Rice Research: Open Access

Research Article Open Access

Breeding for Nitrogen use Efficiency: Lessons from Genomic Prediction Experiments Based on a Broad-based Population of Upland Rice

Joël rakotomalala[1]*, kirsten vom brocke[1,3,5], julien frouin[4,5], david pot[4,5], ravo rabekijana[1,6], alain ramanantsoanirina[1], isabelle ramonta ratsimiala[2], tuong-vi cao[4,5]*

[1]National Center for Applied Research on Rural Development- FOFIFA, Regional Research Station Antsirabe, Antsirabe 110, Madagascar
[2]Faculty of Sciences, University of Antananarivo, Antananarivo 101, Madagascar
[3]CIRAD, UMR AGAP, 101 Antananarivo, Madagascar
[4]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France
[5]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, InstitutAgro, F-34398 Montpellier, France
[6]Highland Production Systems and Sustainability in Madagascar Platform in Partnership-DP SPAD, Antsirabe 110, Madagascar

Abstract

In the context of subsistence farming in Madagascar, upland rice producers have limited access to mineral fertilizers and yields remain very low. Genetic improvement of yield through the nitrogen use efficiency (NUE) component is an avenue explored in breeding programs. A recent GWAS study carried out on an upland rice diversity panel allowed to detect genomic regions involved in NUE variability, nevertheless these regions explained only a moderate part of the total genetic diversity observed in the panel. We investigated the potential of genomic prediction for NUE in order to optimize the FOFIFA-CIRAD upland rice breeding program for this trait. We evaluated the predictive ability of genomic prediction using two validation experiments. The first consisted of a standard cross-validation with 5-fold subdivision of the diversity panel (DP) and the second consisted of an independent experiment involving a breeding population (BP) derived from the DP. The DP was structured into five genetic clusters of different sizes and with some degree of admixture, while the BP was composed of two main clusters. The best predictive ability for NUE was obtained in cross-validation within the DP. The predictive ability in the independent validation experiment was weak (r = 0.25), about three times less than those obtained in the cross-validation. The low kinship between DP and BP, different genetic structures and slightly different LD patterns probably explains the low predictive ability observed across population prediction. Practical implications for the Malagasy upland rice breeding program are discussed.

Keywords: Genomic Prediction; Nitrogen use Efficiency; Upland Rice

Introduction

Rice (*Oryza sativa L.*) is one of the most important food crops worldwide [1]. Its improvement for higher yield in sustainable agriculture systems is also vital to meet energy and nutritional needs of the growing world population [2]. In Madagascar, upland rice cultivation is developing in the high plateau area where soil fertility is generally low [3]. The use of mineral fertilizer is uniformly low because of its high cost in a subsistence farming context [4]. As a consequence, animal manure is often the only source of fertilizer [5]. There is an urgent need to find alternatives to increase yields without massive use of mineral fertilizers. One option consists in improving the nitrogen use efficiency (NUE) of the rice crop [6].

NUE plays a central role in low-input systems [7]. NUE is usually defined as grain yield per unit of N supplied by soil and by fertilizer [8]. NUE and its components are complex quantitative traits and the study of their relationship with other simpler morphological or physiological traits could help understand the mechanisms involved in NUE and identify ways to improve selection efficiency. Genetic variability of both N uptake efficiency and N utilization efficiency has been demonstrated in a large number of crops. A genome-wide association study (GWAS) for 16 NUE-related agronomic traits and yield components using a diversity panel (DP) of 190 mainly japonica varieties and a set of 38,390 SNPs was conducted [10]. Few association signals were identified for NUE corresponding to two haplotype groups and one isolated SNP on chromosomes 6, 7 and 11, explaining 9.5%, 9.6% and 10.4% of phenotypic variation, respectively.

In such a context, genomic selection (GS) should be considered for breeding rice varieties with improved NUE because it encompasses all marker information and can therefore better capture variation that arises from low effect QTLs [11,12] . GS was defind as a combination of genetic markers covering the entire genome with different statistical methods to maximize the efficiency of selection [13]. The principle of GS is (i) to combine molecular markers and phenotypes of individuals within a calibration population where the effects of all markers are estimated simultaneously and then (ii) to predict the genomic estimated breeding value (GEBV) for individuals of the breeding population that are only genotyped [14]. The GEBV is then used to select the best candidates to generate new elite material. Several methods are available to build prediction models [15-17]. They differ in the assumptions about marker effects and the variance of these effects on the observed trait variation. The general conclusion of several empirical studies is that there are no perfect statistical methods [18]. Beyond the statistical methods, the predictive ability of genomic predictions also depends on the characteristics of the target populations (markers density and their distribution in the genome, size and structure of the calibration population, relatedness between calibration and validation populations, respective populations' LD, minor allele frequency) and traits (heritability and correlations between traits) [18-20].

***Corresponding author:** Tuong-Vi CAO, Genetic Improvement and Adaptation of Mediterranean and Tropical Plants Research Unit (UMR AGAP), CIRAD, Avenue Agropolis, Montpellier 34398, France; Tel: +229 52 29 58 24; Email: tuong-vi.cao@cirad.fr

Received July 23, 2021; **Accepted** September 02, 2021; **Published** September 09, 2021

Citation: Rakotomalala J, vom Brocke K, Frouin J, Pot D, Rabekijana R, et al. (2021) Breeding for Nitrogen use Efficiency: Lessons from Genomic Prediction Experiments Based on a Broad-based Population of Upland Rice. J Rice Res 9: 264.

Copyright: © 2021 Rakotomalala J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

GS is expected to accelerate genetic gain for traits such as yield potential, adaptation to climate change constraints and resource use efficiency including nitrogen [21]. GS studies in rice have mainly explored the cross-validation approach within diversity panels for different target traits in more or less advanced generations or lines (Table 1). However, the across-population genomic predictions were explored in rice. The prediction equation was calibrate on diversity panel to predict 97 advanced lines from biparental crosses between 31 founders sampled among 284 DP accessions [20].The breeding population was advanced following a pedigree breeding process until F_5 to F_7. Reasonably high predictive abilities were obtained, i.e. 0.54 for panicle weight and 0.33 for nitrogen index. Then, with calibration population consisted of 228 accessions and the validation population consisted of 95 advanced lines (F5-F7), the predictive ability of GEBV between populations was quite high ranging from 0.43 (FL-As) to 0.48 (CG-As). Such results were obtained because the relationship between the two populations were sufficiently strong [23].

Here, we report an empirical study that evaluates the predictive ability of GS to predict NUE, using an upland rice diversity panel to calibrate the model, and advanced lines from a synthetic multi-parent population to validate the model. We explored the potential of GS using both cross-validation and across population genomic prediction experiments. We sampled 184 accessions from the working collection of the upland rice breeding program, hereafter called diversity panel (DP) [10]. used by and 198 advanced breeding lines, from a synthetic population, hereafter referred to as the breeding population (BP). We (1) evaluated the predictive ability of genomic prediction for yield and NUE through cross-validation within the DP, (2) analyzed the phenotypic characteristics (genetic variance, heritability, correlations between traits) of the BP, in comparison with the DP and (3) assessed the ability of the calibration developped on the DP to predict the breeding value in the BP. We finally discussed our results in the lights of some well-known breeding practices.

Materials and Methods

Plant material

The plant material is composed of two populations of diverse geographic origins (tropical and temperate japonica subspecies).

The first population corresponds to a subsample of the diversity panel (DP) [10].it includes 184 accessions and represents the working collection of the upland rice breeding program of the National Centre of Applied Research and Rural Development (FOFIFA) and French Agricultural Research Centre for International Development (CIRAD) .Eighty accessions of the DP are of Malagasy origin and 104 are of diverse geographical origins, including Brazil, Colombia, Ivory Coast, China and other Asian countries. DP accessions constitute the basis of the bi-parental crosses that form the starting point of the pedigree breeding scheme within the breeding program. The DP is representative of the global diversity of tropical japonica subspecies [10].

The second population, corresponds to advanced lines from a synthetic multi-parent population and is defined as the Breeding Population (BP)composed of 198 $S_{4:6}$ lines representing a sample of a dvanced lines derived from a synthetic population. The population was developed by crossing a sample of an existing synthetic population (PCT 11) segregating for a recessive genic male sterility (ms) gene [18]. and twelve founders from the DP used as males. First, the 12 founders were crossed to male sterile plants of PCT 11 in isolated plots. Fully filled seed were harvested from male sterile plants in order to obtain F1 seed being 100% heterozygous at the ms locus.

At the F_2 generation, equal numbers of seeds from each cross were bulked together. Bulked seed were sown to allow random mating through open pollination. After two cycles of random mating, four generations of single-seed progeny (SSD) were run to obtain several hundred $S_{3:4}$ plants, that constituted the starting point to (1) produce $S_{4:5}$ and $S_{4:6}$ progenies for field phenotyping and (2) sample a single $S_{5:6}$ seed for each $S_{3:4}$ line for the genotyping.

Table 1: Research in Genomic Selection of rice.

Plant material	Phenotypic traits	Number of markers	Applied statistical methods	Precision of GEBV	Main conclusions	References
278 in F_2	GY, TL, TGW and GP	250 000	GBLUP, LASSO and SSVS	GY: 0.09-0.13 TL: 0.21-0.23 TGW: 0.67-0.69 GP: 0.35-0.37	The GBLUP tends to perform better if the trait is largely polygenic, compared to the other two genomic prediction models	[22]
354 in $S_{3:4}$	DFI, PH, PW and GY	8,336	RR-BLUP, GBLUP, BRR, Lasso and BL	DFI: 0.18 - 0.40 PH: 0.46 - 0.62 PW:0.20-0.45 GY: 0.16 – 0.46	The predictive ability of GEBV is affected by the correlation between the calibration and validation population, trait inheritance, prediction models, LD, MAF, and the composition of the calibration population	[18]
284 in F_5-F_7	DFI, PW and NI	43,686	GBLUP, RKHS and BayesB	DFI: 0.51-0.65 PW: 0.54-0.62 NI: 0.52-0.57	The predictive ability of the prediction varies considerably depending on the composition of the calibration population, the trait, LD and the frequency of minor alleles (MAF)	[20]
228 in F_5-F_7	FL-As, CG-As/ FL-As and CG-As	22,370	GBLUP, BayesA and RKHS	FL-As:0.45 CG-As/FL-As:0.33 CG-As:0.53	Excluding the most redundant SNP markers based on LD information had a limited effect on predictive ability. The level of predictive ability was similar to levels reported in the literature.	[23]
1495 in F_1	GY, GN, TGW, PH, GL and GW	102 795	GBLUP	GY: 0.54 GN: 0.62 TGW: 0.54 PH: 0.58 GL: 0.92 GW: 0.87	The accuracy is better when the two populations are linked. The accuracy of the GEBV is of the same order of magnitude as the heritability of traits	[24]

Phenotypic traits: CG-As, dehulled grain; DFI, Flowering date; FL-*As,* Flag Leaf; GL, Grain Length; GN,Grain Number; GP; Grains per Panicle; GW, Grain Width; GY, Grain Yield; NI, Nitrogen Index; PB, Primary Branches per panicle; PH, Plant Height; PL, Panicle Length; PN, Panicles per Plant; PW, Panicle Weight; SB, Secondary Branch;SP,Seeds per Panicle; TGW, Thousand Grain Weight; TL,Tillering; ZN, grain zinc concentration.

