1   *G3: Genes, Genomes, Genetics*

2

3   # Genomic prediction accounting for genotype by environment
4   interaction offers an effective framework for breeding simultaneously
5   for adaptation to an abiotic stress and performance under normal
6   cropping conditions in rice

7

8   Ben Hassen M[†,‡ 1], Bartholomé J[†,‡ 1], Valè G[§], Cao TV[†,‡ 1], Ahmadi N[†,‡]

9   [†] CIRAD, UMR AGAP, F-34398 Montpellier, France.

10   [‡] AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

11   [§] CREA- Council for Agricultural Research and Economics, Research Center for Cereal and

12   Industrial Crops, S. S. 11 to Torino Km 2.5, Vercelli, 13100, Italy

13   [1] These authors contributed equally to the work

14

15

16

17   **Running title**: Genomic selection in rice

18   **Keywords**: Rice, genomic selection, progeny prediction, G×E interaction, alternate wetting and

19   drying (AWD).

20   **Corresponding author**: Nourollah Ahamadi

21   TA-A 108/01, Avenue Agropolis 34398 Montpellier Cedex 5, France

22   phone: +33 4 67 61 57 41 ; email: nourollah.ahmadi@cirad.fr

23

Abstract

Developing rice varieties adapted to alternate wetting and drying water management is crucial for the sustainability of irrigated rice cropping systems. Here we report the first study exploring the feasibility of breeding rice for adaptation to alternate wetting and drying using genomic prediction methods that account for genotype by environment interactions. Two breeding populations (a reference panel of 284 accessions and a progeny population of 97 advanced lines) were evaluated under alternate wetting and drying and continuous flooding management systems. The accuracy of genomic prediction for response variables (index of relative performance and the slope of the joint regression) and for multi-environment genomic prediction models were compared. For the three traits considered (days to flowering, panicle weight and nitrogen-balance index), significant genotype by environment interactions were observed in both populations. In cross validation, prediction accuracy for the index was on average lower (0.31) than that of the slope of the joint regression (0.64) whatever the trait considered. Similar results were found for across population validation (progeny validation). Both cross-validation and progeny validation experiments showed that the performance of multi-environment models predicting unobserved phenotypes of untested entrees was similar to the performance of single environment models with differences in accuracy ranging from -6% to 4% depending on the trait and on the statistical model concerned. The accuracy of multi-environment models predicting unobserved phenotypes of entrees evaluated under both water management systems outperformed single environment models by an average of 30%. Practical implications for breeding rice for adaptation to AWD are discussed.

Introduction

47

48   Rice is the world's most important staple food and will continue to be so in the coming

49   decades. In the future, the necessary increases in rice production to meet demand will have to

50   come mainly from an increase in yield per unit of land, water and other resources (CGIAR

51   Research Program on Rice 2016). At the same time, 15–20 million ha of rice lands will suffer

52   some degree of water scarcity (Tuong and Bouman 2003; Mekonnen and Hoekstra 2016). The

53   predicted increase in water scarcity threatens the sustainability of rice production (Rijsberman

54   2006). It is thus crucial to develop agronomic practices that reduce water use while maintaining

55   or increasing yields. A concomitant challenge is to adapt rice varieties to these water-saving

56   agronomic practices by improving their performance under water-limited conditions.

57   In recent decades, different water management systems have been developed with the aim

58   of reducing water consumption by irrigated rice (Tuong et al. 2005; Yang et al. 2007). Among

59   them, the alternate wetting and drying (AWD) system, in which paddy fields are subjected to

60   intermittent flooding with dry periods managed by soil water potential measurements, is one of

61   the most widely used (Linquist et al. 2015; Lampayan et al. 2015). A meta-analysis of 56 studies

62   comparing AWD with continuous flooding (CF) reported an overall decrease in yield of about

63   5% (Carrijo et al. 2017). However, marked variations were observed mainly depending on the

64   severity of the drying phase (i.e. the soil moisture at the end of each drying cycle) and on soil

65   characteristics (Lampayan et al. 2015; Carrijo et al. 2017). Significant differences in genotypic

66   responses to AWD, measured by changes in grain yield, have also been reported and attributed

67   to modified biomass partitioning (Bueno et al. 2010). Root architectural traits such as the

68   number of nodal roots and root dry weight at a depth of 10-20 cm 22-30 days after transplanting

69   also significantly contribute to yield stability under AWD (Sandhu et al. 2017). Genome wide

70   association analysis using a diversity panel revealed AWD-specific associations for several

71   agronomic traits including days to flowering, plant height, tillering, and panicle and seed

72   characteristics (Volante et al. 2017). Thus, rice adaptation to AWD appears to involve typical

73   complex traits, whose improvement requires genome-wide breeding approaches that account

74   for genotype by environment (G×E) interactions, i.e. the amplitude of the response of the

75   genotypes to a shift from CF management to the AWD system.

76   In plant breeding, G×E interactions are usually assessed in multi-environment trials and

77   expressed as a change in the relative performance of genotypes in different environments, with

78   or without change in the ranking of the genotypes (Freeman 1973). G×E analysis plays a

79   fundamental role in assessing genotype stability, in predicting the performance of untested

80   genotypes and in maximizing response to selection. Statistical methods for assessing G×E

81   interactions and estimating their sizes and opportunities to exploit them are widely discussed in

82   the literature (Freeman 1973; Cooper et al. 1993;; Malosetti et al. 2013; Elias et al. 2016; de

83   Leon et al. 2016). One of the earliest and most widely used methods is linear regression of the

84   performance (often of yield) of the individual genotype on the mean performances of all

85   genotypes evaluated in each test environment (Yates and Cochran 1938). The method, known

86   as *joint regression analysis*, was further formalized by Eberhart and Russel (1966) to enable

87   testing of the significance of deviation of individual regression from the general linear

88   component of G×E. Most evaluations of the effect of the environment on performance

89   undertaken for the purpose of plant breeding rely on multi-environmental field testing that

90   represents target production environments or a target population of environments (Cooper and

91   Hammer 1996). One specific case of G×E experiments is managed-environment trials that aim

92   to <u>assess</u> the effect of particular environmental variables (e.g., abiotic stresses) or cropping

93   practices (e.g. fertilizer, irrigation, etc.) that influence crop performance in the production

94   environment concerned (Cooper and Hammer 1996). A still more specific case of G×E

95   experiments is managed abiotic stress trials that aim to provide a measure of genotypic response

96   to stress based on yield loss under stress compared with under normal conditions. Several indexes

97   have been proposed to evaluate the stress intensity and genotypic response in such experiments,

98   mainly in the context of selection for drought tolerance (Fischer and Maurer 1978; Rosielle and

99   Hamblin 1981; Fischer et al. 2003).

100   With the advent of molecular markers, new G×E analysis methods have been developed

101   based on linear mixed models that connect the differential sensitivity of genotypes to

102   environments to particular regions of the plant genome and to specific biological mechanisms

103   (Malosetti et al. 2004; Boer et al. 2007; van Eeuwijk et al. 2010). More recently, the potential

104   of genomic selection (GS) to accelerate the pace of genetic gains in major field crops has

105   encouraged the development of multi-environment models for genomic prediction. The first

106   statistical framework using a linear mixed model to model G×E for the purpose of genomic

107   prediction was proposed by Burgueño et al. (2012). It extended the single-trait, single-

108   environment genomic best linear unbiased prediction (GBLUP) model to a multi-environment

109   context. Jarquín et al. (2014) proposed a method of modeling interactions between a high-

110   dimensional set of markers and environmental that incorporates genetic and environmental

111   gradients, as random linear functions (reaction norm) of markers and environmental covariates,

112   respectively. Lopez-Cruz et al. (2015) proposed a marker × environment interaction (M×E) GS

113   model that can be implemented using regression of phenotypes on markers or using co-variance

114   structures (a GBLUP-type model). Cuevas et al. (2016) further developed this approach by

115   using a non-linear (Gaussian) kernel to model the G×E: the reproducing kernel Hilbert space

116   with kernel averaging and the Gaussian kernel with the bandwidth estimated using an empirical

117   Bayesian method. Crossa et al. (2016) extended the M×E model using priors that produce

118   shrinkage (Bayesian ridge regression) or variable selection (BayesB), and reported better

119   prediction performances for these models compared to single environment and across-

120   environment models. The latest multi-environment genomic prediction models fall back on a

121   Bayesian approach (Cuevas et al. 2017). Application of these methods to one maize and four

122   wheat CIMMYT data sets showed that models with G×E always have higher prediction ability

123   than single-environment models, regardless of the genetic correlation between environments.

124   The predictive ability of these Bayesian methods was also generally better than that obtained

125   with the G×E models proposed by Lopez-Cruz et al. (2015) and Cuevas et al. (2016), when

126   applied to the same datasets.

127   In the present study, we evaluated the effect of AWD on the performance of two rice

128   breeding populations: a reference panel and a population of advanced lines both genotyped with

129   32 k SNP markers. Our general objective was to explore the feasibility of genomic selection for

130   the adaptation of rice to AWD in the framework of a pedigree breeding scheme. Our specific

131   objectives were to: (i) access expression of the response of the above-mentioned populations to

132   AWD compared to the CF irrigation system, and (ii) compare the performance of different

133   genomic prediction models that include G×E interactions in <u>answering the two well-known</u>

134   <u>issues relevant</u> in breeding programs: predicting unobserved phenotypes of untested lines and

135   predicting unobserved phenotypes of lines that have been evaluated in some environments but

136   not others. The two issues are analyzed in the context of intra-population prediction (cross-

137   validation experiments), and across-populations prediction (progeny-validation), as the

138   population of advanced lines was derived from biparental crosses between some of the members

139   of the diversity panel.

Material and method

140

Field trial and phenotyping

141

142   The plant material used in this study comprised a reference population (RP) of 284

143   accessions belonging to the rice *japonica* subspecies, and a progeny population (PP) of 97

144   advanced (F_5-F_7) inbred lines. The RP is representative of the working collection of the

145   Research Centre for Cereal and Industrial Crops (CREA), Vercelli, Italy. The PP was derived

146   from bi-parental crosses involving 31 accessions of RP, using a pedigree breeding scheme.

