UPEP
Unidad de Posgrados y Educación Permanente

FACULTAD DE
AGRONOMÍA
UNIVERSIDAD DE LA REPÚBLICA

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

CARACTERIZACIÓN Y PREDICCIÓN DE LA INTERACCION GENOTIPO POR AMBIENTE DEL RENDIMIENTO EN EL PROGRAMA DE MEJORAMIENTO DE ARROZ DE INIA

María Inés REBOLLO PANUNCIO

Magíster en Ciencias Agrarias
opción Bioestadística

Julio 2022

CARACTERIZACIÓN Y PREDICCIÓN DE LA INTERACCION GENOTIPO POR AMBIENTE DEL RENDIMIENTO EN EL PROGRAMA DE MEJORAMIENTO DE ARROZ DE INIA

María Inés REBOLLO PANUNCIO

Magíster en Ciencias Agrarias
opción Bioestadística

Julio 2022

Tesis aprobada por el tribunal integrado por Ing. Agr. Dr. Ariel Castro, Ing. Agr. Dra. Lucía Gutiérrez y Bioq. Clín. Dra. Bettina Lado el 26 de julio de 2022. Autora: Ing. Agr. María Inés Rebollo Panuncio. Director Lic. Dr. Juan Eduardo Rosas Caissiols, codirector Ing. Agr. Dr. Ignacio Aguilar García.

AGRADECIMIENTOS

Le agradezco a los tutores y mentores que me acompañaron durante este proceso. En primer lugar, le agradezco a Juan por su generosidad y por todo lo que me enseñó, desafió y acompañó. Le agradezco por ser el tutor que quise tener y mostrarme que una maestría puede ser un espacio de aprendizaje y bienestar a la vez. Le agradezco a Ignacio por su paciencia y simplicidad para explicarlo todo, que me ayudó a perderle el miedo al mundo del mejoramiento animal, lo que enriqueció inmensamente mi aprendizaje. A Gregor Gorjanc y a Lucía Gutiérrez por abrirme las puertas a sus grupos de trabajo y por todo lo que aprendí con ellos. A Ariel Castro y Bettina Lado por sus valiosos aportes a este trabajo como miembros externos del tribunal.

Le agradezco infinitamente a mi familia y amigos por su apoyo fundamental en cada paso de este proceso.

Le agradezco a mis amigos y compañeros de facultad: Pablo, Venancio, Gina, Germán, Sheila y Matías por estar siempre presentes y en tantos sentidos. Les agradezco por sus valiosas reflexiones que me ayudaron a convertir las experiencias desafiantes que se nos plantearon en instancias de crecimiento y fortalecimiento personal.

Le agradezco al Programa de Arroz de INIA y a todos sus integrantes, en especial al equipo de Mejoramiento Genético por los valiosos aportes a este trabajo.

Le agradezco a los otros grupos de trabajo que me acompañaron e integraron a sus espacios. Al grupo de Mejoramiento Genético Animal de INIA por permitirme participar en sus seminarios, instancias que generaron importantes intercambios y enseñanzas. Le agradezco también al grupo de Bio de INIA, al DBEC, al Highlander Lab y al Gutiérrez Lab.

A INIA, CAP, ANII y CSIC por la financiación.

TABLA DE CONTENIDO

Página

RESUMEN

El arroz es uno de los cultivos más importantes en el mundo y en la producción nacional. El Programa de Mejoramiento Genético de Arroz de INIA (PMGA) tiene entre sus objetivos generar cultivares con alto rendimiento y estabilidad frente a condiciones climáticas. La interacción genotipo por ambiente (IGA) determina que el PMGA evalúe sus materiales durante varios años, en distintas localidades y fechas de siembra. Un estudio exhaustivo de la IGA, así como modelos que permitan hacer predicciones en ambientes no evaluados, posibilitarían desarrollar estrategias que minimicen el número de ensayos y el tiempo necesario para obtener nuevos materiales. El objetivo de este trabajo fue caracterizar la IGA en el PMGA y evaluar modelos de regresión aleatoria (MRA) que incorporen variables climáticas (VC) para la predicción del rendimiento. Se utilizó información fenotípica de 23 años de evaluación en el campo de dos poblaciones de mejoramiento (*indica* y *japonica*), 16 VC, pedigrí y marcadores moleculares. La IGA se caracterizó mediante la descomposición de varianza, identificación de megaambientes (ME) y correlaciones entre localidades, fechas de siembra y ME. Se preseleccionaron VC asociadas al rendimiento y a la IGA por mínimos cuadrados parciales y con un procedimiento de selección de variables por pasos, se obtuvieron los MRA con la mejor habilidad predictiva en ambientes evaluados. Estos fueron probados para predecir rendimiento en ambientes no evaluados. La IGA del rendimiento representó más del 10 % de la varianza en ambas poblaciones. Se identificaron cinco ME en *indica* y tres en *japonica*. Las fechas de siembra tempranas en *indica* mostraron importante IGA. Se identificaron VC relacionadas con temperatura, radiación, viento y precipitación afectando el rendimiento y la IGA y variaron para cada población. Los MRA seleccionados mejoraron la habilidad predictiva frente a ambientes evaluados y no evaluados.

Palabras clave: regresión aleatoria, HBLUP, predicción genómica

GRAIN YIELD GENOTYPE BY ENVIRONMENT INTERACTION CHARACTERIZATION AND PREDICTION IN INIA'S RICE BREEDING PROGRAM

SUMMARY

Rice is one of the most important crops in the world and in Uruguay. The National Agricultural Research Institute's Rice Breeding Program (IRBP) aims to obtain cultivars with high yield and stability to climatic conditions. Genotype-by-environment interaction (GEI) determines that IRBP materials are evaluated for several years in different locations and sowing dates. A comprehensive study of GEI and models that allow predicting in untested environments would enable the development of strategies that minimize the number of trials and time needed to obtain new materials. This work aimed to characterize GEI in the IRPB and to evaluate random regression models (RRM) that incorporate climatic variables (CV) for yield prediction. Phenotypic information of 23 years of field evaluation of two populations (*indica* and *japonica*), 16 CV, pedigree and molecular markers were used. GEI was characterized by variance decomposition, identification of mega-environments (MEs), and correlations between locations, sowing dates, and MEs. CV associated with yield and GEI were pre-selected by partial least squares. RRM with the best predictive ability in evaluated environments were obtained with a stepwise forward variable selection procedure. These were tested to predict the performance of IRBP materials in untested environments. Yield GEI was more than 10 % of the total phenotypic variance in both populations. Five MEs were identified in *indica* and three in *japonica*. The early sowing dates in *indica* showed strong GEI. CV related to temperature, radiation, wind, and precipitation affecting yield and GEI were identified and differed for each population. The use of the selected RRMs improved the predictive ability in both tested and untested environments.

