaptation capabilities. In this context of GS and G×E interactions, the use of sparse testing methods in which only a subset of the genotyped individuals is also phenotyped in multiple environments could be attractive to reduce the phenotyping efforts (Jarquín et al. 2017b; Jarquin et al. 2020). Although very promising this strategy that relies on the use of various sites to calibrate the model, thus accounting for the GxE, is mainly dependant on the level of correlation between sites. To ensure a gain in resource allocation for the phenotyping step, certain site similarity should be considered.

Another aspect to consider is the relationship between training and validation populations. Optimizing the training population has previously been shown to improve the predictive ability of genomic selection models (Rincent et al. 2012; Isidro et al. 2015; Akdemir, Rio, and Isidro y Sánchez 2021). Several methods have been developed to optimize the selection of individuals for inclusion in the training set based on the relationship between genotypes in the training set and/or between training and validation sets. The selection of genotypes to be phenotyped and included in the training set has two major interests: it could reduce the number of families to be phenotyped and at the same time, can increase predictive ability of the GS.

Genomic selection can also shorten the length of reproductive cycles and increase genetic gain per unit time by reducing intergenerational time (Heffner et al. 2010; Spindel and Iwata 2018). However, only a few studies report germplasm development based on early genomic selection of promising lines (Mendonça et al. 2020). It is rare to see descriptions of GS applied on breeding populations composed of segregating progenies. A particularity of our plant breeding pipeline is a change from a very heterozygous genetic make-up in the population to a fixed germplasm prior to cultivar release. A potentially important difference in allelic fixation can be found between the calibration and the prediction unless GS is conducted when the germplasm has reached homozygosity, in which case a gain in time is not maximised.

The collaborative rice breeding program between CIAT (International Center for Tropical Agriculture, member of the CGIAR centers) and Cirad (French Agricultural Research Centre for International Development) has developed synthetic populations, managed through recurrent selection (RS) which presents an ideal context for applying GS. The orientation towards the use of population improvement took place in the 90s following observation of the declining crop genetic diversity among improved rice germplasm (Martinez et al. 2014). A recurrent selection scheme consists of three main steps conducted

recurrently and is summarized as follows: i) evaluation of families, ii) selection of the best families, iii) inter-crossing of those best candidates to develop the next generation. In the CIAT-Cirad program, the RS scheme applied to the inbreeder rice was facilitated through the use of a recessive nuclear male-sterility gene (ms-IR36, reviewed in Frouin et al. 2014) segregation in the population. The number of crosses and combinations of crosses among best haplotypes have never been limited during the various cycles of population breeding. At each cycle, about 3000 plants derived from the best candidates (segregating progenies of each candidate), randomly distributed in the field are the parents, either male or female parent, of the new cycle. In addition to the population improvement, at each cycle, the best candidates are selected to enter for the variety development pipeline which is conducted through conventional pedigree breeding. As detailed in Baertschi et al. (2021), the CIAT-Cirad rice breeding program benefits from two distinct locations in Colombia to develop improved populations and inbred lines. At CIAT-HQ (Palmira, PAL) rice is grown under irrigated conditions throughout the year and with limited pathogen pressure. The second site, located in the Llanos of Colombia, farther from CIAT-HQ is at the Colombian National Federation of rice growers (Fedearroz) research station in Santa Rosa, Meta (SRO). The station is amidst one of the most productive rice growing areas in the country where rice is direct seeded and grown under rainfed conditions. This ecosystem is the target production environment for the japonica rice and the research field location in SRO presents the advantage for the breeding program to have a high incidence of diseases, notably blast. Two locations for phenotyping and a proof of concept that GS is feasible on the CIAT-Cirad japonica synthetic population, are the basis for our current research to apply GS for accelerated recurrent GS and optimization of the upland rice breeding scheme. An ideal situation for simplifying the breeding scheme would advocate the prediction of candidates as early as possible for population improvement through recurrent selection and for variety development through pedigree breeding.

The first objective of the present study was to evaluate whether we can effectively apply GP models developed in early generation to select breeding candidates in the target production environment based on GEBVs. The second objective was to evaluate whether GP models including GxE interactions into the breeding program are improving the efficiency of RS. The third objective aimed at defining whether using two generations of progeny testing and combining two evaluation trials per year would improve the efficiency of the breeding program. Finally, a fourth objective was to evaluate whether optimizing the choice of these phenotyped individuals in combination with the two-site two-generation scenario could significantly increase the predictive ability of GS and therefore, the genetic gain of the breeding program.

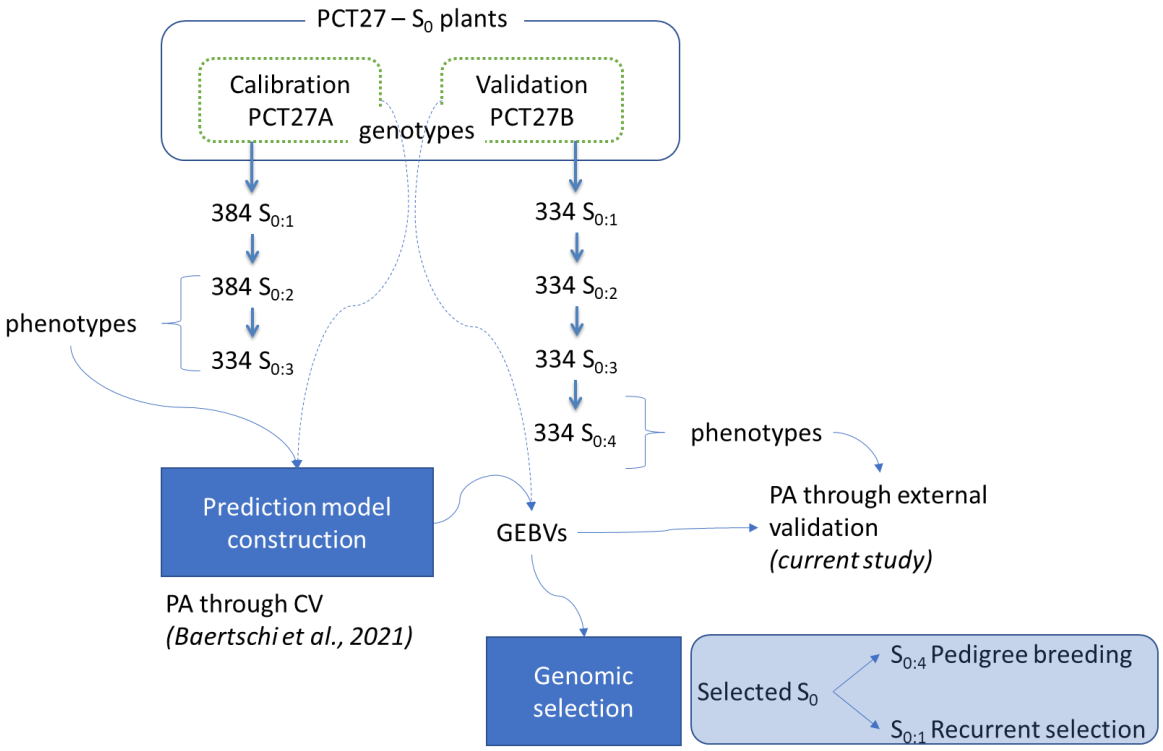


Figure 3-1: Scheme of the GP models and origin of the set used to calibrate and validate the models. From the base population PCT27, two subsets were randomly constituted (PCT27A and PCT27B). Data acquirement, model calibration and model validation through CV within the PCT27A were described in Baertschi et al. (2021).

3.3 Material and Methods

3.3.1 Development of the used population

The training and the validation sets were both derived from a rice synthetic population belonging to the tropical japonica group of rice (*Oryza sativa* L.) as described in Figure 3-1. The population development was earlier described (Grenier et al. 2015; Baertschi et al. 2021). Among a set of approximately one thousand fertile plants extracted from the PCT27 population, 384 were used for calibrating the model (PCT27A), while another set of 334 (PCT27B) was considered for validating the model. All 718 entries were advanced to the $S_{0:2}$, $S_{0:3}$ and $S_{0:4}$ generation by bulk harvesting seeds from 15 to 20 male fertile plants per line per generation as explained in Baertschi et al. (2021). A set of 50 families at the $S_{0:2}$ generation extracted from the set considered for model calibration and designed as "temporal checks" were included in each phenotyping trial to account for the year effect within the site.

3.3.2 Genotyping

Genotyping-by-sequencing (GBS) was performed on the 718 S_0 plants as described in Baertschi et al. (2021). The genetic characterization of the two populations is presented in supplementary Tables and Figures. A total of 9,928 SNP markers fairly well distributed among the 12 rice chromosomes (SFig 3-1) The MAF distribution among the 718 S_0 reflects a population where rare alleles were not depleted, which fits well with long-term objectives of a population breeding program. The degree of allelic

fixation varied greatly between the genotypes but remained relatively low for individuals at the S_0 generation (STable 3-1). Considering the rather large average LD (STable 3-2) and the slow LD decay observed, the average marker density (1 SNP every 40 kb) was considered sufficient so as to allow the capture of all linked QTLs with the SNP matrix in hand. Globally, the whole set of genotypes as two random fractions extracted from a large population were tested for any structuration (SFig 3-2).

3.3.3 Field trial and phenotyping

Field phenotyping was performed at two locations in Colombia from 2017 to 2020. The two sites are described in Baertschi et al. (2021) and consist of the experimental field at CIAT-HQ in Palmira (PAL) located in the Valle del Cauca, Colombia (3.50° N - 76.35° W, 1000 masl) and an experimental site in Santa Rosa (SRO) property of the Fedearroz, located in the Oriental plains of Colombia, in the department of Meta, Colombia (4.03° N - 73.48° W, 300 masl). While PAL location is a surrogate site with irrigation systems freeing rice trials from any constraint on planting time or any severe disease pressure, the SRO site is within a rice growing area, where the crop is cultivated under rainfed conditions during the main cropping season, May to September, and with the natural occurrence of various pathogens such as blast.

Six trials were conducted for four years, using different semesters for each location. Field trials for the $S_{0:2}$, $S_{0:3}$ and $S_{0:4}$ generation were established in PAL on 4 December 2017, 10 December 2018, and 26 December 2019, respectively and in SRO on 12 May 2017, 30 May 2018, and 20 May 2020. PCT27A population was phenotyped at the $S_{0:2}$ and $S_{0:3}$ generations whereas PCT27B population was only phenotyped at the $S_{0:4}$ generation. At each site, the experimental design followed a lattice with 16 blocks and three repetitions and included the 334 families and the 50 $S_{0:2}$ temporal check lines all randomly distributed across the design within each repetition of the two sites and three-year trials. In PAL, the trials were established after transplanting 3-week-old seedlings in a bundled field. The plot size was two rows of 17 plants with 25 cm between plants and between rows. Fertilizer application was split, with NPK nutrients (377 kg/ha urea, 188 kg/ha DAP, 189 kg/ha KCl) added at 25 and 35 days after transplanting. Irrigation was maintained continuously to ensure a 25 cm layer of water in the field until a week prior to the crop maturation period. In SRO, the trials were established by direct sowing of two 4 m-long rows, spaced by 26 cm at a density of 1 gram of seed per linear meter. Split fertilizer application was performed according to the recommended application for growing tropical japonica rice in upland soil conditions (230 kg/ha urea, 217 kg/ha DAP, 150 kg/ha KCl). Phytosanitary treatment was applied in SRO to prevent blast outbreaks.

Four traits were measured following the IRRI Standard Evaluation System (IRRI 2013) on the whole training population including the 50 temporal checks. Flowering date (FL) was expressed as the number of days after crop establishment – being either the date after transplantation (PAL) or sowing (SRO) –

when 50% of the plants within a plot reached anthesis. Plant height (PH) was calculated as the average height measured in centimetres of five plants with their panicle extended. Grain yield (YLD) was obtained by weighing the grains collected within each plot after discarding the plants at the start and end of each plot. For each harvested plot, percent humidity was measured and used to correct the weight of collected grains, expressed in grams per plot, for a relative humidity of 14%. For some plots, due to the low harvest, humidity measurements were not taken but estimated by using other plots from other replicates of the same genotype. The YLD value was neither adjusted for the plot size nor for the count of fertile plants. The grain zinc concentration (ZN), expressed in parts per million (ppm), was measured on a subsample of collected grains polished in Teflon equipment, using energy dispersive X-ray fluorescence spectrometry (X-supreme 8000, Oxford Instrument, Shanghai, CN) available at the CIAT-HQ Nutritional Laboratory. The exact same procedure was used for generations $S_{0:2}$ $S_{0:3}$ and $S_{0:4}$.

The 50 temporal checks were phenotyped as $S_{0:2}$ in all the trials. This allowed measurement of the non-confounded year within site effect on the $S_{0:2}$ and the generation effect in 2018 by analysing the data from the $S_{0:2}$ and $S_{0:3}$ lines (STable 3-3).

3.3.4 Statistical analyses

3.3.4.1 Elementary statistics

The raw data were checked per trial for outliers using the boxplot.stats function of the R package "stats" (R Development Core Team 2018) with a coefficient of 1.5, which means that outliers were identified if the phenotypic values were outside 1.5 time the interquartile range above the upper quartile and below the lower quartile. No outliers were discarded. Variance decomposition was performed using the lmer function of the R package "lme4" (Bates et al. 2015).

To correct for the fixed effects of location, replicate and bloc, best linear unbiased predictions (BLUPs) were estimated for each trait using the lmer function of the R package "lme4" (Bates et al. 2015) using the following model:

$$y_{ijkl} = \mu + Loc_i + Rep_j(Loc_i) + Bl_k\left(Rep_j(Loc_i)\right) + g_l + g_l(Loc_i) + e_{ijkl} \qquad \text{model (3-1)}$$

where y_{ijkl} is the vector of phenotypic values, μ is the overall mean of the phenotypic values, Loc_i is the fixed effect of the location i (PAL or SRO), $Rep_j(Loc_i)$ is the fixed effect of the replicate j (from 1 to 3) within location i, $Bl_k\left(Rep_j(Loc_i)\right)$ is the random effect of the bloc effect k (from 1 to 8) within replicate within location with distribution $Bl \sim N(0, \sigma^2_{Bl})$, g_l is the random effect of the genotype l, $g_l(Loc_i)$ is the random nested effect of the genotype within location with distribution $g \sim N(0, \sigma^2_g)$and e_{ijkl} is the residual considered as a random effect with distribution $e \sim N(0, \sigma^2_e)$. The model was run by generation and the BLUPs values were used for prediction analyses.

Broad sense heritability (H^2) was estimated using the following model:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{g:loc}^2}{NE} + \frac{\sigma_e^2}{NR}}$$ model (3-2)

where σ_g^2 is the genetic variance of the trait under study, $\sigma_{g:loc}^2$ is the genetic by location variance, σ_e^2 is the residual variance, NE is the harmonic mean of the number of locations per genotype and NR is the harmonic mean of the number of replicates per genotype across the two locations. For each trait, correlations of phenotypic values between the two locations were performed using the rcorr function of the R package "Hmisc" (Harrell Jr 2021).

3.3.5 Genomic prediction

Genomic predictions were performed under several scenarios depending on the families included in the training set (TS) and the validation set (VS), as illustrated in Figure 3-2:

1) The first scenario (Uni1) was a cross-validation to estimate the predictive ability of a model calibrated with the genotypes of plants at S_0 generation and the phenotypes of their derived progenies at the generation ($S_{0:4}$) evaluated in a single location (SRO) to predict the values of $S_{0:4}$ families in SRO. In this scenario, the TS consisted in a random draw of 70% of the population PCT27B and the remaining 30% constituted the VS.
2) The second scenario (Uni2) was used to evaluate the suitability of the models when families from the population PCT27A at generation $S_{0:2}$ were used as a TS to estimate the genomic breeding values of all the families of PCT27B at generation $S_{0:4}$. Only one environment (SRO) was included in this scenario.
3) The third scenario (Uni3) was similar to Uni2 except the calibration was performed with PCT27A families at generation $S_{0:3}$.
4) The fourth scenario (Multi1) was performed to highlight the impact of genetic by environment interactions (GxE). Data from two locations (PAL and SRO) from a single generation ($S_{0:4}$) were used. The TS was composed of 100% and 70% of the PCT27B families phenotyped at PAL and SRO, respectively, and the VS was composed of the remaining 30% of the PCT27B families phenotyped in SRO. The families from PCT27B whose phenotypes from SRO were included in the TS and VS were picked by random draw.
5) The last scenario (Multi2) mixed all previous parameters. The potential of genomic prediction was assessed for calibration using data from PCT27A at generation $S_{0:2}$ and $S_{0:3}$ phenotyped in PAL and SRO respectively to predict PCT27B at generation $S_{0:4}$. The TS consisted of the PCT27A families with 100% of $S_{0:2}$ phenotyped in PAL and x% (x%= 25, 50 or 75%) of the $S_{0:3}$ phenotyped in SRO, and the VS included all PCT27B families at $S_{0:4}$ generation phenotyped in SRO. The x% of the $S_{0:3}$ included in the TS were either randomly drawn or selected by an optimisation process, as presented below.

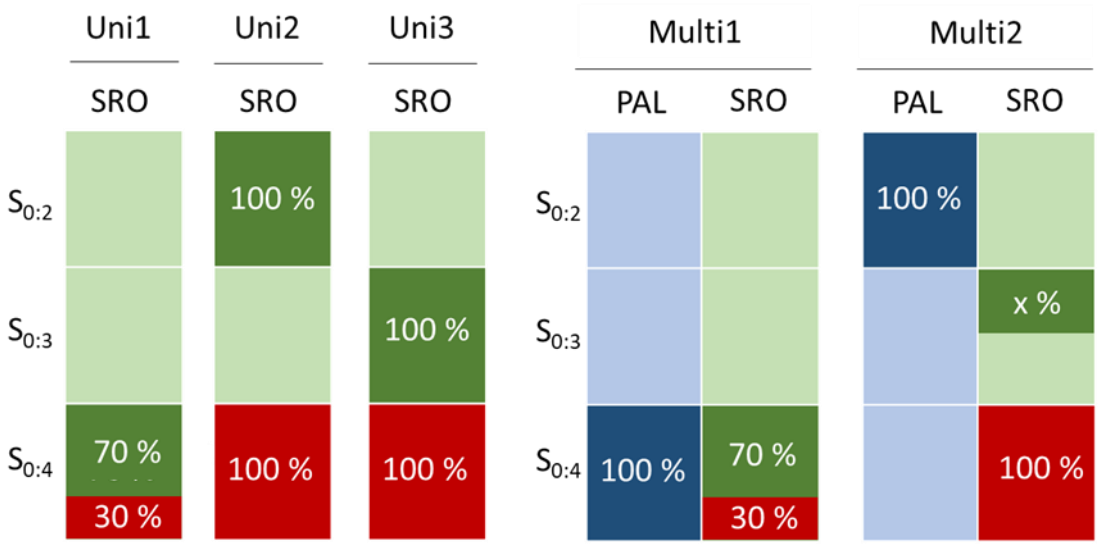


Figure 3-2: The different scenarios of calibration and validation of the GP models to predict the phenotype of the PCT27B at the $S_{0:4}$ generation in Santa Rosa (SRO). The red area represents the validation set (VS), the green and blue represent the training set (TS), from SRO and PAL, respectively. The percentage in the coloured areas represent the fraction of the population used to calibrate or validate the model. The x% of $S_{0:3}$ families phenotyped in SRO included in the TS in Multi2 scenario varies from 25, 50 and 75 %.

All genomic predictions were performed using the R package "BGGE" (Granato et al. 2018) with the following parameters: burn-in = 2,000, nIter = 15,000 and thin = 100.

In the first scenario, two different prediction methods were used and compared: GBLUP (VanRaden 2008) and RKHS (based on the reproductive kernel Hilbert space approach by Gianola and van Kaam (2008)). Because the results obtained from both approaches were similar; the GBLUP was preferred for all the analyses, and RKHS results were not shown in the present study. For Uni1, Uni2 and Uni3 scenarios, the genomic predictions were run using a univariate single-environment model (SM) considering only the main genotypic effects.

In the Multi1 and Multi2 scenarios a GxE interaction random effect was added to the predictive model. To do so, GxE genomic variance matrices were constructed, and genomic prediction performed using a Bayesian linear mixed model. Three different multi-environment models were used in the present study all available in the BGGE package:

i) a multi-environment model (MM) assuming that genetic effects across the environment are constant, and therefore the absence of GxE. In this MM model, a single matrix was constructed, related to the main across-environment effects with the model looking as follow:

$$y_{ij} = \mu + Loc_i + g_j + e_{ij} \qquad \text{model (3-3)}$$

with Loc_i and g_j are as described in model 3-1 with g_j having a variance-covariance structure following $g_j \sim N(0, \sigma_g^2 G)$, G being the genotype relationship matrix from VanRaden (2008);

ii) a multi-environment model (MDs) which is an extension of the model 3-3 including a single random deviation effect of the GxE.

$$y_{ij} = \mu + Loc_i + g_j + g_j(Loc_i) + e_{ij}$$ model (3-4)

the GxE effects following the normal distribution $g_j(Loc_i) \sim N(0, \sigma^2_{GxE}G)$;

iii) a multi-environment, environment-specific, variance GxE deviation model (MDe). This model is the same as model 3-4 with the difference that the environment-specific genetic effects follow the variance-covariance structure $g_j(Loc_i) \sim N(0, \begin{bmatrix} \sigma^2_{PAL}G & 0 \\ 0 & \sigma^2_{SRO}G \end{bmatrix})$ σ^2_{PAL} and σ^2_{SRO} being environment specific variances and G the again the genotype relationship matrix. Full details about these models can be found in Granato et al (2018).

3.3.6 Optimisation methodology

Careful selection of the TS may be relevant to improve the accuracy of GP. Considering the Multi2 scenario including multiple generations and two environments, one of our objectives was to test whether it was possible to reduce the phenotyping effort in SRO in generation $S_{0:3}$. In this scenario, the CDmean-optimality criterion, based on the GBLUP mixed model, was used to select the TS and compared to randomly selected TS. Either twenty-five percent, 50% or 75% of the $S_{0:3}$ phenotyped individuals grown in SRO were included in the TS. This model of optimization was proposed by Rincent et al. (2012), estimating the expected reliability of contrast predictions, defined as the squared correlation between true and predicted contrasts of genetic values. The parameters used were similar to those used for the previous model, adding a value of 1 for the variance ratio λ (λ=(1-h²)/h²) corresponding to a heritability of 0.5. The R TrainSel package (Akdemir, Rio, and Isidro Sanchez 2021) was used for the optimization process with the algorithm parameters as follows: number of iterations for the GA is 200, population size for GA is 300, and number of elite solutions at each iteration is 10.

3.3.7 Model and scenario comparison

For each model and scenario, the predictive ability (PA) was computed as the correlation between predicted and the phenotypic BLUPs adjusted by trial. To ensure that variations in accuracy between models and scenarios were not due to stochastic effects, all predictions were replicated 100 times, allowing the mean and standard deviation of each model to be estimated and compared using all the predictive abilities (100 PA for each model). The model comparisons were performed with a linear model considering the fixed effect of the method used.

3.4 Results

3.4.1 Phenotypic performances

Phenotypic data were collected for two consecutive generations in two separate locations on the same population of S_0 progenies (PCT27A) in 2017 ($S_{0:2}$) and 2018 ($S_{0:3}$) and on another population of S_0 progenies (PCT27B) at the $S_{0:4}$ generation in the same two locations in 2019 and 2020. The phenotypic

Table 3-1: Descriptive statistics for the PCT27B phenotyped at the S0:4 generation in two locations; Palmira (PAL) and Santa Rosa (SRO) with mean, standard deviation (SD), min, max, coefficient of variation (CV) and the phenotypic correlation (Pearson) between locations.

		PCT27B $S_{0:4}$ generation					
Trait[1]	Site	Mean	SD	min	max	CV	Corr
FL	PAL	87.38	3.84	78	96	4.39	0.319
	SRO	81.68	5.73	69	96	7.02	
PH	PAL	120.4	4.98	113.2	128.2	4.14	0.229
	SRO	97.84	8.89	75	121	9.09	
YLD	PAL	759.6	184.1	304.6	1240.1	24.2	0.216
	SRO	137.0	50.1	65.4	270.5	36.6	
ZN	PAL	14.68	1.83	10	19.6	12.5	0.313
	SRO	27.27	3.65	18	37.5	13.4	

[1]Traits are flowering date (FL), plant height (PH), grain yield per plot (YLD) and grain zinc concentration (ZN)

data from PCT27A were presented in Baertschi et al. (2021) and will not be described in the present results.

For all four traits measured on the PCT27B, differences were observed between the two locations (Table 3-1). On average, flowering date (FL) was 6 days earlier and plant height (PH) 20cm shorter at SRO than at PAL. Yields (YLD) were largely reduced (5.5 times lower) at SRO and conversely, grain zinc concentrations (ZN) were 12.6ppm higher at SRO than at PAL. Coefficients of variation (CV) of all traits were higher at SRO than at PAL. Phenotypic correlations were relatively low ranging from 0.216 (for YLD) to 0.319 (for FL).

For each trait measured an analysis of variance components was performed using model 2 (Table 3-2). Surprisingly, the proportion of variance explained by the genotype effect was particularly low for PH, explaining the near-zero H^2. However, distinguishing locations, it appeared that H^2 was negligible for PH measured at PAL, which was not considered except in Multi1 scenario, but high for PH measured at SRO (H^2 = 0.82 when the effects including location were removed). For all other

Table 3-2: Variance decomposition and broad sense heritability (H^2) obtained using Model 2 by trait for the PCT27B at $S_{0:4}$ generation

Trait	Variance component	Variance	Proportion	H^2
FL	Bloc	0.69	3.11	0.67
	Genotype	7.95	35.83	
	Location:Genotype	5.49	24.74	
	Bloc:Rep:Location	0.8	3.61	
	Residuals	7.26	32.72	
PH	Bloc	1.014	2.41	0.02
	Genotype	0.211	0.50	
	Location:Genotype	15.49	36.88	
	Bloc:Rep:Location	0.978	2.33	
	Residuals	24.31	57.88	
YLD	Bloc	237.93	1.311	0.21
	Genotype	1293.41	7.121	
	Location:Genotype	6356.78	35.00	
	Bloc:Rep:Location	264.55	1.46	
	Residuals	10007.7	55.11	
ZN	Bloc	0.042	0.53	0.51
	Genotype	1.854	23.45	
	Location:Genotype	2.438	30.83	
	Bloc:Rep:Location	0.404	5.11	
	Residuals	3.169	40.08	

Table 3-3: Predictive ability (PA, LSmeans ± standard deviation) for the three "Uni Site" scenario combining different make-up of training set and validation set. Within a trait, values followed by different letters are significantly different (p<0.05). The description of the scenarios is in *Figure 2-2*.

Training set	Validation set	Scenario	FL	PH	YLD	ZN
$S_{0:4}$ (70%)	$S_{0:4}$ (30%)	Uni1	0.23 ± 0.08[b]	0.31 ± 0.07[b]	0.39 ± 0.08[a]	0.17 ± 0.08[c]
$S_{0:2}$ (100%)[1]	$S_{0:4}$ (100%)	Uni2	0.31 ± 0.01[a]	0.39 ± 0.01[a]	0.33 ± 0.01[b]	0.32 ± 0.01[a]
$S_{0:3}$ (100%)[1]	$S_{0:4}$ (100%)	Uni3	0.23 ± 0.01[b]	0.25 ± 0.01[c]	0.24 ± 0.01[c]	0.29 ± 0.01[b]

[1]Set derived from the PCT27A

trait combinations, H^2 was moderate, ranging from 0.21 to 0.67, with a lower H^2 for YLD than for the other three traits. The GxE effect explained

a non-negligible part of the variance with an explained proportion ranging from 25% to 37% for the

four traits.

3.4.2 Single generation single site calibrations

The potential of GP was first tested with calibration using single generation data, either by cross-validation (Uni1) or by prediction between population (Uni2 and Uni3) (Table 3-3). For prediction within PCT27B on progenies at the $S_{0:4}$ generation (Uni1), PA ranged from 0.17 for ZN to 0.39 for YLD. The PA using a model calibrated at $S_{0:2}$ (Uni2), was greater than with the $S_{0:3}$ (Uni3) to predict the GEBVs of $S_{0:4}$. The increase in PA using a different set of progenies was significant but moderate for FL and PH (PA increased by 0.08 in Uni2 compared to Uni1) and highly significant for ZN (PA=0.17 ± 0.08 and PA=0.32 ± 0.01 for Uni1 and Uni2, respectively). Differing from the other traits, YLD revealed a higher PA when calibrated directly on $S_{0:4}$ than when the calibration was done on the $S_{0:2}$ of the PCT27A (PA = 0.39 ± 0.08 and 0.33 ± 0.01 for Uni1 and Uni2, respectively).