147   More information on the two populations is provided in Ben Hassen et al. (2017). The two

148   populations were phenotyped separately for two consecutive years at the experimental station

149   of the CREA (45°19'24.00"N; 8°22'26.28"E; 134 m asl.): in 2012 and 2013 for RP and in 2014

150   and 2015 for PP. In each year, the phenotyping experiment included two independent trials

151   corresponding to the two water management systems tested: CF and AWD. For the

152   conventional CF water management system, rice was dry seeded and the field was flooded with

153   10-15 cm of water at the 3-4 leaf stage (typically 30 days after sowing) and maintained flooded

154   until mid-maturity. For the AWD, after initial flooding at the 3-4 leaf stage, the field was

155   subjected to intermittent drying periods. The soil water potential was maintained above -30 kPa

156   by gravity irrigation whenever the soil moisture reached this threshold. The soil water potential

157   was monitored by a set of six tensiometers distributed throughout the field and inserted to a

158   depth of 20 cm. For each population and each year, the two water management systems were

159   conducted in two fields separated by a distance of about 100 m to avoid interference with

160   respect to the water regime. The other soil characteristics were identical (sand 47.8%, loam

161   42.8%, clay 9.4%; pH-H2O 6.4). The experimental design, which was identical in the two

162   conditions, was a complete randomized design with three replicates for RP and a complete

163   randomized block design with three replicates for the PP. The target traits for both RP and PP

164   were days to flowering (FL), panicle weight (PW), and the nitrogen balance index (NI) as

165   described in Ben Hassen et al. (2017).

166   ## Modeling of phenotypic data

167   Phenotypic data for each condition in the RP and the PP were analyzed using mixed models.

168   In order to identify possible outliers among individual data points, a diagnostic analysis based

169   on restricted likelihood distance was implemented, for details see Ben Hassen et al. (2017).

170   This analysis led to the elimination of one accession in the RP in AWD 2012 and 2013

171   experiments, one data point for FL in AWD-2012, and one data point for PW in AWD-2013.

172   The discarded data were considered as missing in the following steps of the analysis. The

173   following mixed models were applied to obtain adjusted means per genotype:

174    $$Y_{ijk}^m = \mu^m + y_i^m + g_j^m + gy_{ji}^m + \varepsilon_{ijk}^m \qquad \text{(RP model 1)}$$

175    $$Y_{ijkl}^m = \mu^m + y_i^m + yr_{ik}^m + g_j^m + gy_{ji}^m + \varepsilon_{ijkl}^m \quad \text{(PP model 1)}$$

176   where Y^m is the observed phenotype for the water management system m; μ^m the overall mean;

177   y^m the year as fixed effect; yr^m the within year replication as fixed effect; g^m the genotype as

178   random effect, gy^m the interaction between genotype and year as random effect; and ε^m the

179   residual. The analysis was performed with the *proc mixed* procedure of SAS 9.2 (SAS Institute,

180   Cary NC, USA); the method of estimation for the variance components was the restricted

181   maximum likelihood (REML). The formula by Holland et al. (2003) was used to estimate broad

182   sense heritability (H^2) as well as the corresponding standard error (SE) for each trait and each

183   water management system in each population:

184    $$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gy}^2}{ny} + \frac{\sigma_e^2}{nr}},$$

185   where σ_g^2, σ_{gy}^2 and σ_e^2 are the variance components associated with the genotype, the interaction

186   between genotype and year and the residual, respectively. ny is the harmonic mean of the

187   number of years per accession and nr, the harmonic mean of the number of plots across years

188   per accession. Conditional coefficients of determination (R²) were also computed using the

189   methodology described by Nakagawa and Schielzeth (2013). The adjusted means per water

190   management system ($\widehat{Y}_j^m = \hat{\mu}^m + \hat{g}_j^m + \widehat{gy}_{ji}^m$) extracted from the model were used as

191   phenotypes in the following steps.

192   For each trait, genetic correlations (r_G) between values measured under the two water

193   management systems were calculated (Cooper and DeLacy 1994; Cooper and Hammer 1996).

194   The confidence interval of r_G was obtained by using Fisher transformation of the estimated

195   correlation ($\hat{Z} = 0.5\,(\ln(1 + \hat{r}_G) - \ln(1 - \hat{r}_G))$, estimating the lower and upper bounds of $\hat{Z}$

196   ($Z_{1,2} = \hat{Z} \pm u_{1-\frac{\alpha}{2}}\,\sqrt{\frac{1}{N_P - 3}}$, with $\alpha = 0.05$, and $N_P = 284$ and 97, for RP and PP respectively),

197   and back transforming the $\hat{z}_1$ and $\hat{z}_2$ bounds into $\hat{r}_1$ and $\hat{r}_2$. The ratio of correlated response to

198   selection under continued flooding (CR_{CF}) and the direct response under alternate watering and

199   drying (DR_{AWD}) was calculated as:

200      $\dfrac{\mathrm{CR}_{CF}}{\mathrm{DR}_{AWD}} = r_G\,\sqrt{\dfrac{H^2_{CF}}{H^2_{AWD}}}$ (Falconer 1989) where r_G is the genotypic correlation defined above,

201   and H^2_{AWD} and H^2_{CF} represent the heritability of the trait under AWD and CF, respectively.

202   In addition to models for each condition, a model gathering data from the two water

203   management systems was also adjusted in order to test the significance of the interaction

204   between water management and genotypes:

205      $Y_{ijkl} = \mu + m_i + y_j + my_{ij} + g_k + gm_{ik} + gy_{jk} + gmy_{ijk} + \varepsilon_{ijkl}$ (RP model 2)

206      $Y_{ijkln} = \mu + m_i + y_j + my_{ij} + myr_{ijl} + g_k + gm_{ik} + gy_{jk} + gmy_{ijk} + \varepsilon_{ijkln}$ (PP model 2)

207   The same notation was used as for the model for each condition with additional fixed and

208   random effects: m the water management as fixed effect; my the water management and year

209   interaction as fixed effect; myr the replication within water management and year as fixed

210   effect; gm the interaction between genotype and water management as random effect; and gmy

211   the interaction between genotype, water management and year as random effect. The analyses

212   were performed with the *proc mixed* procedure of SAS 9.2 (SAS Institute, Cary NC, USA) with

213   REML.

214

Evaluation of genotypic response to water management systems

216 The genotypic response to AWD water management was estimated in two ways using

217 adjusted means. First, an index of relative performance was computed as follows:

218 $I_j = \frac{\hat{Y}_j^{AWD} - \hat{Y}_j^{CF}}{\hat{Y}_j^{CF}}$, where $\hat{Y}_j^{AWD}$ and $\hat{Y}_j^{CF}$ correspond to the adjusted means of accession j under

219 AWD and CF water managements, respectively. This index was also calculated at population

220 level to assess the intensity of stress caused by AWD water management compared to CF: $I =$

221 $\frac{\overline{\hat{Y}^{AWD}} - \overline{\hat{Y}^{CF}}}{\overline{\hat{Y}^{CF}}}$ were $\overline{\hat{Y}^{AWD}}$ and $\overline{\hat{Y}^{CF}}$ are the average performances of all genotypes within each

222 population under AWD and CF, respectively. Second, the slope β_j was computed as defined in

223 the joint regression equation: $\hat{Y}_j^m = \mu_j + \beta_j\,\theta_m + \varepsilon_{jm}$, where $\hat{Y}_j^m$ is the adjusted mean of the

224 genotype j in the water management m; θ_m is the environmental index calculated as the mean

225 value of all genotypes in water management m; μ_j is the intercept of the regression line of

226 genotype j; and ε_{jm} is the residual.

Genotypic data

228 The method used to genotype both RP and PP populations is detailed in Ben Hassen et al.

229 (2017). It relies on the genotyping by sequencing protocol developed by Elshire et al. (2011).

230 Sequencing was performed with a Genome Analyzer II (Illumina, Inc., San Diego, USA). The

231 different steps of analysis (raw data filtering, sequence alignment, SNP calling and imputation)

232 were performed with TASSEL v3.0 and the associated GBS pipeline (Glaubitz et al. 2014). A

233 working set of 32,066 SNPs was obtained with a heterozygosity rate < 5% and minor allele

234 frequency (MAF) > 5%.

Statistical models for genomic prediction

Single environment models

235

236

237   To predict the genomic estimated breeding values within each water management system,

238   hereafter referred to as single environment, two different kernel regression models were used.

239   The first model, which relies on a linear kernel, was the GBLUP as it is one of the most popular

240   methods for genomic prediction (Van Raden 2008). For this model, the kernel matrix (K) was

241   computed as $K = XX'$, X being the centered genotype matrix (-1, 0, 1) with N×P dimension,

242   where N is the number of genotypes and P the number of markers. The second model, which is

243   based on reproducing kernel Hilbert space (RKHS) approaches, used a Gaussian kernel

244     $K(x_i, x_j) = \exp(-h \parallel x_i - x_j \parallel^2)$ to build the kernel matrix between the marker genotype

245   vectors x_i and x_j, where $(i, j) \in \{1, \dots, N\}^2$. The bandwidth parameter h was estimated using

246   the method described by Pérez-Elizalde et al. (2015) based on a Bayesian method that relies on

247   the estimation of the mode of the joint posterior distribution of h and a form parameter φ. We

248   used the R function *margh.fun* provided by Pérez-Elizalde et al. (2015) with a gamma prior

249   distribution for h, with a shape parameter equal to 3, and a scale parameter equal to 1.5.

Multi-environment models

250

251   To predict the genomic estimated breeding values with data from the two water management

252   systems, hereafter referred to as multi-environment prediction, we used the statistical models

253   described above with extensions that integrate environmental effects. In the extended GBLUP

254   model, the effects of *m* environments, and the effects of the P markers are separated into two

255   components: the main effect of the markers for all the environments and the effect of the

256   markers for each environment (Lopez-Cruz et al. 2015). For RKHS, we used two extended

257   models incorporating G×E: RKHS-1 corresponding to the "Empirical Bayesian–Genotype ×

258   Environment Interaction Model" proposed by Cuevas et al. (2016), and RKHS-2 corresponding

259 to the environmental model (3) proposed by Cuevas et al. (2017). Like the extended GBLUP,

260 the first model (RKHS-1) considers the effects of m environments, and the effects of the

261 markers are separated into a main effect for all the environments and an effect specific to each

262 environment:

263 $$y = \mu + u_o + u_E + \varepsilon$$

264 In this mixed model, y is the response vector, μ is the overall intercept, u_o captures the

265 marker information among environments, and u_E accounts for the marker information in each

266 environment. The random effects u_o follow a multivariate normal distribution with mean zero

267 and a variance–covariance matrix $\sigma_{u_o}^2 K_0$, K_0 constructed with the Gaussian kernel from the

268 marker matrix X_0.