Keywords: random regression, HBLUP, genomic prediction

1 INTRODUCCIÓN

El arroz (*Oryza sativa* L.) es uno de los cultivos más importantes en el mundo (FAO, 2018). Uruguay, posicionado como octavo exportador a nivel mundial (FAO, 2020), es reconocido globalmente por la calidad de su producto. El arroz es de alta importancia en la matriz productiva nacional, representando el quinto rubro de exportación (Uruguay XXI, 2018). Cerca del 70 % del arroz uruguayo se cultiva en la región este, mientras que las regiones norte y centro concentran el 20 % y el 10 %, respectivamente, representando diferentes ambientes productivos. Uruguay se encuentra muy bien posicionado mundialmente en términos del promedio país de rendimiento de grano (IndexMundi, 2018), calidad y uniformidad, gracias a la fuerte integración vertical entre agricultores, molineros y al desarrollo de variedades localmente adaptadas (Lanfranco et al., 2018). Sin embargo, existen limitantes ambientales climáticas que generan inestabilidad en la producción (Pérez de Vida y Macedo, 2013). El mayor componente de variación ambiental en las condiciones productivas de arroz en Uruguay es la variabilidad climática interanual (Monteverde et al., 2019), lo que hace que la estabilidad del rendimiento frente a condiciones climáticas variables sea importante para el desarrollo de cultivares de arroz en Uruguay.

El Programa de Mejoramiento Genético de Arroz de INIA (PMGA), que se originó en los 70, ha liderado la generación de variedades de adaptación local durante décadas. Actualmente, los objetivos de mejoramiento del PMGA implican alto rendimiento, calidad molinera, resistencia a enfermedades fúngicas y estabilidad frente a condiciones climáticas adversas. El PMGA maneja dos poblaciones de mejoramiento mayoritarias: *indica* y *japonic*a, que corresponden a tipos de grano y calidad diferentes. Los ambientes objetivo del PMGA abarcan toda la zona de producción arrocera del país. Un proceso de mejoramiento típico realizado por el PMGA puede describirse con los siguientes pasos: 1) hacer aproximadamente 100 cruzamientos cada año con un número variable de parentales seleccionados; 2) autofecundar las progenies durante 5 o 6 generaciones con un método de avance de generaciones tipo pedigrí modificado, donde se selecciona por altura y arquitectura de planta y largo de ciclo desde las primeras generaciones (Bonnecarrère et al., 2019) y 3) una vez obtenidas las líneas fijas, se selecciona por el resto de los objetivos de mejoramiento mediante evaluaciones en ensayos de campo con repeticiones. La selección por rendimiento y estabilidad en los ambientes objetivo se realiza a través de, al menos, 5 años consecutivos de evaluación fenotípica en ensayos de campo en la localidad este principalmente, por representar la mayor zona de producción de arroz del país. Luego, en etapas avanzadas de la selección, se realizan ensayos en el norte y el centro. Adicionalmente, en las evaluaciones finales, los materiales son evaluados en épocas de siembra tardía para aumentar la probabilidad de ocurrencia de condiciones climáticas estresantes. No obstante, muchas veces esto es insuficiente para enfrentar a los genotipos a una adecuada representación de la alta variabilidad climática interanual que enfrenta el cultivo (Roel y Baethgen, 2007). Además, los distintos orígenes y adaptación a las condiciones templadas de los materiales *indica* y *japonica* (Khush, 1997) sugieren que las condiciones climáticas que los afectan podrían no ser las mismas. La evaluación tanto en distintas localidades como en diferentes fechas de siembra se realiza porque se ha constatado existencia de interacción genotipo por ambiente (IGA) con estos factores ambientales (Lavecchia y Méndez, 1999, Pérez de vida et al., 2015). Sin embargo, aún no se ha realizado un estudio exhaustivo que cuantifique este fenómeno.

La IGA consiste en cambios en la respuesta relativa de los materiales frente a diferentes ambientes (Fehr, 1987) y afecta mayormente características de herencia de tipo cuantitativa (Bernardo, 2010). Al ser un fenómeno muy común en los ensayos de evaluación fenotípica de los programas de mejoramiento de plantas, diversas metodologías han sido propuestas para su estudio (Crossa, 2012) y para su manejo: ignorarla, explotarla o reducirla. Ignorarla equivale a seleccionar en base a la media genotípica general a través de los diversos ambientes. Explotarla implica considerarla en modelos que permitan utilizar toda la información disponible para predecir la respuesta de los genotipos a la variabilidad ambiental. Por último, la reducción de la IGA es posible si esta responde a patrones predecibles o repetibles (por ejemplo, localidad o fecha de siembra), en cuyo caso es conveniente seleccionar de forma independiente dentro de cada ambiente. Los megaambientes (ME), por definición, minimizan la IGA dentro del ME (Braun et al., 1996) y diversos estudios han buscado identificar patrones repetibles de IGA mediante la identificación de ME (Gauch y Zobel, 1997, Yan et al., 2000, Lado et al., 2016). Un adecuado entendimiento de la estructura de la IGA en los ambientes objetivo de un programa de mejoramiento permite la toma de decisiones informadas en cuanto a la eficiente asignación de recursos.

Otra estrategia para mejorar la eficiencia de los programas de mejoramiento genético es la selección genómica (SG). La SG es un método de mejoramiento complementario a la selección fenotípica con potencial para contribuir a mejorar características de herencia compleja, como son el rendimiento y la estabilidad, en forma más eficiente (Meuwissen et al., 2001). La SG consiste en utilizar información de marcadores moleculares de todo el genoma, datos fenotípicos y de pedigrí para la predicción de valores de cría genómicos (Aguilar et al., 2010, Mrode, 2014, Bernardo, 2016, Crossa et al., 2017, Schmid y Bennewitz, 2017). La SG permite reducir el ciclo de mejoramiento generando mayor eficiencia y acortando el tiempo de desarrollo de cultivares (Jannink et al., 2010, Crossa et al., 2017). Por ello, ha sido ampliamente adoptada por muy diversos programas de mejoramiento de animales y de plantas en todo el mundo: públicos y privados, grandes y pequeños (Gholami et al., 2021).

La utilización de SG en el marco de un programa de mejoramiento vegetal implica tomar decisiones sobre cómo modelar la IGA. Este es un campo de activa investigación en la comunidad científica del mejoramiento de plantas (Jarquín et al., 2021) y diversas metodologías relacionadas con los modelos mixtos han sido exploradas. Los modelos mixtos consideran efectos fijos y aleatorios y permiten flexibilidad para modelar la estructura de varianzas y covarianzas entre los efectos aleatorios. Entre ellos, lo más utilizado han sido modelos multicarácter que toman el desempeño en cada ambiente como un carácter distinto y modelan la matriz de varianzas y covarianzas entre caracteres con base en información de covariables ambientales, (Jarquín et al., 2014), máxima verosimilitud restringida (Burgueño et al., 2012) o siguiendo métodos de reducción de dimensiones como los modelos de factor analítico (Tolhurst et al., 2019). Sin embargo, estos modelos carecen de capacidad predictiva para ambientes no observados (Heslot et al., 2014). Los modelos de regresión aleatoria (MRA), son modelos mixtos que modelan la respuesta diferencial de cada individuo frente a determinadas covariables, permitendo la predicción en ambientes desconocidos (Buntaran et al., 2021). Los MRA han sido exitosamente utilizados en el mejoramiento animal para modelar la IGA debida al estrés térmico (Ravagnolo y Misztal, 2000). Cuando el número de variables disponibles es alto, aplicar un método de selección de variables resulta ventajoso previo a ajustar este tipo de modelos (Buntaran et al., 2021). La utilización de MRA para variables climáticas seleccionadas representa una estrategia prometedora para explotar la IGA y su utilidad en el marco de la selección en un programa de mejoramiento vegetal aún no ha sido probada.