3.4.3 Genomic selection and GxE interactions

The Multi1 scenario considered one generation but two locations. It was tested to assess the utility of including the GxE interaction in the GP models. Using this scenario, it was possible to estimate the PA of models including a single location (SM), both locations with a location effect (MM), and the GxE interaction effect with a single or two different variances for each of the two locations (MDs and MDe, respectively). The PA obtained with the Multi1 scenario and the four different models are shown in Figure 3-3. From these analyses, it appeared that for FL, PH and ZN, the PA of models using multiple location (MM and MDs) were significantly higher than the model based on a single location (SM) with PA increased by +0.09, +0.04, and +0.1 for FL, PH and ZN, respectively. All three traits responded in broadly the same way: PA using the MM model had the highest values, followed by MDs and MDe. Only for FL and PH the PA values obtained with MM were not significantly different from those of the MDs model (PA for FL were 0.31 and 0.29 for MM and MDs, respectively and PA for PH were 0.34 and

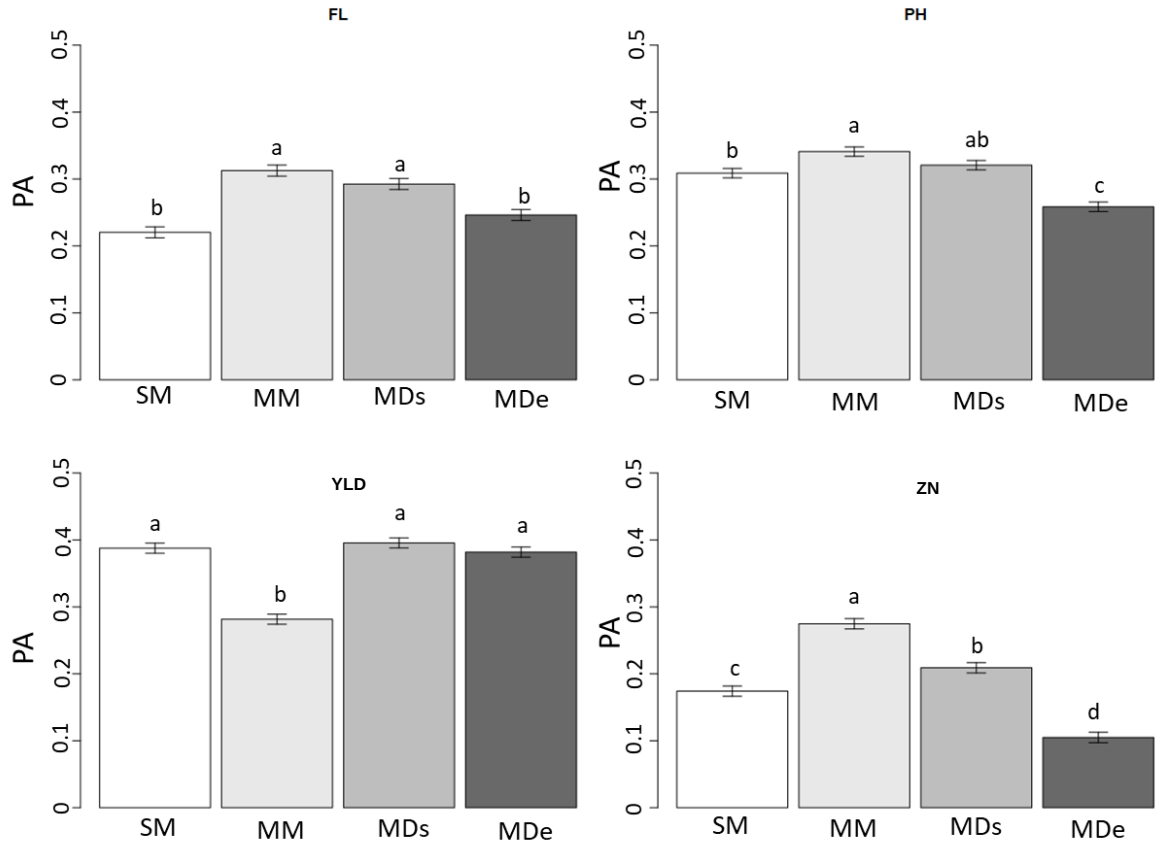


Figure 3-3: Predictive ability (LSmeans with error-bars representing the standard error) within the Multi1 scenario considering the single-site model (SM), the multi-site model without genotype by environment interaction (MM) and the multi-site model including the genotype by environment interaction with similar variances between environments (MDs) or with different variances between environments (MDe). Calibration and validation were performed within the PCT27B population phenotyped at the $S_{0:4}$ generation in Santa Rosa (SRO) for the four traits of interest: flowering date (FL), plant height (PH), grain yield per plot (YLD) and grain zinc concentration (ZN). TS included 100% of the records in Palmira (PAL) and 70% of the records in SRO for all the models except SM where the TS included only 70% of the phenotypes recorded in SRO. For all models VS was 30% of phenotypes in SRO.

0.32 for MM and MDs, respectively). While for FL the PA obtained with MDe (PA=0.25) was not significantly different to the PA obtained in SM (PA=0.22), for PH and ZN the MDe model induced a significant reduction in PA (-0.05 for PH, -0.07 for ZN). Except for YLD, the PA were always higher when two locations were included in the models. For YLD, the effect of including location without GxE interaction (MM model) greatly reduced PA with significantly lower PA (PA=0.28) than with the SM model (PA=0.39). However, including the GxE (MDs and MDe models) did not reduce the PA (PA=0.40 and 0.38 for MDs and MDe, respectively) in comparison to SM model.

3.4.4 Multi-generation and multi-environment genomic selection

We previously showed that for most of the traits, PA were greater when using Uni2 scenario and the consideration of multiple environments tended to increase the PA of the GP models. Therefore, combining these approaches of early-generation prediction using a TS of genetic constitution differing from the VS and multi-environment GP, as presented in the Multi2 scenario, was tested (random part of Figure 3-4). In this scenario, the calibration was performed on PCT27A population. The TS consisted

of 100% of the families phenotyped in PAL at the $S_{0:2}$ generation and 25, 50 or 75% of the families measured at SRO at the $S_{0:3}$ generation. The validation was made, as before, with the phenotypes of the whole population PCT27B at the $S_{0:4}$ generation grown at SRO. With the exception of PH, it appears that the more phenotypes of $S_{0:3}$ families are included in the TS, the higher the PA. For FL and YLD, the best estimates of PA (PA= 0.22 and 0.30, for FL and YLD, respectively) were found with an MDs model including 75% (250 from the 334 individuals) of the $S_{0:3}$ phenotypes. For ZN, the best model (PA= 0.25) also included 75% of the $S_{0:3}$ phenotypes but using the MM model, for which a location effect was included but without GxE interaction. Finally, for the PH trait, the results were completely different, with the best models being MM models (PA = 0.30), regardless of the number of $S_{0:3}$ families included in the TS.

3.4.5 Optimization of the training set

Within the Multi2 scenario, one way to gain PA while keeping the phenotyping effort low would be to optimize the choice of individuals to be included in the TS. As TS optimization method, CDMean was

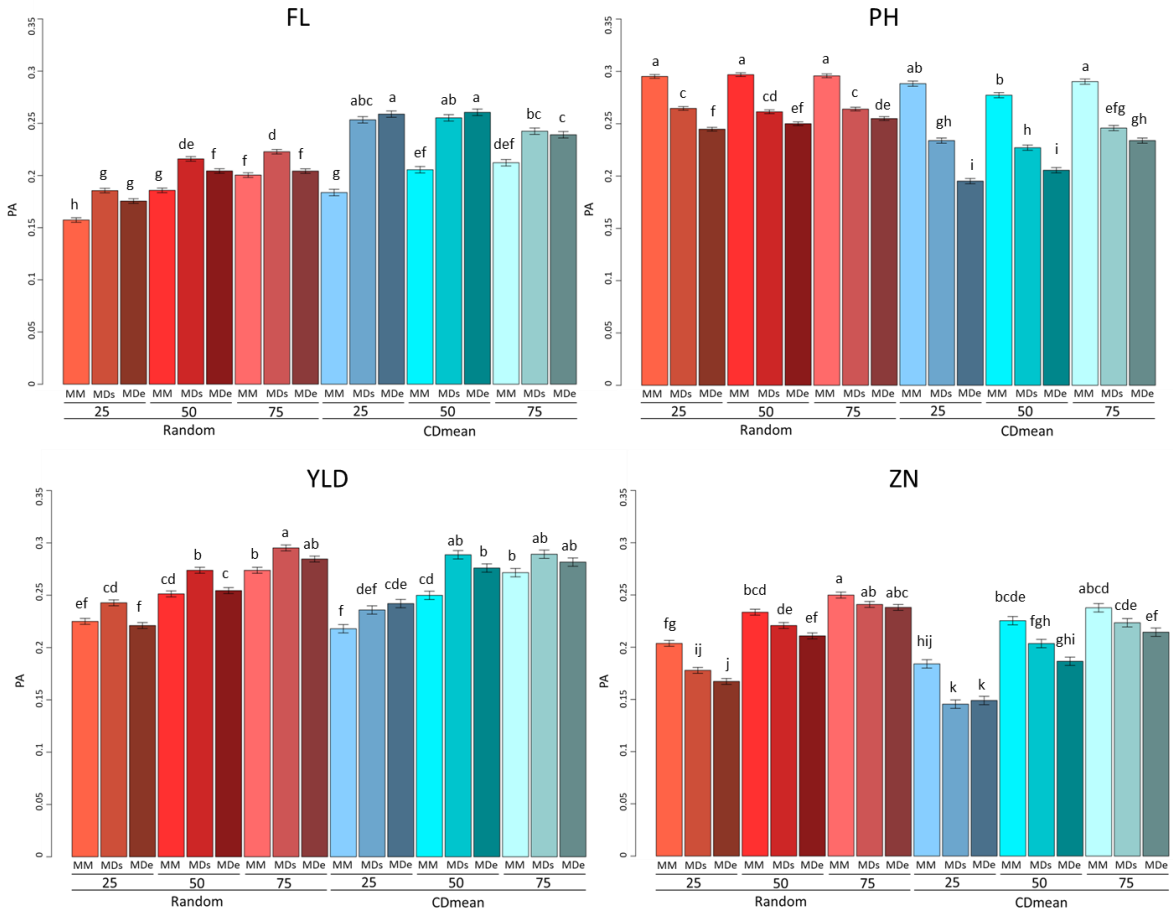


Figure 3-4: Predictive ability (LSmeans with error-bars representing the standard error) for the multi-site model (Multi2 scenario) without G × E interaction (MM), including the G × E interaction with similar variances between environments (MDs) or with different variances between environments (MDe). Validation was performed with the phenotypes of the PCT27B at generation $S_{0:4}$ in Santa Rosa (SRO) for the four traits of interest: flowering date (FL), plant height (PH), grain yield per plot (YLD) and grain zinc concentration (ZN). Within a trait, the letters represent significant differences between estimations

performed to optimize the choice of the $S_{0:3}$ families phenotyped at SRO to be included in the TS (Figure 3-4). Globally, across the three TS sizes and the three GxE models, optimizing the selection of $S_{0:3}$ families to include in the TS increased the PA only for FL (from 0.23 to 0.26) compared to a random selection of the TS.

The superiority of the largest TS (75% of the $S_{0:3}$ families phenotyped at SRO) found in the random sampling holds true only for ZN in the case of optimized sampling of TS. Regarding the impact of the GxE interaction, the results were similar to those obtained when $S_{0:3}$ families were selected by random draw except for FL where MDe appeared similar to the MDs model.

3.4.6 Selection of the best families

The main interest of GP is to estimate with high accuracy among a large set of progenies and with the least amount of phenotyping possible which families would be selected to be candidate for variety development and/or crossed to constitute the next generation. Therefore, it seemed important to evaluate if the different prediction models used would select the same families or not. By ranking the $S_{0:4}$ families according to their GEBVs for a phenotypic performance in SRO, it was possible to define which progenies would be selected if the 10, 20 or 50 best ones, i.e. those with the highest GEBVs, were selected. This analysis was performed by combining all models together and calculating the percentage of times each family was selected (Figure 3-5). For all traits, it appears that a large majority of families were never selected because of their low GEBVs encountered across all 18 models. Moreover, few families were selected in almost all the models used. On average, eight to ten families

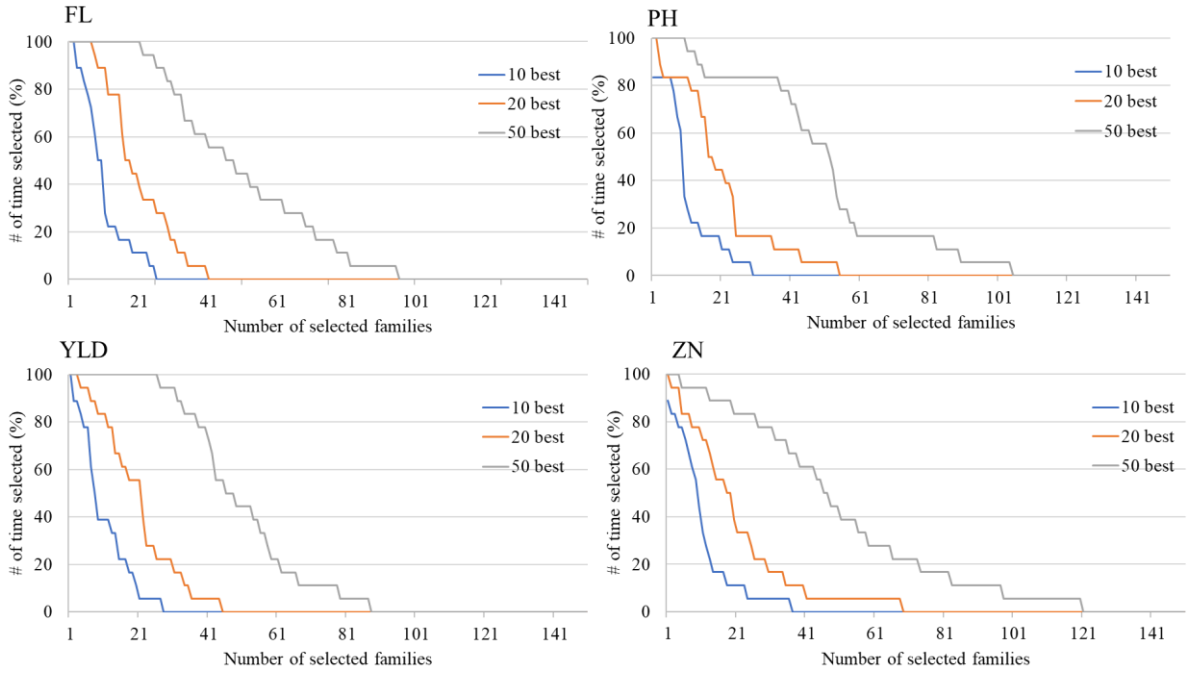


Figure 3-5: Number of time (in %) families were selected across the 18 models of the Multi2 scenario (with random draw or optimization of the S0:3 families included in the TS) when selection threshold was 10 (in blue), 20 (in orange) or 50 (in grey) best according to their estimated GEBVs

were selected in at least 50% of the GP methods (STable 3-4) when the 10 best families were selected. This number increases from 18 to 21 and 47 to 52 depending on the trait considered when the 20 and 50 best families were selected, respectively.

3.5 Discussion

Genomic selection can have a significant impact in terms of improving genetic gain in plant breeding programs (Bernardo and Yu 2007; Heffner, Jannink, and Sorrells 2011; Rutkoski et al. 2012; Sorrells 2015; Grenier et al. 2015). Currently, phenotyping is one of the most challenging and costly activities in breeding programs. Genomic assisted breeding has been advocated as a major player to develop climate-smart and nutrient dense crop cultivars in a cost- and time-efficient manner (Varshney et al. 2021). However, even in the context of genomic selection, it is still important to find a way to reduce phenotyping efforts, however not at the cost of a reduced predictive ability (PA) of prediction models. The constitution of the training set (TS) to calibrate the prediction models have been shown to strongly influence the PA values (Spindel et al. 2015; Berro et al. 2019; Merrick et al. 2022). The overall objective of the present study was to assess whether genomic prediction (GP) could improve the recurrent selection scheme in the actual CIAT-Cirad program. Specifically, we wanted to investigate which TS with which GP models based on the infrastructure of the program would allow the highest PA and the possibility to achieve higher genetic gain. To do this, several scenarios were developed and tested through their ability to predict GEBVs of families of a population (PCT27B) at a specific generation ($S_{0:4}$) phenotyped at one target site (SRO). Thus, different prediction models were considered depending on the inclusion data originating from two locations, the consideration of the G×E interaction, the addition of data from a different set of progenies derived from the original population (PCT27), and the generation, size and composition of the TS.

3.5.1 Predictive ability in a single environment and in a single population

The variance decomposition and PA of the $S_{0:2}$ and $S_{0:3}$ generations of PCT27A have been analysed in Baertschi et al. (2021) and will not be discussed in this study (STable 3-5), except to compare them to the results we obtained on the subset of progenies PCT27B at the $S_{0:4}$ generation. Although the models used to estimate the variances were not exactly similar, the results were close. The PA obtained for the prediction of $S_{0:4}$ families at SRO were relatively low compared to those previously estimated in the literature (reviewed by Ahmadi et al. 2020). Compared to the PA from the cross-validation of Baertschi et al. (2021) on the $S_{0:2}$ and $S_{0:3}$ generations (PCT27A), the PA estimates were lower for the $S_{0:4}$ in PCT27B (STable 3-6). The only exception was the PA for YLD where the values were comparable between the generations of the families evaluated, as well as to estimates made on other selection programs (Spindel et al. 2015; Grenier et al. 2015; Morais Júnior et al. 2018; Yang Xu et al. 2021). For ZN, Baertschi et al. (2021) obtained PA of 0.26 and 0.24 in the $S_{0:2}$ and $S_{0:3}$ generations and the estimate

was 0.17 in the $S_{0:4}$ generation, showing a high decrease in advanced generations. Early generation GP using phenotypes measured at the $S_{0:2}$ generation appeared to be more accurate than using phenotypes of families measured at later generation. One potential explanation could be that the year had a relatively high effect which randomly resulted in higher PA when the calibrations was based on $S_{0:2}$ rather than on $S_{0:3}$. Another potential explanation could be that the model is calibrated with genomic information of S_0 plants, thus the segregating families used to generate the phenotypic dataset are closer to their original parent, and the association between the genomic information and the phenotypic expression in the derived families is better. This observation holds for all traits but YLD for which a potential loss of allelic diversity during the three cycles of generation advances by selfing and bulk harvest could have had a lesser impact than for the more oligogenic traits. Other explanations can also be suggested to explain this phenomenon of reduced PA at more advanced generation, such as the probability of error during the generation advance phase. The unexpected gene flow from mistakenly collected seed from sterile plants segregating in the families could also have induced an error rate in the estimation of the GEBVs on the more advanced generations. Nevertheless, this observation of better PA achieved with phenotypic data from $S_{0:2}$ families is of great interest as it suggests that phenotyping for model calibration could be performed as early as possible in the recurrent selection breeding scheme.

3.5.2 Potential to increase intensity of selection

One of the great potentials of GS lies in its ability to increase the selection intensity (Heffner et al. 2009; Hunt et al. 2018; Cobb et al. 2019; R2D2 Consortium et al. 2021). Phenotypes of PCT27B $S_{0:4}$ families in SRO were predicted with a TS consisting of individuals from another S_0 progeny (PCT27A) at the $S_{0:2}$ generation with better accuracy than with a model using these PCT27B $S_{0:4}$ families in the TS. Therefore, in population breeding with thousands of S_0 plants available, it appears possible to calibrate a model with early generation families ($S_{0:2}$) derived from a set of a few hundred S_0 progenies to predict the value of $S_{0:4}$ families in the rest of the population. The result was surprising given that predicting $S_{0:4}$ at SRO with a model built with the same genetics (PCT27B families) gave a lower PA than using different set of S_0 progeny for calibration (PCT27A families), with additional potential year and GxE interaction effects. One potential reason could that $S_{0:2}$ families are less fixed than $S_{0:4}$, their progeny means are more representative of the PCT27 at large, thus more adequate to predict a different subset of the population. Obviously, biases due to GxE is another potential explanation that cannot be excluded. Nevertheless, the Uni2 scenario seems promising to include GP in the current CIAT-Cirad breeding program, with early calibration of the genomic model based on fewer families than the number of potential candidates for selection (the genotyped S_0 plants).

However, there is still room for improvement, notably in terms of speed and cost of the program. Two phenotyping locations are available, one being a site where rice can be grown all year around, which raises the question of whether the sparse phenotyping could be applied the in the CIAT-Cirad breeding program.

3.5.3 Interest in considering GxE interactions

Although SRO is the target selection site, it is far away from CIAT-HQ, and more complex to manage within the research activities. PAL being a more practical location for conducting field location, our objective was to concentrate the phenotyping efforts on the surrogate site while keeping relevance for the target site. In the $S_{0:4}$ generation for each trait, phenotypic correlations between the two locations were relatively low, and lower than reported in the earlier generation except for YLD (Baertschi et al. 2021). This low phenotypic correlation between location suggests a high genotype-by-environmental effect and makes accurate prediction across sites more difficult. (Hunt et al. 2018). The value of sparse testing – multi-environment trial in which some families are phenotyped in all locations while others are phenotyped in only one location – is clear. It allows a reduction of investments in phenotyping in multiple locations. Instead of evaluating the whole population in each environment, the population is divided in sets each evaluated in another location (Burgueño, de los Campos, et al. 2012; Jarquín et al. 2014a; Lopez-Cruz et al. 2015; Ben Hassen, Bartholome, et al. 2018; Millet et al. 2019a). However, this option, despite its strong economic incentive, must be carefully considered if phenotypic correlation between locations is not high. Therefore, one of the objectives of the present study was to evaluate the possibility to combine a reduced phenotyping effort at SRO and complete phenotyping of families at PAL. The sparse testing design within a population, holds potential in the context of the CIAT-Cirad breeding program as it could be possible to reduce the phenotyping effort at SRO and even increase the PA for FL, PH and ZN. Only for YLD, the PA of the multi-environment model assuming no GxE interaction (MM model) was significantly reduced compared to the single site model (SM) or any of the other models including GxE interactions (MDs and MDe models). This was also the only trait which did not benefited from an evaluation in surrogate site in Baertschi et al. (2021). This confirms that for complex traits with low site correlation, there is no added value in phenotyping in more location to calibrate the GP model. On the contrary, calibration with multisite data improved the PA for FL, PH and ZN, but only in cases where no GxE or GxE with a single variance was considered (MM and MDs). The calibration models considering the environment-specific variance (MDe) were similar to the GP model developed in Baertschi et al. (2021). On a different set of S_0 progenies from the same PCT27 population, larger PA for all the traits were obtained when data from the whole population phenotyped in the surrogate site were included in the calibration. However, while for FL, PH and ZN the calibration model significantly increased the PA in the IMB strategy (334 and 200 families

at PAL and SRO, respectively) including environment specific variance of Baertschi et al. (2021), such was not the case in our study. The PA using MDe model to predict PH and ZN was lower than those achieved with any other models within the Multi1 scenario (Figure 3-3), and to the previously reported study, likely because of a strong reduction in correlation between sites, or a difficulty inherent to the use of the $S_{0:4}$ generation of families, as mentioned earlier (Table 3-3).

Despite this, using only the PCT27B $S_{0:4}$ generation, the multi-environment models (Multi1) allow for an increase in the number of families included in the TS (all the PCT27B, i.e. 334 families phenotyped in PAL) compared to the single-environment model (Uni1) where the TS consisted of 70 % of PCT27B (i.e. 234 families phenotyped in SRO), which may also have an impact on increasing the PA of the models. Combining phenotypes from more locations acquired on early generations of a set of S_0 progenies to predict on a larger set of S_0 genotypes would have a greater impact on genetic gain, as it would increase the intensity of selection, reduce time to selection and potentially, increase the accuracy of predictions.

3.5.4 The inclusion of GxE interactions in a multi-generation model

PAL is an ideal location to produce a large amount of high-quality seeds due to the optimal conditions all year around and thus the lack of stress impacting rice productivity. The use of families phenotyped in the $S_{0:2}$ generation at PAL and a subset of those in the $S_{0:3}$ generation at SRO in the TS (Multi2 scenario) was set as a scenario of interest to test. This sequence was proposed as growing the $S_{0:2}$ during the off season in PAL allows gathering of phenotypic data and production of seeds for the evaluation of $S_{0:3}$ in SRO during the main season. Overall, as described above, the higher the proportion of $S_{0:3}$ at SRO included in the TS, the higher the PA, in line with what is commonly reported that larger TS improves PA (Ahmadi et al. 2020). Regardless of the model and trait, the PA using the Multi2 scenario are always lower than the PA using the Uni2 scenario where the TS consisted of $S_{0:2}$ families grown at SRO, with a reduction of PA ranging from 9.1 to 29 % when the best model is considered in the Multi2 scenario. These relatively strong reductions in the present analyses can be explained by different points already mentioned: i) the low correlations between locations; ii) the fact that $S_{0:4}$ can be further distant from the S_0 plants and some genome/phenotype relations are missed; iii) the presence of environment × year interaction effects.

One way to potentially improve the PA in this Multi2 scenario would be to optimize the choice of $S_{0:3}$ families grown at SRO to be include in the TS. Several studies have demonstrated an improved PA, when the choice of individuals to be included in the TS was optimized using a specific optimization method (Rincent et al. 2012; Akdemir, Sanchez, and Jannink 2015; Akdemir, Rio, and Isidro y Sánchez 2021; Mangin et al. 2019; Isidro y Sánchez and Akdemir 2021) such as CDmean (Rincent et al. 2012), based on the variance of prediction error derived from the realized additive relationship -BLUP model.

Improved PA with optimized TS can thus conduct to reduce the phenotypic effort without reducing the power of GP. Such an optimization of the $S_{0:3}$ included in the TS resulted in a higher PA for FL (maximum a + 0.08 for the MDe models and 25% of $S_{0:3}$ in the TS), but it did not significantly improve the PA of PH, YLD and ZN, regardless of the model and proportion of $S_{0:3}$ considered.

3.5.5 Impact of the GP models on the family ranking

As we have seen, the choice of the genomic prediction models will have an impact on PA values. These PA values are only used to describe the achieved correlations between BLUPs and GEBVs, which may or may not lead to changes in the choice of families to select for the next cycle of recombination (Blondel et al. 2015; Mendonça et al. 2020). Regardless of the method and the model used to predict GEBVs, the objective of all these predictions is to rank individuals and select the best ones to be used as parents for the next cycle of selection.

Overall, the different models had a good ability to select the same individuals at the top of the ranking. The ranking of individuals was relatively similar between the prediction methodologies, which leads to the conclusion that the methodologies and models used to predict GEBVs will not have a substantial impact in the actual genetic gain of the breeding program.

3.5.6 Economic impact of the different scenarios

The scheme is based on two parts: the RS for population improvement and the pedigree breeding for genetic fixation and selection of candidates for variety release. A strategy opted for in this scheme of variety development is to advance the selected families to a relatively good level of genetic fixation ($S_{0:4}$) by bulk harvest in order to maintain the variability within the family, prior to proceed to two generations of pedigree breeding. While advancing generation in PAL, this material is used to calibrate the model. In recurrent selection, the possibility to select quickly the best families through progeny testing will help recycle faster and recombine the best selected candidates to improve the population. Ultimately, we want to have a rapid, easy, and cost-effective way to select the best families for recombination and for generation advance in order to develop new cultivars. This was the rational for testing all the strategies presented in this work for which we compared scenarios based on various uses of surrogate sites, or sparse testing in the two locations. Our findings reveal that PA was greater when performing calibration with the Uni2 scenario, compared to Uni1, Uni3 or any Multi scenario. Yet, in the top 50 best ranked families, 26 to 47% were similar between Uni2 and the best Multi2 scenario including 50% of $S_{0:3}$ at SRO in the TS (STable 3-4). Comparing our five scenarios in terms of time spent for the calibration, while the GP model could be built in 1.5 years for the Uni2, Uni3 and

Table 3-4: Time and cost for each scenario to generate the material (generation advance) and phenotype the training set (TS) to calibrate a genomic prediction (GP) model and produce the generation on which to start pedigree breeding scheme. Cost X$PAL and X$SRO are unit price for the phenotyping of 1200 plots in Palmira (PAL) and Santa Rosa (SRO), respectively.

Scenario	Season*	Generation ¶	TS size	Generation advance in PAL	Phenotyping		GP	Total cost
					PAL	SRO		
Uni1	Yr1-A	$S_{0:1}$ → $S_{0:2}$	100%	0.4X$$_{PAL}$$				1.2X$$_{PAL}$$ + 1X$$_{SRO}$$
	Yr1-B	$S_{0:2}$ → $S_{0:3}$	100%	0.4X$$_{PAL}$$				
	Yr2-A	$S_{0:3}$ → $S_{0:4}$	100%	0.4X$$_{PAL}$$				
	Yr2-B	no activity						
	Yr3-A	$S_{0:4}$ → $S_{0:5}$	100%			1X$$_{SRO}$$	GP	
Uni2	Yr1-A	$S_{0:1}$ → $S_{0:2}$	100%	0.4X$$_{PAL}$$				0.9X$$_{PAL}$$ + 1X$$_{SRO}$$
	Yr1-B	$S_{0:2}$ → $S_{0:3}$	100%	0.4X$$_{PAL}$$				
	Yr2-A	$S_{0:2}$ → $S_{0:3}$	100%			1X$$_{SRO}$$	GP	
	Yr2-B	$S_{0:3}$ → $S_{0:4}$	sel. fam	0.05X$$_{PAL}$$				
Uni3	Yr1-A	$S_{0:1}$ → $S_{0:2}$	100%	0.4X$$_{PAL}$$				0.8X$$_{PAL}$$ + 1X$$_{SRO}$$
	Yr1-B	$S_{0:2}$ → $S_{0:3}$	100%	0.4X$$_{PAL}$$				
	Yr2-A	$S_{0:3}$ → $S_{0:4}$	100%			1X$$_{SRO}$$	GP	
Multi1	Yr1-A	$S_{0:1}$ → $S_{0:2}$	100%	0.4X$$_{PAL}$$				2.2X$$_{PAL}$$ + 1X$$_{SRO}$$
	Yr1-B	$S_{0:2}$ → $S_{0:3}$	100%	0.4X$$_{PAL}$$				
	Yr2-A	$S_{0:3}$ → $S_{0:4}$	100%	0.4X$$_{PAL}$$				
	Yr2-B	$S_{0:4}$ → $S_{0:5}$	100%		1X$$_{PAL}$$			
	Yr3-A	$S_{0:4}$ → $S_{0:5}$	100%			1X$$_{SRO}$$	GP	
Multi2	Yr1-A	$S_{0:1}$ → $S_{0:2}$	100%	0.4X$$_{PAL}$$				1.4X$$_{PAL}$$ + 0.6X$$_{SRO}$$
	Yr1-B	$S_{0:2}$ → $S_{0:3}$	100%		1X$$_{PAL}$$			
	Yr2-A	$S_{0:3}$ → $S_{0:4}$	50%			0.6X$$_{SRO}$$	GP	

* year-semester; ¶ generation planted → generation harvested

Multi2 scenario, it took 3 years for Uni1 and Multi1 scenarios (Table 3-4). This means that the calibration work needed for prediction and selection of the best candidates for RS can be reduced, by phenotyping families as early as possible. In terms of cost, we fixed a reference cost per location as the cost of the trials we conducted for this study, being 1X$$_{PAL}$$ and 1X$$_{SRO}$$ for the 1200 plots in PAL and SRO, respectively. The phenotyping involving families at a more advanced generation will necessarily result in higher cost due to the need for multiplication steps (although still less costly than a phenotyping step (0.4X$$_{PAL}$$)) rather than for an evaluation as no replicated design and no phenotyping is involved) and the evaluation of the advanced generation families (Table 3-4). If phenotyping in the target site is more costly than in the surrogate location either because it involves that the breeders travel and be hosted in a different city, or because it requires a particular management due to high pathogen pressure that would impact the evaluation of grain quality traits, or it implicates some risks due to abiotic constraints, the multi-location (Multi2) strategy can be of interest to cost saving as only

a fraction of the population is phenotyped in the target site. Furthermore, having two sites will allow, if problem occurs, to still have a phenotyping record on the population in one location. In our case, $1X\$_{SRO}$ >> $1X\$_{PAL}$ and Multi2 scenario with one year being enough to phenotype the whole population at $S_{0:2}$ at PAL and 50% of the population at $S_{0:3}$ at SRO was the most economic ($1.4X\$_{PAL}$ + $0.6X\$_{SRO}$), even though the PAs were lower (from 0.23 to 0.30) than with the Uni2 (from 0.31 to 0.39) for the four traits considered. Uni2 was the best scenario in terms of PA and the cost estimated at $0.9X\$_{PAL}$ + $1X\$_{SRO}$. Thus, the question comes whether the saving resulting from using the Multi2 scenario (about 4,000 USD) in comparison with Uni2 is worth it, considering the latter had higher PA, notably for PH (+0.1). Yet the question does not stop at this observation as one should also consider the environmental risks in the target site for losing a season, which will be the only source of phenotypes for calibrating the model. There is also the possibility to further optimize the scheme by phenotyping a reduced set of the population in the two locations with a common fraction in both locations, as in the BAL2 scenario of Baertschi et al. (2021), but this was not tested in the current study.