269 The latter model (RKHS-2), considers that the performances of accessions in different

270 environments are correlated such that there is a genetic correlation between environments that

271 can be modeled with matrices of order $m \times m$, m being the number of environments:

272 $$y = \mu + u + f + \varepsilon$$

273 In this mixed model, y is the response vector, μ is the vector with the intercept of each

274 environment, u the random vector of individual genetic values, f the genetic effects associated

275 with individuals that were not accounted for in component u, and ε the random vector of the

276 error. u, f and ε are independent and normally distributed. For more methodological details

277 concerning the extended GBLUP, RKHS-1 and RKHS-2 statistical models please refer to

278 Lopez-Cruz et al. (2015), Cuevas et al. (2016) and to Cuevas et al. (2017), respectively.

279 ## Implementation of the models

280 Analyses were performed in the R 3.4.2 environment (R Core Team 2017) with the R

281 packages *BGLR* 1.0.5 (Pérez and de los Campos 2014) and *MTM* 1.0.0 (De los Campos and

282   Grüneberg 2018). For both packages, 35,000 iterations for the Gibbs sampler were used. For

283   the inference, 3,000 samples were used after removing the first 5,000 samples (burn-in) and

284   keeping one in ten samples to avoid auto-correlation (thinning). Convergence of Markov chain

285   Monte Carlo algorithm was assessed for all parameters of the models with Gelman-Rubin tests

286   (Gelman and Rubin 1992) using the R-package *coda* 0.19-1 (Plummer et al. 2006).

Assessing genomic prediction accuracy

288   Prediction accuracy for the three traits and their related response to water management

289   (index and slope) were assessed with two different validation schemes. The first scheme used

290   only the RP with random partitions and is referred to hereafter as cross-validation. The second

291   validation scheme used information from the RP to predict the performance of the PP (referred

292   as progeny validation). The details of these two validation schemes are explained below.

Cross-validation within the reference population

294   Different types of random partitions were performed depending on the phenotypic and the

295   genotypic information used in the statistical model. For traits in a single environment and for

296   response variables, 80% of the 284 accessions (i.e. 227 accessions) of the RP were used as the

297   training set and the remaining 20% (57 accessions) was used as the validation set. For multi-

298   environment models, two different methods of cross-validation were applied. The first method

299   (M1), which resembled what was done in the single environment, used 80% of the observations

300   as a training set and the remaining 20% as the validation set and assumed that phenotypic

301   observations for the two environments are available for the individuals composing the training

302   set while no phenotypic data are available for the individuals in the validation set. M1

303   corresponds to the situation when the phenotypes of newly generated individuals have to be

304   predicted based only on their genotypic information (Burgueño et al., 2012). The second

305   method (M2) also used 80% of the observations as a training set and the remaining 20% as the

306   validation set but assumed that at least one observation in one environment was available for

307   the individuals in both the training set and the validation set. M2 corresponds to the situation

308   when phenotypes in one environment have to be predicted with genotypic information and

309   phenotypes from the other environment (Burgueño et al. 2012).

310   One hundred replicates were computed for all random partitioning in the training and

311   validation sets. The prediction accuracy of each partition was calculated as the Pearson

312   correlation coefficient between predictions and phenotypes in the validation set. For multi-

313   environment models, the correlation was calculated within each environment. For each trait

314   (FL, NI and PW) and each statistical model (GBLUP, RKHS-1 and RKHS-2), the same

315   partitions were used to compute the prediction accuracy. The resulting estimates of prediction

316   accuracy were averaged and the associated standard error was calculated.

317   To analyze the sources of variation of the prediction accuracy, the accuracy (r) of each

318   prediction experiment was transformed into a Z-statistic using the equation: $Z =$

319    $0.5\left[\ln(1 + r) - \ln(1 - r)\right]$ and used as a dependent variable in an analysis of variance. A

320   separate analysis was performed for each trait. After estimating the confidence limits and means

321   for Z, these were transformed back to r variables.

Progeny validation across populations

323   For progeny validation, the model was trained on the RP in order to predict the performance

324   of the PP based on genotypic information. Three validation scenarios were evaluated. In the

325   first scenario (S1) only the 31 parental lines were used as the training set. In the second scenario

326   (S2), the CDmean method (Rincent et al. 2012) was used to select 100 accessions in the RP for

327   the training set. In the third scenario (S3), all the RP accessions were included in the training

328   set. In all three scenarios, the validation set was made up of all the PP lines. Like for cross-

329 validation, prediction accuracy was calculated as the Pearson correlation coefficient between

330 predictions and phenotypes in the validation set.

Data Availability

332 The genotypic and phenotypic data are available in the TropGene database in the tab "Studies"

333 as "GS-Ruse": To access the TropGene database go to

334 http://tropgenedb.cirad.fr/tropgene/JSP/interface.jsp?module=RICE).

335

Results

Analysis of the phenotypic variations and responses to water management

338 The partitioning of the observed phenotypic variation into different sources of variation via

339 the mixed model analysis is shown in Table S1. Models were adjusted separately for each

340 population (RP and PP) and each water management system (CF and AWD). Conditional R^2

341 ranged from 0.33 to 0.96, indicating moderate to good fit of the model (Table 1). The lowest R^2

342 values were obtained for NI trait in both populations and both conditions. The highest R^2 values

343 were obtained for FL. Whatever the trait or water management system considered, the genotype

344 contributed significantly to the phenotypic variation in each population. A higher contribution

345 of the genotype effect to the phenotypic variation was observed for FL compared to NI and to

346 a lesser extent to PW. Broad-sense heritability (H^2) tended to confirm this trend (Table 1).

347 Indeed, depending on the population and the condition, H^2 ranged from 0.85 to 0.94 for FL,

348 from 0.75 to 0.90 for PW, and from 0.56 to 0.77 for NI. A slight increase in H^2 was observed

349 in CF water management compared to in AWD for FL and PW in RP. There was no significant

350 difference in PP.

351 The three traits investigated exhibited normal distribution in the RP and PP under both AWD

352 and FC (Figure 1). The AWD water management resulted in medium intensity stress for FL

353 (7.4% and 10.8% for RP and PP, respectively) and NI (-15.6% and -7.6%), and in rather severe

354 stress intensity for PW (-26.6% and -27.9%). On average, both populations flowered

355 significantly later under AWD than CF. The average FL values were 100.3 (102.8) in AWD

356 and 93.4 (92.9) in CF, for RP and (PP). Conversely, significantly lower NI and PW values were

357 observed in AWD compared to CF in both populations. For PW, the average differences

358 between the two water management systems were 89.4 g for RP and 77.7 g for PP. For NI, in

359 addition to differences in the average performance of the two water management systems,

360 significant differences in distribution were also observed between RP and PP, for the extent of

361 diversity, much larger for the RP, and for the frequency of individuals with low NI, much higher

362 in the PP (Figure S1).

363 Partitioning of the phenotypic variation from the two water management systems into

364 different sources of variation revealed the existence of significant interactions between

365 genotypes and water management systems in both RP and PP, for all traits except FL in RP

366 (Table S2). For all traits and populations, the ranking of the individuals was affected by water

367 management and the Spearman's rank correlation coefficients between traits values under the

368 two water management systems were medium to high (Figure 2). As a result, for each trait the

369 ratio of correlated response to selection under FC, relative to direct response to selection under

370 AWD, ranged from medium (0.56 and 0.75 for NI) to very high (0.98 and 0.90 for FL),

371 suggesting indirect selection for adaptation to AWD is feasible (Table 1).

372 The two computed variables (index and slope) characterizing the accessions' response to

373 AWD, revealed a Gaussian distribution for the three phenotypic traits considered (Figure S1).

374 An ANOVA of these computed variables revealed significant genotype effects on the three

375 traits in both RP and PP populations (Table S3). By construction, the correlations between

376   phenotypic values per condition and the slope were higher than those with the index whatever

377   the trait and the population considered. Interestingly, the index behaved differently in each trait

378   (Figure S1). For FL, low correlations were found either with AWD or CF. For NI, higher

379   correlations were found with CF (- 0.51 for RP and - 0.58 for PP) than with AWD (0.39 for RP

380   and 0.13 for PP). For PW, correlations were higher with AWD (0.42 for RP and 0.71 for PP)

381   than with CF (- 0.23 for RP and 0.24 for PP). For the three traits considered, there was almost

382   no correlation between the index and the slope variables (Figure S1): FL (0.12 for RP and 0.17

383   for PP), NI (0.-0.16 for RP and -0.31 for PP) and PW (-0.03 for RP and 0.04 for PP).

Accuracy of genomic prediction for the response variables

Prediction accuracy in the reference population

386   The average prediction accuracies obtained for the two response variables were compared

387   with those obtained for the observed variables in each water management system considered as

388   references (Table 2). The overall mean accuracy for the observed variables (the three traits

389   under the two water management systems), and for the response variables, was 0.54 but the

390   range extended from -0.12 to 0.88, depending on the prediction model, the trait and the type of

391   variable (**Erreur ! Source du renvoi introuvable.**, Table S4). The most significant factor

392   influencing accuracy was the type of variable (Table 2). Indeed, regardless of the trait or the

393   statistical model, accuracy for the index was lower than for the slope: 0.31 against versus 0.64

394   on average (**Erreur ! Source du renvoi introuvable.**). Interestingly, NI, which presented the

395   highest G×E, was the trait with the lowest accuracy for the index (0.17 and 0.21). However,

396   index predictions were less accurate for FL, the trait with the lowest G×E, (0.29 and 0.30) than

397   for PW (0.43 and 0.48) with intermediate G×E. In agreement with the medium to high

398   correlations at phenotypic level, the prediction accuracies for the slope and the variables under

399   each condition were comparable. However, different trends were observed depending on the

400   trait. For FL and PW, accuracies for the slope were closer to accuracies under AWD than under

401   CF. For NI, the opposite was observed. In all cases, slope prediction was as accurate as the best

402   single-environment prediction. The level of accuracy depended secondly on the trait (Table 2).