Experimental site and phenotyping

Field phenotyping was conducted at Ivory in the Mid-West region of Madagascar located at 19° 33'27"S, 46° 24'43"E at 960 m altitude, in red ferralitic soils, especially deficient in nitrogen (N). The DP was phenotyped during two successive rainy seasons, 2014-2015 and 2015-2016, for 16 NUE-related agronomic traits and yield components as described in [10]. The BP was phenotyped for the same traits in 2016-2017 and 2017-2018.

The field experiments consisted in four augmented alpha-lattice designs with two complete replications each. Replications were subdivided into a number of blocks depending on the population size and the dimensions of the available field. The two checks varieties (Nerica 4 and *WAB 880*-1-32-1-1-P2-HB-1) were replicated in each block. In the DP phenotyping experiments, two nitrogen treatments were applied within each block, half of which was under high nitrogen (HN) treatment (120 kg/hm^2 N) and half under low nitrogen (LN) treatment (0 kg/hm^2 N). In year-1, the size of each subplot was 1.8 m × 2.4 m (108 hills or holes made for sowing seeds) for the first replication and 1.4 m × 2.0 m (70 hills) for the second replication. In year-2, the size of each subplot was 1.8 m × 1.6 m (72 hills) in the first replication and1.2 m × 1.6 m (48 hills) in the second replication [10]. In the BP phenotyping experiments, only the LN treatment (0 kg/hm^2 N) was applied. In year-1, the size of each subplot was 1.6 m x 2.0 m (80 hills) for the two repetitions. In year-2, the subplot size was 1.6 m x 2.0 m (80 hills) for the first replication and 1.4 m x 2.0 m (70 hills) for the second replication. Field preparation started with ox plowing followed by hand surfacing of the soil. In the context of subsistence farming, we considered only the LN condition in the analysis because it represents the farmers' current fertilization practices for upland rice. Four to six rice seeds were sown per hill with 20 cm × 20 cm spacing between hills. In the DP, right before sowing, the same base dressing of cattle manure (5 000 kg/hm^2), triple superphosphate (69 kg/hm^2 P_2O_5), potassium sulfate (62.4 kg/hm^2 K_2O) and dolomite (500 kg/hm^2) was applied to each hill in all the plots [10]. In the BP, right before sowing, the same base dressing was applied except for triple superphosphate (150 kg/hm^2 P_2O_5) and potassium sulfate (130 kg/hm^2 K_2O).

The target traits for both DP and BP were time (in days) from sowing to flowering (DFl), thousand grain weight (TGW), grain yield (GY), grain nitrogen content (GNC) and nitrogen use efficiency (NUE). DFl was recorded as the number of days after sowing, when 50% of the plants in the plot has flowered. TGW was calculated from the weight of 200 harvested grains after 72 hours of oven drying at 60°C. GY was recorded by weighing dried grains in grams of each plot and extrapolating in kg/ha. GNC was measured in grinded grains, using near infrared spectrometry (NIRS) and a calibration equation developed by [10]. NUE was estimated as GY/nitrogen supply [8].

Analysis of phenotypic data

For each population, the experiments from the two different years were analyzed together. The whole phenotypic data processing was based on the PROC MIXED procedure (SAS 9.4; SAS Institute Inc, Cary), including a model-based diagnostic analysis to identify and, if necessary, eliminate influential individuals (plots), as described by and [18,20,24].

The same mixed linear model was run on the cleaned data from the DP and the BP. The model was:

$$Y_{cijkl} = \mu + A_l + R_{jl} + T_c + (arb)_{jkl} + g(T)_{ci} + gy(T)_{cil} + e_{cijkl}$$

where:

Y_{cijkl}	observed phenotypic value of genotype i in block k, nested in repetition j of year l,
μ	experiment mean
A_l	effect of year l
R_{jl}	effect of repetition j nested in year l
T_c	effect of check c
$(arb)_{jkl}$	effect of block k nested in repetition j of year l, with $(arb)_{jkl} \sim (0, \sigma^2_{bl})$
$g(T)_{ci}$	effect of genotype i nested in check c, with $g(T)_{ci} \sim (0, \sigma^2_g)$. For plots receiving either check (Nerica 4 or WAB), $g(T)_{ci} = 0$, or $g(T)_{ci} \neq 0$ otherwise
$gy(T)_{cil}$	interaction between genotype i and year l nested in check c, with $gy(T)_{cil} \sim (0, \sigma^2_{gy})$. As for $g(T)_{ci}$, for plots receiving either check, $gy(T)_{cil} = 0$, $orgy(T)_{cil} \neq 0$ otherwise.
e_{cijkl}	residual, with $e_{cijkl} \sim (0, \sigma^2_e)$

The variance components allowed to estimate the broad sense heritability (H^2) [25]:

with n_y, being the harmonic number of years and n_p,being the harmonic number of plots across years, respectively.

Genotyping and genotypic data

Genotypic information for the DP was retrieved [10]. The BP population was genotyped specifically to fulfill the objective of the present study. Seeds of each $S_{4:6}$ progeny originating from each S_4 line of the BP were grown at the Cirad greenhouse (Montpellier, France). DNA was extracted from 30 mg of young leaves of a single $S_{4:6}$ plant for each $S_{3:4}$ line three weeks after seedling [26]. Two hundred ng of the extracted DNA was used with Apek I enzyme to obtain a DNA fragment library to perform genotyping by sequencing (GBS) .The GBS methodology consists in digesting genomic DNAs with a restriction enzyme, ligating adapters comprising barcodes, pooling all these fragments, amplifying by PCR this pool, and estimating the quality of bank before sequencing it [27]. The complete sequence of Nipponbare rice (Os-Nipponbare-Reference-IRGSP-1.0) was used as a reference for alignment with Bowtie2 using default parameters.

The SNP calling was done with Tassel GBS pipeline v5.2.37 without any filtering. The molecular information was stored and managed through a web-based tool, Gigwa-Genotype investigator for genome-wide analyses [28,29]. A first filtering was applied to the datasets from the two populations simultaneously to obtain SNPs, with minor allele frequency (MAF) higher than 1% and maximum missing data below 20%. Missing data were imputed separately per population using BEAGLE 5.0 R40B [30]. then filtered for heterozygosity (Ho ≤ 30%), minor allele frequency (MAF ≥ 1%) and imputation rate (≤ 20%). A total of common 87,089 SNPs to the two populations was finally obtained [30].

Genetic characterization of the populations

Populations were characterized individually using 87,089 SNPs: First, minor allele frequency (MAF), observed heterozygosity (Ho) and the kinship between/within population were calculated with Tassel software version 5.0. Then, diversity tree was drawn on Darwin version 6.0.14 [31]. to provide a visualization of the structure among

and within populations. Next, we used Fst to determine the degree of differentiation between DP and BP. Fst calculation was done with hierfstat [32]. A filter on MAF >5% on the 87,089 SNPs was performed and allowed the selection of 46,055 SNPs. Then, the removal of strictly redundant SNPs allowed to refine the SNP set to 29,847 SNPs to reduce the size of the genotypic matrix and the computation time without information loss. Afterwards, the Landscape and Ecological Association Studies (LEA) package was used to analyze the structure within and among populations [33]. We arbitrary used the probability threshold of 0.6 to assign individuals to the different genetic groups. An individual was assigned to a cluster if the associated probability is greater than 0.6, otherwise it was classified as admixed and placed in cluster 9. Finally, linkage disequilibrium (LD) was calculated with LDcorSV package in the R environment [34,35]. LD was calculated using the 29,847 SNPs dataset.

Genomic prediction methods

Two incidence matrices, of size 46,055 SNPs and 29,847 SNPs, were used for genomic predictions. Two kernel-based methods were used to predict GEBV as they are computationally more efficient than their initial formulation counterparts in the context of "small-n-large-p" problem [16]. One was the Genomic Best Linear Unbiased Predictor method (GBLUP) that utilizes the well-known genomic relationships matrix G, as kernel. The GBLUP method was implemented using the Expectation–Maximization convergence algorithm and the genomic matrix G = XX′, X being the centered genotype matrix containing values of −1, 0 and +1, with N × P dimension, where N is the number of entries and P the number of markers [20,36,37]. The second method was the Reproducing Kernel Hilbert Space method (RKHS), uses a Gaussian kernel to connect the genotypes and phenotypes and is defined as $K = k(x_i,x_j) = \exp(-||x_i - x_j||^2/h)$, where x_i and x_j are the genotypic vectors of individuals i and j respectively, h is a bandwidth parameter which controls how fast the relationship between two genotypes decays as the distance between the corresponding pairs of marker vectors increases [38]. The choice of the bandwidth parameter can be optimized by applying a cross-validation or a Bayesian approach treating h as a random variable. K matrix is able to approximate any arbitrary function including the multiplicative linear function used to model epistatic inheritance [16,39]. The genomic prediction model can be described by the following equation:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zg} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y}$ is an $n\times1$ vector of observation, $\mathbf{X}$ is the incidence matrix for the fixed effects, $\boldsymbol{\beta}$ is the vector of non-genetic fixed effects, the part Xβ includes only the global mean μ and when the structure is taken into account in the model one adds μ plus the structure matrix, $\mathbf{Z}$ is the genotypic matrix of size N (individuals) × P (markers), $\mathbf{g}$ is the vector of random regression coefficients of accessions and $\boldsymbol{\varepsilon}$ is the vector of residuals.

Assessment of genomic prediction ability

Cross-validation experiments: Firstly, cross-validation experiments were performed with 5-fold subdivision of the DP, with 147 accessions in the calibration set and 37 accessions in the validation set. Each cross-validation experiment was replicated 100 times using 100 independent partitions of the accessions into the training set and validation set. For each independent partition, the correlation between the predicted and the observed phenotype was calculated, so as to obtain 100 correlations for each cross-validation experiment. The predictive ability of each cross-validation experiment was computed as the mean value of the 100 correlations. Once a model was calibrated, it was used to predict GEBV of the validating set. The same partition was used for the 5 phenotypic traits considered, the 2 prediction models and the 2 SNP datasets.

Secondly, the predictive ability was specifically estimated for the 12 founders of the BP, using the remaining 172 accessions as the calibration set.

Finally, the cross-validation experiment was also performed within the BP and its predictive ability estimated and compared with the one of DP.

Prediction across-populations: In this experiment, the prediction models were trained with the phenotypic and genotypic data of the 184 DP accessions and then served to predict GEBV for the BP accessions; the correlation between the predicted and the observed phenotypes of the 198 BP accessions was calculated.

Analysis of factors affecting the predictive ability of GEBVs

We were interested in identifying factors and combinations of factors that influence significantly the variation of predictive ability. In the case of the cross-validation experiment, the analysis distinguished two categories of effects: (a) effects due to the controlled factors and (b) effects due to the unpredictable part of the process. The analysis is based on the following ANOVA models:

First, the model (1) decomposed the overall variation into two components: between-scenario ($\boldsymbol{MS_{ms}}$) and within-scenario (e_{mse}). The scenario was the only explanatory factor and the model was run using conventional single factor ANOVA. Four scenarios were defined for each trait (2 prediction methods × 2 SNP datasets). If scenarios differed significantly, then the model (2) decomposed further the overall variation, still using ANOVA. The model (2) decomposed the variation into all possible effects. Because the two-way interaction could represent significant sources of variation, the F tests for main effects and associated p-values could be inflated. To prevent this inflation, the main effects were also tested using the interaction mean square as the error term when the interaction effect was significant.