The inclusion of the GP in our breeding scheme also has to include a model update to ensure that the GP model stays relevant while the population improves through the recurrent cycles of genomic selection. The cost of the breeding program will thus have to include this step of recurrent model update. This is currently being tested with a simulation approach.

Our study revealed that phenotype measured as early as S02 have some predictive ability for later generation phenotypes. Based only on the PA, the best approach is still to only phenotype in the target site. However, considering practical concern such as securing the availability of data for selection, multi-environment calibration might have a place in the breeding program.

3.6 Literature cited

Ahmadi N, Bartholomé J, Cao T-V, Grenier C (2020) Genomic selection in rice: empirical results and implications for breeding. pp 243–258

Akdemir D, Rio S, Isidro Sanchez J (2021a) TrainSel Usage

Akdemir D, Rio S, Isidro y Sánchez J (2021b) TrainSel: An R Package for Selection of Training Populations. Front Genet 12:655287. https://doi.org/10.3389/fgene.2021.655287

Akdemir D, Sanchez JI, Jannink J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. Genetics Selection Evolution 47:38. https://doi.org/10.1186/s12711-015-0116-6

Baertschi C, Cao T-V, Bartholomé J, et al (2021) Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. G3 Genes|Genomes|Genetics. https://doi.org/10.1093/g3journal/jkab320

Bartholomé J, Prakash PT, Cobb JN (2021) Genomic prediction: progress and perspectives for rice improvement. arXiv:210914781 [q-bio]

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:1–48. https://doi.org/10.18637/jss.v067.i01

Ben Hassen M, Bartholome J, Valè G, et al (2018a) Genomic prediction accounting for genotype by environment interaction offers an effective framework for breeding simultaneously for adaptation to an abiotic stress and performance under normal cropping conditions in rice. G3 - Genes Genomes Genetics. https://doi.org/<https://doi.org/10.1534/g3.118.200098>

Ben Hassen M, Cao T-V, Bartholome J, et al (2018b) Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theoretical and Applied Genetics. https://doi.org/<https://doi.org/10.1007/s00122-017-3011-4>

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science 47:1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Berro I, Lado B, Nalin RS, et al (2019) Training Population Optimization for Genomic Selection. The Plant Genome 12:190028. https://doi.org/10.3835/plantgenome2019.04.0028

Bhandari A, Bartholomé J, Cao-Hamadoun T-V, et al (2019) Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLOS ONE 14:e0208871. https://doi.org/10.1371/journal.pone.0208871

Blondel M, Onogi A, Iwata H, Ueda N (2015) A Ranking Approach to Genomic Selection. PLOS ONE 10:e0128570. https://doi.org/10.1371/journal.pone.0128570

Burgueño J, Campos G de los, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Science 52:707–719. https://doi.org/10.2135/cropsci2011.06.0299

Cobb JN, Juma RU, Biswas PS, et al (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theor Appl Genet 132:627–645. https://doi.org/10.1007/s00122-019-03317-0

Crossa J, Beyene Y, Kassa S, et al (2013) Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. G3 Genes|Genomes|Genetics 3:1903–1926. https://doi.org/10.1534/g3.113.008227

Crossa J, Campos G de L, Pérez P, et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724. https://doi.org/10.1534/genetics.110.118521

Crossa J, de los Campos G, Maccaferri M, et al (2016) Extending the Marker × Environment Interaction Model for Genomic-Enabled Prediction and Genome-Wide Association Analysis in Durum Wheat. Crop Science 56:2193–2209. https://doi.org/10.2135/cropsci2015.04.0260

Cuevas J, Crossa J, Montesinos-López OA, et al (2017) Bayesian Genomic Prediction with Genotype × Environment Interaction Kernel Models. G3 (Bethesda) 7:41–53. https://doi.org/10.1534/g3.116.035584

Cuevas J, Crossa J, Soberanis V, et al (2016) Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. The Plant Genome 9:0. https://doi.org/10.3835/plantgenome2016.03.0024

Endelman JB, Atlin GN, Beyene Y, et al (2014) Optimal Design of Preliminary Yield Trials with Genome-Wide Markers. Crop Science 54:48–59. https://doi.org/10.2135/cropsci2013.03.0154

Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics. 4th edition. Longman Scientific & Technical, Burnt Mill, Harlow, United Kingdom.

Frouin J, Filloux D, Taillebois J, et al (2014) Positional cloning of the rice male sterility gene ms-IR36, widely used in the inter-crossing phase of recurrent selection schemes. Molecular Breeding 33:555–567. https://doi.org/10.1007/s11032-013-9972-3

Gianola D, van Kaam JBCHM (2008) Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics 178:2289–2303. https://doi.org/10.1534/genetics.107.084285

Granato I, Cuevas J, Luna-Vázquez F, et al (2018) BGGE: A New Package for Genomic-Enabled Prediction Incorporating Genotype × Environment Interaction Models. G3 (Bethesda) 8:3039–3047. https://doi.org/10.1534/g3.118.200435

Grenier C, Cao T-V, Ospina Y, et al (2015) Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. PLOS ONE 10:e0136594. https://doi.org/10.1371/journal.pone.0136594

Harrell Jr FE (2021) Hmisc: Harrell Miscellaneous. R package version 4.6-0

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. The Plant Genome 4:65–75. https://doi.org/10.3835/plantgenome2010.12.0029

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Science 50:1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. Crop Science 49:1–12. https://doi.org/10.2135/cropsci2008.08.0512

Hunt CH, van Eeuwijk FA, Mace ES, et al (2018) Development of Genomic Prediction in Sorghum. Crop Science 58:690–700. https://doi.org/10.2135/cropsci2017.08.0469

Isidro J, Jannink J-L, Akdemir D, et al (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128:145–158. https://doi.org/10.1007/s00122-014-2418-4

Isidro y Sánchez J, Akdemir D (2021) Training Set Optimization for Sparse Phenotyping in Genomic Selection: A Conceptual Overview. Frontiers in Plant Science 12:

Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1

Jarquin D, Howard R, Crossa J, et al (2020) Genomic Prediction Enhanced Sparse Testing for Multi-environment Trials. G3: Genes, Genomes, Genetics 10:2725–2739. https://doi.org/10.1534/g3.120.401349

Jarquín D, Lemes da Silva C, Gaynor RC, et al (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. Plant Genome 10:. https://doi.org/10.3835/plantgenome2016.12.0130

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3 (Bethesda) 5:569–582. https://doi.org/10.1534/g3.114.016097

Lorenz AJ, Smith K p., Jannink J-L (2012) Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. Crop Science 52:1609–1621. https://doi.org/10.2135/cropsci2011.09.0503

Mangin B, Rincent R, Rabier C-E, et al (2019) Training set optimization of genomic prediction by means of EthAcc. PLOS ONE 14:e0205629. https://doi.org/10.1371/journal.pone.0205629

Martinez CP, Torres EA, Châtel M, et al (2014) Rice Breeding in Latin America. In: Plant Breeding Reviews: Volume 38. https://agritrop.cirad.fr/575285/. Accessed 23 Mar 2022

Mendonça L de F, Galli G, Malone G, Fritsche-Neto R (2020) Genomic prediction enables early but low-intensity selection in soybean segregating progenies. Crop Science 60:1346–1361. https://doi.org/10.1002/csc2.20072

Merrick LF, Herr AW, Sandhu KS, et al (2022) Optimizing Plant Breeding Programs for Genomic Selection. Agronomy 12:714. https://doi.org/10.3390/agronomy12030714

Millet EJ, Kruijer W, Coupel-Ledru A, et al (2019) Genomic prediction of maize yield across European environmental conditions. Nat Genet 51:952–956. https://doi.org/10.1038/s41588-019-0414-y

Morais Júnior OP, Breseghello F, Duarte JB, et al (2018) Assessing Prediction Models for Different Traits in a Rice Population Derived from a Recurrent Selection Program. Crop Science 58:2347. https://doi.org/10.2135/cropsci2018.02.0087

Onogi A, Ideta O, Inoshita Y, et al (2015) Exploring the areas of applicability of whole-genome prediction methods for Asian rice (Oryza sativa L.). Theor Appl Genet 128:41–53. https://doi.org/10.1007/s00122-014-2411-y

R Development Core Team (2018) R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/.

R2D2 Consortium, Fugeray-Scarbel A, Bastien C, et al (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. Frontiers in Genetics 12:629737

Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (Zea mays L.). Genetics 192:715–728. https://doi.org/10.1534/genetics.112.141473

Rutkoski J, Benson J, Jia Y, et al (2012) Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. The Plant Genome 5:. https://doi.org/10.3835/plantgenome2012.02.0001

Sorrells ME (2015) Genomic Selection in Plants: Empirical Results and Implications for Wheat Breeding. In: Ogihara Y, Takumi S, Handa H (eds) Advances in Wheat Genetics: From Genome to Field. Springer Japan, Tokyo, pp 401–409

Spindel J, Begum H, Akdemir D, et al (2015) Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. PLOS Genetics 11:e1004982. https://doi.org/10.1371/journal.pgen.1004982

Spindel J, Iwata H (2018) Genomic Selection in Rice Breeding. In: Sasaki T, Ashikari M (eds) Rice Genomics, Genetics and Breeding. Springer, Singapore, pp 473–496

Tanaka R, Iwata H (2018) Bayesian optimization for genomic selection: a method for discovering the best genotype among a large number of candidates. Theor Appl Genet 131:93–105. https://doi.org/10.1007/s00122-017-2988-z

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Varshney RK, Bohra A, Yu J, et al (2021) Designing Future Crops: Genomics-Assisted Breeding Comes of Age. Trends in Plant Science 26:631–649. https://doi.org/10.1016/j.tplants.2021.03.010

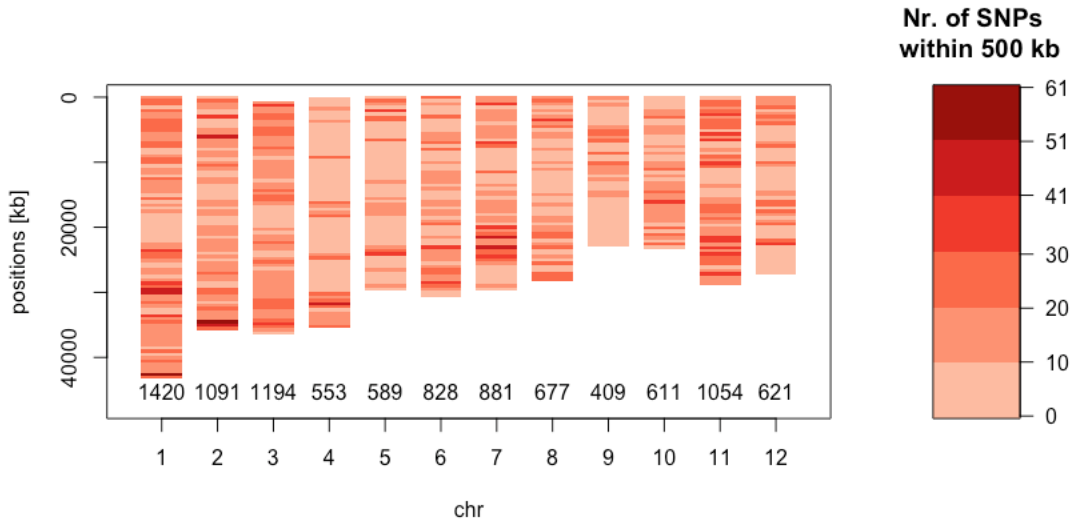
Wang X, Li L, Yang Z, et al (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118:302–310. https://doi.org/10.1038/hdy.2016.87

Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 28:2086–2087. https://doi.org/10.1093/bioinformatics/bts335

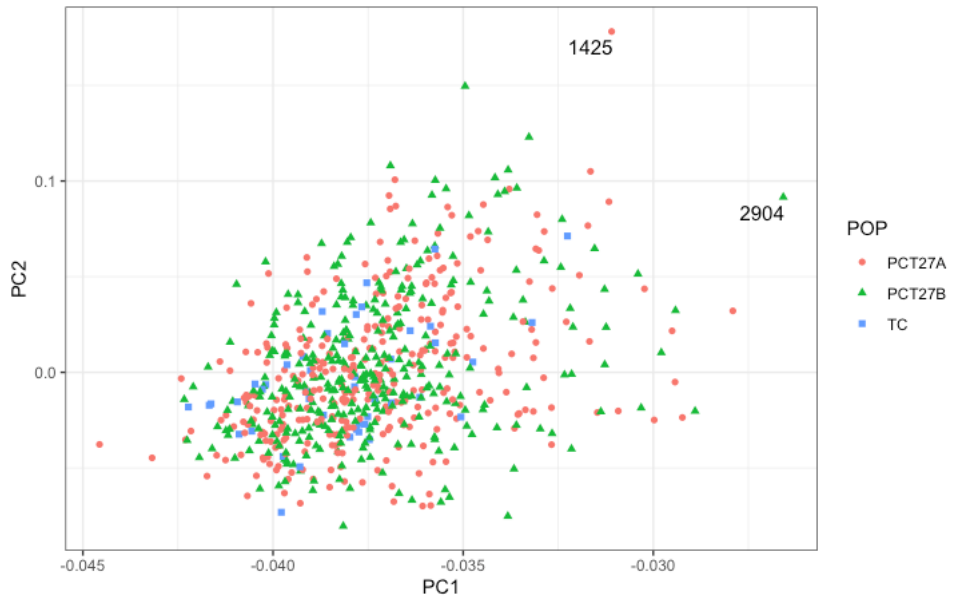
Xu Y, Ma K, Zhao Y, et al (2021) Genomic selection: A breakthrough technology in rice breeding. The Crop Journal 9:669–677. https://doi.org/10.1016/j.cj.2021.03.008

Zhao Y, Gowda M, Liu W, et al (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776. https://doi.org/10.1007/s00122-011-1745-y

3.7 Supplementary Figures



SFig 3-1: Density of SNP markers in the two populations (PCT27A and PCT27B) and the TC set (713 S0 plants) in the 12-chromosome R package Synbreed (Wimmer et al. 2012)



SFig 3-2: Biplot from PCA performed on 7,766 SNP (after prunning) and 713 S0 plants (PLINK). Grouping by colour of PCT27A, PCT27B and the temporal checks (TC belonging to PCT27A)

3.8 Supplementary Tables

STable 3-1: : Genetic characterization of the two TS together (genotypes of the 668 S0 plants). A) Summary information on the distribution, MAF and heterozygosity of the 9 928 SNP loci. B) Observed heterozygosity (Ho) among the 668 genotypes

A

		The rice 12 chromosomes												
	Chr	Os01	Os02	Os03	Os04	Os05	Os06	Os07	Os08	Os09	Os10	Os11	Os12	All
	Size (bp)	43 116 429	35 882 519	35 609 111	35 179 110	29 918 859	30 979 888	29 632 986	28 363 938	22 733 337	23 066 693	28 685 980	27 488 622	370 657 472
	Number	1420	1091	1194	553	589	828	881	677	409	611	1054	621	129 928
	Density	30 408	32 897	30 489	63 999	50 857	37 441	33 659	41 924	55 941	37 906	27 331	44 280	40 594
Distribution of distances (bp) between two adjacent SNP loci	Minimum	1	1	1	1	1	1	1	1	1	1	1	1	1
	1st Quartile	32	34	40	19	16	17	18	32	17	27	12	21	23
	Median	10 143	11 494	10 538	8 953	6 676	7 492	7 123	15 476	9 666	12 314	3 094	10 400	9 082
	Average	30 385	32 920	29 848	63 730	50 882	37 461	33 674	41 958	55 719	37 814	27 242	44 336	37 384
	3rd Quartile	35 317	38 327	39 304	48 947	39 108	38 980	36 728	49 010	64 973	48 416	33 443	53 539	41 156
	Maximum	2 197 892	711 533	471 228	1 951 759	2 209 631	1 196 222	1 499 537	773 794	1 188 875	677 358	528 305	792 342	2 209 631
Distribution of the minor allele frequency (MAF)	Minimum	0.021	0.025	0.022	0.024	0.022	0.019	0.026	0.026	0.022	0.020	0.026	0.024	0.023
	1st Quartile	0.063	0.066	0.072	0.085	0.055	0.103	0.054	0.136	0.090	0.081	0.084	0.094	0.082
	Median	0.135	0.159	0.166	0.187	0.116	0.144	0.155	0.238	0.167	0.121	0.178	0.155	0.160
	Average	0.184	0.201	0.198	0.224	0.170	0.200	0.205	0.245	0.185	0.170	0.219	0.203	0.200
	3rd Quartile	0.274	0.306	0.287	0.385	0.266	0.288	0.332	0.353	0.237	0.229	0.356	0.292	0.300
	Maximum	0.500	0.500	0.500	0.500	0.499	0.496	0.499	0.499	0.497	0.500	0.500	0.497	0.499
Distribution of heterozygosity	Minimum	4%	5%	4%	5%	4%	4%	5%	4%	4%	4%	4%	3%	4.1%
	1st Quartile	11%	12%	13%	16%	11%	18%	10%	22%	17%	14%	14%	17%	14.6%
	Median	23%	26%	27%	30%	20%	24%	26%	35%	27%	21%	28%	26%	26.1%
	Average	27%	30%	29%	34%	28%	29%	29%	34%	31%	26%	30%	30%	29.8%
	3rd Quartile	39%	42%	40%	46%	41%	42%	45%	44%	33%	36%	43%	40%	41.0%
	Maximum	98%	100%	99%	100%	99%	99%	99%	98%	99%	100%	100%	96%	98.9%

B

	Ho
Min.	0.07887
1st Qu.	0.27065
Median	0.29543
Mean	0.29444
3rd Qu.	0.32061
Max.	0.44309

Bin of Ho	Frequency	Percentage	Cumulative frequency	Cumulative percentage
[0.07 - 0.108[	3	0.4	3	0.4
[0.108 - 0.146[	5	0.7	8	1.1
[0.146 - 0.184[	11	1.5	19	2.7
[0.184 - 0.222[	9	1.3	28	3.9
[0.222 - 0.26[	87	12.2	115	16.1
[0.26 - 0.298[	263	36.9	378	53
[0.298 - 0.336[	227	31.8	605	84.9
[0.336 - 0.374[	86	12.1	691	96.9
[0.374 - 0.412[	19	2.7	710	99.6
[0.412 - 0.45]	3	0.4	713	100

STable 3-2: Average linkage disequilibrium (r2) between marker pairs according to chromosomes and the distance between markers, considering loci with MAF >2.5%. In italics are r2 with values less than initial r2/2

	The rice 12 chromosomes													
Distance range (kb) between markers	**Os01**	**Os02**	**Os03**	**Os04**	**Os05**	**Os06**	**Os07**	**Os08**	**Os09**	**Os10**	**Os11**	**Os12**	**Average**	**std**
]0:25]	0.517	0.537	0.622	0.433	0.213	0.447	0.528	0.566	0.638	0.453	0.424	0.497	0.49	0.112
]25:50]	0.428	0.483	0.607	0.476	0.215	0.372	0.468	0.536	0.592	0.406	0.344	0.432	0.447	0.108
]50:75]	0.442	0.478	0.568	0.343	0.358	0.457	0.465	0.451	0.478	0.411	0.363	0.427	0.437	0.063
]75:100]	0.407	0.456	0.582	0.305	0.201	0.437	0.424	0.446	0.44	0.389	0.353	0.424	0.405	0.092
]100:150]	0.384	0.439	0.511	0.313	0.26	0.444	0.355	0.463	0.439	0.38	0.303	0.42	0.393	0.074
]150:200]	0.366	0.416	0.461	0.269	0.472	0.443	0.404	0.46	0.429	0.359	0.289	0.353	0.394	0.067
]200:250]	0.319	0.377	0.448	0.304	0.367	0.397	0.321	0.449	0.401	0.364	0.273	0.38	0.367	0.055
]250:300]	0.308	0.361	0.431	0.295	0.321	0.384	0.331	0.381	0.433	0.389	0.268	0.332	0.353	0.052
]300:400]	0.274	0.323	0.393	0.235	0.309	0.339	0.296	0.357	0.379	0.312	0.234	0.292	0.312	0.05
]400:500]	0.243	0.27	0.376	0.212	0.275	0.329	0.277	0.338	0.348	0.292	0.218	0.23	0.284	0.054
]500:750]	0.193	0.227	0.338	0.204	0.213	0.257	0.225	0.301	0.361	0.257	0.196	0.224	0.25	0.056
]750:1000]	0.158	0.191	0.259	0.159	0.107	0.207	0.176	0.258	0.336	0.2	0.169	0.147	0.197	0.062

STable 3-3: Phenotypic correlations between years for the 50 temporal checks repeated in all trials in SRO. Means are in diagonal.

FL	**$S_{0:2}$**	**$S_{0:3}$**	**$S_{0:4}$**
$S_{0:2}$	86.5	0.53	0.69
$S_{0:3}$		88.1	0.61
$S_{0:4}$			86.4

PH	**$S_{0:2}$**	**$S_{0:3}$**	**$S_{0:4}$**
$S_{0:2}$	128.5	0.72	0.75
$S_{0:3}$		127.2	0.62
$S_{0:4}$			128.4

YLD	**$S_{0:2}$**	**$S_{0:3}$**	**$S_{0:4}$**
$S_{0:2}$	701.7	0.64	0.52
$S_{0:3}$		712.8	0.57
$S_{0:4}$			715.2

ZN	**$S_{0:2}$**	**$S_{0:3}$**	**$S_{0:4}$**
$S_{0:2}$	14.8	0.79	0.80
$S_{0:3}$		15.1	0.71
$S_{0:4}$			14.9

STable 3-4: Number of families selected included in the 10, 20 or 50 best ones according to their estimated GEBVs in all the methods of the Multi2 scenario

	Selection of the 10 best families			**Selection of the 20 best families**			**Selection of the 50 best families**		
Trait	Number of families selected at least once	Number of families selected in at least 50% of the models	Number of families selected in all the models	Number of families selected at least once	Number of families selected in at least 50% of the models	Number of families selected in all the models	Number of families selected at least once	Number of families selected in at least 50% of the models	Number of families selected in all the models
FL	25	10	2	40	18	7	95	48	21
PH	29	9	0	54	18	2	104	52	10
YLD	27	8	1	44	21	3	87	48	26
ZN	36	9	0	68	19	1	120	47	4

STable 3-5: Variance decomposition and broad sense heritability (H^2) obtained using Model 2 by trait and generation

		PCT27A $S_{0:2}$			PCT27A $S_{0:3}$		
Trait	**Variance component**	**Variance**	**Proportion**	**H^2**	**Variance**	**Proportion**	**H^2**
FL	Bloc	0.03	0.11	0.57	0.97	4.23	0.75
	Genotype	9.68	34.57		8.42	36.75	
	Location:Genotype	12.51	44.68		1.92	8.38	
	Bloc:Rep:Location	0.2	0.71		0.66	2.88	
	Residuals	5.58	19.93		10.94	47.75	
PH	Bloc	1.98	2.57	0.68	4.45	7.84	0.77
	Genotype	23.56	30.62		22.25	39.18	
	Location:Genotype	8.77	11.40		4.96	8.73	
	Bloc:Rep:Location	2.02	2.63		<0.001	0.00	
	Residuals	40.62	52.79		25.13	44.25	
YLD	Bloc	22.31	0.12	0.38	845.19	8.77	0.21
	Genotype	2354.9	12.14		516.48	5.36	
	Location:Genotype	3908.6	20.15		1872.05	19.41	
	Bloc:Rep:Location	1498	7.72		392.82	4.07	
	Residuals	11611	59.87		6015.79	62.39	
ZN	Bloc	0.279	4.69	0.53	<0.001	0.00	0.57
	Genotype	1.436	24.14		1.31	26.44	
	Location:Genotype	1.688	28.37		1.278	25.79	
	Bloc:Rep:Location	0.004	0.07		0.287	5.79	
	Residuals	2.542	42.73		2.08	41.98	

STable 3-6: Predictive ability of the different scenarios and models

Scenario	Model	FL	PH	YLD	ZN
Uni1		0.225 ± 0.077	0.309 ± 0.069	0.388 ± 0.079	0.174 ± 0.080
Uni2		0.311 ± 0.005	0.389 ± 0.005	0.333 ± 0.005	0.323 ± 0.006
Uni3		0.229 ± 0.004	0.254 ± 0.006	0.243 ± 0.008	0.293 ± 0.006
Multi1	SM	0.220 ± 0.082	0.309 ± 0.069	0.388 ± 0.079	0.174 ± 0.080
Multi1	MM	0.313 ± 0.082	0.341 ± 0.074	0.282 ± 0.078	0.275 ± 0.076
Multi1	MDs	0.292 ± 0.080	0.321 ± 0.070	0.396 ± 0.072	0.209 ± 0.078
Multi1	MDe	0.246 ± 0.085	0.258 ± 0.069	0.382 ± 0.073	0.105 ± 0.076
Multi2	Random_25_MM	0.157 ± 0.019	0.295 ± 0.011	0.225 ± 0.016	0.204 ± 0.020
Multi2	Random_50_MM	0.186 ± 0.018	0.297 ± 0.013	0.251 ± 0.017	0.233 ± 0.022
Multi2	Random_75_MM	0.200 ± 0.012	0.296 ± 0.008	0.274 ± 0.014	0.250 ± 0.016
Multi2	Random_25_MDs	0.186 ± 0.033	0.265 ± 0.024	0.243 ± 0.038	0.178 ± 0.045
Multi2	Random_50_MDs	0.216 ± 0.024	0.261 ± 0.020	0.274 ± 0.032	0.221 ± 0.035
Multi2	Random_75_MDs	0.223 ± 0.015	0.264 ± 0.013	0.295 ± 0.023	0.241 ± 0.021
Multi2	Random_25_MDe	0.175 ± 0.038	0.245 ± 0.035	0.221 ± 0.055	0.167 ± 0.047
Multi2	Random_50_MDe	0.204 ± 0.030	0.250 ± 0.024	0.254 ± 0.039	0.211 ± 0.041
Multi2	Random_75_MDe	0.204 ± 0.023	0.255 ± 0.017	0.285 ± 0.027	0.238 ± 0.027
Multi2	CDmean_25_MM	0.184 ± 0.012	0.288 ± 0.006	0.218 ± 0.013	0.184 ± 0.012
Multi2	CDmean_50_MM	0.206 ± 0.010	0.277 ± 0.009	0.250 ± 0.013	0.225 ± 0.011
Multi2	CDmean_75_MM	0.212 ± 0.008	0.290 ± 0.006	0.272 ± 0.011	0.238 ± 0.010
Multi2	CDmean_25_MDs	0.253 ± 0.016	0.234 ± 0.015	0.236 ± 0.025	0.145 ± 0.025
Multi2	CDmean_50_MDs	0.255 ± 0.012	0.227 ± 0.009	0.289 ± 0.019	0.203 ± 0.016
Multi2	CDmean_75_MDs	0.242 ± 0.009	0.246 ± 0.008	0.289 ± 0.015	0.223 ± 0.013
Multi2	CDmean_25_MDe	0.259 ± 0.016	0.195 ± 0.020	0.242 ± 0.032	0.149 ± 0.026
Multi2	CDmean_50_MDe	0.261 ± 0.016	0.206 ± 0.014	0.276 ± 0.019	0.187 ± 0.019
Multi2	CDmean_75_MDe	0.239 ± 0.013	0.234 ± 0.009	0.282 ± 0.016	0.214 ± 0.014

Chapter 4 : Rapid genomic recurrent selection as a tool to increase the rate of genetic gain: a simulation study on rice

Avant-Propos

Ce dernier chapitre m'aura permis de me familiariser avec la simulation de programme de sélection, un monde en soit. Il m'a également permis de considérer très en détail le programme de sélection CIAT-Cirad. Ça a également été l'opportunité de collaborer avec Giovanni Eduardo Covarrubias Pazaran et Christian Werner de l'EiB (Exellence in breeding) deux très bons généticiens quantitatifs et spécialistes de la simulation. Ils m'auront aidé à scripter la simulation et nous ont fait de précieuses et pertinentes critiques sur le contenu du chapitre 4. Ce chapitre sera soumis à la revue à BMC Plant Biology.