403   On average, accuracy was higher for FL (0.6) than for PW (0.58) and NI (0.45). The statistical

404   models differed significantly from each other although the effect was small. RKHS performed

405   better than GBLUP in almost all cases with differences in accuracy of up to 0.05. The

406   interactions between factors influencing prediction accuracy were not important, except for the

407   one between the response variable and the trait (Table 2).

Prediction accuracy across populations

409   On average, across generation prediction for both observed and computed response variables

410   was less accurate (0.28) than prediction within the reference population (Table S5). Accuracies

411   ranged from -0.01 to 0.38, with an average of 0.25 for FL, from -0.1 to 0.45, with an average

412   of 0.22, for NI, and from 0.14 to 0.56, with an average of 0.38 for PW, depending on the type

413   of variables (observed variables, index and slope), the scenario and the model (Figure 4).

414   Among these factors, the most influential was again the type of response (Table S5), with the

415   lowest average accuracy of 0.12 for index and the highest average accuracy of 0.35 for slope.

416   The prediction accuracy under the single environment AWD and CF averaged 0.34 and 0.32,

417   respectively. The effect of the scenario came in second, with an average accuracy of 0.27 for

418   S1, 0.22 for S2, and 0.35 for S3. The statistical models GBLUP and RKHS performed similarly

419   on average (accuracy of 0.28) but the range of variation was slightly wider in RKHS (-0.1 to

420   0.56) than in GBLUP (-0.01 to 0.51).

421   Accuracy of genomic prediction using multi-environment models

422   Prediction accuracy in the reference population

423   The focus here was on multi-environment models and the two different cross-validation

424   methods (M1 and M2), using single environment models as the baseline. Average accuracies

425   ranged from 0.47 to 0.96, depending on (in decreasing importance): the trait, the type of model

426   (i.e. single versus multi-environment), the cross-validation strategy, the statistical model and

427   the water management system (Figure 5, Table S6). The average accuracy was of 0.79, 0.56

428   and 0.69 for FL, NI and PW respectively. Whatever the trait or the water management system,

429   multi-environment models with the M1 strategy performed similarly to the single environment

430   model with a decrease of up to 0.02 for GBLUP and up to 0.03 for RKHS-1 and RKHS-2. As

431   expected, the multi-environment models with the M2 strategy outperformed single environment

432   models with an average gain of 0.23 and 0.27 for FL, 0.14 and 0.10 for NI and 0.20 and 0.20

433   for PW in AWD and CF, respectively. These gains in accuracy were in agreement with the level

434   of G×E found for each trait. Among the significant interactions between factors, the trait × cross

435   validation strategy interaction was the most important and corresponded to a scale interaction

436   (Table 3). Among the multi-environment prediction models, RKHS-1 and RKHS-2 performed

437   similarly, with average accuracy of 0.72 and 0.71, respectively, and performed systematically

438   slightly better than GBLUP, with a gain in accuracy of up to 0.04.

439   Prediction accuracy across populations

440   The overall mean accuracy was 0.33, with values ranging from -0.03 up to 0.58 (Figure 6,

441   Table S7), mainly depending on traits and scenarios for the composition of the training set. The

442   average accuracy was of 0.30, 0.27, and 0.44 for FL, NI and PW, respectively. The average

443   accuracy of the three scenarios was 0.32, 0.28 and 0.40 for S1, S2 and S3, respectively. The

444   range of variation in accuracy for the remaining factors (single versus multiple environment,

445 target environment and prediction model) did not exceed 0.03. These latter factors influenced

446 the accuracy mainly in interactive mode.

447

Discussion

Impact of AWD water management system on rice performance

450 The AWD water management implemented in this study (a new cycle of irrigation was

451 triggered when soil water potential reached -30 kPa) resulted in medium intensity stress for FL

452 and NI traits, rather severe stress intensity for PW when evaluated in terms of relative

453 performance. The effects of AWD we observed on PW (-27% on average), are similar to those

454 reported by Carrijo et al. (2017) on yield, in their review of 56 studies with 528 side-by-side

455 comparisons of yield under AWD and CF. These authors reported an average decrease in yields

456 of 5.4%, almost no yield losses under mild AWD (i.e. when soil water potential was kept ≥ -20

457 kPa), and yield losses of 22.6% relative to CF under severe AWD, when the soil water potential

458 went beyond -20 kPa. However, in contrast with our experiment, which pioneered the analysis

459 of genotypic responses to AWD within a diversity panel representing a large share of diversity

460 of one of the sub-species of rice (*O. sativa, japonica*), the majority of the studies included in

461 Carrijo et al.'s (2017) meta-analysis used only a small number of rice varieties and the crop was

462 established by transplanting. Among the few studies reporting on traits other than grain yield,

463 Sudhir et al. (2011) reported crop maturity delay of 5-10% under severe AWD, similar to our

464 results (9% on average).

Genomic prediction of response to AWD

466 The two computed variables (response index and slope of the joint regression) were intended

467 to provide a measurement of G×E for each accession of RP and PP, which could be used as the

468   entry phenotype for genomic prediction. The index, which evaluates tolerance to AWD water

469   management, was very closely correlated with the stress sensitivity and tolerance index

470   proposed by Fischer and Maurer (1978) and (Rosielle and Hamblin 1981), respectively (data

471   not shown). The slope provides a measurement of stability of breeding material along

472   environmental gradients in multi-environment trials (Eberhart and Russell 1966; Lin et al.

473   1986). However, the fact that the environmental index is not independent of the performances

474   of the studied genotypes can introduce a bias in the estimate of the regression parameters

475   (Crossa 1990). Moreover, the percentage of G×E variance explained is often very low, below

476   25% (for a review, see Brancourt-Hulmel et al. 1997). In our case, the number of environments

477   considered, two, was probably too few for a precise estimate of the regression slope for each

478   genotype. On the other hand, the large number of genotypes involved in the estimate of the

479   environmental index (284 for RP and 97 for PP) limited the above-mentioned risk of bias. Given

480   the very high correlations between the computed slopes and the measured phenotypes for the

481   three traits under AWD and CF in both RP and PP populations ($r > 0.9$, except for PW under

482   AWD in PP ($r = 0.73$), it represents a reasonably good single entry phenotype to consider for

483   breeding both for adaptation to AWD and performance under CF.

484   The accuracy of genomic prediction for the response index was significantly lower than for

485   the slope and for the corresponding measured traits under AWD and CF, suggesting limited

486   genetic control of variation for the response index. Similar results were reported by (Huang et

487   al. 2016) for trait stability in wheat. Nevertheless, given the loose correlations between the

488   response index and the measured traits, genomic prediction for the index and the measured trait

489   in CF could be used to select for good performance in both systems.

Genomic prediction using multi-environment data

491 The potential of GS to accelerate the pace of genetic gains in major field crops has been

492 documented by a large number of studies using a simulation approach or experimental data

493 (Crossa et al. 2017; Hickey et al. 2017). In the case of rice, several empirical studies,

494 summarized in Ben Hassen et al. (2017), confirmed this potential. However, the focus of most

495 previous crop genomic prediction studies was on within-environment prediction, based on

496 single environment models. It was recently demonstrated that the accuracy of genomic

497 prediction models that account for G×E is significantly greater than that attained by single

498 environment models (Cuevas et al. 2016; Cuevas et al. 2017; Burgueño et al. 2012; Jarquín et

499 al. 2014; Lopez-Cruz et al. 2015; Heslot et al. 2014). The empirical component of almost all of

500 these studies was based on data from unmanaged multi-environment trials of genotypes across

501 several locations (and often several years), mainly conducted to study G×E and the general

502 stability of the genotype across environments. The multi-environment genomic prediction

503 results we present here stand out among the aforementioned ones because we used data from

504 managed bi-environment trials undertaken to study G×E and genotype adaptation to a specific

505 abiotic constraint, i.e. AWD water management.

506 The level of prediction accuracy obtained in our cross validation experiments in the reference

507 population under the M1 prediction strategy with the multi-environment GBLUP, RKHS-1 and

508 RKHS-2 models, calibrated with data from both AWD and CF water management, was similar

509 to that obtained with their single environment counterparts, calibrated with data from either

510 AWD or CF. These results confirm the power of multi-environment genomic models to predict

511 the performances of untested genotypes using data from multiple trials. Under the M2

512 prediction strategy, the three multi-environment models provided significantly higher

513 prediction accuracy for genotypes that had not been tested in one of the two water management

514 systems than their single-environment counterparts, further confirming the advantages of multi-

515   environment prediction models. In order to challenge the performance of the multi-environment

516   models further, we ran the M1 and M2 strategies with a larger number of untested entrees (40%

517   instead of 20%) in both AWD and CF for M1, in AWD or CF for M2. The results in Figure S2

518   show a very small reduction in prediction accuracy. The average prediction accuracy for the

519   three traits, the two water managements and the three prediction models was 0.59 instead of

520   0.61 for M1, and 0.79 instead of 0.81 for M2. These results suggest the possibility of optimizing

521   the method of evaluation of the lines by targeting a specific set of lines for each condition

522   (Rincent et al. 2017).

523   Lopez-Cruz et al. (2015) reported gains in prediction accuracy of up to 30% with the

524   GBLUP-type multi-environment model compared to an across-environment analysis that

525   ignores G×E, when applied to the wheat grain yield of three sets of advanced lines recorded in

526   three different years under three irrigation regimes. In our case, significant gains in accuracy

527   were observed only with the M2 strategy, and ranged from 17% for NI to 29% for FL. Using

528   wheat and maize data, Cuevas et al. (2016) reported up to 68% higher accuracy for RKHS-1

529   models compared to single environment models and up to 17% compared to GBLUP-G×E.

530   These authors hypothesized that the superiority of the Gaussian kernel models over the linear

531   kernel was due to more flexible kernels that account for small, more complex marker main

532   effects and marker specific interaction effects. In our experiments, RKHS-1 was up to 35%

533   more accurate than single environment GBLUP and up to 10% more accurate than GBLUP-

534   G×E model. On the other hand, we did not observe any notable differences in the prediction

535   accuracy of the RKHS-2 model compared to GBLUP-G×E and RKHS-1, as already reported

536   by Cuevas et al. (2017). This is probably due to the positive correlation between performances

537   under AWD and CF water management systems in our experiments, while the most favorable

538   context for the application the approach developed by Cuevas et al. (2017) is said to be when

539   different types of correlation (positive, zero, or negative) between the environments considered,

540   coexist.