RESULTS

Phenotypic characteristics of the two populations

The variance components and associated statistics are shown for the five traits in the DP and BP in Table 2. There is a highly significant

	(a)		(b)	
$\boldsymbol{PA_{mse}}=$	$\boldsymbol{\mu + MS_{ms}}$	+	e_{mse}	Model (1)
$\boldsymbol{PA_{mse}}=$	$\boldsymbol{\mu + M_s + S_m + MS'_{ms}}$	+	e_{mse}	Model (2)

where:

PA_{mse}	individual value of predictive ability,
μ	overall mean
MS_{ms}	scenario effect which combines all effects
M_s	effect of prediction method, $s \in$ {GBLUP, RKHS}
S_m	effect of SNP number m, $m \in$ {29 847, 46 055}
MS'_{ms}	interaction between prediction method and SNP number
e_{mse}	residual, $e_{mse} \sim (0, \sigma_e)$, $e \in$ {1, 2, 3, …, 100}

Table 2: Variance components and broad sense heritability for five traits in DP and BP populations. The phenotyping was conducted for two years (2014-2015 and 2015-2016 for DP, 2016-2017 and 2017-2018 for BP).

Population	Component	DFI	TGW	GY	GNC	NUE
Diversity panel (DP)	Year[1]	381.55***	7.60**	39.41***	31.35***	0.47NS
	Repetition(Year)[1]	2.17NS	1.16NS	4.07**	18.82***	3.51*
	Check[1]	0.08 NS	0.26 NS	3.42*	0.10NS	3.28*
	Genotype[2]	27.2382***	15.8094***	347768***	0.0099***	28.8524***
	Block(year*Rep)[2]	0.5707*	0.08819NS	146954***	0.0063***	12.4641***
	Genotype*Year[2]	5.9885***	1.3474***	103766*	0.0017*	13.6166**
	H^2	0.836	0.925	0.461	0.471	0.437
	± SE	0.023	0.011	0.063	0.055	0.064
Breeding population (BP)	Year[1]	111.48***	304.35***	548.49***	3.63NS	25.38***
	Repetition(Year)[1]	4.33*	9.80***	13.61***	9.25***	2.69NS
	Check[1]	0.02NS	0.27NS	6.95**	0.23NS	8.56***
	Genotype[2]	37.3015***	9.7258***	44524***	0.0062***	12.0119***
	Block(year*Rep)[2]	2.3134***	0.1159NS	44462***	0.0060***	11.2422***
	Genotype*Year[2]	7.6253***	0NS	8771.42NS	0.0031NS	0.6083NS
	H^2	0.826	0.911	0.333	0.327	0.333
	± SE	0.023	0.011	0.060	0.065	0.058

[1]F- value and test of the fixed effect

[2]Estimate and Wald test of the random effect

H^2 = Broad sense heritability of the family means

SE = Standard Error of H^2

Significance level : NS = Not significant, * = Significant at p-value between 0.05 and 0.01, ** = Significant at p-value between 0.01 and 0.001,*** = Significant at p-value < 0.001

genetic variance in each population for all the traits . The genetic variance in BP for GY is only 12.5 % that of DP. It reaches 40 to 60% for TGW, GNC and NUE. The year effect is highly significant for most traits except NUE in DP and except GNC in BP. The interaction between genotype and year (G × Y) is significant for all traits in DP. It is significant only for DFl in BP.

H^2 is high for DFl and TGW and moderate for GNC, GY and NUE in both populations. In terms of genetic correlations between traits, the two populations harbor comparable correlation structure according to Mantel test (correlation of 0.93 with p-value < 0.05, Table 3). GY and GNC traits were weakly correlated, with r = -0.16 and-0.11 at the 5% threshold, in DP and BP respectively. Conversely, correlations between GY and NUE were highly significant (r = 0.96 in BP to 0.98 in DP with p-value < 0.0001). However, DFl, TGW and GNC were slightly and rather negatively correlated with the other traits in both populations.

Genotypic properties of the two populations

Figure 1 shows 382 genotypes consisting of 184 DP accessions including 12 founders of the BP, and 198 BP accessions. Genotypes from the same origins tended to group together. The majority of the accessions (126) belong to the tropical japonica group, 16 to the temperate japonica group, and 42 accessions are admixed. The founders were distributed over the diversity of the DP. The BP is disseminated within the DP.

The genetic differentiation between DP and BP (estimated by F_{st}) was evaluated at only 0.01, meaning that the differentiation between the two populations is almost null. Most of the genetic diversity was found in the populations. They formed together six genetic clusters using LEA package. Moreover, we artificially assigned most accessions to cluster 9 when any associated probability exceeded 0.6; such accessions were considered highly admixed. This cluster was composed of 96 DP and 115 BP accessions (Figure 2). The DP was under-represented in cluster 1 and fairly evenly distributed among clusters 2 to 6. In contrast, the BP population was better represented in clusters 1, 2 and 5 and under-represented in clusters 3, 4 and 6. Four of the six clusters appeared to be more population specific (clusters 3, 4 and 6, specific to DP and cluster 1 specific to BP). However, they represented only 43.2% of the total accessions (165/382) (Table 4).

The proportion of monomorphic loci calculated on 87,089 SNPs was 2.05% in DP, and 2.45% in BP. It should be noted that the monomorphic rate among the 12 founders (without PCT 11, not genotyped) reaches 34.70%. The analysis of variance showed that DP and BP differed very highly significantly (p value < 0.0001) for both MAF (10.2% for DP and 9.7% for BP) and Ho (1.8% for DP and 3.2% for BP).

Regarding kinship coefficients, the distribution was flatter and more spread out around zero within DP, meaning that some paired accessions were unrelated, while others were highly related (min=-0.597 and max=4.915). In contrast, the kinship coefficients associated with the BP were more concentrated around zero, meaning that a higher proportion of the population was unrelated with another member of the population (min=-0.416, maxi=3.754). When considering paired accessions from the two populations, the same kind of distribution was observed as within the BP population, which means a weak mean relationship between the two populations (Figure S1).

Finally, the average linkage disequilibrium (LD) in the 12 chromosomes for distances between 0 to 25 kb was 0.59 for the DP

Table 3: Genetic correlation between traits in (a) diversity panel and (b) breeding population.

(a) Variables	DFI	TGW	GY	GNC	NUE
DFI	1	-0.191	-0.121	-0.170	-0.155
TGW	-0.191	1	0.064	-0.082	0.068
GY	-0.121	0.064	1	-0.108	0.983
GNC	-0.170	-0.082	-0.108	1	-0.100
NUE	-0.155	0.068	0.983	-0.100	1
(b) Variables	**DFI**	**TGW**	**GY**	**GNC**	**NUE**
DFI	1	-0.302	-0.235	0.117	-0.229
TGW	-0.302	1	0.228	-0.078	0.218
GY	-0.235	0.228	1	-0.160	0.965
GNC	0.117	-0.078	-0.160	1	-0.195
NUE	-0.229	0.218	0.965	-0.195	1

Values in bold are different from 0 with a significance level alpha = 0.05.

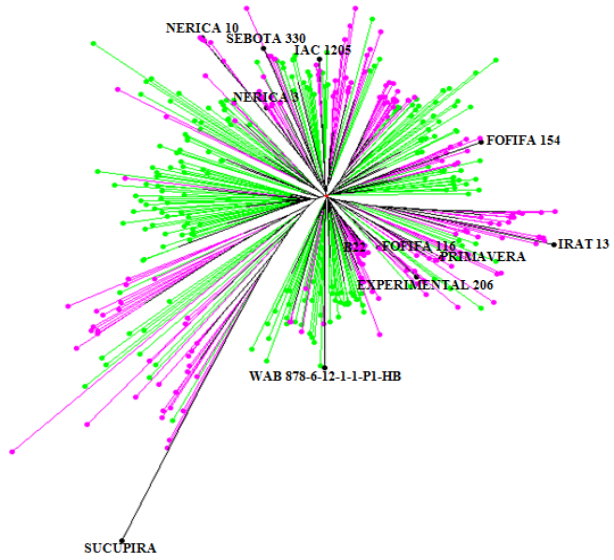


Figure 1: Neighbor joining tree of the 382 accessions based on 87 089 single nucleotide polymorphisms (SNPs).

Unweighted neighbor-joining tree was based on simple matching distances constructed from the genotypic information of 184 accessions from the DP (pink), 12 founders of the BP (black) and 198 genotypes from the BP (green).

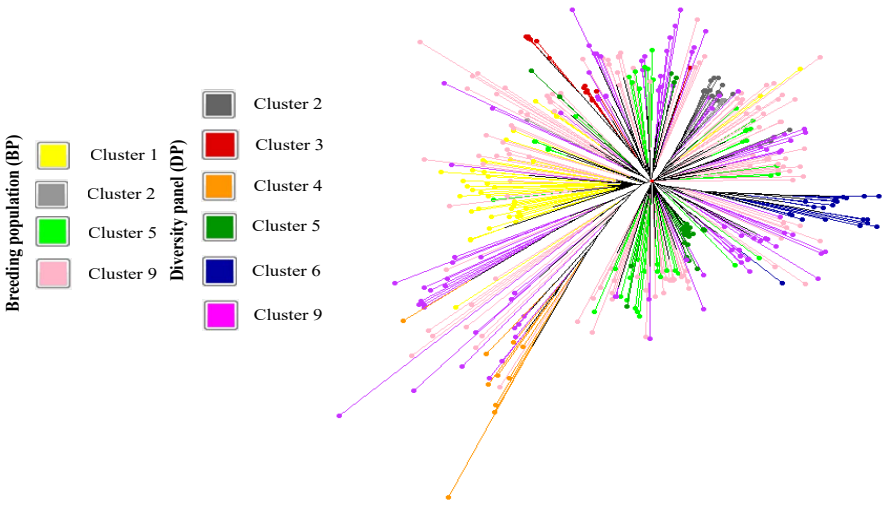


Figure 2: Neighbor joining tree based on the membership of genotypes to clusters. Unweighted neighbor-joining tree was built on the 382 accessions. The diversity tree was built with 87 089 imputed SNPs.

Table 4: Population structure given by the LEA package with the numbers of accessions belonging to the different clusters.

Populations	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 9	Total
BP	39	3			41		115	198
DP		15	13	10	29	21	96	172
Parents			2	1	2	1	6	12
			Nerica 3	Sucupira	B22	IRAT 13	Exp 206	
			Nerica 10		FOFIFA 116		FOFIFA 154	
							IAC 1205	
							SEBOTA 330	
							WAB 878 6-12-1-1-P1-HB	
							PRIMAVERA	

Parental genotype names through the clusters are indicated in the table.