Cédric Baertschi[12], Cécile Grenier[123]J Giovanni Eduardo Covarrubias-Pazaran[4], Christian Werner[4], Jérôme Bartholomé[123],

[1]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France.

[3]Alliance Bioversity-CIAT, A.A.6713, Km 17 Recta Palmira Cali, Colombia

[4] Excellence in Breeding (EiB), CGIAR

4.1 Abstract

Population improvement via recurrent selection has recently regained attention in the plant breeding community with the possible integration of genomic prediction (GP) in the schemes due to the reduction of genotyping costs. For several decades, the CIAT-Cirad rainfed rice (Oryza sativa L.) breeding program has been using a two-parts breeding program with a population improvement based on recurrent selection and a cultivar development following pedigree breeding. More recently, the recurrent selection integrated GP to improve the efficiency of progeny evaluation. In this study, we used simulations to assess the long-term effect of the integration of GP into the CIAT-Cirad breeding program. We investigated the effect of trait architecture (levels of genotype-by-environment interaction (GxE) and dominance) on the performance of two breeding schemes. The current breeding scheme based on a two-year phenotypic evaluation for the training set (BS1) was compared with a strategy based on a one-year evaluation (BS2).

For the recurrent selection part, the observed rate of genetic gain (ΔG) ranged from 1.37% to 5.29% for the two traits under positive selection and from -0.32% to 0.21% for the two traits selected for stability. This variability was mostly associated with the breeding schemes (with BS1 having the greatest gain) and the level of GxE. The level of dominance had little impact on ΔG. As expected, the differences between the two schemes increased in favour of BS1 when the level of GxE increased. The ΔG were lower and a higher interannual variability was found for the candidate varieties at the end of the product development part. The better performances of BS1 were related to the higher accuracies of the genomic prediction models: 0.64 and 0.59 on average for BS1 and BS2, respectively. With the increase of GxE, the accuracies dropped for both schemes with similar intensity.

Population improvement via recurrent selection is an ideal framework for designing efficient short-cycle breeding programs and therefore increases the rate of genetic gain for complex traits. The results from the simulation experiments are currently being used for the optimization of the CIAT-Cirad breeding program toward a faster genetic progress.

4.2 Introduction

Food demand is expected to increase between 45% and 51% by 2050 (van Dijk et al. 2021).The increase in yield observed in the last decades is however not sufficient to cover future demand (Ray et al., 2013). While meeting the food demand will be a challenge for agriculture in the future, it will have to be done while taking into account changing rainfall patterns (Trenberth, 2011), increasing temperatures (Zhao et al., 2017) or new biotic stresses (Bebber et al., 2013) due to climate change. In this race against time to ensure food security, plant breeding has a central role to play. Indeed, genetic improvement has been shown to be an important driver of the increase in plant production over the last decades (Laidig et al., 2014; Piepho et al., 2014). However, the gains from breeding varied greatly depending on the crops and the breeding programs.

Genetic gain is the change in population mean across time following artificial selection. It is influenced by the intensity of selection, its accuracy, the additive genetic variance available in the population under selection and the length of a breeding cycle as formalized in the breeder's equation (Lush, 1937). By looking at its parameters, one finds valuable information on how to address the genetic gain of a breeding program. Recently, much attention has been given to the optimization of public breeding programs with a great wealth of advice and potential leads to increase the rate of genetic gain for yield (Cobb et al. 2019b; Rutkoski 2019). Among the different options to increase the rate of genetic gain, the most promising is the reduction of the breeding cycle length via the integration of genomic prediction (GP). Since its introduction in the 2000 (Whittaker, Thompson, and Denham 2000; Meuwissen, Hayes, and Goddard 2001), GP has gained in popularity among plant breeders due to the drastic reduction in the genotyping costs. The concept of GP is simple: train a statistical model with genotypic and phenotypic information from a reference population and then use the model to predict the performance of selection candidates based only on molecular markers. GP has been tested and validated on multiple crops and types of population and is now a valuable tool in the breeder's toolbox (Crossa et al., 2017; Hickey et al., 2017; Jannink et al., 2010; Lorenz et al., 2011).

By nature, breeding programs are complex and the integration of new methodologies, such as GP, always necessitates evolutions of the breeding schemes. For logistical reasons as well as fear of ending up with a less efficient program, the introduction of new methods and tools are generally slow. Therefore, strong evidence of the positive impact of the integration of new tools are required to plan their integration. Stochastic simulation is a fast and cost-effective tool to test breeding schemes under conditions relevant for the target breeding program. Recent simulation studies have shown that GP is especially interesting in the context of recurrent selection (RS) (Gaynor et al., 2017; Müller et al., 2017). Those works shed a new interest on this now old breeding technique (Hull, 1946) in the plant breeding world, even though it is classically used by animal breeders (Hickey et al., 2017).

The CIAT (Centro Internacional para la Agricultura Tropical) with the Cirad (Centre international de recherche agronomique pour le développement) have together run a breeding program for upland rice for Latin America and the Caribbean for more than 30 years. In 1996, the decision was made to broaden the genetic base from which varieties were derived (Châtel et al., 2005). To reach this objective, the program has been based on two distinct parts: a population improvement part based on RS and a product development part based on pedigree breeding, with the improved population serving as a source of diversity for the pedigree breeding. RS was described by Fehr et al. (1991) as "*the systematic selection of desirable individuals from a population followed by recombination of the selected individuals to form a new population*". The CIAT-Cirad program uses a classical progeny evaluation for the selection of the recurrent parent. In this breeding scheme, selected S_0 plants (as they went through zero generation of selfing) are advanced in generation through selfing and bulking up to $S_{0:2}$ and $S_{0:3}$ for phenotyping (the main traits are: days to flowering, plant height, grain yield, grain zinc content and tolerance to blast). The selection of the S_0 is then based on $S_{0:2}$ and $S_{0:3}$ phenotypes. The crosses are realized with the progeny of the selected S_0 at generation $S_{0:1}$ to generate the new population of S_0 for the next cycle. To facilitate and increase the number of crosses in the RS scheme, the breeding program uses a segregating male sterility gene (*ms*-gene) (Frouin et al., 2019; Singh and Ikehashi, 1981). As the phenotypes are easily visually identified, the breeder can plant a mix of families segregating for this gene and let open pollination happen in the field. Then, by harvesting male sterile plants, one can ensure getting sibs or half-sibs seeds coming from outcrosses. For this reason, the crosses are done with $S_{0:1}$, the bulked progeny from S_0, at this generation, 25% of the individuals are male sterile which allows a sufficient rate of outcrossing. With this strategy, a cycle is completed every four years: ½ year to get the crosses, ½ year to advance to $S_{0:1}$, ½ year to advance to $S_{0:2}$, ½ year to advance to $S_{0:3}$, 1 year to phenotype the $S_{0:2}$ families and 1 year to phenotype the $S_{0:3}$ families. As crosses are done with $S_{0:1}$, the advance in generation is useful only for phenotyping and not to generate the material necessary for the crosses. This shows high potential for improvement if the selection of the parental families could be disconnected from the actual phenotyping as it has been shown that increasing the number of recombination steps per unit of time can increase the genetic gain (Gorjanc et al., 2018).

GP has already been evaluated in the framework of the CIAT-Cirad rice breeding program. Grenier *et al.* (2015) experimented with different genomic prediction models on advanced lines within a single environment as well as with different types of genotypic data. With predictive abilities of 0.31 for grain yield, 0.30 for flowering or 0.54 for plant height, results were considered promising. However, the approach suffered from one major drawback: it was applied relatively late in the program, in generation $S_{2:3}$ for genotyping and $S_{2:4}$ for phenotyping. Based on those first results, the integration of GP was improved by using S_0 genotypes and progeny obtained from $S_{0:2}$ and $S_{0:3}$ families (Baertschi et

al., 2021). This second experiment also took the opportunity to integrate the multi-environment aspect. When new material was predicted for the targeted environment, predictive abilities of 0.25 for yield, 0.37 for zinc concentration, 0.33 for flowering or 0.40 for plant height have been recorded. The conclusions about the utility of a multi-environment approach, however, varied depending on the level of genotype-by-environment interaction associated with each trait.

Encouraged by such results, the CIAT-Cirad breeding program started to implement GP in its population improvement scheme. So far, the approach has been to take the same experimental design as in the full phenotypic selection scheme and use it to train a predictive model as in Baertschi et al. (2021). Two years of phenotyping allow the capture of part of the genotype-by-year interaction, however it requires time for the generation advancement as well as for the phenotyping itself. Also, as the phenotyping is done in two consecutive years on two consecutive generations, there is a risk to confound the year (environment) and the generation effects under this scheme. Additionally, if a supplementary year of phenotyping does not slow down the RS, it will delay the collection of new phenotypic data and add one more recombination event between the newest calibration data and the predicted genotype. This will increase the number of crossing steps and possibly the genetic distance between the calibration and prediction population with potential negative impacts on the prediction accuracy. The simplest and easiest way to reduce the cost of the phenotyping would be to limit it to one generation/year of phenotyping. It would not only halve the phenotyping effort but also reduce the number of crossing events between the newest calibration material and the prediction set. However, there would be no way to account for the year effect on the tested material. Under those conditions, traits showing strong genotype-by-environment interactions are expected to deliver poor calibration data as the year fluctuation would be confounded with the intercept and genetic effect. As the dominance variance evolves through the generations of fixation, they are expected to also influence the accuracy of the phenotyping and hence the predictive ability of our model.

The goal of this study was therefore to assess the long-term performances of the CIAT-Cirad upland rice breeding program: a two-part breeding program integrating a rapid cycling genomic selection component. We compared two breeding schemes combining RS and GP: one based on the current strategy with two generations of phenotyping by cycle to update the GP model (BS1) and a second one with only one generation of phenotyping to represent a scenario where the budget is limited (BS2, Figure 4-1). We investigated the impact of different levels of dominance and genotype-by-environment interactions (GxE) on each breeding scheme in order to identify the conditions that favour one scheme over the other. The simulation parameters were chosen to reflect as closely as possible the CIAT-Cirad breeding program. Four traits (T1, T2, T3 and T4) and their relative level of variance were modelled to represent the four main traits of the program (grain zinc content, grain yield, days to flowering and

plant height). The genetic gain, the accuracy of prediction, the variance component and the allele frequencies were followed to dissect the performances of the two breeding schemes (BS1 and BS2).

4.3 Material and Methods

4.3.1 Breeding scheme description

A schematic representation of the simulated breeding schemes can be found in Figure 4-1 and details on the population sizes used and the number of years necessary for its application are given in STable 4-6 for BS1 and STable 4-7 for BS2.

4.3.1.1 Current breeding strategy: Breeding scheme 1

The breeding scheme 1 (BS1), as well as the alternative BS2 (see below), can be described in three tasks depending on each other in terms of data and material but not necessarily locked together in terms of time. The first task consists in running RS . In its first semester, 500 candidates for selection

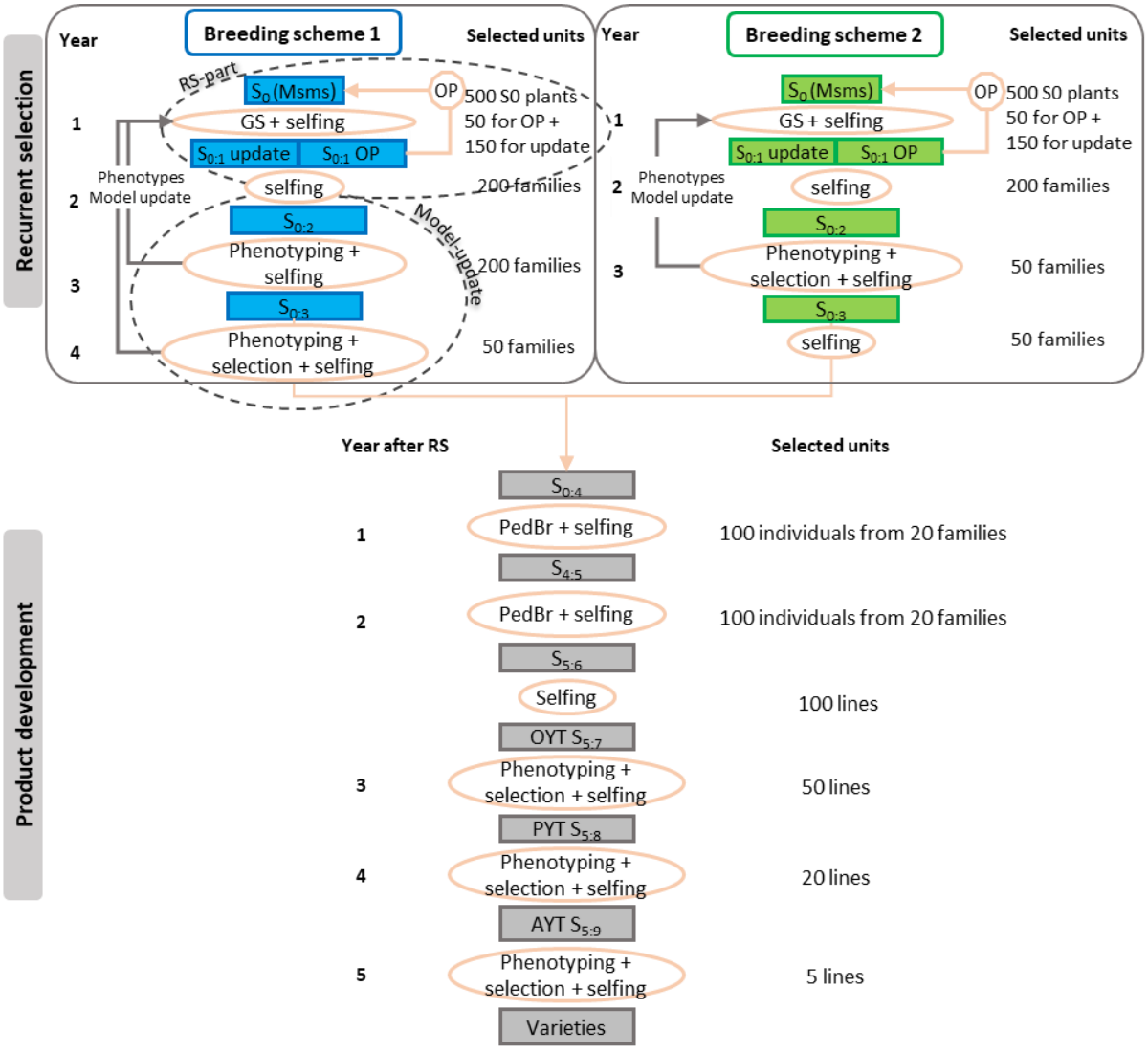


Figure 4-1: Schematic representation of the two breeding schemes. The colored squares represent populations in the field while the orange circles stand for tasks done on those populations. The black arrows represent a flow of information and the orange arrows a flow of material. On the upper part are the two recurrent selection parts and on the lower panel the common product development part.

are randomly sampled in a population of S_0, their genotype is acquired, and their genomic estimated breeding values (GEBV) predicted. Based on the GEBV, the best 50 S_0 are selected for building the recombination set. This set of S_0 and an the additional next best 150 S_0 based on GEBV are used as an update set for the prediction model. All 200 S_0 are selfed and their $S_{0:1}$ progenies are bulked by S_0 to build families. For the recombination set, 60 seeds per family are mixed to create a population for open pollination. The open pollination is made possible in autogamous rice by the presence of a segregating male-sterility gene (see Appendix 1 for details). The male sterile plants can only be fertilized by pollen from other male fertile plants and will carry S_0 seeds from which a new cycle of RS can be started. One cycle of RS needs one semester for the open pollination and one semester for the generation advance from S_0 to $S_{0:1}$. Under those conditions, a RS cycle lasts one year with parental material recycled every year.

We considered that we have a prediction model from the beginning of the population. Under this condition, the RS can be run independently but the model is expected to lose its accuracy over the cycles. For this reason, a second task consisting of the phenotyping for the model-update is connected to the RS. For the 150 S_0 only in the update set, all the $S_{0:1}$ progenies are selfed and bulked to advance to generation $S_{0:2}$ while, for 50 S_0 also in the recombination set, the seeds remaining after the sampling for future recombining are used. In a similar way, part of the $S_{0:2}$ seeds are used to advance to generation $S_{0:3}$ while the rest is kept for phenotyping. The same operation is done on $S_{0:4}$ progenies with one part used for phenotyping and the rest kept for later generation advancements. The phenotyping consisted of measuring the mean value of a plot composed of bulked progeny coming from a single S_0. The exact same design was applied at generation $S_{0:2}$ and $S_{0:3}$. Once the phenotyping was completed, the means of the $S_{0:2}$ and $S_{0:3}$ phenotypes were computed and used with the genotypes from their respective S_0 plant to update the GP model. Based on the phenotypic data as well as on the genotypic data from S_0, the best families were selected as a base population for pedigree breeding and advanced to generation $S_{0:4}$ using the stored $S_{0:3}$ seeds. In theory, generation advance and phenotyping could be done at the same time. However, in practice the generation advance is currently done separately as combining phenotyping and seed production activities would increase too much the risk of mistakes.

With the pedigree breeding starts the third task of the program: the product development. For two consecutive generations, $S_{0:4}$ followed by $S_{4:5}$, the 20 best families and then the five best plants within the family were selected. After the pedigree breeding, the selected plants went through a step of multiplication consisting of one generation of selfing and bulking of progeny to generate enough material for the upcoming field testing. The lines go through three multi-environment yield trials: the observation yield trial (OYT), the preliminary yield trial (PYT) and the advanced yield trial (AYT). The OYT is done in two sites with two replicates and allows the selection of 50 lines based on adjusted

phenotypes. It is followed by the PYT run in three sites with three replicates. At the end of this step 20 lines are selected to go through the AYT. The AYT is realized in five sites with three replicates and allows at the end the selection of five candidate varieties. This represents the scheme as it is currently run and is later referred to as BS1.

4.3.1.2 Alternative breeding strategy: Breeding scheme 2

In an effort to reduce the phenotyping to a single generation, a second scheme (BS2) was tested with the GP model update based on the phenotyping in $S_{0:2}$ only (Figure 4-1). The recurrent selection runs as in BS1 as well as the advance in generation up to $S_{0:2}$. Once in generation $S_{0:2}$, part of the material was used for phenotyping while the rest was kept aside for generation advance. The candidate families for pedigree breeding were selected on their GEBV based on a model calibrated with $S_{0:2}$ phenotypes and S_0 genotypes. The selected families were then advanced to generation $S_{0:3}$ and $S_{0:4}$. The product development started from there and the scheme followed the same steps as BS1 with two generations of pedigree breeding, one generation of multiplication and three generations of yield testing.

4.3.2 Simulation

4.3.2.1 Genome simulation

The genome was simulated to approach a rice genome. It had 12 chromosomes with a physical length of $33.4*10^6$ base pairs, a genetic length 140 cM (Chen et al., 2002) and the default mutation rate of $2.5*10^{-8}$. Each chromosome had 833 SNP and 300 quantitative trait loci (QTL). The effective population size (Ne) was set at 50, according to the real breeding population (Baertschi et al., 2021). The genome simulation was done with the software MaCS (Chen et al., 2009) embedded in AlphaSimR (Gaynor et al., 2021) which was used for later simulation.

4.3.2.2 Initial population

Based on the simulated genome, 80 heterozygous founders were generated. From the 80 founders, 4,500 crosses were realized with ten progeny per cross. From those 45,000 S_0 plants, 300 were randomly sampled to assemble a first synthetic population. The 300 crosses were used to generate again 4,500 random crosses giving each ten progenies generating a new synthetic population. Those successive crossings and random samplings were realized in total five times. Within the offspring from the fifth random crosses, 400 S_0 were randomly sampled to be used as the initial population.

The genome simulation followed by the creation of the initial population was replicated twenty times and the same replicates were used later for both breeding schemes. Details on the population size are given in STable 4-8.

4.3.2.3 Genetic and phenotypic values

Four correlated traits were simulated (T1, T2, T3 and T4). The mean values and total genetic variance in the initial population were chosen to reflect the one observed in the program for zinc concentration

of the grain (T1), the grain yield (T2), the number of days from sowing to flowering (T3) and the plant height (T4) (Table 4-1). Three levels of dominance were tested. AlphaSimR simulates dominance effects as the product of dominance degrees and the absolute additive effect. The dominance degrees are sampled from a normal distribution with custom parameters (Table 4-1). Three different variances were used, while the mean was kept at zero. This means that the probability of a strongly dominant QTL varied among the variance of dominance degree but they were always equally likely to be positive or negative.

Three levels of genotype by environment interaction (GxE) always proportional to the additive variance were simulated (Table 4-1). The heritabilities for the different steps involving phenotyping were set after the values in Table 4-2. More details on the method of trait simulation can be found in the vignette "Traits in AlphaSim" (Gaynor, 2020).

For each phenotyping step, heritabilities are chosen with values close to what could be observed in field experiments of similar design (Table 4-2). AlphaSimR requires an error variance to simulate phenotypic values from genetic values. This error variance σ_ε^2 was computed at each phenotyping steps following the function:

$$\sigma_\varepsilon^2 = \frac{\sigma_g^2}{H_{step}^2} \qquad \text{Eq. 4-1}$$

with σ_g^2 being the true genetic variance and H_{step}^2 the heritability parameter given for each trait at each phenotyping step.

Table 4-1: Parameter for the trait simulation. Number of QTLs is the total number of loci with effects, Mean is the intercept for the genetic value, Genetic variance gives the total genetic variance for each traits, the Mean DD gives the centre of the normal distribution from which the dominance degree for each locus is drawn and the DD variance is used to compute is standard deviation, varGxE is the total genotype by environment interaction and is computed from as *varGxE=(GxS + GxY)*Genetic variance*

	T1	**T2**	**T3**	**T4**
Number of QTLs		3600		
Mean	20	500	100	100
Genetic variance (Var)	7	500	50	50
Mean DD		0		
DD variance		Low = 0.1, High = 0.6		
varGxE = Var x GxE		Low = 1, Medium = 3, High= 5		

Additive genetic correlation	T1	T2	T3	T4
T1	1	-0.1	0.1	-0.1
T2	-0.1	1	-0.2	0.2
T3	0.1	-0.2	1	0
T4	-0.1	0.2	0	1

Step	T1	T2	T3	T4
Progeny testing $S_{0:2}$	0.4	0.2	0.3	0.6
Progeny testing $S_{0:3}$	0.4	0.2	0.3	0.6
Pedigree breeding $S_{0:4}$	0.1	0.1	0.1	0.1
Pedigree breeding $S_{4:5}$	0.1	0.1	0.1	0.1
OYT	0.5	0.4	0.6	0.8
PYT	0.5	0.4	0.6	0.8
AYT	0.5	0.4	0.6	0.8

Table 4-2: Heritabilities set for the different steps of the breeding scheme. Heritability for generation $S_{0:3}$ is relevant only for BS1

4.3.2.4 Genomic prediction models

Recently the CIAT-Cirad program used the genotyping of 400 S_0 and their subsequent phenotyping in $S_{0:2}$ and $S_{0:3}$ to build a strong initial calibration population (Baertschi et al., 2021). Following the same approach, the initial GP models of the simulations (one for each trait) were calibrated on 400 S_0 from the initial population and their phenotypes in $S_{0:2}$ and $S_{0:3}$ for BS1 or $S_{0:2}$ for BS2. Then, two different phases of genomic prediction application were simulated. For the first cycles of the simulations, the GP models were calibrated using the genotypes and the phenotypes of the 400 families from the initial population. These models were only updated after cycle 5 or cycle 4 for BS1 and BS2, respectively. Indeed, the progeny testing phase takes three or four years depending on if $S_{0:2}$ and $S_{0:3}$ generations are evaluated or only $S_{0:2}$ (Figure 4-1). This was done to mimic the transition between the phenotypic selection and the genomic selection. After these first cycles, the GP model was updated at each cycle by adding 200 new S_0 to the calibration sets.

The predictions were based on a random regression best linear unbiased predictor (RRBLUP) with the following model:

$$Y_{ij} = \mu + c_i + g_{ij} + \varepsilon_{ij} \qquad \text{Eq. 4-2}$$

The prediction accuracy was estimated as the correlation between the GEBV and the line ability of the predicted S_0. The line ability of a cross is the "*expected value of all lines which can be derived from it*" (Gallais, 1979). To compute it, 100 double haploid (DH) for each S_0 predicted were generated and their average genetic value (GV) (DHGV) computed. The precision was then measured as $cor(GEBV, DHGV)$.

4.3.2.5 Index definition

The multi-traits selection was based on a selection index. Target gains stated as the number of standard deviations were defined and adjusted using the variance-covariance between the traits by matrix multiplication $R = G^{-1}t$, G^{-1} being the inverse of the trait variance-covariance matrix and t the vector of target change in the population means. The indices are then used in the index $I_G = \beta_R X\hat{u}$ where X is the matrix of allele dosages for the predicted genotypes and $\hat{u}$ is the vector of marker effects (Céron-Rojas and Crossa, 2018). Two different vectors of targets $t' = [t_{T1}, t_{T2}, t_{T3}, t_{T4}]$ were used, one for the selection of parental families used in the recurrent selection part t_RS'= [0.8, 1, 0, 0]

and one for the three different yield trial steps at the end of the pedigree breeding t_PD'= [0.5, 1, 0, 0].

4.3.3 Breeding scheme evaluation

4.3.3.1 Genetic gain

The two breeding schemes were compared for different scenarios considering the levels of GxE and dominance in the set of traits simulated. First the change in population mean was followed at generation S_0. The true genetic values were extracted at each cycle for all the available crosses and not only the candidates S_0 (n = 500) and averaged. Similarly, the true genetic values for the fixed lines obtained at the end of the pedigree breeding part (varieties, n = 5) were extracted and averaged by cycle. Based on those means by cycle, the genetic gain over the 20 cycles were computed as $\mathrm{GG} = \frac{1}{20} * \left(\frac{\overline{GV_{20}} - \overline{GV_0}}{\overline{GV_0}}\right) * \backslash 100$, for $\overline{\mathrm{GV}_0}$ being the mean genetic value of the initial population and $\overline{\mathrm{GV}_{20}}$ at the last cycle (20). Aside from the population mean, the total genetic variance as well as its additive fraction were measured across all cycles for all treatment and breeding schemes.

For the step involving phenotyping the broad sense heritabilities were computed as $H^2 = \frac{\sigma_G^2}{\sigma_P^2}$ to control the simulation. This was done for $S_{0:2}$ and $S_{0:3}$ progeny testing and OYT to get a sense of the precision of the phenotyping at those stages. Only the simulations of BS1 were displayed but the values are expected to be the same no matter the breeding scheme.

4.3.3.2 Evolution of the population diversity and structure

The additive and dominance variance were tracked across the cycles on the S_0 population. The genotypes at the QTLs (n=3600) were extracted from the 500 genotyped S_0 plant and the minor allele frequencies (MAF) for each QTLs were computed at five different cycles (1, 5, 10, 15, 20) during the simulation for all breeding schemes, GxE levels and dominance levels.

A principal component analysis was also run on the combined genotypic matrix of the genotypes of the non-QTL SNPs (n=9,996) of the 500 candidates for prediction at S_0. The outputs were limited to cycles 1, 5, 10, 20 plus the 400 genotypes of the initial population (cycle 0). The results for one replicate of BS1 with low dominance and medium GxE is displayed.

4.4 Results

4.4.1 Genetic gain in a two-part breeding scheme

4.4.1.1 Population improvement and rate of genetic gain

The two breeding schemes (Figure 4-1, BS1 and BS2) were first compared for the evolution of the mean genetic value of the population at S_0. This allowed an assessment of the scheme for population improvement under the different levels of GxE and dominance. Looking at the evolution of the population mean (Figure 4-2), clear trends either for gain (T1 and T2) or for stability (T3 and T4) were

Table 4-3: Genetic gain per cycle. The genetic gain was obtained by standardizing the total population increase on the population mean at cycle 1 and divided by the total number of breeding cycles. The genetic gain was computed on the true breeding value of the S_0 and of the candidate varieties. For the S_0, a cycle is done in one year and it can be seen indistinctly as genetic gain per cycle or per year.

Trait	Dom	Scenario	S_0			Varieties		
			Low GxE	Medium GxE	High GxE	Low GxE	Medium GxE	High GxE
T1	Low	BS1	5.28	4.37	4.35	4.33	3.52	3.52
	High	BS1	5.25	4.20	3.44	4.37	3.28	2.86
	Low	BS2	4.92	3.80	3.33	4.51	3.66	2.86
	High	BS2	4.80	3.68	2.78	3.91	3.25	2.08
T2	Low	BS1	2.41	1.97	1.52	1.72	1.40	1.15
	High	BS1	2.26	1.72	1.62	1.61	1.23	1.18
	Low	BS2	2.30	1.80	1.39	1.68	1.29	1.03
	High	BS2	2.12	1.51	1.37	1.48	1.11	0.93
T3	Low	BS1	-0.09	-0.05	0.08	0.07	0.00	0.05
	High	BS1	-0.32	-0.11	-0.13	-0.25	0.05	-0.03
	Low	BS2	-0.04	-0.06	-0.14	-0.13	-0.24	0.13
	High	BS2	-0.25	-0.12	-0.07	-0.06	0.00	0.07
T4	Low	BS1	0.06	0.11	0.01	0.06	0.13	-0.10
	High	BS1	0.16	0.14	0.15	0.24	0.17	-0.06
	Low	BS2	0.06	0.14	0.00	-0.02	0.26	-0.06
	High	BS2	0.20	0.06	0.08	0.16	0.00	0.01

found. The observed rate of genetic gain (ΔG) ranged from 2.78% to 5.29% for T1, from 1.37% to 2.41% for T2, from -0.32% to 0.08% for T3 and from 0% to 0.21% for T4 (Table 4-3). The variability observed within each trait was related to the breeding scheme, the level of GxE and to a lesser extent the level of dominance (STable 4-1). Differences between the two breeding schemes were observed for T1 with BS1 presenting a significantly higher ΔG than BS2 in average (ΔG = 4.48% for BS1 compared to ΔG = 3.88% for BS2). A similar trend was found for T2 with the average ΔG at 1.92% for BS1 and at 1.75% for BS2. The GxE levels had a significant effect on the ΔG for all traits but T4. As expected, ΔG decreased as the level of GxE increased. For T1, the average ΔG dropped from 5.07% under low GxE to 3.47% under high GxE. Similarly, for T2 the average ΔG went from 2.27% under low GxE to 1.48% under high GxE. While the effect was in general more pronounced under BS2 than under BS1, no trait showed an interaction between the breeding schemes and the GxE levels. Considering the two traits selected for stability, T3 and T4, no significant differences were observed between the schemes nor between the GxE levels (STable 4-1). Despite selection for stability, T3 and T4 means did slightly evolve across the cycle. While the population mean for T3 decreased, the population mean for T4 increased. The dominance levels influenced significantly the population mean for T1, T2 and T3. For T1 and T2, higher dominance level resulted in lower ΔG while for T3, the ΔG were higher when dominance was high (Table 4-3, STable 4-1).