541   The results of our progeny validation experiments did not question the higher prediction

542   accuracy of multi-environment models compared to single environment ones observed in our

543   cross validation experiments in the reference population. However, in progeny validation

544   experiments, the multi-environment models affected prediction accuracy mainly in interaction

545   with other factors, such as the composition of the training set and the trait considered. These

546   results also confirmed the important role of relatedness between the training and the validation

547   set in prediction accuracy. It also confirmed the fact that relatively high accuracy could be

548   achieved using only a rather small share of the RP, the most closely related to the PP as the

549   training set, as reported by Ben Hassen et al. (2017).

550   Finally yet importantly, in both cross validation and progeny validation experiments, the

551   multi-environment approach achieved higher prediction accuracy than the genomic prediction

552   for the response index and the slope of the joint regression. For instance, compared to prediction

553   for slope, the mean advantage of multi-environment prediction was 8% and 10% with GBLUP-

554   G×E and RKHS-1 models, respectively. The advantage reached 25% under the M2 strategy of

555   predicting unobserved phenotypes. In the progeny-validation experiments, the mean advantage

556   was 20% and reached 30% under the S2 scenario of composition of the training set. To our

557   knowledge, this finding has not yet been reported in the literature. It opens new perspectives in

558   breeding for adaptation to AWD and to other abiotic stresses.

559   ## Practical implications for breeding rice for adaptation to AWD

560   "More rice with less water" is vital for food security and for the sustainability of irrigated

561   rice cropping systems (Tuong et al. 2005). AWD water management is one of the most widely

562   used water-saving techniques practiced today (Carrijo et al. 2017). The development of rice

563   varieties adapted to AWD, i.e. with as high yields as the best high yielding variety under CF,

564   would greatly contribute to wider adoption of AWD water management by farmers (Price et al.

565   2013; Volante et al. 2017). Given the genetic diversity we observed for response to AWD within

566   the working collection of the CREA, which represents only a share of the genetic diversity of

567   the rice *japonica* sub-species, one can expect large genetic diversity at the whole species level.

568   The almost identical and high level of broad-sense heritability observed under AWD and CF

569   water management systems demonstrates the feasibility of direct selection for AWD. Such high

570   heritability under managed abiotic stress has already been reported in rice for grain yield under

571   drought stress (Venuprasad et al. 2007; Kumar et al. 2008). However, the adoption of the direct

572   selection option may not be practicable for breeding programs with limited resources, if they

573   also need to continue to breed for CF water management. Moreover, this option would not take

574   full advantage of historical data produced by the breeding program for CF. The high accuracy

575   of multi-environment genomic prediction we observed in the present study, especially in across-

576   environment prediction, paves the way for a new breeding option: conducting simultaneously

577   direct and indirect selection for both AWD and CF. Indeed, as we saw in our M2 strategy, the

578   multi-environment genomic models can boost the predictive power of across-environment

579   predictions, i.e. from CF to AWD and *vice versa*. In this context, the practical question would

580   be the number of selection candidates that need to be phenotyped under the two water

581   management systems relative to the number of candidates that need to be phenotyped under one

582   water management system only. Our results suggest that, for the germplasm and environmental

583   conditions we used and the traits we considered, the percentage of untested candidates under

584   AWD can go up to 40% with no significant negative effect on prediction accuracy as long as

585   they are tested under CF, or *vice versa*. Considering the additional cost reductions that could be

586   obtained by optimizing the size of the training set, as shown by the S1 scenario in our across-

587   generations prediction experiments, it seems possible to add the objective of adaptation to AWD

588 to an existing GS based rice breeding program for CF, with rather limited additional costs. Ben

589 Hassen et al. (2017) showed that rice breeding programs based on pedigree schemes can use a

590 genomic model trained with data from their working collection to predict performances of

591 progenies produced by the conventional pedigree breeding program. Breeding for adaptation to

592 AWD can be integrated in this general scheme. The feasibility of application of this breeding

593 approach to other abiotic stresses deserves further exploration.

594

Authors' contributions

596   NA and GV conceived the study. JB, MBH and TVC analyzed the data. MBH, TVC, JB and

597   NA wrote the manuscript.

598

References

Ben Hassen, M., T.V. Cao, J. Bartholomé, G. Orasen, C. Colombi *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*.

Boer, M.P., D. Wright, L. Feng, D.W. Podlich, L. Luo *et al.*, 2007 A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, With an Example in Maize. *Genetics* 177 (3):1801-1813.

Brancourt-Hulmel, M., V. Biarnès-Dumoulin, and J. Denis, 1997 Points de repère dans l'analyse de la stabilité et de l'interaction génotype-milieu en amélioration des plantes. *Agronomie* 17 (4):219-246.

Bueno, C.S., E. Pasuquin, B. Tubaña, and T. Lafarge, 2010 Improving sink regulation, and searching for promising traits associated with hybrids, as a key avenue to increase yield potential of the rice crop in the tropics. *Field Crops Research* 118 (3):199-207.

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa, 2012 Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Science* 52 (2):707-719.

Carrijo, D.R., M.E. Lundy, and B.A. Linquist, 2017 Rice yields and water use under alternate wetting and drying irrigation: A meta-analysis. *Field Crops Research* 203:173-180.

CGIAR Research Program on Rice, 2016 Annual Report 2016. International Rice Research Institute (IRRI), Los Banos, Philippines.

Cooper, M., I. DeLacy, and R. Eisemann, 1993 Recent advances in the study of genotype× environment interactions and their application to plant breeding, pp. 116-131 in *Focused Plant Improvement: Towards Responsible and Sustainable Agriculture. Australian Plant Breeding Conference Gold Coast, Queensland*.

Cooper, M., and I.H. DeLacy, 1994 Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* 88 (5):561-572.

Cooper, M., and G.L. Hammer, 1996 *Plant adaptation and crop improvement*. Wallingford: CAB INTERNATIONAL.

Crossa, J., 1990 Statistical Analyses of Multilocation Trials, pp. 55-85 in *Advances in Agronomy*, edited by N.C. Brady. Academic Press.

Crossa, J., G. de los Campos, M. Maccaferri, R. Tuberosa, J. Burgueño *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.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín *et al.*, 2017 Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *TRENDS in Plant Science* 22 (11):961-975.

Cuevas, J., J. Crossa, O.A. Montesinos-López, J. Burgueño, P. Pérez-Rodríguez *et al.*, 2017 Bayesian Genomic Prediction with Genotype × Environment Interaction Kernel Models. *G3: Genes|Genomes|Genetics* 7 (1):41-53.

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).

de Leon, N., J.-L. Jannink, J.W. Edwards, and S.M. Kaeppler, 2016 Introduction to a special issue on genotype by environment interaction. *Crop Science* 56 (5):2081-2089.

De los Campos, G., and A. Grüneberg, 2018 MTM (Multiple-Trait Model) package.

Eberhart, S.A., and W.A. Russell, 1966 Stability Parameters for Comparing Varieties. *Crop Science* 6 (1):36-40.

Elias, A.A., K.R. Robbins, R.W. Doerge, and M.R. Tuinstra, 2016 Half a Century of Studying Genotype × Environment Interactions in Plant Breeding Experiments. *Crop Science* 56 (5):2090-2105.

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.

Falconer, D.S., 1989 *Introduction to quantitative genetics*. Burnt Mill, Harlow, Essex, England; New York: Longman, Scientific & Technical ; Wiley.

Fischer, K.S., R. Lafitte, S. Fukai, G. Atlin, and B. Hardy, 2003 *Breeding rice for drought-prone environments*. Los Baños (Philippines): International Rice Research Institute.

Fischer, R., and R. Maurer, 1978 Drought resistance in spring wheat cultivars. I. Grain yield responses. *Australian Journal of Agricultural Research* 29 (5):897-912.

Freeman, G.H., 1973 Statistical methods for the analysis of genotype-environment interactions. *Heredity* 31 (3):339-354.

Gelman, A., and D.B. Rubin, 1992 Inference from iterative simulation using multiple sequences. *Statistical science*:457-472.

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.

Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink, 2014 Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics* 127 (2):463-480.

Hickey, J.M., T. Chiurugwi, I. Mackay, W. Powell, and C.B.P.W.P. Implementing Genomic Selection in, 2017 Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* 49 (9):1297-1303.

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

Huang, M., A. Cabrera, A. Hoffstetter, C. Griffey, D. Van Sanford *et al.*, 2016 Genomic selection for wheat traits and trait stability. *Theoretical and Applied Genetics* 129 (9):1697-1710.

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. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 127 (3):595-607.

Kumar, A., J. Bernier, S. Verulkar, H.R. Lafitte, and G.N. Atlin, 2008 Breeding for drought tolerance: direct selection for yield, response to selection and use of drought-tolerant donors in upland and lowland-adapted populations. *Field Crops Res* 107.

Lampayan, R.M., R.M. Rejesus, G.R. Singleton, and B.A.M. Bouman, 2015 Adoption and economics of alternate wetting and drying water management for irrigated lowland rice. *Field Crops Research* 170:95-108.

Lin, C.S., M.R. Binns, and L.P. Lefkovitch, 1986 Stability Analysis: Where Do We Stand? *Crop Science* 26 (5):894-900.

Linquist, B.A., M.M. Anders, M.A.A. Adviento-Borbe, R.L. Chaney, L.L. Nalley *et al.*, 2015 Reducing greenhouse gas emissions, water use, and grain arsenic levels in rice systems. *Global Change Biology* 21 (1):407-417.

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: Genes|Genomes|Genetics* 5 (4):569-582.

Malosetti, M., J.-M. Ribaut, and F.A. van Eeuwijk, 2013 The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology* 4.

Malosetti, M., J. Voltas, I. Romagosa, S.E. Ullrich, and F.A.V. Eeuwijk, 2004 Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137:139-145.

Mekonnen, M.M., and A.Y. Hoekstra, 2016 Four billion people facing severe water scarcity. *Science Advances* 2.

Nakagawa, S., and H. Schielzeth, 2013 A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2):133-142.