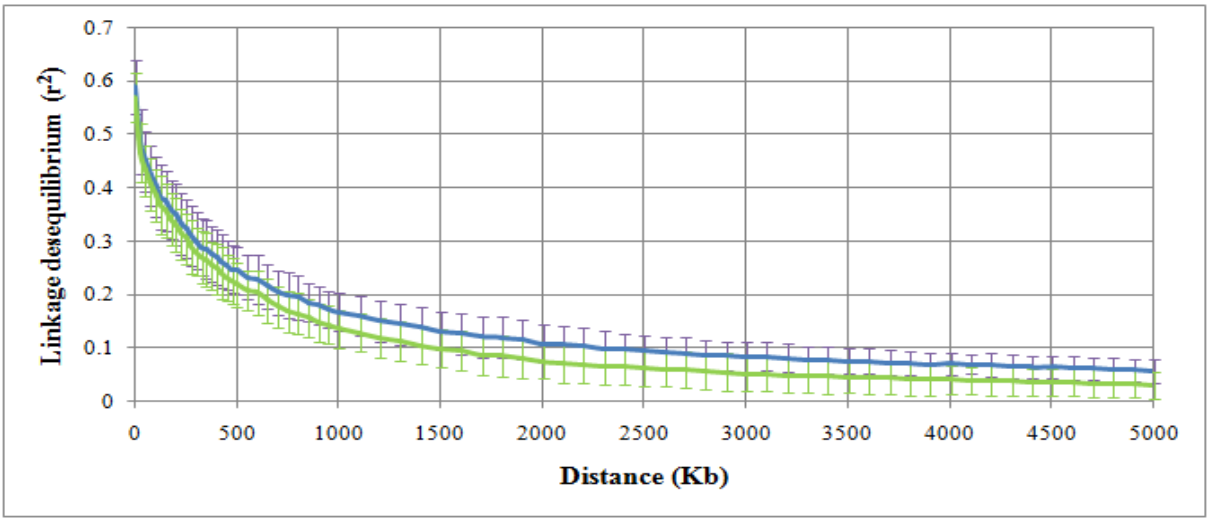


Figure 3: Patterns of decay in linkage disequilibrium in diversity panel (DP, purple) and in the breeding population (BP, green). The curve represents the average r^2 among the 12 chromosomes and the bars represent the associated standard deviation across chromosomes.

and 0.57 for the BP (Figure 3). The r^2 reached half its initial value at a distance of 275-300 kb for both populations. The r^2 was equal to 0.2, for a distance of 750 kb in the DP and 600 kb in the BP. The r^2 reached 0.1, for pairwise distances of 2,200 kb in DP and 1,400 kb in BP. The LD decreases slower in DP than in BP.

Predictive ability of genomic prediction in cross-validation experiment

Each data point on the figure represents the average predictive ability (AvPA) of a scenario while each vertical bar represents the corresponding standard deviation obtained over the 100 replications. Overall, AvPA ranged from 0.46 to 0.76. There were some differences of AvPA between traitsWith strong heritability on the TGW trait (0.925), its AvPA was very high. (Figure 4).

Differences between scenarios were significant only for NUE (Table 5), for which the prediction model effect was significant at p-value < 0.0001 and the interaction between model and SNP number was significant at p-value = 0.022. The interaction was then used as error term to test the main effects; this resulted in no significant difference between the two genomic prediction models anymore. For all traits, a low fraction of total variation was explained by the linear model (R^2 between 0.3% and 5.6%), meaning that most variation was driven by the random sampling of the cross-validation process.

Positive relationships (0.46 for RHKS with p-value = 0.21 and 0.54 for GBLUP with p-value = 0.16) were observed between average predictive ability and trait heritability but not significant (Figure S2).

When the GEBV of the 12 founder accessions was predicted using model trained with the remaining 172 accessions of the DP (Figure 5), high predictive ability, ranging from 0.59 to 0.86 were obtained.

On the other hand, the cross-validation experiment within BP showed very low AvPA for all traits (Figure S3).

Predictive ability of genomic prediction across-populations using DP as calibration set

Four scenarios were considered in this non-replicated genomic prediction experiment, by combining two genomic prediction models and two SNP matrices (Figure 6). The breeding population prediction ability (BPPA) was defined as the prediction ability when BP was to be predicted using the calibration developed on the DP. BPPA varied from 0.06 to 0.27. The RKHS model resulted in slightly higher BPPA than GBLUP. The across populations prediction was much less accurate than cross validation within the DP.

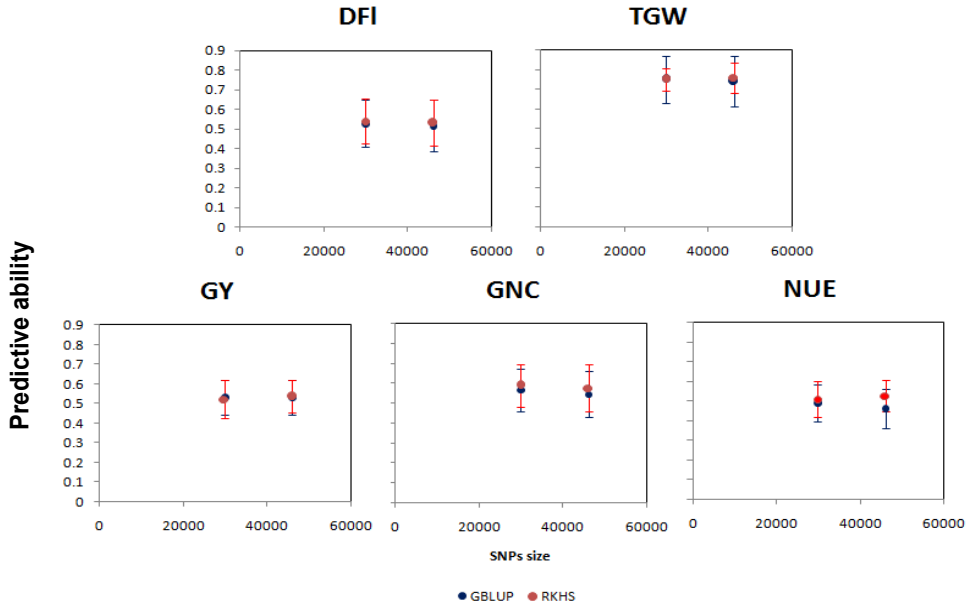


Figure 4: Predictive ability in a 5-fold cross validation experiment in the DP for five traits with two statistical methods (GBLUP, RKHS) and two SNP numbers. 100 repetitions were applied.

									Default F test		
Trait	**R²**	**CV**	**RMSE**	**Mean**	**Source**	**DF**	**SS**	**MS**	**F value**	**p-value**	
FL	0.003	27.841	0.168	0.604	Scenarios	3	0.036	0.012	0.420	0.737	NS
					Error	396	11.213	0.028			
					Corrected Total	399	11.249				
					Model	1	0.023	0.023	0.820	0.365	NS
					SNP number	1	0.012	0.012	0.430	0.513	NS
					Model*SNP number	1	0.000	0.000	0.020	0.902	NS
GNC	0.019	26.080	0.173	0.664	Scenarios	3	0.226	0.075	2.510	0.059	NS
					Error	396	11.887	0.030			
					Corrected Total	399	12.113				
					Model	1	0.167	0.167	5.550	0.019	*
					SNP number	1	0.057	0.057	1.880	0.171	NS
					Model*SNP number	1	0.002	0.002	0.080	0.775	NS
GY	0.003	20.866	0.125	0.599	Scenarios	3	0.016	0.005	0.340	0.794	NS
					Error	396	6.181	0.016			
					Corrected Total	399	6.197				
					Model	1	0.003	0.003	0.160	0.689	NS
					SNP number	1	0.005	0.005	0.320	0.573	NS
					Model*SNP number	1	0.009	0.009	0.550	0.459	NS
NUE	0.056	22.502	0.125	0.555	Scenarios	3	0.364	0.121	7.790	<.0001	***
					Error	396	6.173	0.016			
					Corrected Total	399	6.538				
					Model	1	0.277	0.277	17.780	<.0001	***
					SNP number	1	0.005	0.005	0.330	0.564	NS
					Model*SNP number	1	0.082	0.082	5.250	0.022	*
TGW	0.012	16.451	0.164	0.997	Scenarios	3	0.128	0.043	1.590	0.192	NS
					Error	396	10.652	0.027			
					Corrected Total	399	10.780				
					Model	1	0.047	0.047	1.750	0.187	NS
					SNP number	1	0.002	0.002	0.080	0.774	NS
					Model*SNP number	1	0.079	0.079	2.930	0.088	NS

Table 5: ANOVA of factors affecting variation of the AvPA in the cross validation experiment within the diversity population (DP) considering 20 scenarios. Sources of variation were: models (GBLUP and RKHS) and SNP numbers (29 847 and 46 055). R^2coefficient of determination.

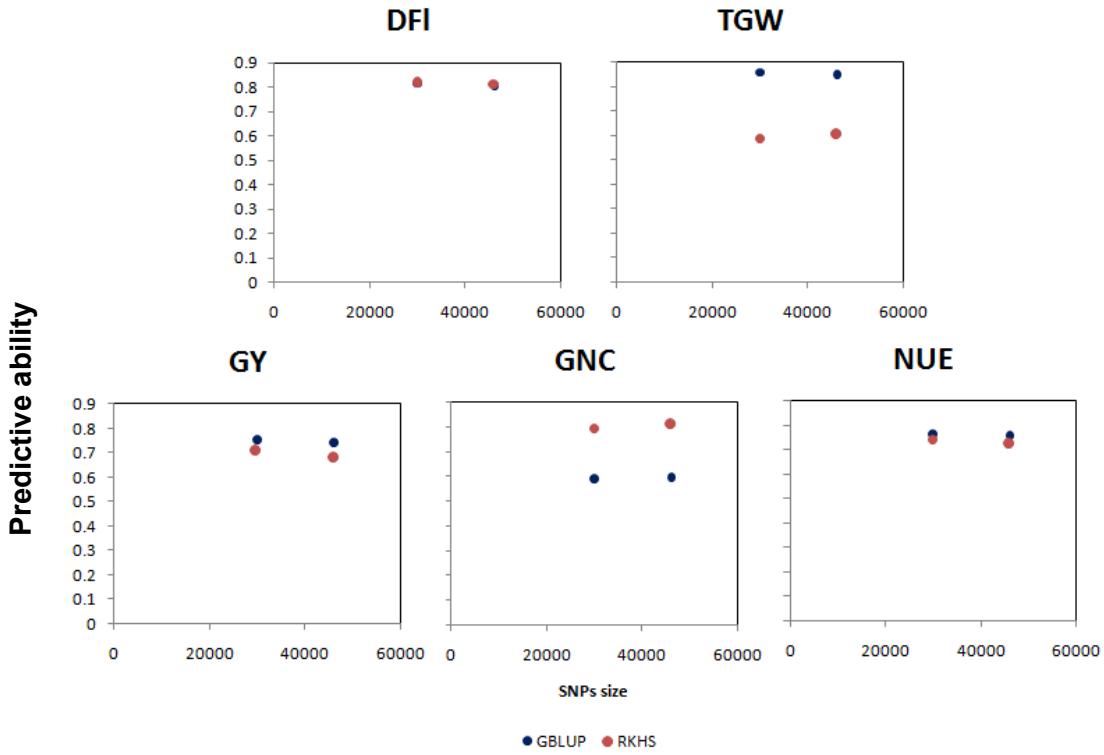


Figure 5: Predective ability intra-DP using 172 DP lines without BP founders to calibrate the model and 12 founders of the BP to validate; for five traits with two statistical methods (GBLUP, RKHS) and two SNP numbers.