The differences between the two schemes did increase in favour of BS1 with increasing GxE. This was especially visible with T1. Initially small in the first cycles, the differences increased to reach a

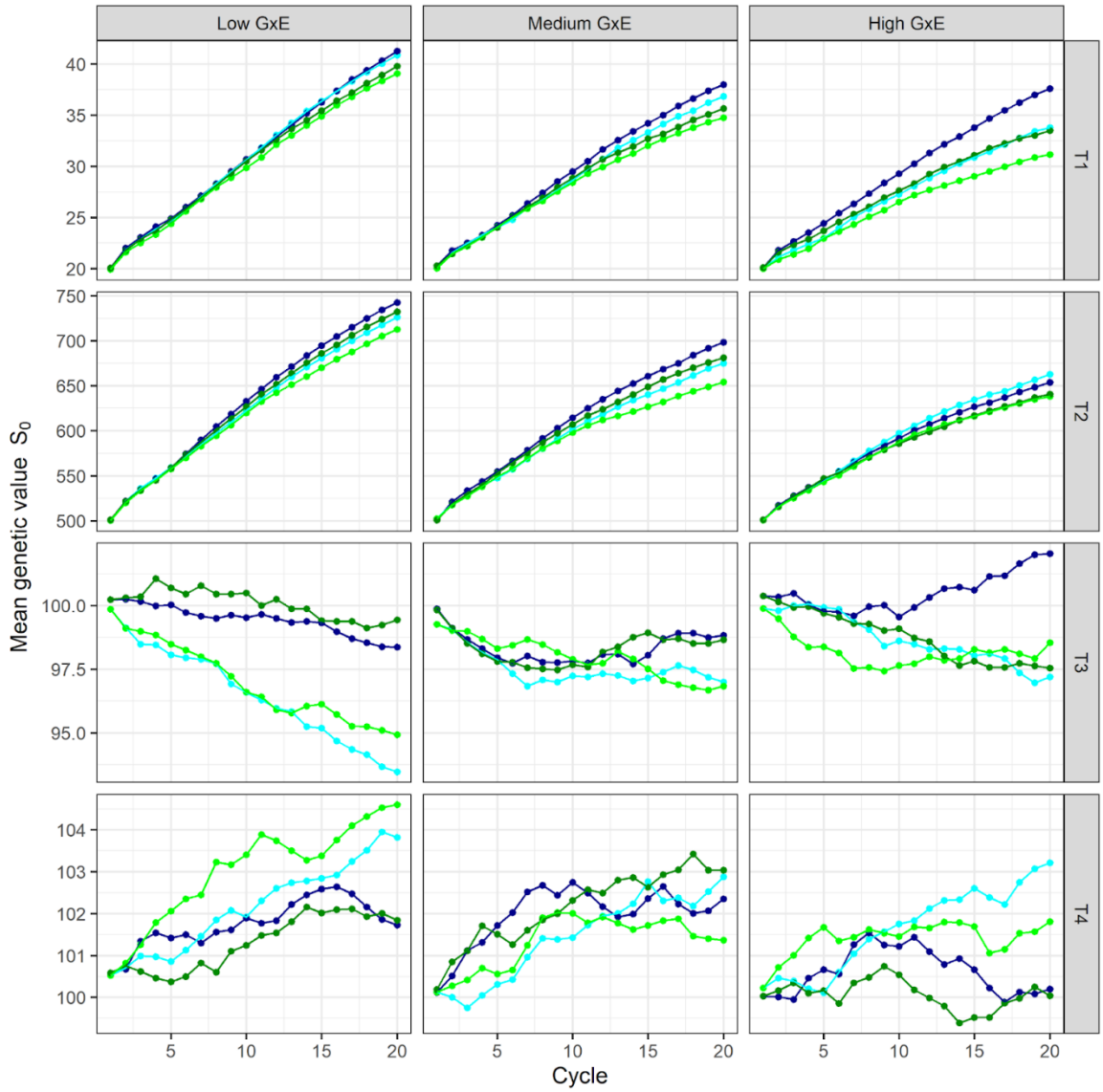


Figure 4-2: Evolution of the population mean at S_0. The mean genetic value computed on all the available crosses S0 (n=4500) and averaged on 20 replicates. The results for BS1 are in blue while the results for BS2 are in green. The levels of dominance are identified by the color darkness, darker colors standing for low dominance and brighter color for high dominance. The three levels of GxE are as columns and each row is a different trait (see Table 4-1 for trait definition).

maximum after 20 cycles. The population mean under BS1 was at this point 9% higher than the population mean under BS2 at high GxE while only 4% higher at low GxE. Differences between schemes were smaller for T2 with the average differences between the schemes across all GxE and dominance levels ranging from 1% to 4% at the maximum at cycle 20. The population mean of T3 and T4 behaved more erratically and no tendency in scheme performances due to GxE or dominance could be observed.

4.4.1.2 Genetic gain of the candidate varieties

Differences between the breeding schemes were also investigated on fixed lines at the end of the product development part. Compared to the ΔG obtained for the RS, the ΔG of fixed lines were lower and a higher interannual variability was found (Figure 4-3, Table 4-3). Indeed, the trends in genetic

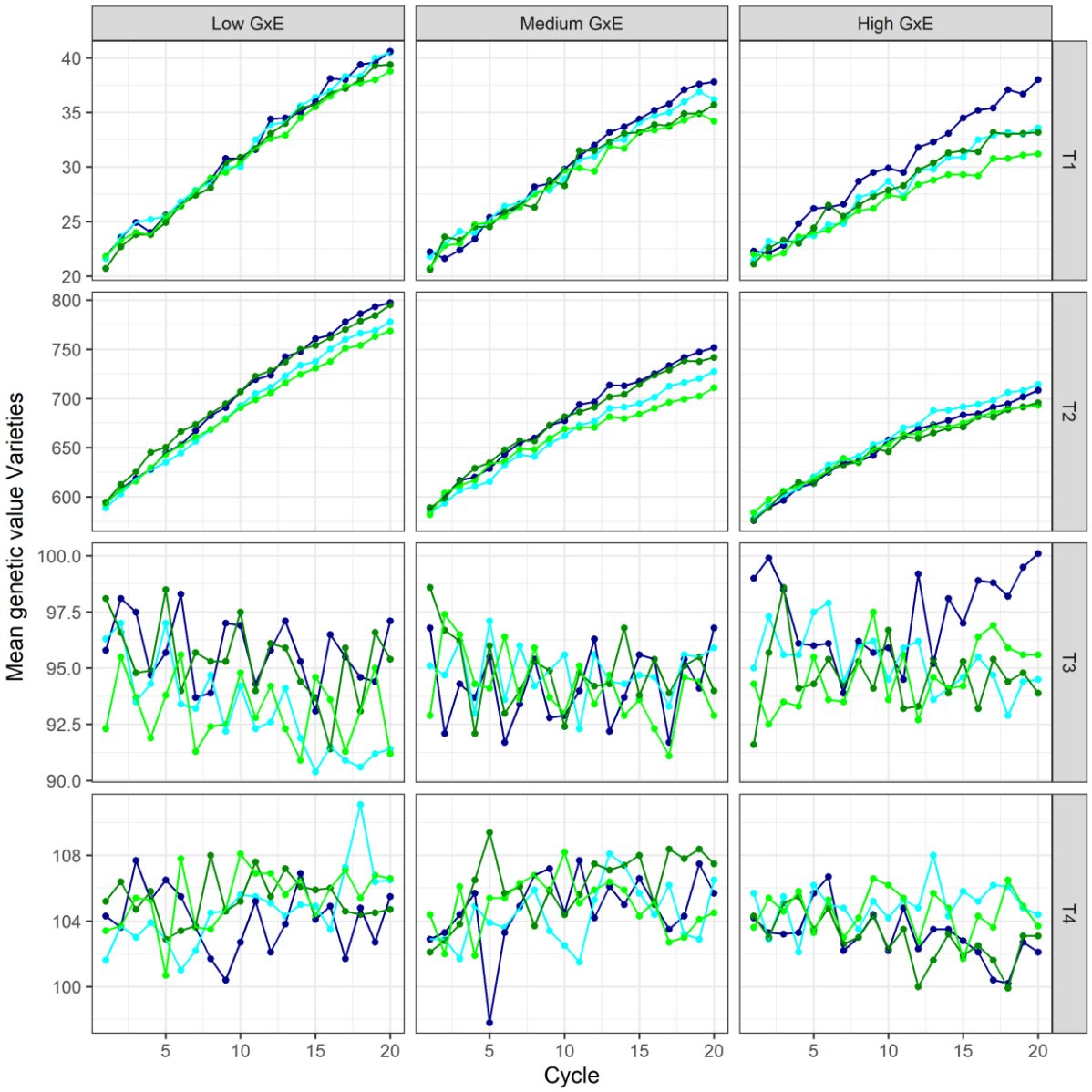


Figure 4-3: Evolution of the mean of the varieties at the end of the product development. The mean genetic value computed on all the available crosses variety (n=5) and averaged on 20 replicates. The results for BS1 are in blue while the results for BS2 are in green. The levels of dominance are identified by the color darkness, darker colors standing for low dominance and brighter color for high dominance. The three levels of GxE are as columns and each row is a different trait (see Table 4-1 for trait definition).

mean presented a less continuous increase for T1 and T2 and a saw-tooth profile for T3 and T4. A clear increase was found for traits under selection with ΔG ranging from 2.08% to 4.51% for T1 and from 0.93% to 1.72% for T2 (Table 4-3). The values ranged from -0.25 to 0.13% for T3 and from -0.10% to 0.26% for T2. As for the RS part, an increase in GxE levels significantly reduced the ΔG for T1 and for T2. Despite the absence of significant interaction between breeding schemes and GxE levels, the drop in ΔG due to the increase in GxE was stronger for both traits under BS2 (STable 4-2). For T1, no significant effect of the breeding scheme was found on ΔG. However, this average trend hid an increase of the difference between BS1 and BS2 when the level of GxE increased. For low GxE, the largest gain (ΔG = 4.51%) was observed under the BS2 (4.51% for BS2 compared to 4.33% for BS1). For the higher

level of GxE, the inverse was found: 3.52% for BS1 and 2.86% for BS2. Contrary to T1, the difference between breeding schemes was significant for T2, with an average ΔG of 1.38% for BS1 against 1.25% for BS2 (STable 4-2). For T2, the strongest gains were observed for BS1 under low GxE and low dominance (ΔG = 1.72%) which correspond to a ΔG of 1.68% under BS2.

Dominance levels had a significant effect on ΔG for T1 and T2. Under low dominance, the average ΔG of T1 was 3.84% and dropped to 3.41% when dominance was high. Similarly, for T2 the average ΔG went from 1.38% under low dominance to 1.26% under high dominance.

4.4.1.3 Comparison between recurrent selection and product development

The final values of the population and of the fixed lines were very close, sometimes even lower for the fixed lines with differences varying between a reduction in mean of -6% of the population value or an increase of 9%. The pedigree breeding allowed a slightly better selection for T2 with a gain between population and fixed lines mean of 5-7%. For T1 and T2, differences between population and fixed lines means were not influenced by the scheme. The response of the ΔG of the RS part and of the candidate varieties to the GxE and dominance were rather similar and both BS1 and BS2 showed lower ΔG on the candidate varieties than on the RS part. Proportionally to the RS ΔG, differences were about half as large under T1 than under T2, where ΔG of at varieties-level were about 70% of the one observed for the population improvement. For T3 and T4, the differences between improvement of the population and improvement of the varieties mean are much less homogenous across the GxE and dominance levels as well as across the schemes with the relative difference in ΔG taking values between -300% and +200%. As we wanted to select T3 and T4 for stability, we see that our breeding scheme performed best at selecting for populations with stable T3 and T4 mean rather than for fixed lines with stable values.

4.4.2 Role of genomic prediction on breeding scheme performance

4.4.2.1 Prediction accuracies

All traits showed a similar trend in terms of the evolution of prediction accuracies (Figure 4-4). The initial accuracies were high, almost at 0.8 for T1, T2 and T4 and around 0.7 for T3 when calibration data were from still closely related genotypes. Then, up to the third cycle under BS2 and up to the fourth cycle under BS1 the accuracy dropped. After the models were updated five or six times, accuracies got back to their initial values for T1, T2, T4 and even higher than the initial accuracy for T3.

For T1, the highest accuracies were obtained under low GxE for both breeding schemes. BS1 reached 0.72 under low dominance and 0.71 under high dominance. For the low GxE as well BS2 was at 0.69 and 0.65 for low and high dominance respectively. With increasing GxE, the accuracies dropped for both schemes with also increasing differences between the two. The lowest accuracies were for both schemes under high GxE and dominance. The picture was very similar for T2 with smaller differences

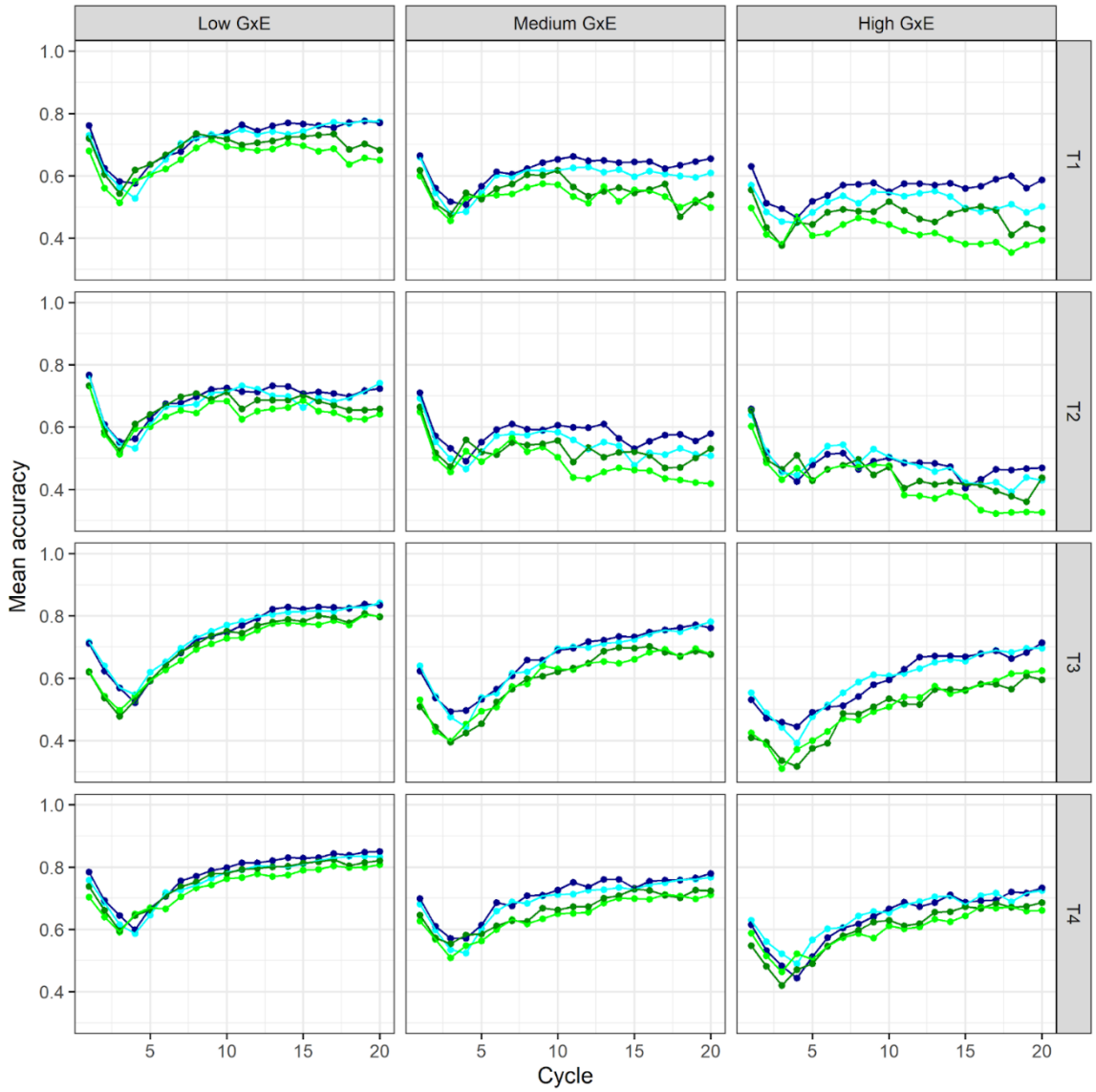


Figure 4-4: Prediction accuracy for the line value of S_0. The accuracy was computed as the correlation between the GEBV and the mean genetic value of 100 double haploid lines developed from the S0 sampled for genomic selection (n=500). The single replicate accuracies were first Fisher transformed before the average was computed on the 20 replicates and the results back transformed. The blue lines stand for BS1 and the green lines for BS2. The levels of dominance are identified by the color darkness, darker colors standing for low dominance and brighter color for high dominance. The three levels of GxE are as columns and each row corresponds to a different trait (see Table 4-1 for trait definition).

between BS1 and BS2. Again, average accuracies were at the highest under low GxE with 0.69 for BS1 and 0.66 for BS2. They dropped to 0.48 for BS1 under high GxE and 0.45 for BS2. Differences between the two dominance levels were smaller than for T1. For both T1 and T2, the accuracy of the differences between the schemes, the GxE levels and the dominance levels increased with the cycle. The average accuracy for T3 was higher than for T1 and T2. Although, it started slightly lower than the other traits, the accuracy increased more in the last cycles pulling the average upward. When the GxE increased so did the differences between BS1 and BS2. More importantly, the difference was visible from the first cycle and stayed more or less constant across all simulated cycles. T4 was similar to T3 as the last cycles

showed accuracies higher than the initial ones. The differences between the breeding schemes were less pronounced than under T3 especially when the GxE was high.

Looking at the mean accuracies across all cycles, the two breeding schemes delivered significantly different value for all traits with accuracy for BS1 systematically higher than the one for BS2 (STable 4-4, STable 4-3). The levels of GxE had significant effects on the accuracy for all traits with accuracy dropping when levels of GxE increased. Nevertheless, no interactions with the breeding scheme were observed. Dominance level had a significant impact on accuracies for T1 and showed a p-value of 0.06 for T2.

4.4.2.2 Precision of the phenotyping

The broad sense heritabilities (H^2) were in all cases below the set heritabilities (Figure 4-5). They dropped as the GxE increased but did not respond to the different levels of dominance. Higher heritabilities also showed stronger response to the GxE. It can be observed between traits but also within traits. Indeed, higher H^2 were used for OYT steps, which showed a systematically larger response to GxE than the two progeny testing steps.

4.4.3 Impact of rapid recurrent selection on population diversity

The proportion of fixed loci increased similarly in the two breeding schemes from about 5% in the first cycle to 50% of all positions in the cycle 20 (Figure 4-6). The level of fixation was almost linear with five cycles of selection and recombination fixing about 10% of the remaining polymorphic QTLs. No

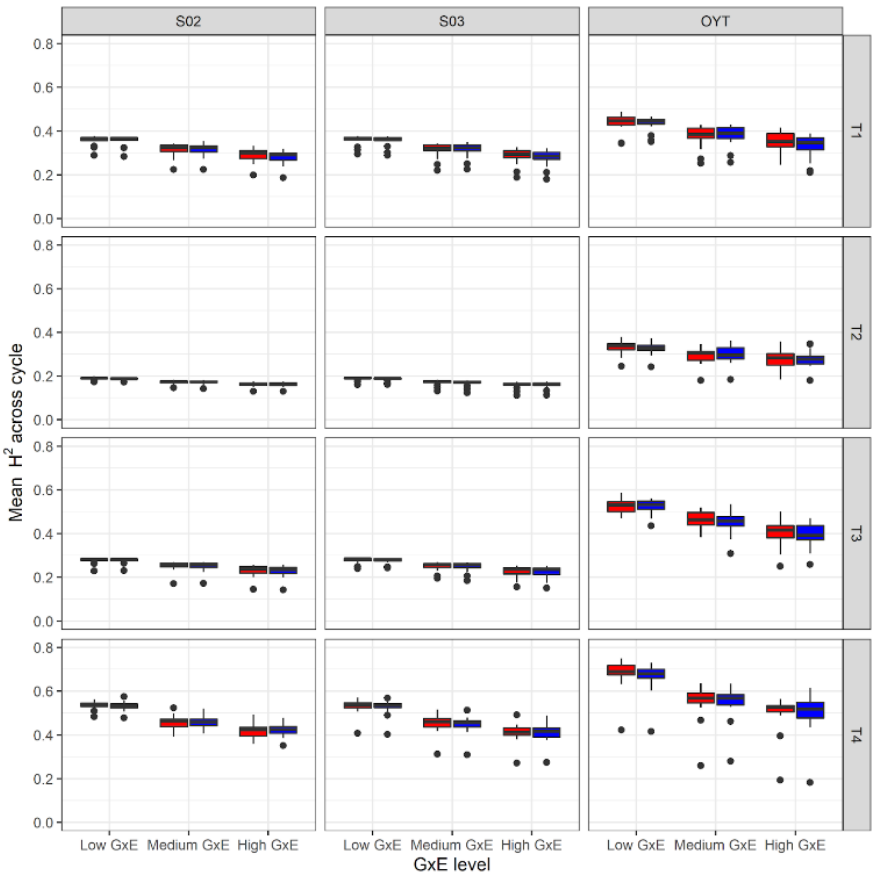


Figure 4-5: Broad sense heritability computed at the $S_{0:2}$ and $S_{0:3}$ progeny testing and OYT under BS1. The GxE levels are on the x-axis and the levels of dominance are identified by the boxplot colors (red=low, blue=high). Rows correspond to the four simulated traits (see Table 4-1 for trait definition).

differences could be observed between the breeding scheme nor between the GxE and the dominance levels.

We looked at the other categories of MAF for the same cycles and treatments. The distribution of proportion of position with $0 < MAF \leqslant 0.5$ were uniform with only the level decreasing as more position were fixed (data not shown).

A principal component analysis was also run on the genomic data of the same five cycles (Figure 4-7). A clear evolution could be observed. While the population at cycle 1 did not differ from the initial population, four additional cycles of selection clearly differentiated the populations as cycles 1 and 5 show. The distancing of the populations continued at each cycle while the variability was simultaneously reduced as the cloud of points was reduced.

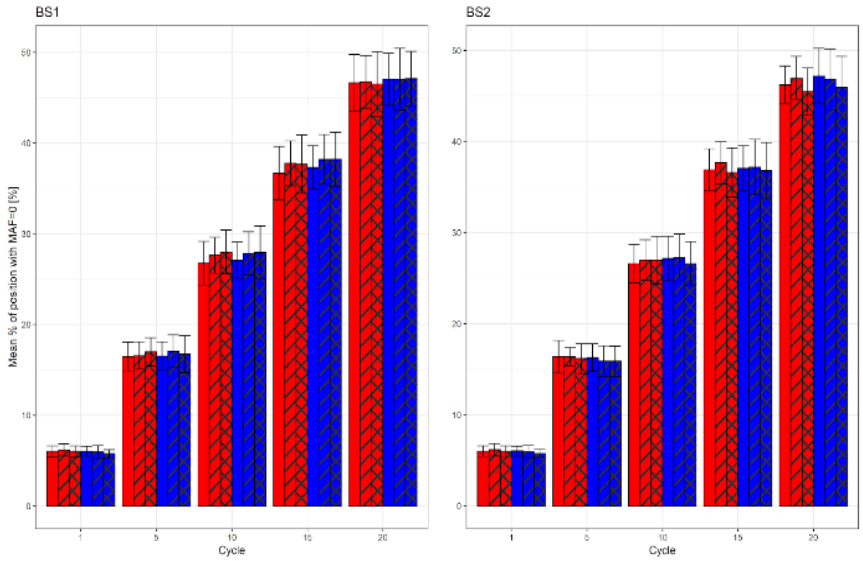


Figure 4-6: Evolution of the proportion of fixed QTL across the cycle. The percentage of fixed QTLs at cycle 1, 5, 10, 15, 20. BS1 is on the left panel and BS2 on the right one. Plain bars are for low GxE, hatched bars for medium GxE and crosshatched bars for high GxE. Low Dominance are in red and high dominance in blue

Table 4-4: Average change by cycle in additive and dominance variance expressed as percent of the initial variances

Trait	Dom	Scheme	Additive variance			Dominance variance		
			Low GxE	Medium GxE	High GxE	Low GxE	Medium GxE	High GxE
T1	Low	BS1	-2.68 (0.47)	-2.49 (0.37)	-2.52 (0.41)	-2.49 (0.39)	-2.45 (0.43)	-2.31 (0.50)
	High	BS1	-2.57 (0.34)	-2.54 (0.64)	-2.37 (0.66)	-2.59 (0.29)	-2.41 (0.40)	-2.40 (0.48)
	Low	BS2	-2.77 (0.40)	-2.45 (0.58)	-2.44 (0.53)	-2.69 (0.49)	-2.36 (0.49)	-2.24 (0.41)
	High	BS2	-2.61 (0.44)	-2.43 (0.55)	-2.26 (0.36)	-2.48 (0.40)	-2.32 (0.32)	-2.32 (0.40)
T2	Low	BS1	-2.79 (0.38)	-2.73 (0.36)	-2.55 (0.40)	-2.59 (0.41)	-2.43 (0.33)	-2.35 (0.47)
	High	BS1	-2.79 (0.41)	-2.63 (0.51)	-2.37 (0.53)	-2.53 (0.36)	-2.52 (0.27)	-2.33 (0.41)
	Low	BS2	-2.87 (0.49)	-2.53 (0.48)	-2.52 (0.56)	-2.62 (0.52)	-2.32 (0.48)	-2.25 (0.65)
	High	BS2	-2.91 (0.36)	-2.47 (0.61)	-2.30 (0.62)	-2.45 (0.45)	-2.46 (0.36)	-2.27 (0.42)
T3	Low	BS1	-2.42 (0.39)	-2.22 (0.54)	-2.20 (0.58)	-2.46 (0.38)	-2.33 (0.43)	-2.23 (0.58)
	High	BS1	-2.32 (0.47)	-2.14 (0.36)	-2.32 (0.33)	-2.64 (0.26)	-2.39 (0.37)	-2.26 (0.41)
	Low	BS2	-2.34 (0.47)	-2.32 (0.60)	-2.30 (0.46)	-2.37 (0.36)	-2.43 (0.48)	-2.20 (0.56)
	High	BS2	-2.31 (0.49)	-2.32 (0.56)	-2.11 (0.52)	-2.41 (0.29)	-2.44 (0.37)	-2.23 (0.56)
T4	Low	BS1	-2.49 (0.36)	-2.18 (0.34)	-2.13 (0.48)	-2.55 (0.46)	-2.44 (0.34)	-2.16 (0.53)
	High	BS1	-2.25 (0.59)	-2.20 (0.48)	-2.08 (0.61)	-2.50 (0.35)	-2.43 (0.38)	-2.23 (0.51)
	Low	BS2	-2.42 (0.57)	-2.13 (0.54)	-2.01 (0.57)	-2.57 (0.39)	-2.30 (0.69)	-2.29 (0.53)
	High	BS2	-2.29 (0.54)	-2.21 (0.59)	-2.14 (0.51)	-2.45 (0.38)	-2.42 (0.30)	-2.28 (0.35)

The rate of change in variance, for its additive as for its dominance component, was relatively uniform across all traits, breeding schemes and GxE and dominance levels (Table 4-4). On average, T1 lost 2.54% of its initial additive variance and 2.45% of its dominance variance at each cycle. T2 lost respectively 2.65% and 2.43% of its additive and dominance variance per cycle. Each cycle, T3 lost 2.31% of its additive variance and 2.38% of its dominance variance and T4 2.24% of its additive variance and 2.41% of its dominance variance.

For all traits, the levels of GxE had a significant effect on both the rate of additive and dominance variance loss. In all cases the reductions of the variance were smaller under larger GxE levels.

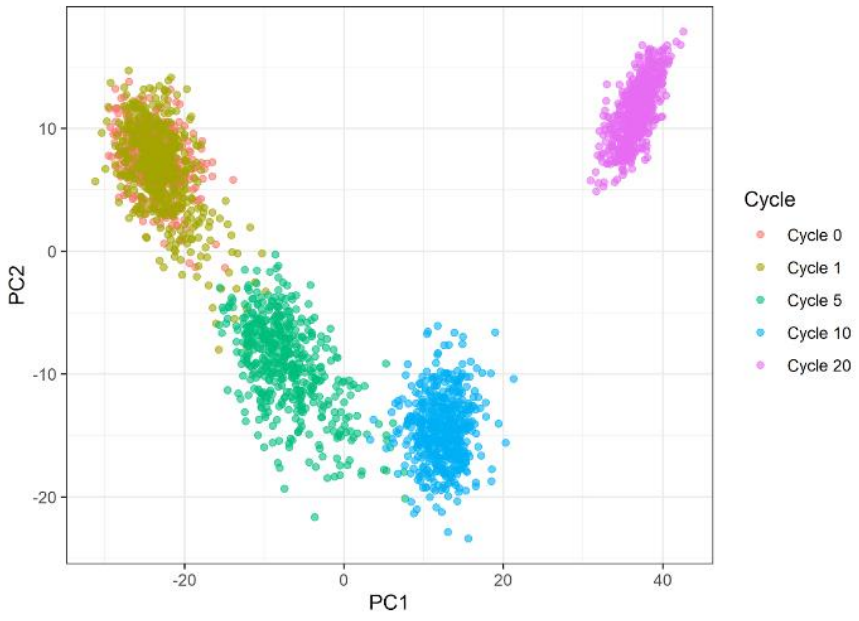
Neither the breeding schemes nor the set dominance variance influenced the evolution of the variance component (STable 4-5).