1.1 HIPÓTESIS

a) Las combinaciones de localidades y fechas de siembras del PMGA se agrupan en distintos megaambientes.

b) Las variables climáticas que afectan el rendimiento en *indica* y *japonica* no son las mismas.

c) La predicción genómica del rendimiento mejora al modelar la respuesta diferencial de cada genotipo a covariables climáticas que afectan el rendimiento.

1.2 OBJETIVOS

1.2.1 Objetivo general

Caracterizar la IGA en el PMGA y evaluar MRA que incorporen variables climáticas para contribuir al mejoramiento del rendimiento de arroz en Uruguay.

1.2.2 Objetivos específicos

a) Caracterizar la IGA en el PMGA.

b) Seleccionar variables climáticas que afecten el rendimiento.

c) Evaluar la habilidad predictiva de MRA que incorporen las variables climáticas identificadas para la predicción del rendimiento en ambientes conocidos y desconocidos.

2 GENOTYPE BY ENVIRONMENT INTERACTION CHARACTERIZATION AND ITS MODELLING WITH RANDOM REGRESSION TO CLIMATIC VARIABLES IN TWO RICE BREEDING POPULATIONS[1]

2.1 RESUMEN

La interacción genotipo por ambiente (IGA) es uno de los principales desafíos en el mejoramiento genético de plantas. Su caracterización necesaria para decidir estrategias de mejoramiento adecuadas. Los modelos de regresión aleatoria (MRA) que incluyen variables ambientales seleccionadas representan un enfoque prometedor para manejar la IGA en la predicción genómica (PG) tanto para predecir en ambientes evaluados como no evaluados y su utilidad aún no se ha demostrado. Utilizamos datos fenotípicos, climáticos, genealógicos y genómicos de dos programas subtropicales de mejoramiento público que manejan las poblaciones de arroz *indica* y *japonica*. Primero, caracterizamos la IGA para el rendimiento en grano (RG) con un conjunto de herramientas: estimación de componentes de varianza, definición de megaambientes (ME) y correlación entre localidades, períodos de siembra y ME. Luego, identificamos las variables climáticas más influyentes relacionadas con RG y las usamos en MRA para PG *single-step*. Finalmente, evaluamos la capacidad predictiva de estos modelos para la predicción en años y ambientes evaluados y no evaluados. Nuestros resultados sugieren gran IGA en ambas poblaciones, aunque más grande en *indica* que en *japonica*. En *indica*, los períodos de siembra temprana mostraron un patrón de IGA diferentes a los demás. Variables climáticas relacionadas con temperatura, radiación, viento y precipitación afectando el RG y su IGA difirieron en cada población. Los modelos de regresión aleatoria con covariables climáticas seleccionadas mejoraron la capacidad predictiva en escenarios de predicción de años y ambientes evaluados y no evaluados.

---

[1] Artículo a publicar en *Crop Science*.

2.2 SUMMARY

Genotype by environment interaction (GEI) is one of the main challenges in plant breeding. A complete characterization of it is necessary to decide on proper breeding strategies. Random regression models (RRM) that include selected environmental variables represent a promising approach to deal with GEI in genomic prediction (GP). They enable to predict for both tested and untested environments, but their utility remains to be shown. We used phenotypic, climatic, pedigree, and genomic data from two public subtropical rice breeding programs, one manages the *indica* population and the other manages the *japonica* population. First, we characterized GEI for grain yield (GY) with a set of tools: variance component estimation, mega-environment (ME) definition, and correlation between locations, sowing periods, and MEs. Then, we identified the most influential climatic variables related to GY and its GEI and used them in RRM for single-step GP. Finally, we evaluated the predictive ability of these models for GY prediction in tested and untested environments. Our results suggest large GEI in both populations while larger in *indica* than in *japonica*. In *indica,* early sowing periods showed crossover GEI with other sowing periods. Climatic variables related to temperature, radiation, wind, and precipitation affecting GY were identified and differed in each population. Random regression models with selected climatic covariates improved the predictive ability in both tested and untested years and environments.

2.3 INTRODUCTION

Breeding for stability (i.e., creating varieties that are resilient, tolerant to biotic and abiotic stresses and of use for a wide range of environments) is extremely complex and has long been an objective in crop breeding programs (Finlay & Wilkinson, 1963). Furthermore, climate change's negative impacts (Lobell & Field, 2007) have increased the plant breeding scientific community's interest in breeding for abiotic stress tolerance and crop resilience with a diversity of techniques (Tester & Langridge, 2010). One of the many challenges plant breeders face that can be even worse under a climate change scenario is the existence of genotype by environment interaction (GEI), and proper handling of it is necessary to effectively breed for stability (Crossa & Cornelius, 1997; Piepho, 1997, 1998; Burgueño et al., 2008; Malosetti et al., 2013).

GEI is a very common phenomenon in the context of plant breeding programs that has a strong effect on quantitative traits (Bernardo, 2010). Fehr (1987) defined it as "changes in the relative performance of genotypes across different environments". In a typical plant breeding program, performance and adaptation of potential cultivars are tested across several locations and years. These represent different environments and conform multi-environment trials (MET) where variable amounts of GEI may be present (van Eeuwijk et al., 2016). Identifying high-yielding and stable genotypes in MET datasets has long been a challenge primarily due to the existence of GEI with crossover interaction. Strategies proposed to cope with GEI include ignoring it by selecting based on mean performance, reducing it by targeting more and smaller homogeneous subgroups of environments, and exploiting it by modelling it and identifying cultivars best suited to specific environments or stable across different environments (Bernardo, 2010). Several models have been proposed to characterize GEI, including fixed linear models (e.g., factorial regression), bilinear models (e.g., PLS: partial least squares), linear-bilinear fixed models (e.g., additive main effect and multiplicative interaction models, AMMI; genotype plus GEI models, GGE; or site regression models, SREG), linear mixed models (e.g., multi-trait models), and linear-bilinear mixed models (e.g., factor analytic) (Crossa, 2012). Other authors included complementing ways to characterize GEI like the partition of variance components and mega-environment (ME) characterization (Lado et al., 2016). The combination of different ways of characterizing GEI in the desired target population of environments is necessary for a breeding program to make an informed decision on their proper strategy.