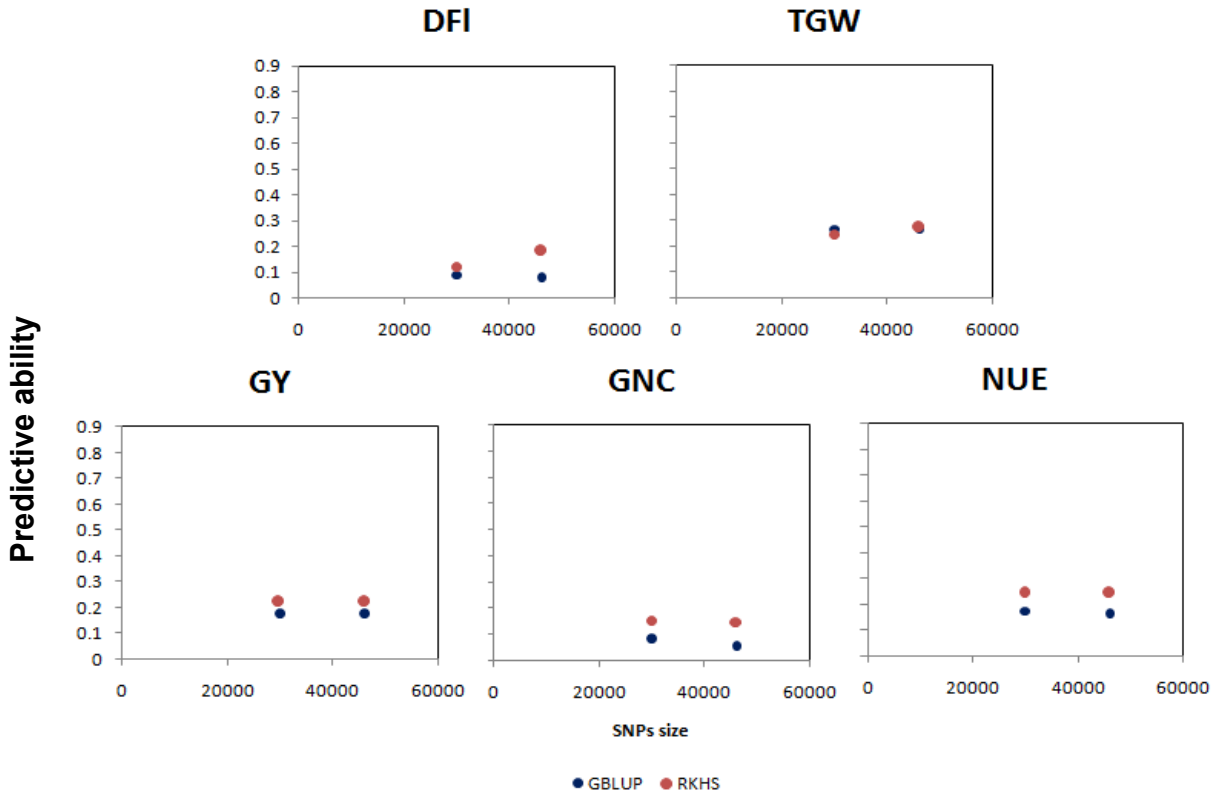


Figure 6: Predective ability across populations using two statistical methods (GBLUP and RKHS). The DP was used as the calibration set and the BP as validation set; for five traits with two statistical methods (GBLUP, RKHS) and two SNP numbers.

Discussions

Reducing fertilizer inputs and improving nitrogen use efficiency (NUE) are the main objectives of plant nutrition research [40]. These two challenges contribute to preserving the environment and improving sustainable and productive agriculture [41]. There have been continuous improvements of NUE over the years along with increases in crop yield as in wheat, maize, barley or rice [7,42-45].

In rice, GWAS results explain only a small proportion of the genetic variability available for NUE [10]. In this study, we explored the potential of genomic selection to improve NUE prediction with the aim to include this strategy in rice breeding programs targeting this trait.

Three research points will be discussed in this section for the improvement of NUE and Yield related traits in upland rice: (i) the potential of genomic prediction to select good candidates for initiating new breeding populations, (ii) the phenotypic and genotypic characteristics of the breeding population (BP) that directly descended from a subset of DP and (iii) the predictive ability of the BP obtained from the diversity panel (DP).

Predictive ability of genomic prediction in the diversity panel

The diversity panel considered in our study is a broad-based panel. The vast majority of varieties belong to the japonica subspecies. Most of them belong to the tropical and temperate japonica groups and are already adapted to local conditions [10]. Moreover, The moderate to high predictive ability obtained by cross validation within the diversity panel highligt the relevance of the genomic calibration developped to identify new NUE donors within the tropical and temperate janopinca groups.

The first five genotypes with good GY in the DP are G3158 (EXP 304), G3156 (EXP 302), G3229 (SCRID090 72-3-1-3-5-1--), G3228 (SCRID090 60-1-1-2-4-1-2), and G3203 (NERICA 3). Four of these lines were derived from the FOFIFA/CIRAD breeding program and one from Africa rice. The lines that exhibited good performance for NUE are G3158 (EXP 304), G3203 (NERICA 3), G3223 (PCT 4\SA\4\1>330-2-3-2-M 5-4-3-1-5), G3156 (EXP 302), and G3229 (SCRID090 72-3-1-3-5-1). Three of these lines were from the FOFIFA/CIRAD breeding program, while PCT4 line was from CIAT. The genetic correlation between the two traits (GY and NUE) is strong (0.98).

Our results from cross-validation with DP are consistent with those found by other authors working on different self-pollinated species such as rice, wheat, sorghum and cross-pollinated maize [20,46-50].

The predictive ability of the GEBV followed the same trend as the trait heritability, [18]. Both genomic prediction models are well suited to use relationships between accessions to construct efficient prediction models, as shown by several authors [51-53].

Phenotypic and genotypic comparison of the two populations

Phenotypic variation was largely influenced by accession, year and accession × year (G × Y) interaction, within each population. We found high significant genetic variance in the two populations at p-value < 0.001 in the two populations, for almost all traits.

We recall that DP is a working collection which is an artificial population because it contains elite lines from several selection programs and historical lines from the Malagasy program. The BP was developed from DP by selecting a subsample of accessions (12 out of 184) and recombining them during 2 generations of random mating (using a thirteenth founder, PCT 11, as female) followed by 4 generations of SSD for generation advancement. The 12 founders were retained primarily for their adaptation to midlands conditions. This sampling first reduced the genetic diversity, with a nearly 17-fold higher rate of monomorphic loci (34.70%) than in the entire DP (2.05%). Fortunately, the PCT 11 founder, which has contributed half of the BP genome, was a broad-based synthetic population PCT 11 re-expanded the BP genetic base and reduced the monomorphic rate to as low as 2.45% .

Both populations showed very weak genetic differentiation (Fst = 0.01).. However, the structure and kinship analyses showed significant differences between them: DP presented a stronger structure than the BP, and the distribution of kinship coefficients in DP was flatter and less centered around zero than in BP. These parameters are known to impact the LD patterns and predictive ability [54].

Across-population prediction and practical implications for the rice breeding programs

When the DP was used as a calibration population to predict GEBV of the BP, the prediction abilities ranged from 0.06 to 0.27, meaning that the mean prediction ability for each trait was drastically reduced compared to cross-validation within DP. This reduction could probably be explained by the weak relationships between the two population considered their contrasting genetic structures. The slightly different LD patterns have contributed a theoretical basis to the importance of relationships between populations in the predictive ability of genomic prediction. Indeed, many experimental results also showed that the prediction ability of GS should be based primarily on close relationships between the calibration and validation data sets [55-58].

In our study, the synthetic PCT 11 population alone contributed half of the genome because it is the only female parent, whereas the 12 male parents contributed theoretically only 4.2% each. PCT 11 was not genotyped in the study even though six inbred lines derived from it were included in the DP and genotyped (Table S1). In addition, phenotyping was performed on $S_{4:5}$ and $S_{4:6}$ progenies directly derived from the respective $S_{3:4}$ reference lines, while genotyping was performed two generations later, by sampling a single plant per $S_{4:6}$ progeny. These complex procedures reduced the genetic relationships between genotyped plants and phenotyped progenies within BP, as well as between BP and DP, which potentially explained the weak predictive ability of genomic prediction in the BP and across population.

In upland rice, good performances of across populations genomic prediction could be obtained in the context of either (1) biparental crosses and pedigree selection to predict advanced F_5-F_7 lines [20,23] or (2) synthetic populations with a broad genetic base and early generations selection of the target population [18]. Simulation and empirical studies of GS showed sufficient accuracies to generate rapid gains in early generations selection, because all individuals to be predicted are linked with a sufficient number of close ancestors [59]. Beyond those early generations selection, allele frequency changes, recombination, and inbreeding make analytical prediction of gain impossible [60-63].

Conclusion

BP can be valuable for a sustainable genetic improvement of NUE. It is hoped that genomic prediction models can be as promising as in previous case studies if the genotypic and phenotypic data are referred to the same genetic unit ($S_{3:4}$ seeds). We propose to genotype again the S_6 generation, using a bulk of $S_{4:6}$ progeny to infer the $S_{3:4}$ genotype. Then, prediction equations can be reconstructed to predict the GEBV of $S_{3:4}$ plants, based on phenotypes of their respective $S_{4:5}$ and $S_{4:6}$ progenies.

This will allow to select and recombine the best BP accessions to develop a new population. The ms alleles that still segregate should make seed management and handling easier and should be used routinely to optimize recombination within the population. Such strategy is already used in recurrent selection programs. Considering the whole breeding process, GS has great potential to accelerate the genetic gain per unit of time and cost through early phenotyping and selection within segregating units before actual phenotypes are measured. In this perspective, prediction models could be calibrated as soon as $S_{0:2}$ are produced.

For the time being, breeders can already use the results of cross-validation on the upland rice diversity panel in the upland rice breeding program. In the DP, we have detected genotypes that have high levels of NUE.

Acknowledgements

Firstly, the authors are grateful to the field staff of the National Center for Applied Research on Rural Development FOFIFA in Madagascar who contributed to the data collection. Secondly, the authors thank Nourollah AHMADI and Gilles TROUCHE from CIRAD for their support in defining the research subject and designing the structure of this article. This research was financially supported by the Agropolis Fondation-GeneRice project, grant 1605-019 and the CGIAR research program on Rice- (CRP Rice), grant C19386.