4.5 Discussion

4.5.1 Simulation as a tool to optimize breeding strategy

The optimization of breeding strategies via simulations is an important component of breeding program modernization (Sun et al., 2011). Deterministic simulations such as the one proposed by (Atlin and Econopouly, 2022) can give interesting information to rank breeding schemes. However, this type of simulation relies on the assumption at the base of the breeder's equation (the infinitesimal genetic model, independent QTL, constant additive variance) and it does not consider the evolution of the population with the time. This limits the conclusion that one can draw from this type of approach. Recently, several tools have been developed to perform stochastic simulations and help breeders evaluate alternative breeding schemes. Among them, AlphaSimR (Gaynor et al., 2021) , ADAM-Plant (Liu et al., 2019), BSL (Yabe et al., 2017) or MoBPS (Pook et al., 2020) allowed the users to design breeding schemes in silico and test different hypothesis. The advantage of such an approach is the

Figure 4-7: Principal component analysis on the S0 genotype at cycle 0, 1, 10, 20. Only the results for the high GxE high Dom and of one replicate are shown



possibility to model complex breeding schemes and access different properties of the population under selection. In the present study, the parameters were chosen to reflect the characteristics of the CIAT-Cirad breeding population. Indeed, the effective population size, the mean and the variance components associated with each trait, and the correlations between traits were defined based on the data from the most recent population. However, we were constrained to do some simplifications. For example, the male sterility gene was not directly simulated. Instead, the fraction of fertile and male sterile plants was computed based on simple Mendelian inheritance at each generation. As we did not choose a position and defined it as our ms-gene and we did not simulate any potential linkage drag around this gene. We did not simulate a specific ms-gene because first we had no control on the effect of QTLs that would be simulated around it and second, the drag around the gene had been shown to be at worst negligible (Frouin et al., 2014).

Another simplification is forced by the AlphaSimR. If one wants to simulate traits with defined correlation, all the traits will have exactly the same of QTLs. Therefore, if one QTL is fixed because it is favourable for one trait, the same QTL will also be fixed for any other traits. This will automatically negatively impact the variability of the other traits, even the one we choose not to select. This is of importance as one goal of RS is to improve some trait of interest while keeping a reservoir of variability for future selection.

4.5.2 Expected rate of genetic gain

Maximizing the rate of genetic gain in a given context (crop, resources, breeding objectives) is an important objective of every breeding program. Even if there are multiple ways to assess the performance of a breeding program, the genetic gain for a particular trait (usually associated with the productivity) is the main indicator (Ceccarelli, 2015; Rutkoski, 2019). In the present study, we found rates of genetic gain ranging from 1.37% to 2.41% for the trait representing grain yield. These values are medium to high compared to the values reported in the literature on real data. Using data from the irrigated rice breeding program at the International Rice Research institute, Juma et al. (2021) reported a rate of genetic gain for grain yield of 0.23% per year. For upland rice, few studies have been conducted to estimate the rate of genetic gain for grain yield. They reported a wide range of annual gain: 0.67% for advanced lines (Breseghello et al., 2011), and up to 2.68% with segregating material ($F_{3:4}$) (Barros et al., 2018). Interestingly, Morais Júnior et al. (2017) reported a gain of 1.96% per year for a breeding program based on population improvement through RS. These results highlighted that the expected genetic gain estimated with our simulation are realistic even if in our case the breeding cycle is only one year with the use of GP and that in the majority of the studies based on real data breeding cycles are longer (3-6 years). Hence, we could have expected higher genetic gain in our context. Indeed, shortening the breeding cycle is the most efficient way to improve the rate of genetic gain (Cobb et al.,

2019). Different simulation studies have proposed to go even further by achieving two to three cycles per year with the use of GP (Muleta et al., 2018). In the field, rice is grown between 90 days to 120 days depending on the genetic parameters as well environment parameters. Considering logistics, and under tropical conditions two generations per year is a routine for most breeding programs when water is available. However, in this context, the program must genotype two cohorts per year. In addition, the effect of the increased number of cycles per year on the genetic gain is limited if no measures are taken to limit the loss of variability (Gaynor et al., 2017; Gorjanc et al., 2018). When only a small number of parents were used in the population improvement, the increase in genetic gain was possible only by passing from one to two cycles per year. Any increase in the number of cycles resulted in a loss of genetic gain after 20 years (Gorjanc et al., 2018). In addition, selection of parents only on their performances strongly reduces the efficiency of converting variability into genetic gain and larger number of cycles per unit of time even increase this effect (Gorjanc et al., 2018). Finally, the drop in accuracy in the first cycles also showed that predictions are less reliable the more crossing events there are between the genotypes of the calibration set and the predicted genotypes. The benefit from more than one RS cycle per year might be compensated by the loss in accuracy. This need, however, to be further investigated.

4.6 Literature cited

Atlin, G.N., Econopouly, B.F., 2022. Simple deterministic modeling can guide the design of breeding pipelines for self-pollinated crops. Crop Sci. 62, 661–678. https://doi.org/10.1002/csc2.20684

Baertschi, C., Cao, T.-V., Bartholomé, J., Ospina, Y., Quintero, C., Frouin, J., Bouvet, J.-M., Grenier, C., 2021. Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. G3 GenesGenomesGenetics 11, jkab320. https://doi.org/10.1093/g3journal/jkab320

Barros, M.S., Morais Júnior, O.P., Melo, P.G.S., Morais, O.P., Castro, A.P., Breseghello, F., 2018. Effectiveness of early-generation testing applied to upland rice breeding. Euphytica 214, 61. https://doi.org/10.1007/s10681-018-2145-z

Bebber, D.P., Ramotowski, M.A.T., Gurr, S.J., 2013. Crop pests and pathogens move polewards in a warming world. Nat. Clim. Change 3, 985–988. https://doi.org/10.1038/nclimate1990

Bernardo, R., 1994. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Sci. 34, 20–25. https://doi.org/10.2135/cropsci1994.0011183X003400010003x

Breseghello, F., de Morais, O.P., Pinheiro, P.V., Silva, A.C.S., da Maia de Castro, E., Guimarães, É.P., de Castro, A.P., Pereira, J.A., de Matos Lopes, A., Utumi, M.M., de Oliveira, J.P., 2011. Results of 25 Years of Upland Rice Breeding in Brazil. Crop Sci. 51, 914–923. https://doi.org/10.2135/cropsci2010.06.0325

Ceccarelli, S., 2015. Efficiency of Plant Breeding. Crop Sci. 55, 87–97. https://doi.org/10.2135/cropsci2014.02.0158

Cerón-Rojas, J.J., Crossa, J., 2019. Efficiency of a Constrained Linear Genomic Selection Index To Predict the Net Genetic Merit in Plants. G3 Bethesda Md 9, 3981–3994. https://doi.org/10.1534/g3.119.400677

Châtel, M., Ospina, Y., Rodriguez, F., Lozano, V.H., 2005. CIRAD/CIAT Rice Project: Population Improvement and Obtaining Rice Lines for the Savannah Ecosystem, in: Population Improvement: A Way of Exploiting the Rice Genetic Resources of Latin America. Rome, pp. 237–254.

Chen, G.K., Marjoram, P., Wall, J.D., 2009. Fast and flexible simulation of DNA sequence data 7.

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N.K., Tyagi, A.K., Soderlund, C., Dean, R.A., Wing, R.A., 2002. An Integrated Physical and Genetic Map of the Rice Genome. Plant Cell 14, 537–545. https://doi.org/10.1105/tpc.010485

Cobb, J.N., Juma, R.U., Biswas, P.S., Arbelaez, J.D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., Ng, E.H., 2019. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theor. Appl. Genet. 132, 627–645. https://doi.org/10.1007/s00122-019-03317-0

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh,

R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci. 22, 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Fehr, W.R., Fehr, E.L., Jessen, H.J., 1991. Principles of cultivar development. W.R. Fehr, Ames, Iowa.

Frouin, J., Filloux, D., Taillebois, J., Grenier, C., Montes, F., de Lamotte, F., Verdeil, J.-L., Courtois, B., Ahmadi, N., 2014. Positional cloning of the rice male sterility gene ms-IR36, widely used in the inter-crossing phase of recurrent selection schemes. Mol. Breed. 33, 555–567. https://doi.org/10.1007/s11032-013-9972-3

Frouin, J., Labeyrie, A., Boisnard, A., Sacchi, G.A., Ahmadi, N., 2019. Genomic prediction offers the most effective marker assisted breeding approach for ability to prevent arsenic accumulation in rice grains. PLoS One 14. https://doi.org/10.1371/journal.pone.0217516

Gallais, A., 1979. The concept of varietal ability in plant breeding. Euphytica 28, 811–823. https://doi.org/10.1007/BF00038955

Gaynor, C., 2020. Traits in AlphaSimR 7.

Gaynor, R.C., Gorjanc, G., Bentley, A.R., Ober, E.S., Howell, P., Jackson, R., Mackay, I.J., Hickey, J.M., 2017. A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. Crop Sci. 57, 2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gaynor, R.C., Gorjanc, G., Hickey, J.M., 2021. AlphaSimR: an R package for breeding program simulations. G3 GenesGenomesGenetics 11. https://doi.org/10.1093/g3journal/jkaa017

Gorjanc, G., Gaynor, R.C., Hickey, J.M., 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor. Appl. Genet. 131, 1953–1966. https://doi.org/10.1007/s00122-018-3125-3

Grenier, C., Cao, T.-V., Ospina, Y., Quintero, C., Châtel, M.H., Tohme, J., Courtois, B., Ahmadi, N., 2015. Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. PLOS ONE 10, e0136594. https://doi.org/10.1371/journal.pone.0136594

Hickey, J.M., Chiurugwi, T., Mackay, I., Powell, W., Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants, 2017. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat. Genet. 49, 1297.

Hull, F.G., 1946. Recurrent selection for specific combining ability in corn. J. Am. Soc. Agron. 37, 134–146. https://doi.org/10.2134/agronj1945.00021962003700020006x

Jannink, J.-L., Lorenz, A.J., Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9, 166–177. https://doi.org/10.1093/bfgp/elq001

Juma, R.U., Bartholomé, J., Thathapalli Prakash, P., Hussain, W., Platten, J.D., Lopena, V., Verdeprado, H., Murori, R., Ndayiragije, A., Katiyar, S.K., Islam, M.R., Biswas, P.S., Rutkoski, J.E., Arbelaez, J.D., Mbute, F.N., Miano, D.W., Cobb, J.N., 2021. Identification of an Elite Core Panel as a Key Breeding Resource to Accelerate the Rate of Genetic Improvement for Irrigated Rice. Rice 14, 92. https://doi.org/10.1186/s12284-021-00533-5

Laidig, F., Piepho, H.-P., Drobek, T., Meyer, U., 2014. Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. Theor. Appl. Genet. 127, 2599–2617. https://doi.org/10.1007/s00122-014-2402-z

Liu, H., Tessema, B.B., Jensen, J., Cericola, F., Andersen, J.R., Sørensen, A.C., 2019. ADAM-Plant: A Software for Stochastic Simulations of Plant Breeding From Molecular to Phenotypic Level

and From Simple Selection to Complex Speed Breeding Programs. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.01926

Lorenz, A.J., Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., Smith, K.P., Sorrells, M.E., Jannink, J.-L., 2011. Genomic Selection in Plant Breeding, in: Advances in Agronomy. Elsevier, pp. 77–123. https://doi.org/10.1016/B978-0-12-385531-2.00002-5

Lush, J.L., 1937. Animal breeding plans, Iowa State College Press. ed. Ames, Iowa.

Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Morais Júnior, O.P., Breseghello, F., Duarte, J.B., Morais, O.P., Rangel, P.H.N., Coelho, A.S.G., 2017. Effectiveness of Recurrent Selection in Irrigated Rice Breeding. Crop Sci. 57, 3043–3058. https://doi.org/10.2135/cropsci2017.05.0276

Muleta, K.T., Pressoir, G., Morris, G.P., 2018. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. G3 GenesGenomesGenetics g3.200932.2018. https://doi.org/10.1534/g3.118.200932

Müller, D., Schopp, P., Melchinger, A.E., 2017. Persistency of Prediction Accuracy and Genetic Gain in Synthetic Populations Under Recurrent Genomic Selection. G3 GenesGenomesGenetics 7, 801–811. https://doi.org/10.1534/g3.116.036582

Piepho, H.-P., Laidig, F., Drobek, T., Meyer, U., 2014. Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. Theor. Appl. Genet. 127, 1009–1018. https://doi.org/10.1007/s00122-014-2275-1

Pook, T., Schlather, M., Simianer, H., 2020. MoBPS - Modular Breeding Program Simulator. G3 GenesGenomesGenetics 10, 1915–1918. https://doi.org/10.1534/g3.120.401193

Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE 8, e66428. https://doi.org/10.1371/journal.pone.0066428

Rutkoski, J.E., 2019. Estimation of Realized Rates of Genetic Gain and Indicators for Breeding Program Assessment. Crop Sci. 59, 981–993. https://doi.org/10.2135/cropsci2018.09.0537

Singh, R.J., Ikehashi, H., 1981. Monogenic Male-sterility in Rice: Induction, Identification and Inheritance 1. Crop Sci. 21, 286–289. https://doi.org/10.2135/cropsci1981.0011183X002100020020x

Slavov, G.T., Davey, C.L., Bosch, M., Robson, P.R.H., Donnison, I.S., Mackay, I.J., 2018. Genomic index selection provides a pragmatic framework for setting and refining multi-objective breeding targets in Miscanthus. Ann. Bot. https://doi.org/10.1093/aob/mcy187

Sleper, J.A., Bernardo, R., 2018. Genomewide Selection for Unfavorably Correlated Traits in Maize. Crop Sci. 58, 1587–1593. https://doi.org/10.2135/cropsci2017.12.0719

Sun, X., Peng, T., Mumm, R.H., 2011. The role and basics of computer simulation in support of critical decisions in plant breeding. Mol. Breed. 28, 421–436. https://doi.org/10.1007/s11032-011-9630-6

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. 108, 20260–20264. https://doi.org/10.1073/pnas.1116437108

Trenberth, K., 2011. Changes in precipitation with climate change. Clim. Res. 47, 123–138. https://doi.org/10.3354/cr00953

Whittaker, J.C., Thompson, R., Denham, M.C., 2000. Marker-assisted selection using ridge regression. Genet. Res. 75, 249–252. https://doi.org/10.1017/S0016672399004462

Yabe, S., Iwata, H., Jannink, J.-L., 2017. A Simple Package to Script and Simulate Breeding Schemes: The Breeding Scheme Language. Crop Sci. 57, 1347. https://doi.org/10.2135/cropsci2016.06.0538

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J., Ruane, A.C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., Asseng, S., 2017. Temperature increase reduces global yields of major crops in four independent estimates. Proc. Natl. Acad. Sci. 114, 9326–9331. https://doi.org/10.1073/pnas.1701762114

4.7 Appendix 1

To simulate the presence of the *ms*-gene in the population we adapted the number of possible crosses and number of phenotyped plants per plot considering the fraction of the possible genotypes expected at each generation assuming Mendelian inheritance.

We know that S_0 carry either the genotype [Msms] or [msms], as the mother is necessarily male sterile hence [msms]. As the $S_{0:1}$ come from the selfing of a S_0, it can only be a S_0 with the genotype [Msms] and hence they follow the classic proportion for the selfing of a heterozygous 25:50:25 for [MsMs], [Msms] and [msms]. The same logic was followed for generation $S_{0:2}$ further but considering that the selfed plant harvested could be either [MsMs] or [Msms].

For the recombination steps we assumed that only a certain number of crosses were possible in the population considering that only 25% of the population could be allo-fecundated, that the crosses only occurred between direct neighbours and that we expect that each plant has 6 male fertile neighbours (75% of 8). For the phenotyping steps, the maximum number of plants per plot was multiplied by the expected fraction of male fertile plants to simulate a phenotyping on a reduced number of plants. This was done up to generation $S_{0:4}$ (90% of male fertile plants). In more advanced generations it was considered as negligible and phenotyping plots were always considered fully fertile.

4.8 Supplementary Tables

STable 4-1: Analyze of variance on the mean genetic gain per cycle for population S_0

Var	Df	T1			T2			T3			T4		
		Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)
Scenario	1	21.755	15.476	0	1.682	7.374	0.007	0.003	0.032	0.858	0.01	0.074	0.787
GxE	2	52.779	37.545	0	13.127	57.564	0	0.278	2.551	0.08	0.094	0.707	0.494
Dom	1	5.793	4.121	0.044	1.069	4.688	0.031	0.83	7.609	0.006	0.282	2.111	0.148
Scenario:GxE	2	1.02	0.726	0.485	0.032	0.142	0.867	0.099	0.909	0.404	0.021	0.155	0.856
Scenario:Dom	1	0.162	0.115	0.735	0.06	0.262	0.609	0.156	1.433	0.233	0.028	0.208	0.649
GxE:Dom	2	2.431	1.729	0.18	0.496	2.175	0.116	0.145	1.331	0.266	0.137	1.024	0.361
Scenario:GxE:Dom	2	0.268	0.19	0.827	0.012	0.051	0.95	0.13	1.193	0.305	0.028	0.209	0.812
Residuals	228	1.406			0.228			0.109			0.134		

STable 4-2: ANOVA on the mean genetic gain per cycle for varieties

Var	Df	T1			T2			T3			T4		
		Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)
Scenario	1	3.02	1.28	0.26	0.96	6.87	0.01	0.01	0.04	0.85	0.03	0.08	0.78
GxE	2	43.88	18.54	0.00	6.33	45.15	0.00	0.40	1.19	0.31	0.86	2.62	0.08
Dom	1	10.79	4.56	0.03	0.87	6.21	0.01	0.06	0.19	0.66	0.11	0.32	0.57
Scenario:GxE	2	2.70	1.14	0.32	0.06	0.44	0.65	0.39	1.15	0.32	0.10	0.31	0.73
Scenario:Dom	1	1.63	0.69	0.41	0.08	0.59	0.44	0.60	1.77	0.18	0.16	0.48	0.49
GxE:Dom	2	1.68	0.71	0.49	0.12	0.88	0.42	0.33	0.96	0.38	0.43	1.32	0.27
Scenario:GxE:Dom	2	0.12	0.05	0.95	0.02	0.14	0.87	0.20	0.58	0.56	0.16	0.48	0.62
Residuals	228	2.37			0.14			0.34			0.33		

STable 4-3: ANOVA on the mean prediction accuracy

Var	Df	T1			T2			T3			T4		
		Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)	Mean Sq	F value	Pr(>F)
Scenario	1	0.28	23.29	0.00	0.14	13.70	0.00	0.27	30.79	0.00	0.10	14.61	0.00
GxE	2	0.84	69.20	0.00	0.92	89.82	0.00	0.64	72.22	0.00	0.43	63.34	0.00
Dom	1	0.06	4.86	0.03	0.04	3.49	0.06	0.00	0.00	0.98	0.01	1.17	0.28
Scenario:GxE	2	0.01	0.92	0.40	0.00	0.37	0.70	0.01	1.56	0.21	0.00	0.65	0.52
Scenario:Dom	1	0.00	0.07	0.80	0.00	0.23	0.63	0.00	0.00	0.99	0.00	0.13	0.72
GxE:Dom	2	0.00	0.31	0.74	0.00	0.28	0.76	0.00	0.02	0.98	0.00	0.64	0.53
Scenario:GxE:Dom	2	0.00	0.12	0.89	0.00	0.07	0.94	0.00	0.08	0.92	0.00	0.12	0.89
Residuals	228	0.01			0.01			0.01			0.01		

STable 4-4: Summary statistics for the prediction accuracies across all 20 cycles

GxE	Dom	Scenario	T1				T2				T3				T4			
			mean	SD	max	min	mean	SD	max	min	mean	SD	max	min	mean	SD	max	min
Low GxE	Low	BS1	0.72	0.07	0.78	0.58	0.69	0.06	0.77	0.55	0.74	0.1	0.84	0.52	0.78	0.07	0.85	0.6
Low GxE	Low	BS2	0.69	0.05	0.74	0.54	0.67	0.05	0.73	0.52	0.71	0.1	0.81	0.48	0.76	0.07	0.82	0.6
Low GxE	High	BS1	0.71	0.07	0.78	0.53	0.68	0.06	0.76	0.53	0.74	0.09	0.84	0.55	0.76	0.07	0.84	0.59
Low GxE	High	BS2	0.65	0.05	0.72	0.51	0.64	0.05	0.73	0.51	0.7	0.1	0.8	0.5	0.74	0.06	0.81	0.59
Medium GxE	Low	BS1	0.62	0.05	0.66	0.51	0.58	0.04	0.71	0.49	0.66	0.09	0.77	0.49	0.71	0.07	0.78	0.57
Medium GxE	Low	BS2	0.55	0.04	0.62	0.47	0.52	0.04	0.66	0.47	0.6	0.1	0.7	0.4	0.66	0.06	0.73	0.55
Medium GxE	High	BS1	0.59	0.05	0.66	0.48	0.54	0.05	0.69	0.47	0.66	0.1	0.78	0.44	0.69	0.07	0.77	0.52
Medium GxE	High	BS2	0.54	0.03	0.6	0.46	0.48	0.06	0.65	0.42	0.59	0.09	0.7	0.4	0.64	0.06	0.71	0.51
High GxE	Low	BS1	0.56	0.04	0.63	0.46	0.48	0.05	0.66	0.4	0.59	0.09	0.71	0.44	0.63	0.09	0.73	0.44
High GxE	Low	BS2	0.47	0.04	0.56	0.38	0.45	0.06	0.65	0.36	0.49	0.09	0.61	0.32	0.6	0.08	0.69	0.42
High GxE	High	BS1	0.51	0.03	0.57	0.45	0.48	0.06	0.64	0.39	0.59	0.09	0.7	0.39	0.65	0.07	0.72	0.49
High GxE	High	BS2	0.42	0.04	0.5	0.35	0.42	0.07	0.6	0.32	0.5	0.09	0.62	0.31	0.6	0.06	0.67	0.46

STable 4-5: ANOVA on the difference between initial and final variance for the additive component and for the dominance component

Variance	**Variable**	**Df**	**T1**			**T2**			**T3**			**T4**		
			Sum Sq	**F value**	**P value**	**Sum Sq**	**F value**	**P value**	**Sum Sq**	**F value**	**P value**	**Sum Sq**	**F value**	**P value**
Additive	Scenario	1	0.07	0.30	0.58	0.12	0.51	0.48	0.01	0.04	0.84	0.03	0.09	0.76
	GxE	2	2.76	5.76	0.00	6.74	14.30	0.00	0.61	1.28	0.28	3.09	5.64	0.00
	Dom	1	0.55	2.31	0.13	0.43	1.80	0.18	0.13	0.53	0.47	0.06	0.21	0.65
	Scenario:GxE	2	0.30	0.62	0.54	0.76	1.61	0.20	0.46	0.96	0.39	0.00	0.01	0.99
	Scenario:Dom	1	0.03	0.14	0.71	0.00	0.01	0.93	0.03	0.14	0.71	0.21	0.78	0.38
	GxE:Dom	2	0.37	0.77	0.46	0.50	1.06	0.35	0.01	0.03	0.97	0.75	1.36	0.26
	Scenario:GxE:Dom	2	0.00	0.00	1.00	0.03	0.06	0.94	0.49	1.03	0.36	0.04	0.07	0.93
	Residuals	228	54.68			53.74			54.39			62.46		
Dominance	Scenario	1	0.10	0.58	0.45	0.23	1.21	0.27	0.10	0.55	0.46	0.00	0.00	1.00
	GxE	2	2.52	7.03	0.00	2.50	6.55	0.00	2.43	6.43	0.00	3.12	7.83	0.00
	Dom	1	0.00	0.00	0.97	0.00	0.00	1.00	0.19	1.02	0.31	0.00	0.00	0.99
	Scenario:GxE	2	0.23	0.64	0.53	0.05	0.12	0.89	0.54	1.43	0.24	0.28	0.71	0.49
	Scenario:Dom	1	0.19	1.05	0.31	0.00	0.01	0.92	0.06	0.32	0.57	0.00	0.00	0.95
	GxE:Dom	2	0.24	0.66	0.52	0.49	1.28	0.28	0.07	0.19	0.83	0.21	0.52	0.60
	Scenario:GxE:Dom	2	0.28	0.79	0.46	0.08	0.21	0.81	0.06	0.15	0.86	0.16	0.41	0.67
	Residuals	228	40.79			43.53			43.00			45.41		

STable 4-6: Details of the BS1 breeding scheme.

Year	Semester	Step	Generation in field	Nb families	Plant per families	Selected families	Plant produced per selected families	Selected on
1	A	Cross	S01	50	60	500	1	random sampling of crosses
	B	Self S0	S0	500	1	50 + 150	34 * 3/4 for adv. 60 for cross	GEBV
2	A	Self S01	S01	200	34	all	34 for adv. 348 for phenot.* 5/6	
	B	Self S02	S02	200	34	all	34 for adv. 348 for phenot. * 9/10	
3	A	Pheno PAL S02	S02	200	174	none		
	B	Pheno SRO S02	S02	200	174	none		
4	A	Pheno PAL S03	S03	200	174	none		
	B	Pheno SRO S03	S03	200	174	100	34	S02 and S03 phenotype + S0 genotypes (GBLUP)
5	A	Self S03	S03	100	34	all	78	
	B	Pedigree breeding	S04	100	78	20	5	visual phenotyping
6	A							
	B	Pedigree breeding	S45	100	78	20	5	visual phenotyping
7	A	Multiplication	S56	100	34	all	816	
	B	OYT	S57	100	408			
8	A	OYT	S57	100	408	50	1836	adjusted phenotypes
	B	PYT	S58	50	918			
9	A	PYT	S58	50	918	20	10200	adjusted phenotypes
	B	AYT	S59	20	5100			
10	A	AYT	S59	20	5100	5		

STable 4-7: Details of the BS2 breeding scheme

Year	Semester	Step	Generation in field	Nb families	Plant per families	Selected families	Plant produced per selected families	Selected on
1	A	Cross S01	S01	50	60	500	1	random sampling of crosses
	B	Self S0	S0	500	1	50 + 150	34 * 3/4 for adv. 60 for cross	GEBV
2	A	Self S01	S01	200	34	all	34 for adv. 348 for phenot.* 5/6	
	B	Pheno SRO S02	S02	200	174	none		
3	A	Pheno PAL S02	S02	200	174	100	34	S02 and S03 phenotype + S0 genotypes (GBLUP)
	B	Self S02	S02	100	34	all	34	
4	A	Self S03	S03	100	34	all	78	
	B	Pedigree breeding	S04	100	78	20	5	visual phenotyping
5	A							
	B	Pedigree breeding	S45	100	78	20	5	visual phenotyping
6	A	Multiplication	S46	100	34	all	816	
	B	OYT	S47	100	408			
7	A	OYT	S47	100	408	50	1836	adjusted phenotypes
	B	PYT	S48	50	918			
8	A	PYT	S48	50	918	20	10200	adjusted phenotypes
	B	AYT	S49	20	5100			
9	A	AYT	S49	20	5100	5		

STable 4-8: Details for the population sizes at the different simulated steps on the initial population.

Year	Semester	Step	Generation in field	Nb families	Plant per families	Selected families	Plant per selected families	Selected on
1	A	Cross	S01	50	60	400	1	random sampling of crosses
	B	Self S0	S0	400	1	all	34 for adv. 60 for cross	
2	A	Self S01	S01	400	34	all	34 for adv. 492 for phenot.	
	B	Self S02	S02	400	34	all	34 for adv 492 for phenot.	
3	A	Pheno S02	S02	400	246	none	--	
	B	Pheno S02	S02	400	246	none	--	
4	A	Pheno S03	S03	400	246	none	--	
	B	Pheno S03	S03	400	246	50	60	random sampling

Chapter 5 : General discussion

5.1 Lesson learned

The field experiments of chapter 2 and 3 showed us that prediction can be realized using S_0 genotypes and phenotypes from generation $S_{0:2}$ and $S_{0:3}$, however with PA not exceeding 0.5 for any traits. In general, multi-environment calibration did not outperform single-environment calibrations and in some cases, had even lower PA. The traits did show some differences both in general PA and in response to multi-environment calibration (Table 5-1). One multi-environment approach which took advantage of the two sites was the imbalance approach (IMB) where each line predicted in SRO had phenotypic data in PAL, with the strongly depending on the traits and their the site correlation (STable 2-9, p.77). However, when single generation data were replaced by multi-generation data in the same approach (Multi2) and $S_{0:4}$ was predicted PA did not compete with single-environment approaches (Chapt. 3). From the different GxE variance-covariance structures none was systematically better for all traits.

Considering only the PA, the general conclusion was that multi-environment GP was not necessary in our situation, as we specifically aimed at predicting SRO and that PAL were, at best, of little utility despite the large phenotyping effort (BAL1), and at worst, detrimental for the PA (Multi2). Considering the cost as well as some practical aspects of the breeding program, Multi2 could however be interesting despite its slightly lower PA.

From the simulation we learned that $S_{0:2}$ phenotypes or the mean of $S_{0:2}$ and $S_{0:3}$ phenotypes can be medium to average proxies for the line ability. At least under the simple conditions of a simulation. Those results are slightly contradictory with the one obtained in chapter 3, where PA were low for $S_{0:2}$, $S_{0:3}$ (Table 5-1) but even lower when the mean of $S_{0:2}$ and $S_{0:3}$ was predicted (data not shown). We also learned that two generations of phenotyping improve the accuracy for traits with medium heritability (T1) more than for the one with low heritability (T2). In any case, BS2, could not compete with BS1 under the most favourable conditions. The most important result from the simulation is that *forward prediction* – or the prediction of current breeding cycles with data from older cycles – is possible.