Genomic prediction (GP) is a breeding approach that has been proven successful and adopted to increase genetic gain in different animal and plant breeding programs (de los Campos et al., 2013). GP uses phenotype, whole-genome molecular markers, and pedigree data, if available, to predict breeding and genotypic values (Bernardo, 2016; Crossa et al., 2017; Aguilar et al., 2010). Its wide adoption across both big and small, private, and public breeding programs has been possible due to a dramatic decrease in the cost of genotyping (Crossa et al., 2017), which in some cases is close to or even less than the cost of phenotyping (Bernardo, 2008). GP is especially useful when the trait to improve is complex and controlled by many quantitative trait loci (QTL), each with a small effect (Asoro et al., 2013; Beyene et al., 2015; Bernardo, 2016). This is the case in most plant breeding programs that aim to get higher-yielding varieties, among other objectives. A specific challenge of applying GP in plant breeding programs is the need to deal with GEI (Burgueño et al., 2012).

Accounting for GEI in GP models has been implemented in different ways and it still represents an area of active research (Jarquín et al., 2021). Mixed models including some kind of (co)variance structure between environments can be used for that purpose (Burgueño et al., 2012). Several structures of the (co)variance matrix between environments can be used: identity, diagonal, compound symmetry, factor analytic, unstructured, and environmental kinship (Bustos-Korts et al., 2015; Malosetti et al., 2016).

Environmental kinship structures can be defined by taking advantage of available environmental covariates (Jarquín et al., 2014). All models based on observed covariance among environments are limited to explaining *a posteriori* variation, lacking the ability to predict a new climatic scenario (Heslot et al., 2014). Random regression models (RRM), which provide a genotype-specific response to each regressor factor can be used to model GEI in a wide range of environments (Schaeffer, 2004) and even make predictions in untested environments (Buntaran et al., 2021). Due to high availability of environmental variables, many are highly correlated and do not need to be included in the model (Bustos-Korts et al., 2015). Variable selection has been proven useful in improving the performance of models with environmental variables (Neyhart et al., 2022). RRMs including selected environmental variables represent a promising strategy to deal with GEI in GP to predict for both tested and untested environments and its utility remains to be shown.

The Uruguayan National Agricultural Research Institute's rice breeding program (IRBP) aims to breed high-yielding, stable varieties. IRBP separately manage germplasm of *indica* and *japonica* subspecies in two breeding populations that are bred for different objectives and target markets and have different reactions to climatic conditions (Yoshida, 1981). IRBP tests its candidate fixed genotypes for at least five years because the inter-annual climatic variability is high in Uruguayan climatic conditions. Also, three testing locations are used. Recently, IRBP's MET information of the last 23 years of field phenotypic evaluation, as well as pedigree and genomic information have been gathered in a database that is easily accessible (Rebollo et al., unpublished data, 2022). This dataset with guaranteed GEI is interesting to test questions regarding RRMs for modelling GEI in two breeding populations.

This work aims to characterize GEI in IRBP historical dataset, evaluate different ways of grouping environments, identify climatic covariates that affect yield, and evaluate the use of RRM with selected climatic covariates models for GP in untested environments to contribute to breeding for stability. We specifically aimed to answer the following questions: how much GEI is present in IRBP dataset? Do *indica* and *japonica* rice populations show the same GEI patterns? Which climatic covariates are associated with yield and yield GEI in each population? How much of the environmental effect can be accounted for with a RRM that includes selected climatic covariates? Is a RRM with climatic covariates model a good option to predict in tested environments? And in untested ones?

2.4 MATERIALS AND METHODS

We used phenotypic, climatic, pedigree, and genomic data from two breeding populations of rice (*indica* and *japonica*) of IRBP, from two public subtropical rice breeding programs. First, we characterized GEI for grain yield (GY) with a set of tools: variance component estimation, ME definition, and correlation between locations, sowing periods, and MEs. Then, we identified the most influential climatic variables for GY and its GEI and used them in RRM for GP using phenotypic information of genotyped and non-genotyped individuals. Finally, we evaluated the predictive ability of these models for GY prediction in tested and untested environments. Given the strong population structure between *indica* and *japonica* (Zhao et al., 2011), and their known differential reaction to climatic conditions (Yoshida, 1981), GEI characterization, climatic variable selection and GY prediction were performed on a population basis.

2.4.1 Data

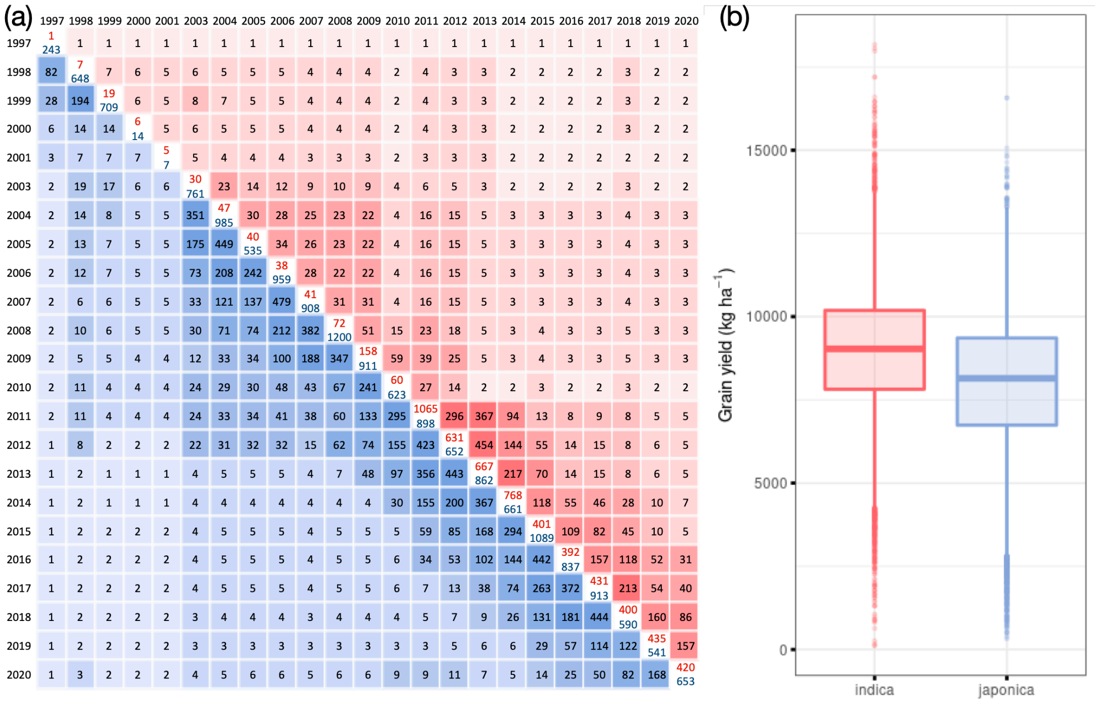
2.4.1.1 Phenotypic Data

MET data for GY (in kg ha^{-1}) and flowering date (FD, date when half of the plants reached anthesis) were retrieved from IRBP's historical MET dataset (Rebollo et al., unpublished data, 2022). Trials had a randomized complete block experimental design, with two blocks for the first evaluation year, three blocks for the second and third evaluation years, and four blocks for the final evaluation stage. Plot area was 2.4 m^2. All trials were conducted under irrigation conditions with current production management standards. Only trials with heritability on a genotype-mean basis (H^2) for GY > 0.1 were selected. H^2 was calculated as $H^2 = \sigma_g^2 \,/\, \left(\sigma_g^2 + \sigma_e^2/r\right)$ where σ_g^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of blocks. Variance components for H^2 calculation were obtained by fitting the following model by trial:

$$y_{ij} = \mu + g_i + b_j + e_{ij} \qquad [1]$$

where y_{ij} is the GY score, μ is the overall mean, g_i is the random effect of the ith genotype, independent and identically distributed (IID) with $g_i \sim N(0, \sigma_g^2)$, where b_j is the random effect of the jth block, IID with $b_j \sim N(0, \sigma_b^2)$, where σ_b^2 is the block variance and e_{ij} is the model random residual, IID with $e_{ij} \sim N(0, \sigma_e^2)$. This model assumes no covariance between g_i, b_j, and e_{ij} effects. Models were fitted using the *lme4* R package (Bates et al., 2015) and variance components were estimated via REML.

The selected dataset resulted in 851 trials conducted during 23 years (from 1997 to 2020) in three locations: the East (E, Paso de la Laguna Experimental Unit, Treinta y Tres, 33.27 S, 54.17 W), the main IRBP location with 748 trials, the North (N, Paso Farías Experimental Field, Artigas, 30.54 S, 57.26 W) with 91 trials, and the Center (C, Pueblo del Barro Experimental Field, Cerro Largo, 31.93 S, 55.38 W) with 12 trials. A total of 3,796 *indica* and 9,817 *japonica* elite breeding genotypes were evaluated. The dataset was highly unbalanced, with 12 to 1,967 (mean of 1,077) genotypes under evaluation each year, and pairs of years sharing from two to 993 genotypes. At least one genotype of each population was shared between all years (Figure 1a). Mean GY was 8,966 kg ha^{-1} for *indica* and 7,967 kg ha^{-1} for *japonica*, and the variable distribution for each population is presented in Figure 1b.



	1997	1998	1999	2000	2001	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1997	1 243	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1998	82	7 648	7	6	5	6	5	5	5	4	4	4	2	4	3	3	2	2	2	2	3	2	2
1999	28	194	19 709	6	5	8	7	5	5	4	4	4	2	4	3	3	2	2	2	2	3	2	2
2000	6	14	14	6 14	5	6	5	5	5	4	4	4	2	4	3	3	2	2	2	2	3	2	2
2001	3	7	7	7	5 7	5	4	4	4	3	3	3	2	3	3	3	2	2	2	2	2	2	2
2003	2	19	17	6	6	30 761	23	14	12	9	10	9	4	6	5	3	2	2	2	2	3	2	2
2004	2	14	8	5	5	351	47 985	30	28	25	23	22	4	16	15	5	3	3	3	3	4	3	3
2005	2	13	7	5	5	175	449	40 535	34	26	23	22	4	16	15	5	3	3	3	3	4	3	3
2006	2	12	7	5	5	73	208	242	38 959	28	22	22	4	16	15	5	3	3	3	3	4	3	3
2007	2	6	6	5	5	33	121	137	479	41 908	31	31	4	16	15	5	3	3	3	3	4	3	3
2008	2	10	6	5	5	30	71	74	212	382	72 1200	51	15	23	18	5	3	4	3	3	5	3	3
2009	2	5	5	4	4	12	33	34	100	188	347	158 911	59	39	25	5	3	4	3	3	5	3	3
2010	2	11	4	4	4	24	29	30	48	43	67	241	60 623	27	14	2	2	3	2	2	3	2	2
2011	2	11	4	4	4	24	33	34	41	38	60	133	295	1065 898	296	367	94	13	8	9	8	5	5
2012	1	8	2	2	2	22	31	32	32	15	62	74	155	423	631 652	454	144	55	14	15	8	6	5
2013	1	2	1	1	1	4	5	5	5	4	7	48	97	356	443	667 862	217	70	14	15	8	6	5
2014	1	2	1	1	1	4	4	4	4	4	4	4	30	155	200	367	768 661	118	55	46	28	10	7
2015	1	2	2	2	2	4	5	5	5	4	5	5	5	59	85	168	294	401 1089	109	82	45	10	5
2016	1	2	2	2	2	4	5	5	5	4	5	5	6	34	53	102	144	442	392 837	157	118	52	31
2017	1	2	2	2	2	4	5	5	5	4	5	5	6	7	13	38	74	263	372	431 913	213	54	40
2018	1	2	2	2	2	3	4	4	4	3	4	4	4	5	7	9	26	131	181	444	400 590	160	86
2019	1	2	2	2	2	3	3	3	3	3	3	3	3	3	5	6	6	29	57	114	122	435 541	157
2020	1	3	2	2	2	4	5	6	6	5	6	6	9	9	11	7	5	14	25	50	82	168	420 653