References

1. Peng B, Li J, Kong DY, He LL, Li MG, et al. (2018) Genetic Improvement of Grain Quality Promoted by High and New Technology in Rice. J Agri Sci 11:81.
2. Agarwal P, Parida SK, Raghuvanshi S, Kapoor S, Khurana P, et al. (2016) Rice Improvement Through Genome-Based Functional Analysis and Molecular Breeding in India. Rice 9.
3. Raboin LM, Randriambololona T, Radanielina T, Ramanantsoanirina A, Ahmadi N, et al. (2014) Upland rice varieties for smallholder farming in the cold conditions in Madagascar's tropical highlands. Field Crops Research 169: 11-20.
4. Minten B, Randrianarisoa JC, Barrett CB (2007) Productivity in Malagasy rice systems: wealth-differentiated constraints and priorities: Productivity in Malagasy rice systems. Agric Econmic 37: 225–237.
5. Alvarez S, Rufino MC, Vayssières J, Salgado P, Tittonell P, et al. (2014) Whole-farm nitrogen cycling and intensification of crop-livestock systems in the highlands of Madagascar: An application of network analysis. Agricultural Systems 126: 25–37.
6. Huang S, Zhao C, Zhang Y, Wang C (2018) Nitrogen Use Efficiency in Rice, in: Amanullah, Fahad, S. (Eds.), Nitrogen in Agricultur Updates. InTech.
7. Tiong J, Sharma N, Sampath R, MacKenzie N, Watanabe S, et al. (2021) Improving Nitrogen Use Efficiency Through Overexpression of Alanine Aminotransferase in Rice, Wheat, and Barley. Frontiers in Plant Sci 12.
8. Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization1. Agr J 74: 562.
9. Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Botany 58: 2369–2387.
10. Rakotoson T, Dusserre J, Letourmy P, Frouin J, Ratsimiala IR, et al. (2021) Genome-Wide Association Study of Nitrogen Use Efficiency and Agronomic Traits in Upland Rice. Rice Sci 28: 379–390.
11. Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for Crop Improvement. Crop Sci 49: 1.
12. Goddard ME, Hayes BJ (2007) Genomic selection: Genomic selection. J of Animal Breeding and Gen 124: 323–330.
13. Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829. PMID:11290733
14. Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9: 166–177.
15. Gonzalez-Recio O, Coffey MP, Pryce JE (2014) On the value of the phenotypes in the genomic era. J Dair Sci 97: 7905–7915.
16. Jacquin L, Cao TV, Ahmadi N (2016) A Unified and Comprehensible View of Parametric and Kernel Methods for Genomic Prediction with Application to Rice. Frontiers in Genetics 7.
17. Xavier A, Muir WM, Rainey KM (2016) Assessing Predictive Properties of Genome-Wide Selection in Soybeans. G3; Gene |Genome |Genetics 6: 2611–2616.
18. Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, et al. (2015) Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. PLOS ONE 10: e0136594.
19. Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, et al. (2012) Comparison Between Linear and Non-parametric Regression Models for Genome-Enabled Prediction in Wheat. G3; Genes Genomes Genetics 2: 1595–1605.
20. Ben Hassen M, Cao TV, Bartholomé J, Orasen G, Colombi C, et al. (2018) Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theo and App Gene 131: 417–435.
21. Atlin GN, Cairns JE, Das B (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Glo Fd Sec 12: 31–37.
22. Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proceedings of the National Academy of Sciences 111: 12456–12461.
23. Frouin J, Labeyrie A, Boisnard A, Sacchi GA, 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: e0217516.
24. Cui Y, Li R, Li G, Zhang F, Zhu T, et al. (2019) Hybrid breeding of rice via genomic selection. Plant Biotech J.
25. Holland JB, Nyquist WE, Cervantes-Martínez CT (2010) Estimating and Interpreting Heritability for Plant Breeding: An Update, in: Janick, J. (Ed.) Plant Breeding Reviews. John Wiley & Sons, Inc., Oxford, UK, Pp: 9–112.
26. Cormier F, Mournet P, Causse S, Arnau G, Maledon E, et al. (2019) Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. Ecology and Evolution 9: 5617–5636.
27. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6: e19379.
28. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, et al. (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9: e90346.
29. Sempéré G, Pétel A, Rouard M, Frouin J, Hueber Y, et al. (2019) Gigwa v2-Extended and improved genotype investigator. Giga Science 8.
30. Browning BL, Browning SR (2016) Genotype Imputation with Millions of Reference Samples. Am J Human Gen 98:116–126.
31. Perrier X, Jacquemoud-Collet JP (2006) DARwin software.
32. Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes 5: 184–186.
33. Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies. Methods in Ecology and Evolution 6: 925–929.
34. Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Ther and Appl Gene 38: 226–231.
35. Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, et al. (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108: 285–291.
36. VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91: 4414–4423.
37. Bhandari A, Bartholomé J, Cao-Hamadoun TV, Kumari N, Frouin J, et al. (2019) Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLOS ONE 14: e0208871.
38. Gianola D, van Kaam JBCHM (2008) Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics 178: 2289–2303.

39. Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. Genetics 201: 759–768.

40. Lammerts van Bueren ET, Struik PC (2017) Diverse concepts of breeding for nitrogen use efficiency. A review. Agronomy for Sustainable Development 37.

41. Masclaux-Daubresse C, Daniel-Vedele F, Dechorgnat J, Chardon F, Gaufichon L, et al. (2010) Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. Annals of Botany 105: 1141–1157.

42. Ortiz-MonasterioR JI, Sayre KD, Rajaram S, McMahon M (1997) Genetic Progress in Wheat Yield and Nitrogen Use Efficiency under Four Nitrogen Rates. Crop Sci 37: 898–904.

43. Sadras VO, Lawson C (2013) Nitrogen and water-use efficiency of Australian wheat varieties released between 1958 and 2007. Eur J Agron 46: 34–41.

44. Ciampitti IA, Vyn TJ (2012) Physiological perspectives of changes over time in maize yield dependency on nitrogen uptake and associated nitrogen efficiencies: A review. Field Crops Research 133: 48–67.

45. Fernie AR (2020) A push, and a pull, to enhance nitrogen use efficiency in rice. The Plant J 103: 5–6.

46. Spindel J, Begum H, Akdemir D, Virk P, Collard B, 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.

47. Xu Y, Wang X, Ding X, Zheng X, Yang Z, et al. (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11.

48. Juliana P, Singh RP, Braun HJ, Huerta-Espino J, Crespo-Herrera L, et al. (2020) Genomic Selection for Grain Yield in the CIMMYT Wheat Breeding Program—Status and Perspectives. Frontiers in Plant Sci 11.

49. Sapkota S, Boyles R, Cooper E, Brenton Z, Myers M, et al. (2020) Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. Crop Sci 60: 132–148.

50. Allier A, Teyssèdre S, Lehermeier C, Charcosset A, Moreau L (2020) Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. Theoretical and Applied Genetics 133: 201–215.

51. Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, et al. (2014) The impact of population structure on genomic prediction in stratified populations. Theoretical and Applied Genetics 127: 749–762.

52. Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, et al. (2015) Training set optimization under population structure in genomic selection. Theoretical and Applied Genetics 128: 145–158.

53. Rio S, Mary-Huard T, Moreau L, Charcosset A (2019) Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. Theoretical and Applied Genetics 132: 81–96.

54. de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D (2013) Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. PLoS Genetics 9: e1003608.

55. Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. Genetics 194: 597–607.

56. Bernal-Vasquez AM, Möhring J, Schmidt M, Schönleben M, Schön CC, et al. (2014) The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. BMC Genomics 15, 646.

57. Ma P, Huang J, Gong W, Li X, Gao H, et al. (2018) The impact of genomic relatedness between populations on the genomic estimated breeding values. J Animal Sci and Biotech 9.

58. Lorenz A, Nice L (2017) Training Population Design and Resource Allocation for Genomic Selection in Plant Breeding, in: Varshney, R.K., Roorkiwal, M., Sorrells, M.E. (Eds.), Genomic Selection for Crop Improvement. Springer International Publishing, Cham, Pp: 7–22.

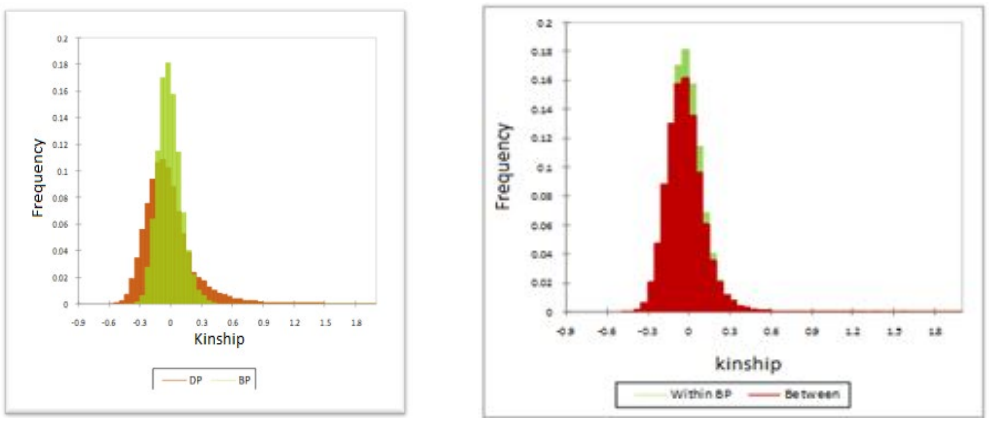
59. Auinger HJ, Schönleben M, Lehermeier C, Schmidt M, Korzun V, et al. (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (Secale cereale L.). Theoretical and Applied Genetics 129: 2043–2053.

60. Jannink JL (2010) Dynamics of long-term genomic selection. Genetics Selection Evolution 42, 35.

61. Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.) . Plant Sci 242: 23–36.

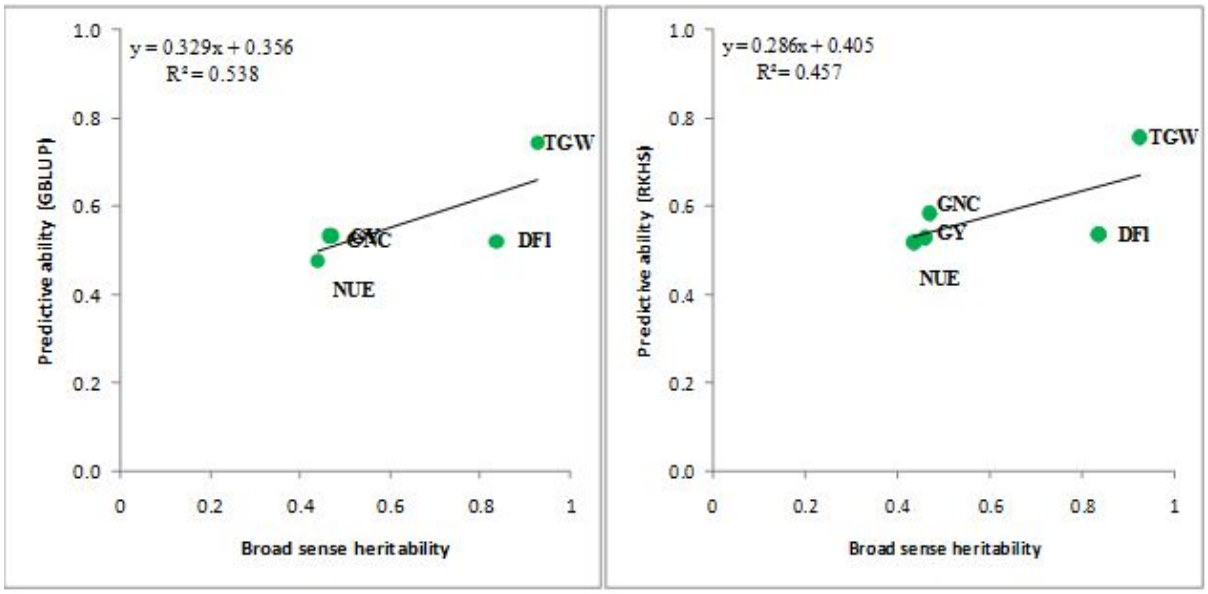
62. DoVale JC, Carvalho HF, Sabadin F, Fritsche-Neto R (2021) Reduction of genotyping marker density for genomic selection is not an affordable approach to long-term breeding in cross-pollinated crops (preprint). Genetics.

63. Xu Y, Ma K, Zhao Y, Wang X, Zhou K, et al. (2021) Genomic selection: A breakthrough technology in rice breeding. The Crop Jrnl 9: 669–677.