Pérez-Elizalde, S., J. Cuevas, P. Pérez-Rodríguez, and J. Crossa, 2015 Selection of the Bandwidth Parameter in a Bayesian Kernel Regression Model for Genomic-Enabled Prediction. *Journal of Agricultural, Biological, and Environmental Statistics* 20 (4):512-532.

704 Pérez, P., and G. de los Campos, 2014 Genome-Wide Regression and Prediction with the BGLR
705   Statistical Package. *Genetics* 198 (2):483-495.
706 Plummer, M., N. Best, K. Cowles, and K. Vines, 2006 CODA: convergence diagnosis and output
707   analysis for MCMC. *R News* 6:7-11.
708 Price, A.H., G.J. Norton, D.E. Salt, O. Ebenhoeh, A.A. Meharg *et al.*, 2013 Alternate wetting and drying
709   irrigation for rice in Bangladesh: Is it sustainable and has plant breeding something to offer?
710   *Food and Energy Security* 2 (2):120-129.
711 R Core Team, 2017 R: A Language and Environment for Statistical Computing. URL http://www.R-
712   project.org/, R Foundation for Statistical Computing, Vienna, Austria.
713 Rijsberman, F.R., 2006 Water scarcity: Fact or fiction? *Special Issue on Water Scarcity: Challenges and*
714   *Opportunities for Crop Science* 80:5-22.
715 Rincent, R., E. Kuhn, H. Monod, F.-X. Oury, M. Rousset *et al.*, 2017 Optimization of multi-environment
716   trials for genomic selection based on crop models. *Theoretical and Applied Genetics*.
717 Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel *et al.*, 2012 Maximizing the Reliability of
718   Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of
719   Methods in Two Diverse Groups of Maize Inbreds (Zea mays L.). *Genetics* 192 (2):715-728.
720 Rosielle, A.A., and J. Hamblin, 1981 Theoretical Aspects of Selection for Yield in Stress and Non-Stress
721   Environment. *Crop Science* 21 (6):943-946.
722 Sandhu, N., S.R. Subedi, R.B. Yadaw, B. Chaudhary, H. Prasai *et al.*, 2017 Root Traits Enhancing Rice
723   Grain Yield under Alternate Wetting and Drying Condition. *Frontiers in Plant Science* 8 (1879).
724 Sudhir, Y., G. Gill, E. Humphreys, S.S. Kukal, and U.S. Walia, 2011 Effect of water management on
725   dry seeded and puddled transplanted rice. Part 1: Crop performance. *Field Crops Research* 120
726   (1):112-122.
727 Tuong, P., B.A.M. Bouman, and M. Mortimer, 2005 More Rice, Less Water—Integrated Approaches
728   for Increasing Water Productivity in Irrigated Rice-Based Systems in Asia. *Plant Production*
729   *Science* 8 (3):231-241.
730 Tuong, T.P., and B.A.M. Bouman, 2003 Rice Production in Water-scarce Environments, pp. 368 in
731   *Water Productivity in Agriculture:Limits and Opportunities for Improvements*.
732 van Eeuwijk, F.A., M.C.A.M. Bink, K. Chenu, and S.C. Chapman, 2010 Detection and use of QTL for
733   complex traits in multiple environments. *Current opinion in plant biology* 13 (2):193-205.
734 Van Raden, P.M., 2008 Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*
735   91 (11):4414-4423.
736 Venuprasad, R., N. Zenna, R.I. Choi, M. Amante, P.S. Virk *et al.*, 2007 *Identification of marker loci*
737   *associated with tungro and drought tolerance in near-isogenic rice lines derived from*
738   *IR64/Aday Sel. IRRN 32.1.*
739 Volante, A., F. Desiderio, A. Tondelli, R. Perrini, G. Orasen *et al.*, 2017 Genome-Wide Analysis of
740   japonica Rice Performance under Limited Water and Permanent Flooding Conditions. *Frontiers*
741   *in Plant Science* 8 (1862).
742 Yang, J., K. Liu, Z. Wang, Y. Du, and J. Zhang, 2007 Water-saving and high-yielding irrigation for
743   lowland rice by controlling limiting values of soil water potential. *Journal of Integrative Plant*
744   *Biology* 49 (10):1445-1454.

745

746

Acknowledgments

Conflict of Interest

751

752   The authors declare that they have no conflict of interest.

753

Acknowledgments

Supplementary figures and tables

Figure S1: Correlation matrix between performance in each condition (AWD: alternate wetting and drying and CF: continuous flooding) and response variables (response index and slope of the joint regression) for the three traits considered: days to flowering (FL), nitrogen-balance index (NI), and panicle weight (PW). The reference (RP) and progeny (PP) populations are in green and grey, respectively.

Figure S2: Single environment and multi-environment (M1 and M2) prediction accuracies in cross validation experiments with 40% of untested entries in the reference population obtained with three statistical models (GBLUP, RKHS-1, RKHS-2). Continuous flooding and alternate wetting and drying water management conditions are in blue and orange, respectively. Three traits are presented: days to flowering (FL), nitrogen balance index (NI) panicle weight (PW). The letters in each panel represent the results of Tukey's HSD comparison of means and apply to each panel independently. The means differ significantly (p-value < 0.05) if two boxplots have no letter in common.

Table S1: Variance components and the associated statistic (F-value for fixed effects and Z-value for random effects) of days to flowering (FL), nitrogen balance index (NI), and panicle weight (PW). Separate analysis of each population and each water management system (alternate wetting and drying – AWD and continuous flooding – CF).

Table S2: Variance components and the associated statistic: F-value for fixed effects and Z-value for random effects) of days to flowering (FL), nitrogen balance index (NI), and panicle weight (PW). Separate analysis of each population pooled over water management conditions.

Table S3: Variance components for the joint regression for days to flowering (FL), nitrogen balance index (NI), and panicle weight (PW). Results are shown for the reference and progeny populations.

Table S4: Mean genomic prediction accuracies in the reference population for the response variables (index and slope) and the performance within each condition (AWD and CF). The results for days to flowering (FL), nitrogen balance index (NI) and panicle weight (PW) are presented. Two statistical models (GBLUP and RKHS) were used.

Table S5: Genomic prediction accuracies for across population validation for the response variables (index and slope) and the performance within each condition (AWD and CF). The scenarios used to define the training set are S1 (only the parents), S2 (100 individuals of the RP selected with CDmean) and S3 (the whole RP). Results for days to flowering (FL), nitrogen balance index (NI) and panicles weight (PW) are presented. Two statistical models (GBLUP and RKHS) were used.

789   **Table S6:** Mean genomic prediction accuracy of the performance within each condition (AWD
790   and CF) using single or multi-environment models in the reference population. For multi-
791   environment models, two methods of cross-validation were used: M1 and M2. Results for days
792   to flowering (FL), nitrogen balance index (NI) and panicle weight (PW) are presented. Two
793   statistical models (GBLUP, RKHS) were used in single environment prediction and three
794   (GBLUP, RKHS-1 and RKHS-2) in multi-environment prediction.

795   **Table S7:** Genomic prediction accuracies of the performance within each condition (AWD and
796   CF) using single or multi-environment models for across population validation. The scenarios
797   used to define the training set are S1 (only the parents), S2 (100 individuals of the RP selected
798   with CDmean) and S3 (the whole RP). Results for days to flowering (FL), nitrogen balance
799   index (NI) and panicle weight (PW) are presented. Two statistical models (GBLUP, RKHS)
800   were used in single environment prediction and three (GBLUP, RKHS-1 and RKHS-2) in multi-
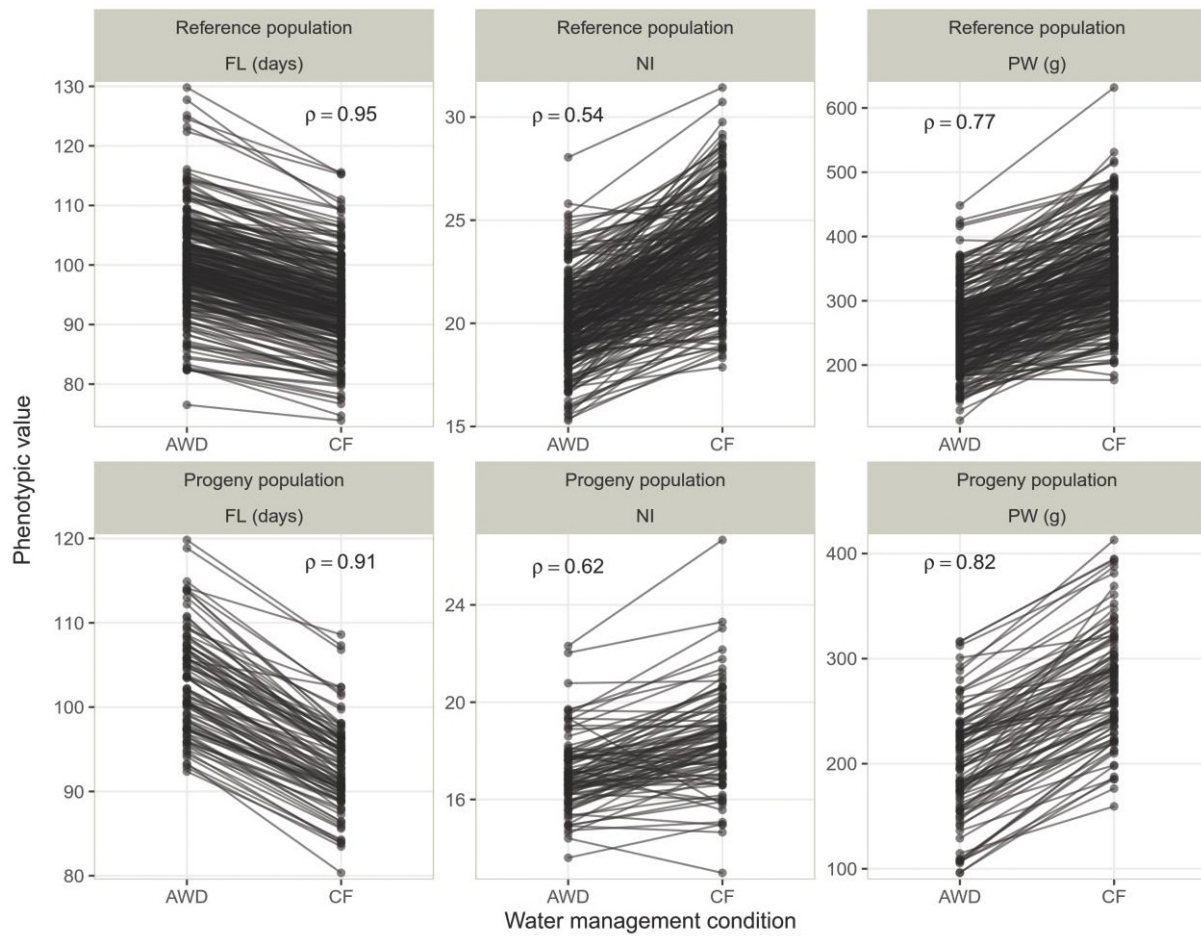801   environment prediction.