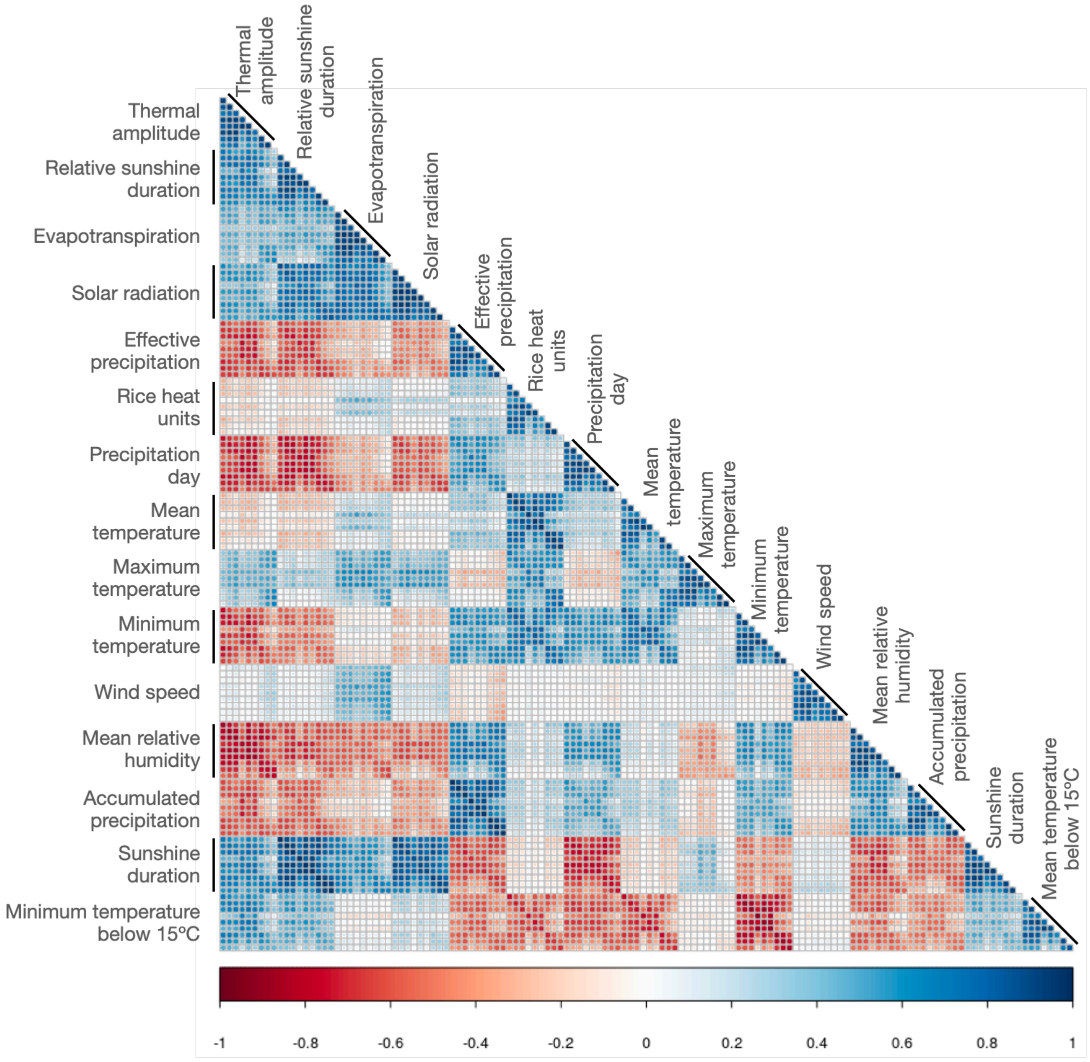
Figure 1. (a) Genotypic connectivity measured as the number of shared genotypes between years for *indica* (above diagonal, red gradient) and for *japonica* (below diagonal, blue gradient) and per year total number of genotypes for *indica* (diagonal, upper red number) and *japonica* (diagonal, lower blue number), and distribution of grain yield (kg ha^{-1}) for *indica* and *japonica* populations (b).

2.4.1.2 Climatic Data

Daily data of the E location for 16 climatic variables related to temperature, precipitation, solar radiation, humidity, and wind (Table 1) were retrieved from INIA GRAS agroclimatic data bank (http://www.inia.uy/gras/Clima/Banco-datos-agroclimatico). Nine time-windows were defined to find the most critical period associated with response to abiotic stresses at the reproductive stage. Time-windows comprised the following intervals before and after FD: 10-10, 15-15, 20-20, 10-20, 20-10,10-30, 30-10, 5-35, 35-5. A total of 144 combinations of climatic covariates and time windows (CTW) were produced. Pearson's correlation between CTW within and between some climatic variables was high (Supplemental Fig. S1). Mean daily values of the respective CTW were computed for each plot and CTWs with zero estimated variance were removed from the analysis.

Table 1. Climatic variables used in this study.

Climatic variable	**Unit**	**Abbreviation**	**Description**
Accumulated precipitation	mm	AccumPp	Daily accumulated precipitation (9 to 9 hs)
Effective precipitation	mm	EffPp	Daily precipitation minus surface runoff
Evapotranspiration	mm	Evap	Penman-Monteith's evapotranspiration value
Ground frost	-	Frost	1 if temperature was below 0 °C at 0.05 m above ground level and 0 if not
Maximum temperature	°C	MaxTemp	Maximum daily temperature
Mean relative humidity	%	RelH	Average of daily relative humidity
Mean temperature	°C	MeanTemp	Average temperature (0 to 24 hs), imputed as (MaxTemp + MinTemp) / 2 when missing
Minimum temperature	°C	MinTemp	Minimum daily temperature
Minimum temperature below 15°C	-	MinT15	1 if minimum temperature was below 15 °C and 0 if not
Precipitation day	-	PpDay	1 if it rained and 0 if not
Relative sunshine duration	%	RelSun	Relationship between the observed sunshine duration value and the maximum for that latitude and day of the year
Rice heat units	°C	HeatU	Daily average temperature minus 10 °C
Solar radiation	cal cm^{-2} day	SolRad	Armstrong's solar radiation value
Sunshine duration	hs	SunHs	Number of hours with direct solar radiation
Thermal amplitude	°C	ThermAmpl	Maximum – minimum temperature
Wind speed	2 m km^{-1} 24 h	Wind	Distance covered by wind (km) at a height of 2 m above ground from 9 to 9 hs



Supplemental Fig. S1. Correlation between combinations of climatic variables and time windows used for climatic variable selection. Within each variable, time windows (days before-days after flowering date) are ordered as follows: 10-10, 15-15, 20-20, 5-35, 10-20, 10-30, 20-10, 30-10, and 35-5.

2.4.1.3 Pedigree Data

All available pedigree records of the genotypes under phenotypic evaluation and their ancestors up to the 6th generation, including number of selfing generations, were retrieved from the IRBP database (Rebollo et al., unpublished data, 2022). This consisted of 20,628 records.

2.4.1.4 Genomic Data

We used genomic information of 395 *indica* and 570 *japonica* advanced breeding genotypes and cultivars commonly used as checks, obtained by genotyping-by-sequencing with the ApeKI enzyme (Elshire et al., 2011). Sequences were obtained in two batches. The first batch consisted of 601 genotypes that were in field evaluations in 2011 and their sequencing was performed at the Biotechnology Resource Center at Cornell University with an Illumina 100bp single-end sequencer (Quero et al., 2018; Rosas et al., 2018). The second batch corresponds to 364 genotypes that were phenotyped in contrasting climatic conditions or widely used as parents and for which there was seed available at INIA's gene bank. This batch was sequenced in 2020 with an Illumina 150bp paired-end sequencer at the University of Wisconsin-Madison DNA Sequencing Facility. Raw sequences of the two batches were processed together on a population basis with *TASSEL* 5.0 pipeline (Bradbury et al., 2007) and SNPs were aligned to the Nipponbare reference genome MUS version 7.0 (Kawahara et al., 2013) using *bwa* (Li & Durbin, 2009). Loci with minor allele frequency ≥ 0.03, ≤ 50% of missing data, and ≤ 15% of observed heterozygosis were retained. Missing data were imputed with *BEAGLE* 5.4 (Browning et al., 2018) for both datasets separately. The final *indica* and *japonica* marker matrices consisted of 50,854 and 23,614 SNP, respectively and were transformed to numeric coding (0, 1, and 2 for homozygotes for the major allele, heterozygotes, and homozygotes for the minor allele, respectively) for analyses.

2.4.2 Genotype by Environment Interaction Characterization

We used the following strategies to characterize GEI: variance component estimation, ME definition through the GGE model and its augmented biplot graphical representation, and Pearson's correlations between locations, sowing periods, and MEs.