Table 5-1: Predictive abilities from the different experiment conducted in chapter 2-3-4. For BAL1 (multi-environment CV with balanced representation of both sites while SIN stands for CV within SRO) the PA from the calibration done with 200 genotypes is presented and a GBLUP approach (Chap. 2). For Multi2 (Cal on PAL $S_{0:2}$ et SRO $S_{0:3}$, val on $S_{0:4}$) the best model of each trait with 50% of genotypes phenotyped in SRO are represented while Uni1(CV within $S_{0:4}$), Uni2 (Cal. on $S_{0:2}$ val. on $S_{0:4}$) and Uni3(Cal. on $S_{0:3}$ val. on $S_{0:4}$) had only one value per trait to choose from (Chapt. 3). For BS1 and BS2 the results are the one at the highest GxE and dominance levels (Chapt. 4). Careful, as not all the chapter followed the same order for the traits, they were ordered following the more common order seen in chapter 2 and 3.

	SIN_ S02	**SIN_ S03**	**BAL1_ S02**	**BAL1_ S03**	**Uni1**	**Uni2**	**Uni3**	**Multi2**	**BS1**	**BS2**
T3/FL	0.32	0.40	0.31	0.41	0.23	0.31	0.23	0.22	0.59	0.50
T4/PH	0.47	0.46	0.47	0.46	0.31	0.39	0.25	0.30	0.65	0.60
T2/YLD	0.39	0.35	0.41	0.39	0.39	0.33	0.24	0.27	0.48	0.42
T1/ZN	0.35	0.31	0.34	0.32	0.17	0.32	0.29	0.23	0.51	0.42

5.2 Limitations of the approach

5.2.1 Suspicion of generation effect on the prediction

Fifty temporal checks were used to trace year and generation effects as well as their interaction effects on the phenotypes. They highlighted a strong year effect but neither a generation effect nor a GxY effect (STable 2-5, p.75). Based on those results combined with the absence of systematic changes in PA between the generations, we concluded in chapter 2 that the differences between $S_{0:2}$ and $S_{0:3}$ were only due to year effect on the phenotype. However, we observed in chapter 3 a systematic decrease in PA when $S_{0:4}$ was predicted with phenotypes from increasingly more advance generation ($S_{0:2}$, $S_{0:3}$ and $S_{0:4}$). This could be a sign that mistakes accumulate at each generation advance making the prediction between generations harder.

As long as we cannot conclude on the reason why $S_{0:2}$ predict $S_{0:4}$ better than $S_{0:3}$ does and even better then $S_{0:4}$ itself, we cannot suggest an appropriate calibration approach. To exclude completely the year effect from the experimental setup, the calibration generation and the validation generation should be grown at the same time. We could use the 50 temporal checks in 2018 to predict $S_{0:3}$ with $S_{0:2}$. The population would be small but at least the PA would only be influenced by the generation.

5.2.2 No test with crossing event between calibration and prediction set

The greatest expectation for GP in our RS scheme is based on the forward prediction. As recombinations occur when the $S_{0:1}$ are crossed, it might change the LD between the marker and the QTL thus negatively affect the PA. As our prediction stayed always within the same RS cycle we could not assess the effect of recombination on the accuracy on real data. Time was clearly the limiting factor here. It took already three full years to collect the data for chapter 2 and 3, without counting the time necessary to produce the material. Assuming we run a crossing in November 2022 based on genomic selection of S01 families, the first validation data would be ready in October 2025. The results of the simulation were also rather comforting for the potential of forward prediction and the literature shows examples where models trained on parents or on former generations worked (Ben Hassen et al. 2017; Hickey et al. 2014; Bernal-Vasquez et al. 2017) so we can be confident on the feasibility of this approach. However, substantial biases have been reported under multi-generations prediction (Michel et al. 2016), hence the necessity to test the forward prediction on the field.

5.2.3 No realistic simulation of our experimental design

One limitation of our simulation experiment was the oversimplification of the phenotyping design. In both predictions based on field data, we focused heavily on the effect of the two different sites but in the simulation, the only parameters we controlled to simulate the phenotyping were the GxE variance, the probability that one simulated phenotype represents a theoretical target site, the residual variance and the number of replicates. If it allows giving an across sites heritability, it is not possible to set site

correlations. One option would have been to simulate the two sites as two traits thus doubling the number of traits. We would have still lacked the effect of the sterility in SRO but this is a possible approach.

5.2.4 Difficulty to simulate realistic traits

Another point is how accurately we simulated the trait. T1 is supposed to simulate a Zn grain concentration-like trait from our field experiment but the amount of field data we had, two years and two sites at the time, were somewhat scarce to ensure a realistic simulation. The same goes for T2 that we wanted close to the yield per plot, T3 close to the flowering date and T4 to the plant height.

The genetic assumption behind the control also had to be simplistic and, in some regards, wrong. As discussed in chapter 4, all four traits had not only the same number of QTLs but all QTLs had pleiotropic effects on four traits. This is forced by AlphaSimR if wants to be able to define correlations between the traits. This, of course, has consequences on the trait simulation and on their evolution across the cycle. Traits not under selection will progressively loose variability as the good QTL for the selected traits get fixed.

Also, the accuracies we had are rather optimistic considering what we could reach with real data even within cycle. The GxE was probably under-estimated. We did base our GxE levels on the observed data however. One reason is probably that we used the true line ability as a reference and not a phenotype in a defined environment where noise can still be present in the adjusted phenotypic values.

5.2.1 Effect of the ms-gene

On aspect of the CIAT-Cirad upland rice breeding program that was little discussed but that is very unique in rice breeding is the use of the ms-gene. It allows more crosses with less work, which is of importance for programmes with few resources. Despite the absence of documentation, some criticise this system for favouring some ideotypes as the father. Tall plants will spread their pollen further and it will drop on the other plants while pollen from short plants will have less chance to reach the stigma of plants that have panicles above them. The early and late flowering genotypes are also expected to cross less with the rest of the population. Alleles being responsible of earliness or lateness would progressively be negatively selected and all the population would converge toward the mean earliness. There is also the risk of genetic drag by using the ms-gene. If no drag was observed around the ms-IR36 gene by Frouin et al. (2014) our genotyping showed an increase heterozygosity around the position of the gene (SFig 5-1).

For those two reasons, the rice breeding program from the EMBRAPA (Brazilian agronomical state research) abandoned the ms-gene and went back to manual crosses (EMBRAPA staff, personal communication). Under those conditions, all ideotypes can have an equal chance, providing they have the expected qualities. Also, they have control on both the male and the female of the crosses, while

only the female can be selected in the currently used scheme at CIAT-Cirad and only based on visual assessment within the crossing plot. If it requires more work it allow to increase the genetic gain (Rutkoski 2019). It is said in the presentation of the breeding scheme that the candidates for selection, phenotypic or genotypic, are randomly selected. But the human eye tends to spot the unusual. Even if I haven't done this step personally, I do expect the "prettier" plants to be selected rather than a completely random sampling. This is not all bad as it is a good opportunity to cull-out the population for some high heritability trait like disease sensibility or grain colour.

5.3 Future evolution of the program

The CIAT-Cirad program will continue on path toward the implementation of genomic prediction in its RS scheme. The definitive phenotyping approach still need to be clarified. The program might in the short future focus the phenotyping on $S_{0:2}$ only but we know now that SRO will remain central in the selection and calibration process. Secondary traits showing good correlation between the sites like FL and PH will probably be phenotyped in PAL while YLD and ZN will still rely heavily on SRO.

A good strategy to clearly trace year and genotype by year interaction will have to be implemented to reduce to a minimum the confounded effect in the design.

The whole thesis was focalized on four traits, however many other traits such as grain dimension, amylose content or chalkiness are already measured routinely. They are good candidates for genomic prediction and will go through the same CV as presented in chapt.2 and 3 in the near future to decide which strategy to adopt for their calibration.

5.4 Perspective

If the results obtained in this thesis answer some questions, they also raised many more. I address a few of the improvement the breeding program could try to implement as well as some future researches to help to design the optimal breeding scheme.

5.4.1 Improvement of the GP

We used a simple approach to model GxE in our GP model. Many different models with a more sophisticated modelling of the environmental variance-covariance exist. Several have been suggested using either factor analytic (Burgueño, Campos, et al. 2012; Oakey et al. 2016) and could be used in our scheme. Another possible improvement for the GP that would be immediately available is the use of multi-variate prediction model (Calus and Veerkamp 2011; Ward et al. 2019; Ben-Sadoun et al. 2020; Fernandes et al. 2018).

Another resource that could easily be tapped in are the meteorological data. Already in 2014, Jarquìn et al. (2014b) used environmental covariates to construct a relationship matrix for the environments of a multi-environment trial. This would give us the opportunity to make better use of multi-year data

and even predict in theoretical or never observed environments (Jarquín et al. 2017b). More sophisticated approaches have since been tested such as the one from Millet et al. (2019b). It would however not be easily used for the CIAT-Cirad program as it relies partly on data from a phenotyping platform.

While, all those approaches are still based on mixed models, prediction could later take a more radical turn and use the deep learning prediction approach (Montesinos-López et al. 2016; Montesinos-López et al. 2021). This will be possible only when sufficient data are accumulated over the breeding cycles.

Changing the genotyping platform could also bring some improvement in the breeding scheme. Using array-based genotyping with specific markers for blast sensitivity for example could allow an early culling based on some eliminatory criterion. Custom genotyping chips with markers for major QTLs of quantitative traits would also allow us to use some markers with those QTLs as fixed effect (Zhang et al. 2014; Bernardo 2014; Bhandari et al. 2019; Ahmadi et al. 2020).

The training set used in chapter 4 was assembled by phenotyping the progeny of the S_0 showing the highest GEBV. One could argue that, as they are the one that will be crossed together to generate the improved population, they would be the most closely related to the future genotype to predict and hence be the better choice (Habier, Fernando, and Dekkers 2007; Hickey et al. 2014). On the other hand, this might reduce the coverage of the genetic space by the calibration set (Bustos-Korts et al. 2016). This would be interesting to test and could be very easily implemented in the program.

During the simulation, we also accumulated the calibration data without considering the utility of the oldest one. It is known that using calibration population containing genotypes that are genetically distant form the genotypes to predict can be detrimental to the accuracy (Lorenz and Smith 2015). Across the simulated cycles, we did follow the mean realized additive relationship of all genotypes. We could see the relationship drop during the three or four cycles without update but as soon as we started to update the calibration set the mean additive relationship stagnate or even increased again (SFig 5-2). We could not see a detrimental effect on the accuracy be we did not test it in detail but improvements are possible here. At one point it might also be necessary just to shorten the computation time.

Different methods were used to assemble the best calibration set. We tested the CDmean in chapter 3, but many other methods exist. They can be targeted – optimized for a specific set of genotypes – or untargeted. Some aim at capturing the diversity as efficiently as possible (Bustos-Korts et al. 2016; T. Guo et al. 2019) while others aim at maximizing reliability of the prediction (Akdemir, Sanchez, and Jannink 2015) or the reliability of the contrast between prediction candidate and individuals in the calibration set (Rincent et al. 2012). The selection could be applied on data from each cycle separately or on the complete data set assembled across the cycle.

Also concerning the ideal calibration set, the CDmean approach used in chapter 3 was not tailored for multi-environment set. Ben-Sadoun et al. (2020) extended the CDmean for a multi-trait context. We could consider a single trait in two sites as two different traits and, *a priori*, easily transfer their approach to our multi-environment case. Unfortunately, we had not time to test this approach. Additionally, it is not adapted for forward prediction as the CDmean optimizes a calibration set considering a defined set to be predicted.

5.4.2 Toward an optimization of the breeding scheme

The field experiments allowed us to test several calibration sets using different combinations of the generations and sites we had at our disposal. One calibration approach that gave good PA was the IMB (STable 2-9). This is however of little use for forward prediction as the high PA were due to the available data in the surrogate PAL site. It inspired us, however, a design to test ZN in SRO. As previously mentioned, SRO is valuable among other reasons for its high blast pressure. To accurately measure gain Zn concentration, we need healthy plants. On another hand, we want to measure YLD under high disease pressure. An imbalance approach could be integrated in the scheme where some line would be protected against blast while the rest would experience normal disease pressure. This would simplify later selection steps as the families would be already assessed under blast pressure. However, we do not know how those data would perform in a forward prediction model.

This thesis has been the opportunity to test several GP scenarios and breeding schemes, but optimization still needs to be done. For the schemes chosen on those preliminary results, we need now to find an optimal repartition of the financial and logistical resources on the population improvement part as well as the product development part. I could not so far study much about optimization approaches but for such a complex model a grid search is probably not feasible as the number of parameters to considered are highly dimensional. Nevertheless, the scripts used for the previous simulation were developed to be easily modulable. A few predefined size combinations could easily be tested to, in the worst case, at least find a local optimum. We will also need to further dissect fix and variable cost of phenotyping at the different sites and depending on the planned work (phenotyping or generation advance) as already started in chapter 3 to be sure to stay within the boundaries of the program.

The use of an index is rather straight forward but the choice of the weights is not. We ran our simulation with defined genetic gain targets in a number of standard deviations rather than by giving economic weights (Pešek and Baker 1969). If we chose those target values to represent the relative importance of T1/ZN and T2/YLD in the breeding program, the values were still arbitrary. More work must be done here. Alternative methods to use selection index in case of unknown economic weights are available for genomic prediction (Cerón-Rojas et al. 2008) but were not tested. Additionally, we

also used index selection for the yield trial. This is probably unrealistic because, for example, a really good T1/ZN value could make us keep a line that has T2/YLD below the best actual cultivar, and that no farmer would use. A culling approach is probably more appropriate and should also be tested.

A common simplification seen in most simulations of breeding schemes is the use of a closed population (Gaynor et al. 2017; Müller, Schopp, and Melchinger 2017; Muleta, Pressoir, and Morris 2018). Our simulation experiment was not different in that regard. Nevertheless, after 20 cycles of simulation, the variability of the population was reduced but not exhausted yet, despite the GP being based on RRBLUP, which has a tendency to select for related individuals (Ramasubramanian and Beavis 2020). The program might still want to take advantage of exotic material. In this case the effect on the GP and on the population itself could be investigated through simulation and method such as the one suggested by Allier et al. (2020) could be used.

We are especially concerned about the population variability as it is the source of variability for the pedigree breeding. So, it is important to preserve it as much as possible. We selected the S_0 families for the crosses based on their predicted progeny performance. Another approach for the selection of the recombining family could be made to balance genetic gain and fixation such as with optimal contribution selection (Woolliams et al. 2015).

If we start to consider the diversity that the crosses generate, one could also rethink the adequacy of the ms-gene. As previously mentioned, some ideotype might be favoured as father compared to others that could have a negative impact on the diversity. It was already observed that the plants showing very early or very late flowering tended to disappear and the flowering time would become more homogenous.

5.5 Literature cited

Ahmadi, Nourollah, Tuong-Vi Cao, Julien Frouin, Gareth J. Norton, and Adam H. Price. 2020. Genomic Prediction of Arsenic Tolerance and Grain Yield in Rice. Contribution of Trait-Specific Markers and Multi Environment Models, September. https://doi.org/10.1101/2020.09.28.316356.

Akdemir, Deniz, Julio I. Sanchez, and Jean-Luc Jannink. 2015. Optimization of Genomic Selection Training Populations with a Genetic Algorithm. Genetics Selection Evolution 47 (1): 38. https://doi.org/10.1186/s12711-015-0116-6.

Allier, Antoine, Simon Teyssèdre, Christina Lehermeier, Laurence Moreau, and Alain Charcosset. 2020. Optimized Breeding Strategies to Harness Genetic Resources with Different Performance Levels. BMC Genomics 21 (1). https://doi.org/10.1186/s12864-020-6756-0.

Ben Hassen, Manel, T. V. Cao, J. Bartholomé, G. Orasen, C. Colombi, J. Rakotomalala, L. Razafinimpiasa, et al. 2017. Rice Diversity Panel Provides Accurate Genomic Predictions for Complex Traits in the Progenies of Biparental Crosses Involving Members of the Panel. Theoretical and Applied Genetics 131 (2): 417–35. https://doi.org/10.1007/s00122-017-3011-4.

Ben-Sadoun, S., R. Rincent, J. Auzanneau, F. X. Oury, B. Rolland, E. Heumez, C. Ravel, G. Charmet, and S. Bouchet. 2020. Economical Optimization of a Breeding Scheme by Selective Phenotyping of the Calibration Set in a Multi-Trait Context: Application to Bread Making Quality. Theoretical and Applied Genetics 133 (7): 2197–2212. https://doi.org/10.1007/s00122-020-03590-4.

Bernal-Vasquez, Angela-Maria, Andres Gordillo, Malthe Schmidt, and Hans-Peter Piepho. 2017. Genomic Prediction in Early Selection Stages Using Multi-Year Data in a Hybrid Rye Breeding Program. BMC Genetics 18 (1): 51. https://doi.org/10.1186/s12863-017-0512-8.

Bernardo, Rex. 2014. Genomewide Selection When Major Genes Are Known. Crop Science 54 (1): 68–75. https://doi.org/10.2135/cropsci2013.05.0315.

Bhandari, Aditi, Jérôme Bartholomé, Tuong-Vi Cao-Hamadoun, Nilima Kumari, Julien Frouin, Arvind Kumar, and Nour Ahmadi. 2019. Selection of Trait-Specific Markers and Multi-Environment Models Improve Genomic Predictive Ability in Rice. PLOS ONE 14 (May): e0208871. https://doi.org/10.1371/journal.pone.0208871.

Burgueño, Juan, Gustavo de los Campos, Kent Weigel, and José Crossa. 2012. Genomic Prediction of Breeding Values When Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers. Crop Science 52 (2): 707–19. https://doi.org/10.2135/cropsci2011.06.0299.

Bustos-Korts, Daniela, Marcos Malosetti, Scott Chapman, Ben Biddulph, and Fred van Eeuwijk. 2016. Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. G3; Genes|Genomes|Genetics 6 (11): 3733–47. https://doi.org/10.1534/g3.116.035410.

Calus, Mario PL, and Roel F Veerkamp. 2011. Accuracy of Multi-Trait Genomic Selection Using Different Methods. Genetics Selection Evolution 43 (1). https://doi.org/10.1186/1297-9686-43-26.

Cerón-Rojas, J. Jesús, Fernando Castillo-González, Jaime Sahagún-Castellanos, Amalio Santacruz-Varela, Ignacio Benítez-Riquelme, and José Crossa. 2008. A Molecular Selection Index

Method Based on Eigenanalysis. Genetics 180 (1): 547–57. https://doi.org/10.1534/genetics.108.087387.

Fernandes, Samuel B., Kaio O. G. Dias, Daniel F. Ferreira, and Patrick J. Brown. 2018. Efficiency of Multi-Trait, Indirect, and Trait-Assisted Genomic Selection for Improvement of Biomass Sorghum. Theoretical and Applied Genetics 131 (3): 747–55. https://doi.org/10.1007/s00122-017-3033-y.

Frouin, Julien, Denis Filloux, James Taillebois, Cécile Grenier, Fabienne Montes, Frédéric de Lamotte, Jean-Luc Verdeil, Brigitte Courtois, and Nourollah Ahmadi. 2014. Positional Cloning of the Rice Male Sterility Gene Ms-IR36, Widely Used in the Inter-Crossing Phase of Recurrent Selection Schemes. Molecular Breeding 33 (3): 555–67. https://doi.org/10.1007/s11032-013-9972-3.

Gaynor, R. Chris, Gregor Gorjanc, Alison R. Bentley, Eric S. Ober, Phil Howell, Robert Jackson, Ian J. Mackay, and John M. Hickey. 2017. A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. Crop Science 57 (5): 2372–86. https://doi.org/10.2135/cropsci2016.09.0742.

Guo, Tingting, Xiaoqing Yu, Xianran Li, Haozhe Zhang, Chengsong Zhu, Sherry Flint-Garcia, Michael D. McMullen, et al. 2019. Optimal Designs for Genomic Selection in Hybrid Crops. Molecular Plant 12 (3): 390–401. https://doi.org/10.1016/j.molp.2018.12.022.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics 177 (4): 2389–97. https://doi.org/10.1534/genetics.107.081190.

Hickey, John M., Susanne Dreisigacker, Jose Crossa, Sarah Hearne, Raman Babu, Boddupalli M. Prasanna, Martin Grondona, et al. 2014. Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. Crop Science 54 (4): 1476–88. https://doi.org/10.2135/cropsci2013.03.0195.

Jarquín, Diego, José Crossa, Xavier Lacaze, Philippe Du Cheyron, Joëlle Daucourt, Josiane Lorgeou, François Piraux, et al. 2014. A Reaction Norm Model for Genomic Selection Using High-Dimensional Genomic and Environmental Data. Theoretical and Applied Genetics 127 (3): 595–607. https://doi.org/10.1007/s00122-013-2243-1.

Jarquín, Diego, Cristiano Lemes da Silva, R. Chris Gaynor, Jesse Poland, Allan Fritz, Reka Howard, Sarah Battenfield, and Jose Crossa. 2017. Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. The Plant Genome 10 (2). https://doi.org/10.3835/plantgenome2016.12.0130.

Lorenz, Aaron J., and Kevin P. Smith. 2015. Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. Crop Science 55 (6): 2657–67. https://doi.org/10.2135/cropsci2014.12.0827.

Michel, Sebastian, Christian Ametz, Huseyin Gungor, Doru Epure, Heinrich Grausgruber, Franziska Löschenberger, and Hermann Buerstmayr. 2016. Genomic Selection across Multiple Breeding Cycles in Applied Bread Wheat Breeding. Theoretical and Applied Genetics 129 (6): 1179–89. https://doi.org/10.1007/s00122-016-2694-2.

Millet, Emilie J., Willem Kruijer, Aude Coupel-Ledru, Santiago Alvarez Prado, Llorenç Cabrera-Bosquet, Sébastien Lacube, Alain Charcosset, Claude Welcker, Fred van Eeuwijk, and François

Tardieu. 2019. Genomic Prediction of Maize Yield across European Environmental Conditions. Nature Genetics 51 (6): 952–56. https://doi.org/10.1038/s41588-019-0414-y.

Montesinos-López, Osval A., Abelardo Montesinos-López, José Crossa, Fernando H. Toledo, Oscar Pérez-Hernández, Kent M. Eskridge, and Jessica Rutkoski. 2016. A Genomic Bayesian Multi-Trait and Multi-Environment Model. G3; Genes|Genomes|Genetics 6 (9): 2725–44. https://doi.org/10.1534/g3.116.032359.

Montesinos-López, Osval Antonio, Abelardo Montesinos-López, Paulino Pérez-Rodríguez, José Alberto Barrón-López, Johannes W. R. Martini, Silvia Berenice Fajardo-Flores, Laura S. Gaytan-Lugo, Pedro C. Santana-Mancilla, and José Crossa. 2021. A Review of Deep Learning Applications for Genomic Selection. BMC Genomics 22 (1): 19. https://doi.org/10.1186/s12864-020-07319-x.

Muleta, Kebede T., Gael Pressoir, and Geoffrey P. Morris. 2018. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. G3; Genes|Genomes|Genetics, December, g3.200932.2018. https://doi.org/10.1534/g3.118.200932.

Müller, Dominik, Pascal Schopp, and Albrecht E. Melchinger. 2017. Persistency of Prediction Accuracy and Genetic Gain in Synthetic Populations Under Recurrent Genomic Selection. G3; Genes|Genomes|Genetics 7 (3): 801–11. https://doi.org/10.1534/g3.116.036582.

Oakey, Helena, Brian Cullis, Robin Thompson, Jordi Comadran, Claire Halpin, and Robbie Waugh. 2016. Genomic Selection in Multi-Environment Crop Trials. G3; Genes|Genomes|Genetics 6 (5): 1313–26. https://doi.org/10.1534/g3.116.027524.

Pešek, J, and R J Baker. 1969. Desired Improvement In Relation To Selection Indices. Canadian Journal of Plant Science, no. 49: 803–4.

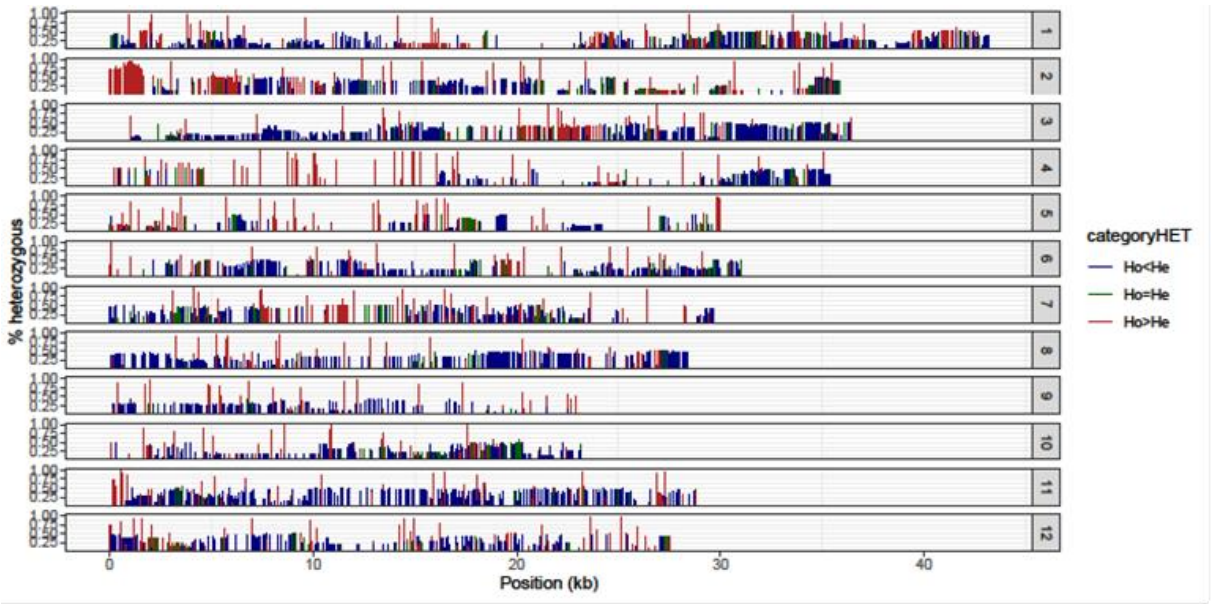
Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V.M. Rodríguez, et al. 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (Zea Mays L.). Genetics 192 (2): 715–28. https://doi.org/10.1534/genetics.112.141473.

Ward, Brian P., Gina Brown-Guedira, Priyanka Tyagi, Frederic L. Kolb, David A. Van Sanford, Clay H. Sneller, and Carl A. Griffey. 2019. Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. Crop Science 59 (2): 491–507. https://doi.org/10.2135/cropsci2018.03.0189.

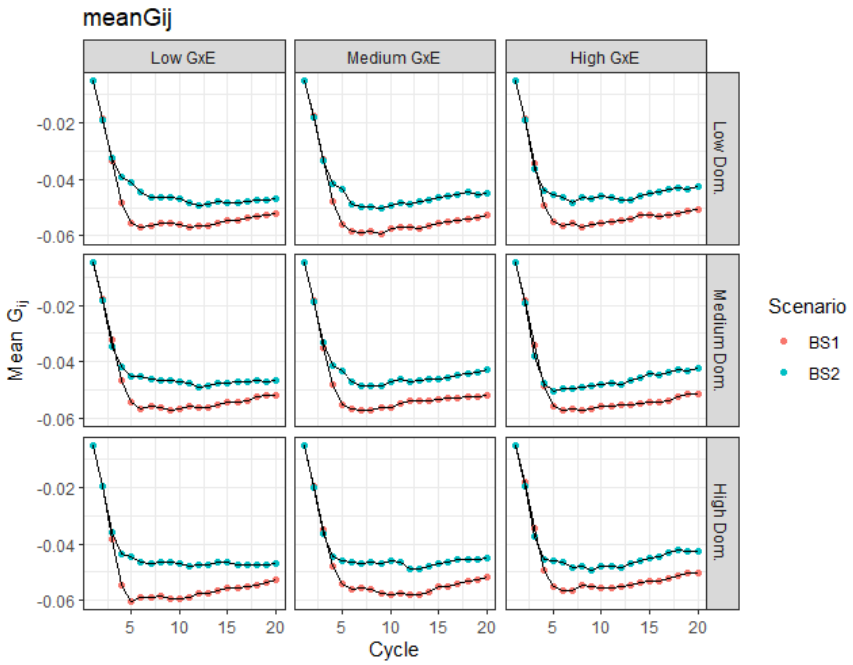
Woolliams, J. A., P. Berg, B. S. Dagnachew, and T. H. E. Meuwissen. 2015. Genetic Contributions and Their Optimization. Journal of Animal Breeding and Genetics 132 (2): 89–99. https://doi.org/10.1111/jbg.12148.

Zhang, Zhe, Ulrike Ober, Malena Erbe, Hao Zhang, Ning Gao, Jinlong He, Jiaqi Li, and Henner Simianer. 2014. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. Edited by Xiaodong Cai. PLoS ONE 9 (3): e93017. https://doi.org/10.1371/journal.pone.0093017.

5.6 Supplementary Figures



SFig 5-1: Comparison of expected (He) and observed (Ho) heterozygosity in the OCT27



SFig 5-2: Mean realized relationship between the genotypes in the calibration and validation population

Résumé de la thèse en français

Intégration de la prédiction génomique précoce dans un schéma de sélection récurrente: exemple du programme de sélection du riz pluvial du CIAT-Cirad

Introduction

Contexte général

Selon les estimations, la demande alimentaire mondiale va augmenter entre 45% et 51% d'ici 2050 (van Dijk et al. 2021). Deux tiers des besoins en calories de la population humaine sont couverts par seulement quatre cultures : le riz (*Oryza sativa* L.), le blé (*Triticum* sp.), le maïs (*Zea mays* L.) et le soja (*Glycine max* (L.) Merr). Jusqu'à maintenant la production mondiale combinées aux stocks a toujours permis de couvrir la demande pour ces quatre cultures (FAO 2021), mais les futurs besoins ne seront peut-être pas assurés (Ray et al. 2013).