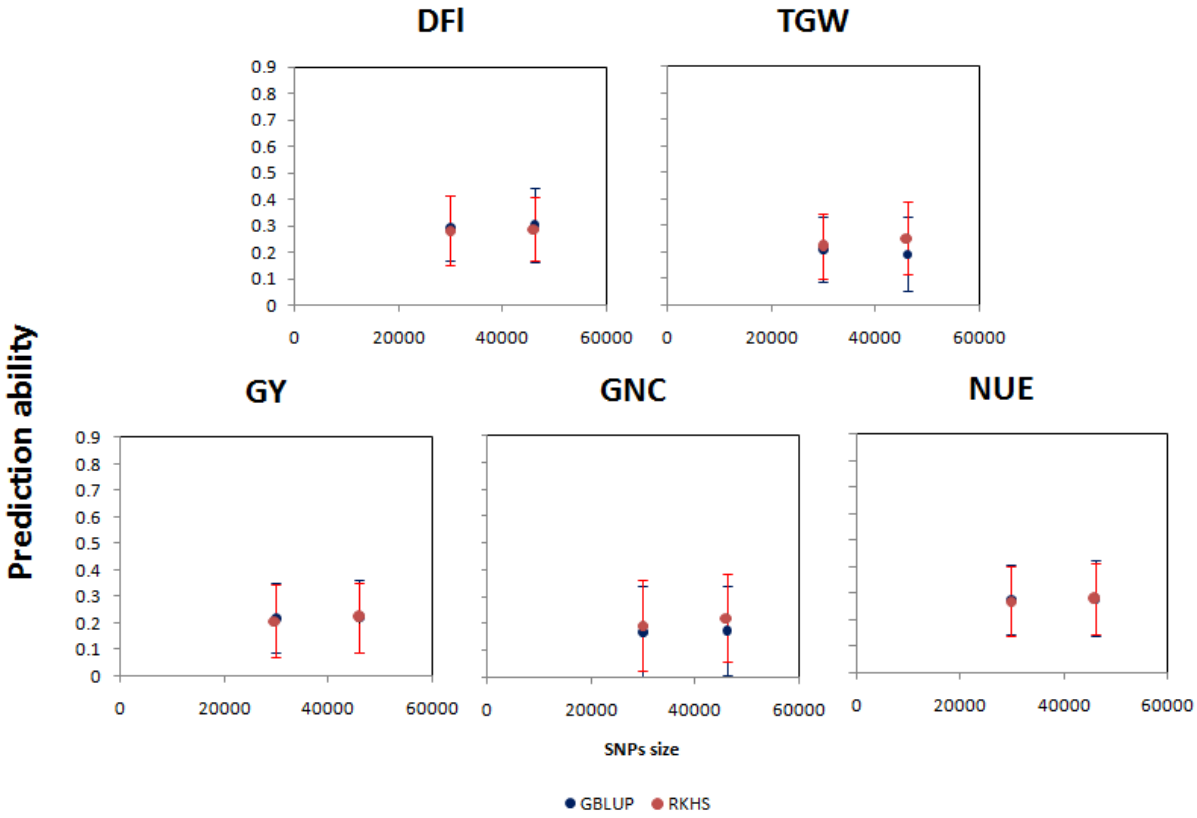


(A) **(B)**

Supplemental Figure 1: Relationship between the diversity panel and the breeding population (a) and the kinship within BP and between the two populations (DP and BP) (b) The comparison of kinship distributions within the BP and between the two populations (DP and BP) was highly significant with p-value < 0.0001.



Supplemental Figure 2: Relationship between predictive ability in the DP and their heritability (a) GBLUP model, (b) RKHS model. Both correlations were statistically not significant with p-value = 0.16 for the GBLUP model and a p-value = 0.21 for the RKHS.



Supplemental Figure 3: Results of predictive ability in a 5-fold cross validation experiment in the breeding population (BP) for 5 phenotypic traits obtained with two statistical methods (GBLUP, RKHS). Once the model was calibrated, it was used to predict the genetic value of the validation set, and a correlation was estimated between observed phenotype and predicted phenotypes. This process was repeated 100 times. The mean correlation gives the predictive ability.

Population	GID	Genotype	Country	Research center	Group	Founder of BP
Diversity panel (DP)	G3122	C 537B	Madagascar	FOFIFA-CIRAD	temperate-japonica	
	G3123	C507 1373-1-b-2- -	Madagascar	FOFIFA-CIRAD	temperate-japonica	
	G3124	C630 139-46-2-3-3-b-1-1-1	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3125	C630 38-4-1-b-3-2-1-b-b	Madagascar	FOFIFA-CIRAD	0	
	G3127	CAIAPO	Brazil		tropical-japonica	
	G3128	CHA LOY OE	Thailand		temperate-japonica	
	G3130	CIRAD 141	Brazil	CIRAD	tropical-japonica	
	G3131	CIRAD 392	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3132	CIRAD 394	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3134	CIRAD 409	Brazil	CIRAD	tropical-japonica	
	G3135	CIRAD 447	Brazil	CIRAD	tropical-japonica	
	G3136	CIRAD 488	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3137	CNA 4123	Brazil	EMBRAPA	tropical-japonica	
	G3139	CNA 4137	Brazil	EMBRAPA	tropical-japonica	
	G3140	CNA4196	Brazil	EMBRAPA	tropical-japonica	
	G3141	CNA-IREM 190	Brazil	EMBRAPA	tropical-japonica	
	G3142	CT 13582-15-5-M	Colombia	CIAT	tropical-japonica	
	G3143	Cuiabana	Brazil		tropical-japonica	
	G3144	CURINCA	Brazil		0	
	G3145	DANGREY	Bhutan		tropical-japonica	
	G3146	Daniela	Brazil		tropical-japonica	
	G3147	DOURADO PRECOCE	Brazil		tropical-japonica	
	G3148	EARLY MUTANT IAC 165	Brazil		tropical-japonica	
	G3149	Estrela	Brazil		temperate-japonica	
	G3150	EXP 003	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3151	Exp 006	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3152	EXP 011	Madagascar	FOFIFA-CIRAD	temperate-japonica	
	G3153	EXP 013	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3154	Exp 202	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3155	EXP 206	Madagascar	FOFIFA-CIRAD	tropical-japonica	1
	G3156	EXP 302	Madagascar	FOFIFA-CIRAD	tropical-japonica	

G3157	EXP 303	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3158	EXP 304	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3159	EXP 401	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3160	EXP 409	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3161	EXP 910	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3162	F152.06.33.53 13-1-5-1-1	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3163	F154.3G.04.12.10 1	Madagascar	FOFIFA-CIRAD	tropical-japonica	1
G3164	FOFIFA 116	Madagascar	FOFIFA-CIRAD	tropical-japonica	1
G3165	FOFIFA 151	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3167	FOFIFA 167	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3168	FOFIFA 168	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3169	FOFIFA 171	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3170	FOFIFA 172	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3171	FOFIFA 173	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3172	FOFIFA 180	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3173	FOFIFA 181	Madagascar	FOFIFA-CIRAD	temperate-japonica	
G3174	FOFIFA 62	Madagascar	FOFIFA-CIRAD	tropical-japonica	
G3175	GUARANI	Brazil		tropical-japonica	
G3176	HD 1-4	France	CIRAD	tropical-japonica	
G3177	IAC 1205	Brazil	IAC	0	1
G3179	IR 53236-275-1	Philippines	IRRI	tropical-japonica	
G3180	IR 66421-105-1-1	Philippines	IRRI	0	
G3181	IRAT 109	Brazil	CIRAD	tropical-japonica	
G3182	IRAT 112	Brazil	CIRAD	tropical-japonica	
G3183	IRAT 13	Brazil	CIRAD	temperate-japonica	1
G3184	IRAT 134	Brazil	CIRAD	tropical-japonica	
G3185	IRAT 212	Brazil	CIRAD	tropical-japonica	
G3186	IRAT 234	Brazil	CIRAD	tropical-japonica	
G3187	IRAT 265	Brazil	CIRAD	tropical-japonica	
G3188	IRAT 367	Brazil	CIRAD	tropical-japonica	
G3189	IRAT 380	Brazil	CIRAD	tropical-japonica	
G3190	IREM 239	Brazil		tropical-japonica	
G3191	KUROKA	Japan		tropical-japonica	

	G3192	luluwini 22M	Brazil		tropical-japonica	
	G3193	Munumliguero	Brazil		tropical-japonica	
	G3194	NABESHI	Taiwan		temperate-japonica	
	G3195	NERICA 1	Ivory Coast	AFRICARICE	tropical-japonica	
	G3197	NERICA 11	Ivory Coast	AFRICARICE	0	
	G3198	NERICA 12	Ivory Coast	AFRICARICE	tropical-japonica	
	G3199	NERICA 13	Ivory Coast	AFRICARICE	tropical-japonica	
	G3200	NERICA 16	Ivory Coast	AFRICARICE	tropical-japonica	
	G3201	NERICA 18	Ivory Coast	AFRICARICE	tropical-japonica	
	G3202	NERICA 2	Ivory Coast	AFRICARICE	tropical-japonica	
	G3203	NERICA 3	Ivory Coast	AFRICARICE	tropical-japonica	1
	G3204	NERICA 5	Ivory Coast	AFRICARICE	tropical-japonica	
	G3205	NERICA 6	Ivory Coast	AFRICARICE	tropical-japonica	
	G3206	NERICA 7	Ivory Coast	AFRICARICE	tropical-japonica	
	G3207	NERICA 8	Ivory Coast	AFRICARICE	tropical-japonica	
	G3209	PCT 11 MAD2007\0\0 14-1-1-1-3-3-2	Colombia	CIAT	tropical-japonica	
	G3210	PCT 11 MAD2007\0\0 28-3-3-5-5-5	Colombia	CIAT	tropical-japonica	
	G3211	PCT 11 MAD2007\0\0 3-3-1-3-2-2-4	Colombia	CIAT	tropical-japonica	
	G3212	PCT 11 MAD2007\0\0 3-5-5-2-1-4-4	Colombia	CIAT	tropical-japonica	
	G3213	PCT 11 MAD2007\0\0 50-1-1-1-5-5-5	Colombia	CIAT	tropical-japonica	
	G3214	PCT 11 x CNA7 42-3-2	Colombia	CIAT	tropical-japonica	
	G3215	PCT 11 x CNA7 73-2-5	Colombia	CIAT	tropical-japonica	
	G3216	PCT 11\0\0\2\Bo\2\1>181	Colombia	CIAT	tropical-japonica	
	G3217	PCT 4 Mad2007\0\1 18-2--1-5-2-3	Colombia	CIAT	0	
	G3218	PCT 4\0\0\1>5-M-1-6	Colombia	CIAT	0	
	G3219	PCT 4\SA\1\1\.SA\2\1>746-1-5-4-1 5-5-1-1-1	Colombia	CIAT	tropical-japonica	
	G3220	PCT 4\SA\1\1>975-M-2-M-3 2-5-5-1-1	Colombia	CIAT	tropical-japonica	
	G3221	PCT 4\SA\4\1>1076-2-4-1-5	Colombia	CIAT	0	
	G3222	PCT 4\SA\4\1>330-1-4-5-1-M 1-1-1-1-2	Colombia	CIAT	tropical-japonica	
	G3223	PCT 4\SA\4\1>330-2-2-3-2-M 5-4-4-3-1-5	Colombia	CIAT	tropical-japonica	
	G3224	PCT 5\PHB\1\0.PHB\1.PHB\1.PHB\1>78-2--6-2-M	Colombia	CIAT	tropical-japonica	
	G3226	SCRID036 4-1-1-5-M	Madagascar	FOFIFA-CIRAD	0	