Figures and Tables

802
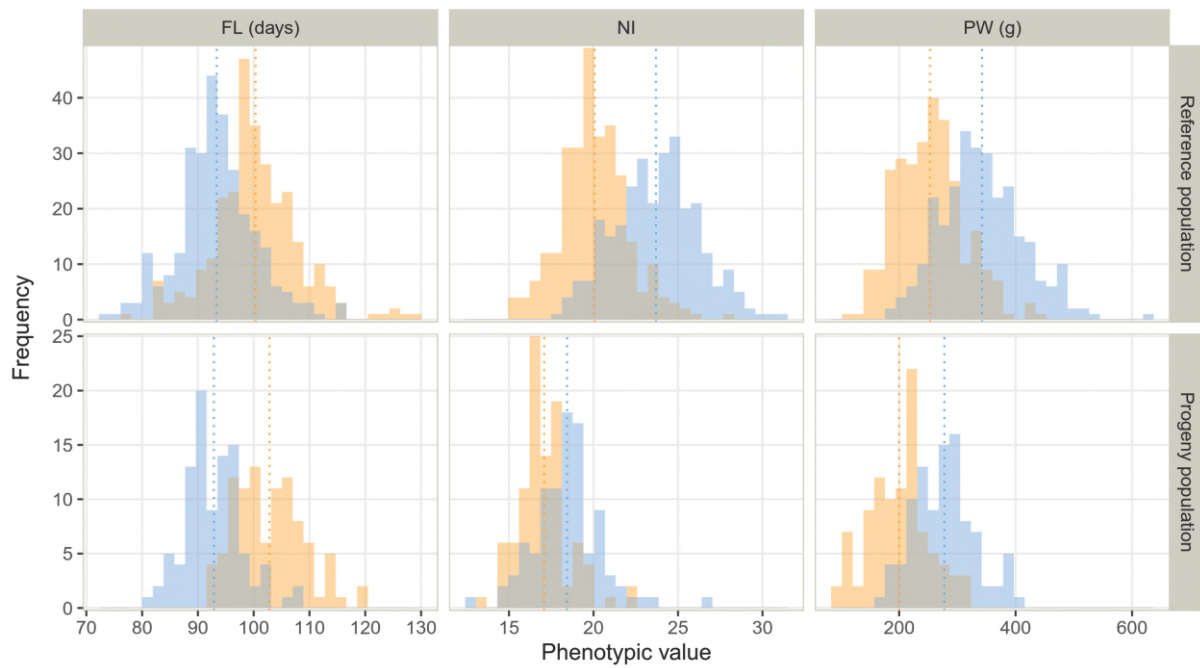
803

804 **Figure 1**: Distribution of adjusted phenotypic values of days to flowering (FL), nitrogen
805 balance index (NI) and panicle weight (PW) within the reference and progeny populations in
806 continuous flooding (blue) and alternate wetting and drying (orange) conditions.

807

**Figure 2**: Reaction norm between the two conditions (AWD: alternate wetting and drying and CF: continuous flooding) for all the genotypes of the two populations (the reference population and the progeny population).The three traits are represented: days to flowering (FL), nitrogen balance index (NI) and panicle weight (PW). Spearman's rank correlation coefficient (ρ) is indicated in each panel.

815
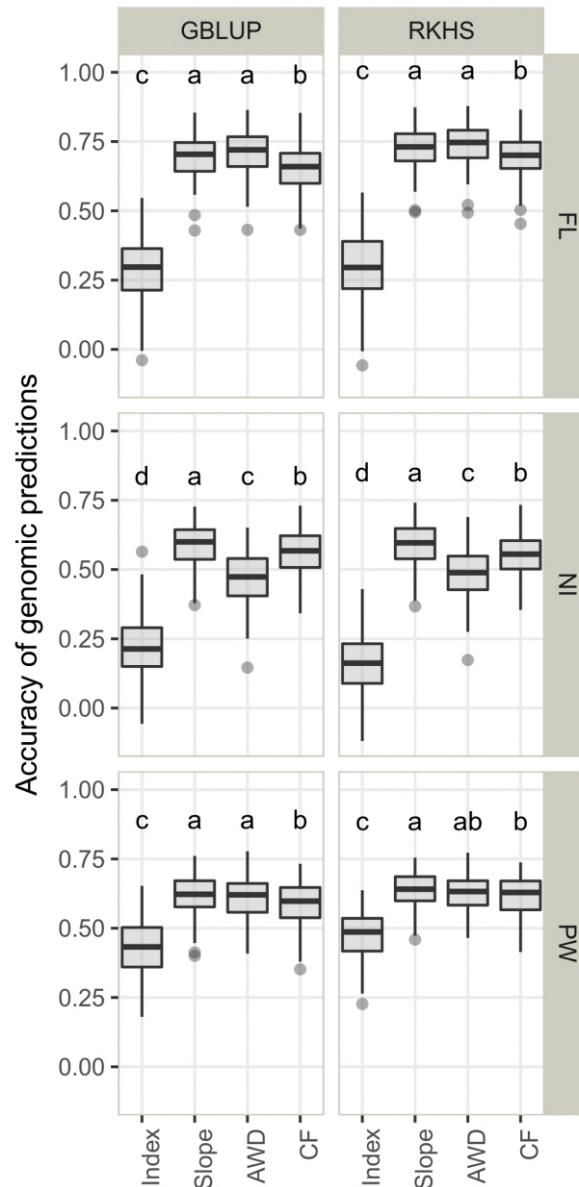
816

817 **Figure 3**: Accuracy of genomic prediction in cross validation experiments within the reference
818 population obtained with two statistical models (GBLUP, RKHS) for the response variables
819 (index and slope) and the performance within each condition (AWD and CF). The three traits
820 are presented: days to flowering (FL), nitrogen balance index (NI) and panicle weight (PW).
821 The letters in each panel represent the results of Tukey's HSD comparison of means and apply
822 to each panel independently. The means differ significantly (p-value < 0.05) if two boxplots
823 have no letter in common.

824
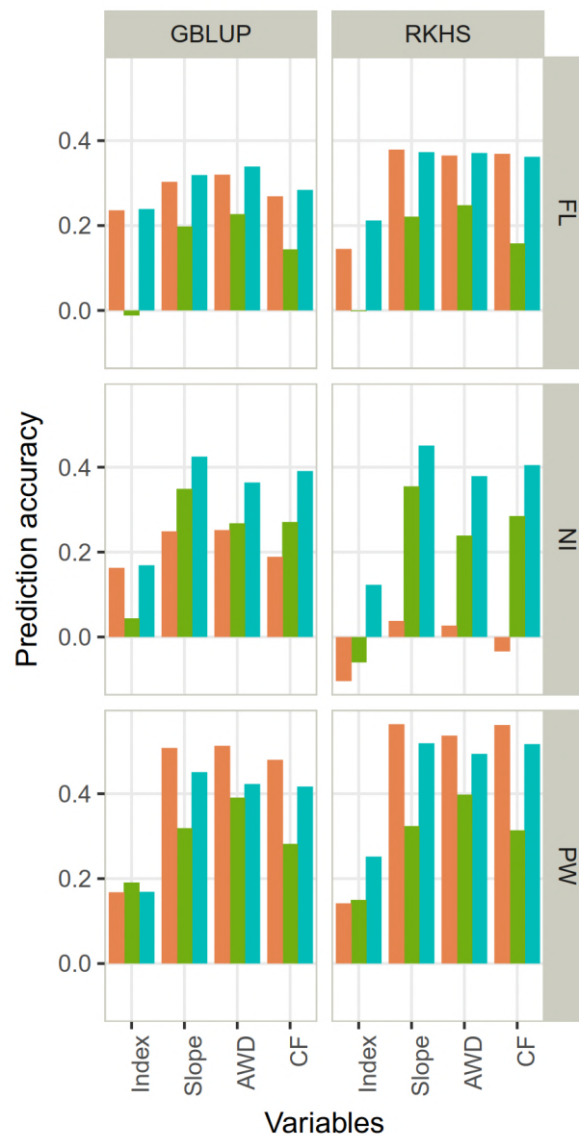
Figure 4: Accuracy of genomic prediction in across population validation for the response variables (index and slope) and the performance within each condition (AWD and CF) obtained. Two statistical models (GBLUP, RKHS) and three traits (days to flowering (FL), nitrogen balance index (NI) and 100 panicle weight (PW)) were studied. The scenarios used to define the training set are in color: orange (S1: only the parents), green (S2: 100 individuals of the RP selected with CDmean) and blue (S3: the whole RP).