C'est principalement l'augmentation des rendements qui a permis à l'offre de continuer à couvrir la demande malgré sa hausse constante ces dernières décennies (Fischer, Byerlee, et Edmeades 2014). L'augmentation du rendement a été en partie due au rétrécissement de l'écart de rendement grâce à de meilleures pratiques agricoles mais c'est surtout l'augmentation du rendement potentiel qui en a été à l'origine (Fischer, Byerlee, et Edmeades 2014).

Le rendement potentiel est le rendement d'une variété dans son environnement cible si elle est cultivée sans contraintes biotiques et abiotiques. Ceci sous-entend que le maintien ou l'augmentation du rendement potentiel dépend de la disponibilité de variétés adaptées à des environnements précis. Le changement climatique risque fortement de rendre d'anciennes variétés obsolètes pour certaines régions à cause de l'augmentation des températures ou des changements dans les régimes de précipitations qui en résulteront (IPCC 2022).

Les sélectionneurs ont donc la tâche de développer des variétés qui seront adaptées aux futures conditions. Les gains annuels en rendement potentiel ont été, ces dernières années, d'environ 0.5-0.8% pour le riz, le blé et le soja et de 1.1% pour le maïs (Fischer, Byerlee, et Edmeades 2014) mais ces progrès ne permettront pas de couvrir les besoins futurs (Ray et al. 2013). Encore pire, un ralentissement de l'augmentation du rendement potentiel a été observé ces dernières années. Les sélectionneurs ont la responsabilité de fournir aux paysans des variétés qui permettront d'assurer la future sécurité alimentaire. Pour ce faire un important travail de modernisation des programmes de sélection doit être entrepris pour s'assurer que l'augmentation du rendement potentiel continuent à couvrir l'augmentation de la demande pour les principales cultures alimentaires.

La sélection variétale chez les plantes

La sélection variétale est l'art de créer des plantes adaptées aux besoins des humains. Elle a progressivement évolué d'une pratique intuitive, lors de la domestication, en une réelle science avec la redécouverte du travail de Mendel et la fondation de la génétique quantitative par Fisher (Fisher 1919). Le rôle d'un sélectionneur est (i) de créer de la variabilité génétique, souvent en croisant deux parents élites, ensuite (ii) de sélectionner les meilleurs descendants du croisement et finalement (iii) de synthétiser les meilleures descendances en un nouvelles variété (Bernardo 2008).

La sélection variétale peut être appréhendée selon deux angles différents: soit le type de variété développée soit la méthodologie appliquée pour obtenir ladite variété. La méthodologie choisie va dépendre du type de variété voulu et du système de reproduction de l'espèce. Les différents types de variétés existant sont les populations synthétiques ou composite, les lignées pures, les hybrides et les variétés clonales. Deux systèmes de reproduction existent dans les plantes supérieures: l'allogamie et l'autogamie. Les plantes allogames sont principalement ou exclusivement, selon l'espèce, fécondées par le pollen d'autres plantes (de la même espèce). C'est le cas du maïs mais aussi du seigle (*Secale cereale* L.) par exemple. Les espèces autogames au contraire favorisent l'autofécondation et seulement une petite proportion des fécondations sont des croisements entre plantes. Le riz, le blé, le soja ou encore l'orge (*Hordeum vulgare* L.) en font partie.

Dans la pratique, le travail du sélectionneur consiste à favoriser l'autofécondation ou l'allofécondation selon s'il veut augmenter la variabilité ou l'homogénéité de la population sur laquelle il travaille.

Un des outils qui permet au sélectionneur de pousser une espèce autogame à se comporter en partie comme une allogame et donc de favoriser les croisements est l'utilisation de gènes nucléiques de stérilité mâle. La gestion d'une population ségrégant pour un gène de stérilité mâle dépendra du degré de dominance de celui-ci. En bref, le sélectionneur devra s'assurer que des phénotypes mâles fertiles et mâles stériles sont présents dans la population de travail ainsi toutes les plantes étant mâle stérile se comporteront de façon allogame et seront exclusivement allofécondées. Ce type de gène existe chez le blé, le sorgho (*Sorghum bicolor*) et le riz dont on parlera plus en détail.

Je vais présenter les trois méthodes de sélection importantes pour ce travail mais ils en existent beaucoup d'autres (Fehr, Fehr, et Jessen 1991).

La première et sans doute la plus simple est la sélection massale. Le travail consiste ici à identifier les phénotypes d'intérêt à l'intérieur d'une population présentant de la variabilité. Les descendances des différentes plantes sont mélangées et semées comme une nouvelle population à la saison suivante. Cette méthodologie est facile à appliquer mais fonctionne surtout pour les caractères présentant une héritabilité élevée.

La seconde est la sélection généalogique. Comme pour la sélection massale, on commencera par sélectionner des plantes parmi une population comportant de la diversité. Ensuite, plutôt que de

mélanger les semences, la généalogie de chaque lignée sera conservée et permettra au sélectionneur de sélectionner non seulement sur la base de la performance propre d'une plante, mais aussi relativement à sa famille et à ses ancêtres. Cette méthode est beaucoup plus adaptée pour des caractères à faible héritabilité.

La troisième méthode que nous aborderons plus en détail est la sélection récurrente. Elle consiste en la sélection systématique d'individus supérieurs dans une population, suivie d'une étape de recombinaison entre les individus sélectionnés pour générer une nouvelle population avancée. Elle est particulièrement adaptée pour les espèces allogames mais peut être appliquée sur des espèces autogames, entre autres en utilisant un gène de stérilité mâle.

Evaluation et amélioration d'un programme de sélection

Un programme de sélection s'évalue, notamment sur son gain génétique(Ceccarelli 2015), soit l'évolution de la moyenne de la population en sélection à travers les cycles de sélection. On peut soit mesurer le gain génétique réalisé en regardant l'évolution des moyennes, soit l'estimer en utilisant l'équation du sélectionneur $\Delta G_t = \frac{k r_{xg} \sigma_g}{L}$ (Lush 1937). Cette équation permet de savoir sur quels paramètres d'un programme influer pour en augmenter le gain génétique. On peut en augmenter l'intensité de sélection k, augmenter la corrélation entre les valeurs utilisées pour la sélection et les valeurs génétiques réelles r_{xg}, augmenter la racine carrée de la variabilité génétique de la population σ_g ou réduire la longueur du cycle de sélection L.

Simulation des programmes de sélection

Les méthodes pour influencer ces paramètres dans la direction permettant une augmentation du gain génétique peuvent être testées en champs. Néanmoins, il est souvent impossible de le faire pour des raisons logistiques ou financières. Une alternative aux essais en plein champs sont les simulations stochastiques.

Une simulation stochastique introduit l'incertitude en échantillonnant des valeurs d'entrée à partir d'une fonction de distribution de probabilité qui représente des processus stochastiques. L'utilisateur peut fixer les paramètres de la distribution mais des valeurs différentes seront échantillonnées à chaque fois que la simulation est exécutée. Dans le contexte de la simulation d'un schéma de sélection, ces processus stochastiques sont, par exemple, la probabilité qu'un allèle soit hérité, la probabilité qu'une plante soit échantillonnée ou la distribution de l'erreur aléatoire pendant le phénotypage, pour n'en citer que quelques-uns. En choisissant correctement les paramètres de la fonction de distribution, les simulations peuvent fournir des données précieuses sur les performances à long terme de programmes de sélection spécifiques dans des conditions particulières et aider à la prise de décision.

Dans la dernière décennie, de nombre software ont été développés spécifiquement pour la simulation stochastique de schéma de sélection (e.g. ADAM-plant (Liu et al. 2019), AlphaSimR (Gaynor, Gorjanc,

and Hickey 2021), BSL (Yabe, Iwata, and Jannink 2017), HaploSim (Coster and Bastiaansen 2010), MoBPS (Pook, Schlather, and Simianer 2020), QU-Gene (Podlich and Cooper 1998), …).

Sélection génomique

Le concept de la sélection génomique est relativement simple. On commence par entrainer un modèle statistique avec une population de référence dont on connait les génotypes et les phénotypes. Ceci nous permet d'estimer l'effet de chaque marqueur génétique et ainsi d'estimer la valeur génétique et de sélectionner des individus sur l'unique base de leur génotype. Elle a été développée dans les années 2000 (Whittaker, Thompson, et Denham 2000; Meuwissen, Hayes, et Goddard 2001) mais a énormément gagnée en popularité ces dernières années grâces à l'apparition de méthodes de génotypage bon marché (Crossa et al. 2017). Elle peut permettre l'amélioration des gains génétiques en influençant tous les paramètres de l'équation du sélectionneur. Le génotypage étant souvent moins cher que le phénotypage, de plus grandes populations peuvent être évaluées et l'intensité de sélection augmentée (k). Elle peut dans certain cas améliorer la précision de sélection (r_{xg}) et également améliorer l'intégration de nouvelle diversité dans une population de sélection (σ_g). Finalement, elle permet de découpler sélection et phénotypage et ainsi de réduire fortement la durée des cycles de sélection (L).

Le programme de sélection CIAT-Cirad

Depuis plus de 30 ans, le CIAT et le Cirad conduisent conjointement un programme d'amélioration chez riz pluvial pour l'Amérique latine et les Caraïbes. Contrairement à la majorité des programmes d'amélioration variétale pour le riz, le programme CIAT-Cirad a toujours été centré sur l'amélioration de ses populations par sélection récurrente.

La sélection récurrente consiste en une succession de cycles comprenant une étape d'évaluation des familles issues d'une population en amélioration, suivie par la sélection des meilleures familles basée sur cette évaluation et finalement la génération d'une population améliorée en recombinant les familles sélectionnées, constituant ainsi le début d'un nouveau cycle de sélection. L'évaluation des familles est traditionnellement pratiquée en avançant en génération la descendance des croisements candidats à la sélection, soit des plantes uniques en S_0, (S pour désigner le nombre de cycle d'autofécondation, donc ici une plante n'ayant subi aucune autofécondation et donc directement dérivée d'un croisement) jusqu'à des générations plus avancées ($S_{0:2}$ ou $S_{0:3}$). La moyenne de l'ensemble de la descendance pour chaque croisement est mesurée aux générations avancées afin d'obtenir une valeur sur descendance (Figure 1-9, p.28). Ce travail d'avancée en génération puis d'évaluation prend du temps, et les valeurs sur descendance ne sont pas disponibles avant au mieux 1.5 ans et au pire 4 ans selon le nombre de génération et d'environnement nécessaire au phénotypage. Le programme entend dans un futur proche améliorer son schéma de sélection récurrent en y

intégrant la prédiction génomique pour la sélection des parents à recombiner. Le potentiel de la prédiction génomique en terme d'augmentation des gains génétiques dépendra de son influence sur les différents paramètres de l'équation du sélectionneur dans le contexte donné du schéma de sélection CIAT-Cirad.

L'objectif général de ma thèse est de proposer différentes voies d'amélioration du programme de sélection basées sur la prédiction génomique. Cet objectif est abordé selon trois axes. Tout d'abord, le potentiel de la prédiction génomique sur les données issues de la population de sélection sera testé par validation croisée à l'intérieur des générations des descendances des croisements à prédire et en intégrant les données de phénotypage provenant des deux sites à disposition. Dans un second temps, des validations externes seront réalisées sur des descendances n'ayant pas contribuées à la calibration des modèles et impliquant des données multi-générations et multi-sites pour développer une approche de calibration qui utilise au mieux les données déjà générées par le programme de sélection. Finalement, le programme sera simulé pour évaluer les prédictions entre cycles de sélection et comparer deux schémas de sélection basés sur la sélection génomique.

Matériels et méthodes

Expérience en plein champs

Acquisition des données

Deux expériences en pleins champs ont été conduites sur deux sous-ensembles d'une même population, les PCT27A et PCT27B. Les deux sous-ensembles ont été génotypé en S_0. La PCT27A a ensuite été avancée en génération $S_{0:2}$ puis $S_{0:3}$ par bulk et phénotypée à ces deux générations dans deux sites Palmira (PAL) et Santa Rosa (SRO). Les deux sites font partie du dispositif d'essai du programme CIAT-Cirad pour l'évaluation sur descendances des candidats à la sélection. A PAL les essais sont logistiquement simples à réaliser car conduits en conditions irriguées dans une région avec une faible pression des maladies du riz et sur le site du CIAT. SRO de son côté est une station expérimentale éloignée du CIAT, non irrigué et donc où la culture est possible uniquement à la saison des pluies. De plus, la station est située dans un point chaud pour les maladies du riz telle que la pyriculariose (*Magnaporthe grisea*). SRO, bien que plus difficile à gérer, représente l'environnement cible du programme. Quatre caractères ont été mesurés dans chacun des quatre essais : le nombre de jours entre le semis et la floraison (FL), la hauteur de plante (PH), le rendement par parcelle (YLD) et la concentration en zinc du grain (ZN). La PCT27B a été avancée jusqu'en génération $S_{0:4}$ et a été phénotypée uniquement à cette génération pour les mêmes caractères et dans les mêmes sites que la PCT27A.

Prédiction génomique

Dans un premier temps la population PCT27A a été utilisée pour tester la validation intra-population une approche mono-environnement avec une calibration exclusivement sur des données SRO. Dans un deuxième temps, des calibrations intégrant les deux environnements ont été évaluées. Les prédictions génomiques ont été réalisées indépendamment pour chaque génération et suivant la procédure proposée par (Lopez-Cruz et al. 2015).

Plusieurs approches de validation croisées ont été utilisées pour tester différents cas de figures concernant la représentation des deux sites dans le set de calibration (Figure 2-2, p.48). Dans l'approche mono-site (SIN_{SRO}) les modèles ont été calibré sur les phénotypes de s familles pour $s \in \{25, 50, 100, 200\}$. Les approches BAL1 et BAL2 utilisent soit des données des deux sites pour toutes les lignées de calibration (BAL1), ou une combinaison de lignées phénotypées uniquement à PAL, uniquement à SRO et dans les deux sites. Finalement, pour IMB les modèles ont été calibrés sur l'ensemble des familles à PAL (n=334) complétés par les phénotypes de s familles à SRO.

La validation externe a consisté en la calibration de modèles de prédiction sur des données PCT27A pour la prédiction de génotypes de la population PCT27B selon plusieurs scénarios(Figure 3-2, p.89). Les scénarios Uni1, Uni2 et Uni3 utilisent des phénotypes $S_{0:2}$, $S_{0:3}$ ou $S_{0:4}$ à SRO pour prédire $S_{0:4}$ à SRO. Le scénario Multi1 utilise une combinaison de données PAL et SRO en $S_{0:4}$ pour prédire SRO en $S_{0:4}$. Finalement, Multi2 utilisée des données phénotypiques $S_{0:2}$ à PAL et $S_{0:3}$ à SRO pour prédire $S_{0:4}$ à SRO. Pour les approches multi-environnement différentes structures de variance-covariance ont été utilisées pour modéliser le GxE, MM représentant un modèle multi-environnement sans effet GxE, MDs avec une variance unique pour les effets GxE et MDe avec une variance spécifique pour chaque environnement. Plus de détails sont donnés dans (Granato et al. 2018).

Les précisions de prédiction (PA, *predictive ability*) ont été mesurées avec la corrélation entre les phénotypes ajustés par essai à SRO et les prédictions génomiques pour SRO.

Simulation

Toutes les simulations ont été réalisées avec le package R AlphaSimR (Gaynor, Gorjanc, et Hickey 2021). Les mêmes dix populations initiales ont été utilisés comme points de départ pour les deux scénarios. Pour chaque population, 80 fondateurs ont été générés avec des déséquilibres de liaison et des fréquences alléliques représentant une taille efficace de population de 50. Ils ont été utilisés comme base pour cinq cycles de recombinaisons et de échantillonnages aléatoire. À la fin des cinq cycles, 400 lignées ont été échantillonnées et leur génotype en S_0 ainsi que leurs phénotypes en $S_{0:2}$ et $S_{0:3}$ ont été utilisés comme base pour la première calibration des modèles de prédiction génomique.

Quatre caractères ont été simulés avec différentes moyennes et décomposition de la variance génétique pour ressembler au nombre de jours jusqu'à la floraison, à la hauteur de plante, au rendement et à la concentration en zinc du grain. Les scénarios ont été choisis pour représenter deux

intégrations de la prédiction génomique directement applicables dans le schéma actuel de sélection récurrente (Figure 4-1, p 121). Dans les deux cas, la prédiction se fait sur les génotypes de candidats à la sélection échantillonnés aléatoirement dans la population en génération S_0. Un modèle a été calibré en amont à l'aide des 400 familles également génotypées en S_0 et phénotypées en $S_{0:2}$ et $S_{0:3}$ pour le scénario BS1 ou seulement en génération $S_{0:2}$ pour le scénario BS2. A chaque cycle, les 200 familles avec les meilleures GEBV sont phénotypées aux générations $S_{0:2}$ ou $S_{0:2}$ et $S_{0:3}$ selon le scénario pour mettre le modèle de prédiction à jour. De plus les 50 meilleures familles sont également croisées entre elles pour démarrer un nouveau cycle.

Résultats et Discussion

Prédiction intra-population

Des PA moyennes relativement similaires ont été obtenues pour tous les caractères avec l'approche SIN_{SRO} à travers les tailles de set s et les méthodes de calibration (PA = 0.30, 0.33, 0.27 et 0.24 pour FL, PH, YLD and ZN, respectivement). La taille de set *s* a eu un effet très fort sur les PA, les grands sets permettant systématiquement de meilleures PA, même pour les caractères à l'architecture génétique plus simple tel que FL ou PH (Figure 2-5, p.55).

Aucune différence flagrante n'a été notée entre les deux années qui représentent des prédictions basées sur des phénotypes obtenus dans deux générations différentes.

En comparant les approches intégrant les deux sites (PAL et SRO), BAL1 et BAL2 à SIN_{SRO}, on voit que l'utilité des données des différents sites dans la prédiction de SRO dépends fortement de la corrélation entre sites pour le caractère prédit. Pour FL, PH et ZN, un fort gain de PA peut être observé en comparaison à SIN_{SRO} quand les phénotypes de 334 familles à PAL (IMB) (Figure 2-6, p.57). Ce gain en PA diminue progressivement quand la taille des sets augmente et ne dépasse la corrélation phénotypique entre site qu'avec les plus grands sets de calibration.

Pour YLD, la combinaison de données PAL et SRO dans la calibration ne permet pas de gain de PA par rapport à l'approche mono-site. La prédiction génomique permet néanmoins d'obtenir des prédictions avec des précisions supérieures à la corrélation phénotypiques entre les sites même avec la plus petite taille de set de calibration.

Prédiction inter-population

D'une manière générale, la précision de prédiction a été meilleure avec des modèles mono-site comparé au modèle multi-site (Table 0-1). Uni2 est pour tous les caractères à l'exception de YLD l'approche avec la meilleure PA. Etonnamment, Uni2 et Uni3, calibrés sur des données PCT27A en $S_{0:2}$ et $S_{0:3}$ respectivement, prédisent $S_{0:4}$ mieux que Uni1 calibré sur des données $S_{0:4}$. Plusieurs causes sont envisagées. La première est que des allèles ont été perdu durant le l'avancement en génération suite à une sélection naturelle. YLD étant hautement polygénique, la perte de certains allèles a eu moins

d'effet sur le phénotype finalement que pour les autres caractères à l'architecture génétique plus simple. Des effets aléatoires pourraient également être à l'origine de la bonne PA de Uni2. Une autre explication possible est que l'avancement en génération a causé des erreurs telles que des mélanges de semences ou la récolte de plantes mâle stérile. La génétique des plantes $S_{0:4}$ ne correspondant pas complètement à celle attendue pour des descendants des génotypes S_0 utilisés pour la prédiction. Dans tous les cas, il est très intéressant de savoir que $S_{0:2}$ est un bon prédicteur pour $S_{0:4}$, ceci même entre populations.

Le modèle Multi2 a été moins performant que les modèles mono-sites malgré un grand nombre de données supplémentaires (Table 0-1, Figure 3-2, p.89). On sait que la performance des modèles multi-environnement est liée aux corrélations entre les environnements intégrés dans la calibration (Lopez-Cruz et al. 2015; Cuevas et al. 2016) or les corrélations phénotypiques observées entre les différents environnements de calibration et l'environnement prédit étaient relativement basses.

Les différentes structures de variance-covariance testées ont montré des résultats variables selon les caractères. Dans aucun des cas, MDe ne permet de meilleures PA, MM étant la meilleure approche pour FL et YLD, et MDs la meilleure approche pour PH et ZN. Bien que les PA varient entre les structures de variance-covariance, les individus sélectionnés sont sensiblement les mêmes d'une structure à l'autre, le rang n'étant pas très différent entre les modèles (Figure 3-5, p.95).

Simulation

Les deux schémas de sélection BS1 et BS2 ont été comparés sur leurs gains génétiques par cycle de sélection récurrent et en fin de sélection généalogiques ainsi que sur leur PA. Pour tous les caractères, les gains en sélection récurrente et en sélection généalogique ont été plus élevés pour BS1. Cela s'explique facilement par le plus grand nombre de données et donc une meilleure estimation de la valeur en lignée pour BS1. Il est important de noter que lorsque l'héritabilité était faible, comme pour le cas de T2 (Table 4-2, p.125), les différences de gain entre schéma étaient plus faibles (Table 4-3, p.127).

Table 0-1 : Résumé des différentes précisions de prédiction obtenues pour les modèles mono-site (MS) Uni1, Uni2 et Uni3 et le modèle multi-site Multi2 75%. Le modèle multi-site a été testé avec trois différentes structurations de l'interaction génotype environnement, soit sans interaction (MM), soit avec un effet aléatoire d'interaction avec un variance unique (MDs) soit avec un effet aléatoire d'interaction avec une variance par site (MDe). La description des schémas de validation se trouve dans la *Figure 3-2, p.89*

Scenario	Model	FL	PH	YLD	ZN
Uni1	MS	0.225 ± 0.077	0.309 ± 0.069	0.388 ± 0.079	0.174 ± 0.080
Uni2	MS	0.311 ± 0.005	0.389 ± 0.005	0.333 ± 0.005	0.323 ± 0.006
Uni3	MS	0.229 ± 0.004	0.254 ± 0.006	0.243 ± 0.008	0.293 ± 0.006
Multi2	MM	0.200 ± 0.012	0.296 ± 0.008	0.274 ± 0.014	0.250 ± 0.016
Multi2	MDs	0.223 ± 0.015	0.264 ± 0.013	0.295 ± 0.023	0.241 ± 0.021
Multi2	MDe	0.204 ± 0.023	0.255 ± 0.017	0.285 ± 0.027	0.238 ± 0.027

Les niveaux de GxE ont également eu un impact négatif significatif sur le gain génétique sur les trois caractères avec l'héritabilité la plus faible T4 n'étant pas affecté. D'un autre coté les PA de tous les caractères ont été négativement influencées par l'augmentation du GxE.

L'intensité de sélection appliquée durant les 20 cycles simulés n'a pas épuisé la variabilité de la population bien qu'elle ait perdu entre 2 et 3 pourcents de sa variabilité initiale à chaque cycle. Après 20 cycles un peu plus de 50% des QTLs n'étaient pas encore fixés.

Conclusion

Finalement, la prédiction génomique devrait permettre une réduction de la durée d'un cycle de sélection, ainsi que de l'effort de phénotypage. Le programme CIAT-Cirad continuera sur la voie de la mise en œuvre de la prédiction génomique dans son schéma d'amélioration de population. L'approche de phénotypage définitive doit encore être clarifiée. Le programme pourrait, dans un avenir proche, concentrer le phénotypage sur les $S_{0:2}$ uniquement mais nous savons maintenant que le site SRO restera central dans le processus de sélection et de calibration. Les différences entre caractères suggèrent que des compromis devront être faits pour accommoder les besoins de phénotypages de chacun. Les caractères montrant une bonne corrélation entre les sites, comme FL et PH, pourront probablement être phénotypés dans PAL, tandis que YLD et ZN devront toujours être phénotypés à SRO, du moins pour une partie de la population. Une bonne stratégie devra encore être mise en place pour réduire au minimum les effets confondus dans le dispositif expérimental tels que l'effet essai et l'effet génération ainsi que pour tracer clairement les interactions génotypes par année.

Résumé

L'amélioration des populations par sélection récurrente a récemment regagné l'attention de la communauté des sélectionneurs grâce à la possibilité de réduire la durée du cycle de sélection en utilisant la prédiction génomique (PG) plutôt que des tests sur descendance. Pendant plusieurs décennies, le programme de sélection récurrente du riz pluvial (*Oryza sativa* L.) CIAT-Cirad a utilisé un schéma de sélection en deux parties (i) l'amélioration de la population basée sur la sélection récurrente, (ii) le développement de cultivars par la sélection généalogique. Récemment, des efforts ont été faits pour mettre en œuvre la PG afin de raccourcir considérablement le cycle de sélection.

L'objectif de cette thèse était d'évaluer le potentiel de la PG pour la sélection récurrente, de tester des stratégies de calibration des modèles en utilisant les dispositifs expérimentaux existants pour ensuite mettre en place une sélection génomique précoce des parents recombinants.

Une population de sélection récurrente a été génotypée à la génération S_0 avant d'être divisée en deux sous-populations : la PCT27A et la PCT27B. La PCT27A a été avancée en génération $S_{0:2}$ et $S_{0:3}$ et phénotypée à ces générations tandis que le PCT27B a été avancé jusqu'à la $S_{0:4}$ et phénotypé. Quatre caractères ont été mesurés dans un site cible et un site de substitution. La précision des modèles de PG a été estimée à l'aide de plusieurs scénarios et modèles, en fonction de la présence d'un ou deux environnements de culture, d'une ou plusieurs générations de phénotypage, de la présence d'une interaction génétique x environnement (GxE), de la taille et composition de la population de calibration. Tout d'abord, nous avons évalué par validation croisée des modèles intra-population. Ensuite, nous avons testé différents modèles pour prédire du matériel avancé dans une population de validation externe. Pour compléter ces résultats, une étude de simulation a été réalisée pour évaluer l'effet à long terme de l'intégration de la PG dans le programme de sélection. Dans cette dernière étude, les effets de trois niveaux de GxE et de deux niveaux de dominance ont été évalués dans deux schémas de sélection basés soit sur une génération de phénotypage, soit sur deux générations.

Les calibrations sur deux sites n'ont pas été plus performantes que celles réalisées uniquement sur le site cible. Les données PCT27A du site cible en $S_{0:2}$ et $S_{0:3}$ ont pu être utilisées pour prédire les références $S_{0:4}$ en PCT27B. Lorsque les calibrations confondaient l'effet de la génération et du site, les précisions étaient plus faibles. Il ne semble donc pas approprier de calibrer sur des données de deux générations obtenues dans deux sites. Les simulations ont permis de mettre en évidence que la prédiction *forward*, utiliser pour la sélection récurrente, était possible avec les deux schémas de sélection, la calibration sur deux générations étant systématiquement meilleure que celle sur une seule génération. Avec l'augmentation des interactions GxE, les précisions ont chuté pour les deux schémas de façon similaire. Seul le niveau de GxE a eu un impact sur la capacité de prédiction et le gain génétique.

En conclusion, il est possible de dire que la PG peut remplacer les tests sur descendance dans la sélection récurrente. L'utilité des différents sites et des différentes générations dans la calibration dépend des caractères prédits et des compromis devront être faits lors de la conception du schéma de sélection pour obtenir des prédictions aussi précises que possibles pour les différents caractères dans les limites financières et logistiques imposées programme. Les résultats de l'étude de simulation et des expériences sur le terrain vont permettre l'amélioration du programme de sélection CIAT-Cirad en vue d'un progrès génétique plus rapide tout en conservant un coût économique relativement stable.

Abstract

Population breeding through recurrent selection has recently regained attention in the plant breeding community with the new possibility to reduce the breeding cycle length by selecting on genomic prediction (GP) rather than progeny testing. For several decades, the CIAT-Cirad rainfed rice (*Oryza sativa* L.) breeding program has been using a two-parts breeding scheme with population improvement based on recurrent selection and a cultivar development following pedigree breeding. More recently, effort have been made to implement GP to shorten the breeding cycle.

The objective of this thesis was to evaluate the potential of GP in CIAT-Cirad program, test calibration strategies using the existing infrastructure to later implement early genomic selection of recombinant parents.

A population was genotyped at generation S_0 before being divided in two subpopulations: the PCT27A and the PCT27B. PCT27A was advance to generation $S_{0:2}$ and $S_{0:3}$ and phenotyped at those generations while PCT27B was advanced up to $S_{0:4}$ and phenotyped. Four traits were measured in a target site and a surrogate evaluation site. The predictive ability of the GP models was estimated using several scenarios and models, according to the presence of one or two growing environments, one or several phenotyping generations, the presence of genetic by environment interaction and the size and composition of the training set.

First, we assessed by cross-validation the GP in a single population. Then, we used the same data to predict more advanced material in an external validation population. To complete the field experiments, a simulation study was realized to assess the long-term effect of the integration of GP into the breeding program and the response of calibration scenario to dominance and genotype-environment interaction (GxE) variance.

In this last study, the effect of three levels of GxE and two levels of dominance were assessed on two breeding schemes based on either two generations of phenotyping or a single generation.

Two-sites calibrations never strongly outperformed single site calibrations. Early generation PCT27A phenotypes from target site could be used to predict generation $S_{0:4}$ from PCT27B. However, when the calibration confounded generation and site effects, the precisions were lower. Hence it seems so far unappropriated to phenotype different generations in different sites. The simulations allowed to highlight that forward prediction, which is the base of recurrent selection, is possible with either breeding schemes, calibration with two generations being systematically better than single generation one. With the increase of GxE, the accuracies dropped for both schemes with similar intensity. Only the level of GxE had an impact on predictive ability and genetic gain.

To conclude, GP can replace progeny testing in recurrent selection. The utility of the different sites and different generations in the calibration depend on the traits predicted and compromises will have to be done when the breeding scheme will be design to reach the best possible accuracy for each trait, while staying within the program financial and logistical limitations. Those results from the simulation and field experiments will be valuable for the improvement of the CIAT-Cirad breeding program toward a faster genetic progress and more sound use of resources.