	G3227	SCRID090 148-1-2-4-5-4-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3228	SCRID090 60-1-1-2-4-1-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3229	SCRID090 72-3-1-3-5-1--	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3230	SCRID090 89-1-5-4-2-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3231	SCRID091 10-1-3-2-5-3-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3232	SCRID091 11-1-4-3-2-4-3	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3233	SCRID091 15-2-2-1-1-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3234	SCRID091 24-3-2-2-3-5-4	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3235	SCRID111 1-4-3-3-5-5-4	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3236	SCRID128 1-3-4-2-4-4	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3237	SCRID128 18-5-4-4-5-3	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3238	SCRID271 67-3-3	Madagascar	FOFIFA-CIRAD	0	
	G3239	SCRID128 21-3-1-1-1-3	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3240	SCRID136 20-1-1-1	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3241	SCRID139 18-2-4-1-1-3-1	Madagascar	FOFIFA-CIRAD	0	
	G3242	SCRID139 9-1-5-2-4-4-1	Madagascar	FOFIFA-CIRAD	0	
	G3244	SCRID195 11-4-4-2-4-3	Madagascar	FOFIFA-CIRAD	0	
	G3246	SCRID195 67-1-1-2-2	Madagascar	FOFIFA-CIRAD	0	
	G3247	SCRID195 A1-3-4-2-4-3	Madagascar	FOFIFA-CIRAD	0	
	G3248	SCRID195-1-5-3	Madagascar	FOFIFA-CIRAD	0	
	G3249	SCRID200 15-4-2-4-1	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3250	SCRID222 122-4-3-3	Madagascar	FOFIFA-CIRAD	0	
	G3251	SCRID222 134-1-1-2	Madagascar	FOFIFA-CIRAD	0	
	G3252	SCRID222 164-1-1-4	Madagascar	FOFIFA-CIRAD	0	
	G3253	SCRID241 1-1-1-1	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3254	SCRID242 22-1-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3255	SCRID243 52-1-1-4	Madagascar	FOFIFA-CIRAD	0	
	G3256	SCRID251 25-2-1-2	Madagascar	FOFIFA-CIRAD	0	
	G3257	SCRID251 95-1-1-3	Madagascar	FOFIFA-CIRAD	0	
	G3258	SCRID252 18-1-2-4	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3259	SCRID253 5-2-2-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3260	SCRID254 85-3-2-3	Madagascar	FOFIFA-CIRAD	0	

	G3261	SCRID260 19-2-1-2	Madagascar	FOFIFA-CIRAD	0	
	G3262	SCRID264 69-1-2	Madagascar	FOFIFA-CIRAD	0	
	G3263	SCRID271 12-1-3	Madagascar	FOFIFA-CIRAD	0	
	G3264	SCRID271 37-1-1	Madagascar	FOFIFA-CIRAD	0	
	G3265	SCRID273 17-1-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3266	SCRID273 25-1-3	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3267	SCRID274 30-1-3	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3268	SCRID275 13-1-5	Madagascar	FOFIFA-CIRAD	0	
	G3269	SCRID275 72-5-5	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3270	SCRID278 148-5-1	Madagascar	FOFIFA-CIRAD	0	
	G3271	SCRID278 151-5-1	Madagascar	FOFIFA-CIRAD	0	
	G3272	SCRID278 42-2-3	Madagascar	FOFIFA-CIRAD	0	
	G3273	SCRID292 116-4-2	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3274	SCRID292 24-2-5	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3275	SCRID6 4-3-M	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3278	SEBOTA 330	Brazil	CIRAD	tropical-japonica	1
	G3279	SEBOTA 337	Brazil	CIRAD	0	
	G3280	SEBOTA 400	Brazil	CIRAD	0	
	G3281	SEBOTA 401	Brazil	CIRAD	0	
	G3283	SEBOTA 403	Brazil	CIRAD	0	
	G3284	SEBOTA 404	Brazil	CIRAD	0	
	G3285	SEBOTA 405	Brazil	CIRAD	tropical-japonica	
	G3286	SEBOTA 406	Brazil	CIRAD	tropical-japonica	
	G3287	SEBOTA 408	Brazil	CIRAD	0	
	G3288	SEBOTA 409	Brazil	CIRAD	0	
	G3289	SEBOTA 410	Brazil	CIRAD	0	
	G3290	sucupira	Brazil		0	1
	G3291	TRES MESES	Brazil		tropical-japonica	
	G3292	WAB 450-11-1-P28-1-HB	Ivory Coast	AFRICARICE	tropical-japonica	
	G3293	WAB 450-25-2-9-4-1-B-HB	Ivory Coast	AFRICARICE	tropical-japonica	
	G3294	WAB 56-125	Ivory Coast	AFRICARICE	tropical-japonica	
	G3295	WAB 56-50	Ivory Coast	AFRICARICE	tropical-japonica	

	G3296	WAB 706-3-4-K4-KB-1	Ivory Coast	AFRICARICE	tropical-japonica	
	G3297	WAB 758 1-1-HB-4	Ivory Coast	AFRICARICE	tropical-japonica	
	G3298	WAB 759-54-2-3-HB-2B	Ivory Coast	AFRICARICE	0	
	G3299	WAB 775-95-2-2-HB-1/CIRAD 409-3 1-2-5-3-1	Ivory Coast	AFRICARICE	tropical-japonica	
	G3300	WAB 788-18-2-2-HB-2/PCT-4\SA\1\1>721-M-2-M-4-M-2-M-5-M-1	Ivory Coast	AFRICARICE	tropical-japonica	
	G3301	WAB 878-6-12-1-1-P1-HB	Ivory Coast	AFRICARICE	tropical-japonica	1
	G3302	WAB 891SG26	Ivory Coast	AFRICARICE	tropical-japonica	
	G3303	WAB 891SG9	Ivory Coast	AFRICARICE	tropical-japonica	
	G3304	YANGKUM RED	Bhutan		temperate-japonica	
	G3305	yunlu 64	China	YAAS	tropical-japonica	
	G3306	yunlu 65	China	YAAS	tropical-japonica	
	G3307	YUNLU 7	China	YAAS	temperate-japonica	
	G3308	yunlu N°50	China	YAAS	0	
	G3309	Yunlu48	China	YAAS	tropical-japonica	
	B22	B22	Brazil	EMBRAPA	tropical-japonica	1
	C409	126-C409-8-1-2	Colombia	CIAT	tropical-japonica	
	CNA4136	CNA 4136	Brazil	EMBRAPA	tropical-japonica	
	F159	FOFIFA 159	Madagascar	FOFIFA-CIRAD	tropical-japonica	
	IAC25	IAC 25	Brazil	IAC	tropical-japonica	
	N10	NERICA 10	Ivory Coast	AFRICARICE	tropical-japonica	1
	N9	NERICA 9	Ivory Coast	AFRICARICE	tropical-japonica	
	PRIMA	PRIMAVERA	Brazil	IAC	tropical-japonica	1
	SEB402	SEBOTA 402	Brazil	CIRAD	tropical-japonica	
	CHHD	Chhomrong dhan	Nepal		tropical-japonica	
Breeding population (BP)	G2662		Madagascar	FOFIFA-CIRAD	0	
	G2664		Madagascar	FOFIFA-CIRAD	0	
	G2666		Madagascar	FOFIFA-CIRAD	0	
	G2667		Madagascar	FOFIFA-CIRAD	0	
	G2669		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2671		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2672		Madagascar	FOFIFA-CIRAD	0	
	G2674		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2677		Madagascar	FOFIFA-CIRAD	0	

	G2678		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2679		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2681		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2682		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2685		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2686		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2687		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2688		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2689		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2691		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2692		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2693		Madagascar	FOFIFA-CIRAD	0	
	G2696		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2699		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2700		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2702		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2703		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2704		Madagascar	FOFIFA-CIRAD	0	
	G2707		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2710		Madagascar	FOFIFA-CIRAD	0	
	G2712		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2715		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2717		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2719		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2720		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2722		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2723		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2726		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2729		Madagascar	FOFIFA-CIRAD	0	
	G2735		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2736		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2738		Madagascar	FOFIFA-CIRAD	tropical-japonica	

	G2739		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2740		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2742		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2743		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2744		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2747		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2748		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2751		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2752		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2753		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2754		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2755		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2757		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2758		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2761		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2764		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2765		Madagascar	FOFIFA-CIRAD	0	
	G2768		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2769		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2772		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2776		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2778		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2779		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2783		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2784		Madagascar	FOFIFA-CIRAD	0	
	G2795		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2797		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2799		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2801		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2802		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2803		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2809		Madagascar	FOFIFA-CIRAD	tropical-japonica	

	G2812		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2815		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2819		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2821		Madagascar	FOFIFA-CIRAD	0	
	G2825		Madagascar	FOFIFA-CIRAD	0	
	G2826		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2827		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2828		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2830		Madagascar	FOFIFA-CIRAD	0	
	G2831		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2833		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2835		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2838		Madagascar	FOFIFA-CIRAD	0	
	G2841		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2844		Madagascar	FOFIFA-CIRAD	0	
	G2847		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2849		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2851		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2854		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2855		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2856		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2857		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2859		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2860		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2861		Madagascar	FOFIFA-CIRAD	0	
	G2863		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2864		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2865		Madagascar	FOFIFA-CIRAD	0	
	G2866		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2870		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2871		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2872		Madagascar	FOFIFA-CIRAD	tropical-japonica	

	G2873		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2881		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2882		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2886		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2889		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2894		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2896		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2898		Madagascar	FOFIFA-CIRAD	0	
	G2900		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2901		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2902		Madagascar	FOFIFA-CIRAD	0	
	G2904		Madagascar	FOFIFA-CIRAD	0	
	G2905		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2906		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2909		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2910		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2911		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2912		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2914		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2918		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2921		Madagascar	FOFIFA-CIRAD	0	
	G2922		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2923		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2925		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2926		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2927		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2928		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2929		Madagascar	FOFIFA-CIRAD	0	
	G2932		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2935		Madagascar	FOFIFA-CIRAD	0	
	G2937		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2938		Madagascar	FOFIFA-CIRAD	tropical-japonica	

	G2940		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2942		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2944		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2946		Madagascar	FOFIFA-CIRAD	0	
	G2947		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2948		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2950		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2951		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2956		Madagascar	FOFIFA-CIRAD	0	
	G2957		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2959		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2960		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2964		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2965		Madagascar	FOFIFA-CIRAD	0	
	G2967		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2968		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2969		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2972		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2978		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2980		Madagascar	FOFIFA-CIRAD	0	
	G2983		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2986		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2989		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2992		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2995		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G2996		Madagascar	FOFIFA-CIRAD	0	
	G2999		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3000		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3002		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3003		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3015		Madagascar	FOFIFA-CIRAD	0	
	G3016		Madagascar	FOFIFA-CIRAD	tropical-japonica	

	G3023		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3028		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3029		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3032		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3039		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3045		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3046		Madagascar	FOFIFA-CIRAD	0	
	G3050		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3052		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3053		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3055		Madagascar	FOFIFA-CIRAD	0	
	G3056		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3059		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3064		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3068		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3069		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3071		Madagascar	FOFIFA-CIRAD	0	
	G3072		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3074		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3076		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3078		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3081		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3082		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3084		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3087		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3088		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3090		Madagascar	FOFIFA-CIRAD	tropical-japonica	
	G3092		Madagascar	FOFIFA-CIRAD	0	
	G3093		Madagascar	FOFIFA-CIRAD	tropical-japonica	

Supplemental Table 1: The 184 accessions in the diversity panel and 198 genotypes in breeding population with their main characteristics.