834

835

Figure 5: Single environment and multi-environment (M1 and M2) prediction accuracy in cross validation experiments in the reference population obtained with three statistical models (GBLUP, RKHS-1, RKHS-2). Continuous flooding and alternate wetting and drying water management conditions are in blue and orange, respectively. The three studied traits are presented: days to flowering (FL), nitrogen balance index (NI) and panicle weight (PW). The letters in each panel represent the results of Tukey's HSD comparison of means and apply to each panel independently. The means differ significantly (p-value < 0.05) if two boxplots have no letter in common.
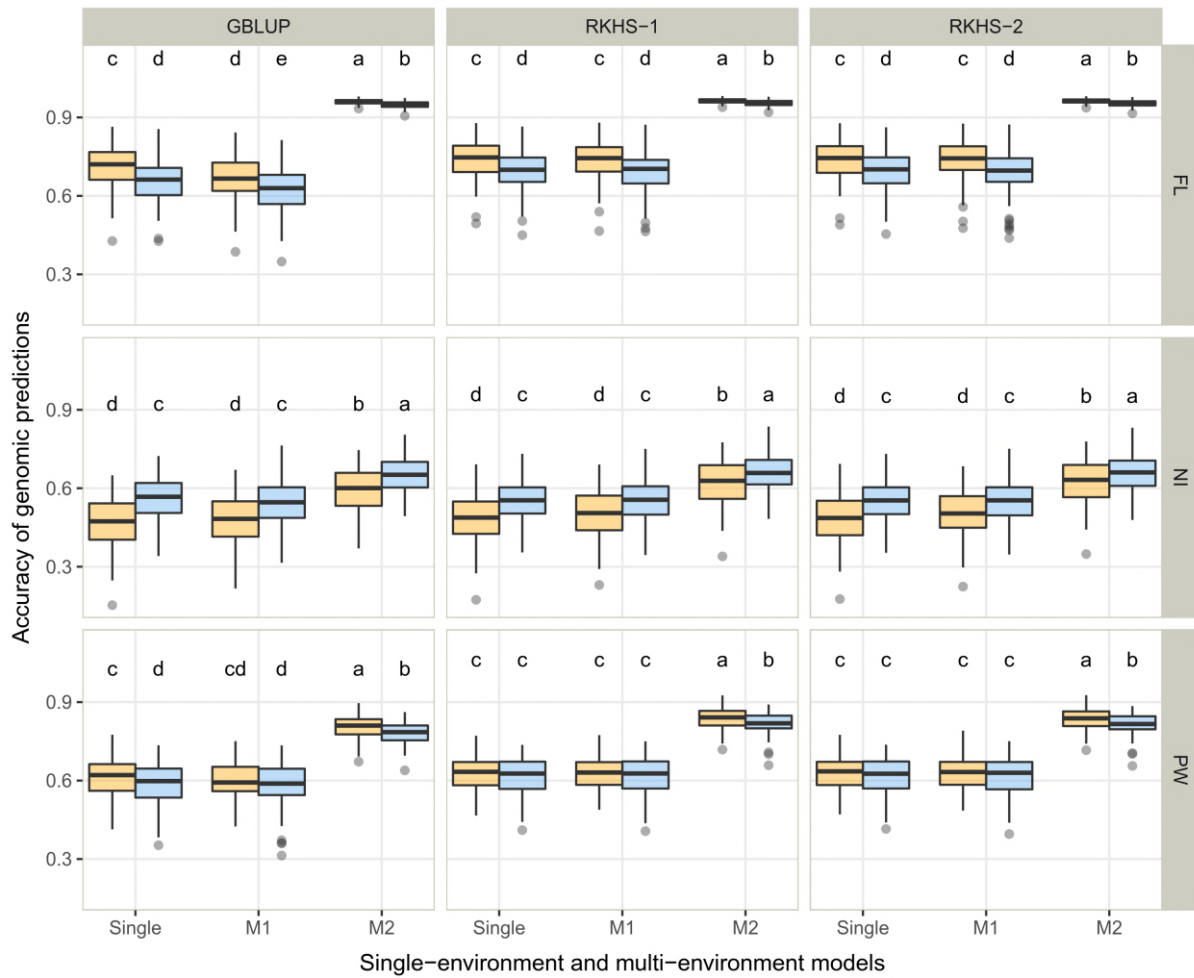
844

Figure 6: Single environment and multi-environment prediction accuracy in across population validation experiments obtained with three statistical models (GBLUP, RKHS-1, RKHS-2). Continuous flooding and alternate wetting and drying water management conditions are in blue and orange, respectively. The scenarios used to define the training set are represented by the different shades of orange or blue: light (S1: only the parents), intermediate (S2: 100 individuals of the RP selected with CDmean) and dark (S3: the whole RP).The three studied traits are presented: days to flowering (FL), nitrogen balance index (NI) and panicle weight (PW).

854 **Table 1**: Sources of phenotypic variation and derived summary statistics of days to flowering
855 (FL), nitrogen balance index (NI) and panicle weight (PW) in two populations of rice (reference
856 RP and progeny PP) conducted in two consecutive seasons under two water management
857 systems (AWD and CF).

Pop	Trait	System	Mean	SD	Σ^2_{Fixe} (1)	Variances accounted by the random effects(2)			Total phenotypic variance	R^2_{Cond} (3)	$H^2(SE)$ (4)	$\hat{r}_G[\hat{r}_1; \hat{r}_2]$ (5)	$\frac{CR}{DR}$ (6)
						G	**Y x G**	**R**					
RP	FL	AWD	100.3	7.8	44.12	57.68	10.90	11.28	123.98	0.91	0.89 (0.01)	0.955 [0.943;0.964]	0.98
		CF	93.4	7.0	8.43	47.78	4.36	5.95	66.52	0.91	0.94 (0.01)		
	NI	AWD	20.1	2.0	0.91	4.99	1.22	14.71	21.83	0.33	0.61 (0.05)	0.589 [0.508;0.661]	0.56
		CF	23.7	2.5	1.50	6.17	4.09	16.75	28.50	0.41	0.56 (0.05)		
	PW	AWD	252.9	57.9	720.96	3435.39	949.48	3142.66	8248.49	0.62	0.76 (0.03)	0.773 [0.722;0.816]	0.82
		CF	342.3	71.1	119.98	5088.95	850.38	2437.24	8496.55	0.71	0.85 (0.02)		
PP	FL	AWD	102.8	6.1	40.94	35.15	8.17	11.78	96.04	0.88	0.85 (0.03)	0.897 [0.850;0.930]	0.90
		CF	92.9	5.2	27.97	23.20	7.38	2.27	60.81	0.96	0.85 (0.03)		
	NI	AWD	17.1	1.5	1.55	3.03	0.00	5.32	9.90	0.46	0.76 (0.04)	0.731 [0.622;0.812]	0.75
		CF	18.4	2.0	2.63	4.12	0.70	3.72	11.16	0.67	0.80 (0.04)		
	PW	AWD	199.9	51.3	889.23	2487.80	466.32	522.24	4365.59	0.88	0.88 (0.02)	0.848 [0.781;0.896]	0.86
		CF	277.6	53.0	258.26	2698.52	415.49	554.00	3926.27	0.86	0.90 (0.02)		

858 (1) Variance accounted for by the season effect: Season effect: 2012 *versus* 2013 for the reference population and
859 2014 *versus* 2015 for the progeny population.
860 (2) Random effects: G: accession, Y x G: Season x Accession, R: Residual
861 (3) R^2_{Cond} : Conditional coefficient of determination
862 (4) $H^2(SE)$: Broad sense heritability for single environment analysis
863 (5) Pearson correlations between adjusted means of accessions under AWD and CF.
864 (6) Ratio of correlated response in CF to direct response in AWD.
865

bioRxiv preprint doi: https://doi.org/10.1101/257808; this version posted January 31, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

866 **Table 2**: Analysis of factors that influence the prediction accuracy of response variables in the
867 reference population. The effects of the type of response (index, slope, AWD, CF), the trait
868 (FL, NI and PW), the statistical model (GBLUP and RKHS) and their interactions were
869 evaluated.

R²	CV	RMSE	Mean	Source	DF	SS	MS	FValue	ProbF
Model 1: Only main effects
0.648	23.617	0.152	0.642	Model	6	101.489	16.915	734.86	<0.0001
				Error	2393	55.082	0.023		
				Corrected Total	2399	156.570			
				Response	3	77.532	25.844	1122.78	<0.0001
				Trait	2	23.571	11.785	512.01	<0.0001
				S model	1	0.386	0.386	16.76	<0.0001
Model 2: Main effects and interactions
0.732	20.681	0.133	0.642	Model	23	114.633	4.984	282.38	<0.0001
				Error	2376	41.937	0.018		
				Corrected Total	2399	156.570			
				Response	3	77.532	25.844	1464.21	<0.0001
				Trait	2	23.571	11.785	667.71	<0.0001
				S model	1	0.386	0.386	21.86	<0.0001
				Response*Trait	6	12.456	2.076	117.61	<0.0001
				Trait*S model	2	0.433	0.217	12.27	<0.0001
				Response*S model	3	0.073	0.024	1.38	0.2459
				Response*Trait*S model	6	0.182	0.030	1.72	0.1126

870 R²: Coefficient of determination; CV: Coefficient of variation; RMSE: Root mean square error; Mean: Intercept
871 value of the transformed accuracy (Z); DF: Degree of freedom; SS: Sum of squares; MS: Mean square.

872

873

874 **Table 3**: Analysis of factors that influence the variation in prediction accuracy in the reference
875 population using multi-environment models. The effects of the statistical model (GBLUP,
876 RKHS-1 and RKHS-2), the trait (FL, NI and PW), the cross-validation strategy (M1 and M2)
877 and the target condition (AWD and CF) and their interactions were evaluated.

R²	CV	RMSE	Mean	Source	DF	SS	MS	FValue	ProbF
Analysis with only main effects
0.723	24.163	0.221	0.914	Model	7	687.496	98.214	2014.66	<0.0001
				Error	5392	262.858	0.049		
				Corrected Total	5399	950.354			
				CV strategy	2	362.879	181.439	3721.86	<.0001
				Trait	2	320.946	160.473	3291.78	<.0001
				S model	2	3.352	1.676	34.38	<.0001
				Target condition	1	0.319	0.319	6.55	0.0105
Analysis with main effects and all first-order interactions
0.899	14.640	0.134	0.914	Model	25	854.176	34.167	1909.11	<.0001
				Error	5374	96.178	0.018		
				Corrected Total	5399	950.354			
				CV strategy	2	362.879	181.440	10138.0	<.0001
				Trait	2	320.946	160.473	8966.54	<.0001
				S model	2	3.352	1.676	93.65	<.0001
				Target condition	1	0.319	0.319	17.83	<.0001
				CV strategy*Trait	4	157.483	39.371	2199.87	<.0001
				Target condition*Trait	2	7.811	3.906	218.23	<.0001
				Trait*S model	4	0.783	0.196	10.94	<.0001
				Target condition*CV strategy	2	0.300	0.150	8.37	0.0002
				CV strategy*S model	4	0.300	0.075	4.20	0.0022
				Target condition*S model	2	0.003	0.002	0.09	0.9169

878 R²: Coefficient of determination; CV: Coefficient of variation; RMSE: Root mean square error; Mean: Intercept
879 value of the transformed accuracy (Z); DF: Degree of freedom; SS: Sum of squares; MS: Mean square.