2.4.2.1 Variance Components

First, we characterized the magnitude of GEI with variance components by fitting the following model:

$$y_{ijkmno} = \mu + g_i + a_j + l_k + s_m + \ b_{n(o)} + ga_{ij} + gl_{ik} + gs_{im} + e_{ijkmno} \text{ [2]}$$

where y_{ijkmno} is the GY raw score, μ is the overall population mean, g_i is the random effect of the ith genotype, IID with $g_i \sim N(0, \sigma_g^2)$, where σ_g^2 is the genotypic variance, a_j is the random effect of the jth year, IID with $a_j \sim N(0, \sigma_a^2)$, where σ_a^2 is the year variance, l_k is the random effect of the kth location, IID with $l_k \sim N(0, \sigma_l^2)$, where σ_l^2 is the location variance, s_m is the random effect of the mth sowing period, IID with $s_m \sim N(0, \sigma_s^2)$, where σ_s^2 is the sowing period variance, $b_{n(o)}$ is the random effect of the nth block nested within the oth trial, IID with $b_{n(o)} \sim N(0, \sigma_b^2)$, where σ_b^2 is the block variance, ga_{ij} is the random effect of the interaction between the ith genotype and the jth year, IID with $ga_{ij} \sim N(0, \sigma_{ga}^2)$, where σ_{ga}^2 is the genotype by year variance, gl_{ik} is the random effect of the interaction between the ith genotype and the kth location, IID with $gl_{ij} \sim N(0, \sigma_{gl}^2)$, where σ_{gl}^2 is the genotype by location variance, gs_{im} is the random effect of the interaction between the ith genotype and the mth sowing period, IID with $gs_{im} \sim N(0, \sigma_{gs}^2)$, where σ_{gs}^2 is the genotype by sowing period variance, and e_{ijkmno} is the model random residual, IID with $e_{ijkmno} \sim N(0, \sigma_e^2)$, where σ_e^2 is the residual variance. Trials' sowing periods were defined as: S1 (early, trials sown until the 15th of October), S2 (intermediate-early, trials sown from the 16th of October to the 31st of October), S3 (intermediate-late, trials sown from the 1st to the 15th of November), and S4 (late, trials sown from the 16th of November on). The number of trials for each category was, 96, 347, 295, and 86 for S1 to S4, respectively, while 27 could not be assigned due to missing sowing date information. Models were fitted using the *lme4* R package (Bates et al., 2015) and variance components were estimated via REML.

2.4.2.2 Mega-environment Definition

Second, the GGE model and its graphical augmented biplot representation were used further characterize GEI by the definition of MEs. As this analysis requires complete information, trials were grouped together into year, location, and sowing period combinations (YLSs), genotypes present in more than 10 YLSs and YLSs with more than 10 genotypes were retained, and missing information was imputed with a factor analytic model of order 2 following Neyhart et al. (2021). The data set with these filters resulted in 52 *indica* genotypes in 52 YLSs (32, 13, and 7 from locations E, N, and C and 17, 16, 15, and 5 from sowing periods 1 to 4, respectively), and 67 *japonica* genotypes in 44 YLSs (34, 9, and 1 from locations E, N, and C and 10, 13, 14, 7 from sowing periods 1 to 4, respectively), and had 71.6% and 74.6% of missing information in *indica* and *japonica,* respectively. The adjusted model was:

$$y_{ijkl} = \mu + e_i + g_{ij} + b_{k(l)} + e_{ijkl} \quad [3]$$

where y_{ijk} is the GY raw score, μ is the overall mean, e_i is the fixed effect for the ith YLS, g_{ij} is the random effect of the jth genotype in the ith YLS with $g_{ij} \sim N(0, \boldsymbol{G})$ where $\boldsymbol{G}$ was modeled as $\boldsymbol{G} = \mathbf{I} \otimes \boldsymbol{\Gamma}$, where is the Kronecker product, $\mathbf{I}$ an identity matrix that represents the covariance among genotypes, and $\boldsymbol{\Gamma}$ is the covariance of genotypes across YLS that is modeled with a factor analytic structure (Smith et al., 2001) of order 2: $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ where $\boldsymbol{\Lambda}$ is a matrix of 2 loadings for YLSs, and $\boldsymbol{\Psi}$ is a diagonal matrix of residual within-YLS genetic variances, $b_{j(k)}$ is the random effect of the kth block nested within the lth trial, IID with $b_{k(l)} \sim N(0, \sigma_b^2)$, where σ_b^2 is the block variance, and e_{ijk} is the model random residual with $e_{ijk} \sim N(0, \boldsymbol{R})$, where $\boldsymbol{R}$ is a diagonal (co)variance structure that allows YLS specific residual variance. Variance components and $\boldsymbol{\Lambda}$ elements were estimated via restricted maximum likelihood and BLUPs of genotypic scores for each YLS were obtained from those estimates. This was performed with the *ASReml-R* R package (Butler et al., 2017). The GGE model fit and its graphical augmented biplot representation were performed with the *GGEBiplots* R package (Dumble, 2022). The ME definition was made by the graphical interpretation of the biplot representation of the GGE model. For the sake of balance in the number of YLSs in each ME in both populations, we combined MEs with only one YLS with the adjacent ME.

2.4.2.3 Correlations

Last, we computed Pearson's correlations between all pairs of YLS for the whole dataset (i.e., ungrouped) and between and within locations, sowing periods, their combinations, and MEs. This was performed with the adjusted genotype means of imputed dataset we used for the ME definition. Within the C location correlation for *japonica* could not be calculated because there was only one YLS.

2.4.3 Climatic Variable Selection

For the climatic variable selection, PLS was used given the high correlation between and within CTWs. Briefly, PLS performs the decomposition of the matrix with predictors ($\boldsymbol{X}$) and the matrix with responses (Y) simultaneously: $\boldsymbol{X} = \boldsymbol{TP}' + \boldsymbol{E}_x$ and $\boldsymbol{Y} = \boldsymbol{UQ}' + \boldsymbol{E}_y$ where $\boldsymbol{T}$ and $\boldsymbol{U}$ are latent variables matrices, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are latent variables loadings, $\boldsymbol{E}_x$ and $\boldsymbol{E}_y$ are error matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively and the latent variables orientation is selected to maximize the covariance between $\boldsymbol{T}$ and $\boldsymbol{U}$. In this study, $\boldsymbol{X}$ was the CTWs matrix for the mean FD of each genotype in each environment defined as combination of year and sowing date, and $\boldsymbol{Y}$ was alternatively one of the following three: the adjusted genotype means for each environment ($\widehat{G_{ij}}$), the genotype effect in a specific environment ($\widehat{g_{ij}}$), or the GEI effect ($\widehat{gxe_{ij}}$). $\widehat{G_{ij}}$ sums up for the genetic, environmental and GEI effect, $\widehat{g_{ij}}$ the genetic and GEI effects, and $\widehat{gxe_{ij}}$ only the GEI effect. They were defined as follows:

We first fit the following model per environment:

$$y_{ijk} = \mu + g_i + b_{j(k)} + e_{ijk} \quad [4]$$

where y_{ijk} is the GY score, μ is the overall mean, g_i is the fixed effect of the i^{th} genotype, $b_{j(k)}$ is the random effect of the j^{th} block nested within the k^{th} trial, IID with $b_{j(k)} \sim N(0, \sigma_b^2)$, where σ_b^2 is the block variance and e_{ijk} is the model random residual, IID with $e_{ijk} \sim N(0, \sigma_e^2)$, where σ_e^2 is the error variance. This model assumes no covariance between $b_{j(k)}$, and e_{ijk} effects. From this model, estimated effects for each i^{th} genotype for each j^{th} environment ($\widehat{g_{ij}}$) and adjusted genotype means ($\widehat{G_{ij}} = \widehat{\mu_j} + \widehat{g_i}$) were obtained. Then, to estimate GEI effects ($\widehat{gxe_{ij}}$), $\widehat{G_{ij}}$ from each j^{th} environment was explicitly modelled as:

$$\widehat{G_{ij}} = \mu + g_i + e_j + gxe_{ij} \qquad [5]$$

where μ is the overall mean, g_i is the effect of the i^{th} genotype, e_i is the effect of the j^{th} environment, and gxe_{ij} is the random residual effect for the interaction between the i^{th} genotype in the j^{th} environment, IID with $gxe_{ij} \sim N(0, \sigma_{gxe}^2)$ where σ_{gxe}^2 is the genotype by environment variance.

Only trials with associated climatic variables (i.e., trials from the E location) were used here and environment grouping in year and sowing date combination resulted in 187 environments. Of these, 185 environments had information corresponding to *indica* and 175 to *japonica*. The range of genotypes under evaluation per environment was 1 to 363 with a mean of 30 for *indica* and 1 to 418 with a mean of and 93 for *japonica*. The number of shared genotypes under evaluation per environment ranged from 0 to 152 genotypes mean of 2 in *indica* and from 0 to 266 with a mean of 2 in *japonica*.

This was performed using base R (R Core Team, 2021) and *lme4* R package (Bates et al., 2015) under the RStudio environment (RStudio Team, 2020). PLS was performed with the *mdatools* R package (Kucheryavskiy, 2020). To determine the number of latent variables, complete cross-validation was performed. VIP scores and selectivity ratio were calculated and the first three variables of each index, as well as the variables in common between the first 15 variables of each index were selected. When more than one window of the same climatic variable were selected, the one with the higher index value was kept.

2.4.4 Genomic Predictions

2.4.4.1 Model Selection

We incorporated CTWs that the PLS method identified as relevant for $\widehat{G_{ij}}$ and $\widehat{gxe}$ response variables (G and GxE models, respectively) in a single-step RRM by a stepwise forward selection procedure. Briefly, stepwise forward selection involves starting with a baseline model with no covariates (BaseG model for G models and BaseGxE model GxE models) and then testing the addition of each covariate based on the predictive ability. Predictive ability was estimated as the Pearson's correlation between the predicted values and the corresponding $\widehat{G_{ij}}$, based on a random 7-fold cross-validation with 1000 iterations. Wilcoxon rank-sum tests were performed to compare the predictive ability of each model to the previous (less complex) one following Brzozowski et al. (2022). In this manner, the variable whose inclusion gave statistically significant improvement in the prediction ability was added to the model. This stepwise process was repeated for covariates and their interactions in the order in which they improved the baseline model predictive ability until none improved the model fit. FullG and FullGxE models were defined as the last models of the stepwise forward selection procedure that improved prediction predictive ability for the G and GxE models, respectively.

BaseG model consisted of a standard single-step genomic BLUP model for the mean performance of genotypes:

$$y_{ij} = \mu + g_i + e_{ij} \qquad [6]$$

where y_{ij} is the vector of adjusted genotype means for the ith genotype ($\widehat{G_{ij}}$ previously defined) and the jth environment, μ is the overall mean, g_i is the random effect of the ith genotype with $g_i \sim \mathrm{N}(0, \mathbf{H}\sigma_g^2)$ where σ_g^2 is the genotypic variance and **H** is the (co)variance matrix that combines the additive relationship matrix from the pedigree with modifications to account for selfing generations (Rebollo et al., 2020) and the realized genomic relationship matrix (VanRaden, 2008) among genotypes according to Aguilar et al. (2010), e is a random vector of residual effects, IID with $e \sim \mathrm{N}(0, \sigma_\mathrm{e}^2)$ where σ_e^2 is the error variance. This model assumes no covariance between g_i and e_{ij} effects.

BaseGxE model added E_j, the random environmental main effect:

$$y_{ij} = \mu + g_i + E_j + e_{ij} \qquad [7]$$

E_j is IID with $E_j \sim \mathrm{N}(0, \sigma_\mathrm{E}^2)$ where σ_E^2 is the environmental variance and there is no covariance between E_j, g_i, and e_{ijk} effects.

G models consisted of single-step RRM that depending on the stepwise selection stage, incorporated c_{nij} effects, corresponding to the n^th^ selected CTWs covariates for the $\widehat{G_{ij}}$ effect or their interaction as:

$$y_{ij} = \mu + g_{0i} + g_{ni}c_{nij} + e_{ij} \qquad [8]$$

where y_{ij} is the adjusted genotype mean for the i^th^ genotype (G_i previously defined) in the j^th^ environment, μ is the overall mean, g_{0i} and g_{ni} are the random regression coefficients, attributed to the random additive genetic effect of the i^th^ genotype for the intercept, and the response to each unit change in c_{nij}, respectively, with $\begin{matrix} g_{0i} \\ g_{ni} \end{matrix} \sim \mathrm{MVN}(0, \boldsymbol{\Sigma})$. $g_{\mathrm{n}i}$ effects and $\boldsymbol{\Sigma}$ were expanded according to the number of CTWs and interactions being added, for instance, for 1 CTW, where $\boldsymbol{\Sigma} = \mathbf{H} \otimes \begin{bmatrix} \mathrm{Var}(g_0) & Cov(g_0, g_1) \\ Cov(g_1, g_0) & \mathrm{Var}(g_1) \end{bmatrix}$, where **H** is the same (co)variance matrix previously defined for the previous models, $\otimes$ is the Kronecker product. e_{ijk} is the model random residual, IID with $e_{ijk} \sim N(0, \sigma_e^2)$, where σ_e^2 is the error variance.

GxE models are also single-step RRMs, but they differ from G models in that c_{nij} covariates are the CTWs or interactions between CTWs selected for the $\widehat{gxe_{ij}}$ response variable and that E_j which is the main environmental effect as in the BaseGxE model is also included:

$$y_{ij} = \mu + g_{0i} + g_{\mathrm{ni}} c_{nij} + E_j + e_{ij} \quad [9]$$

where all other effects are the same as previously defined for the G models.

Predictions were performed by using the software BLUPF90 (Misztal et al., 2018) with modifications to account for selfing generations (Rebollo et al., 2020). Variance components were estimated by REML once with the full dataset for each model.

2.4.4.2 Prediction in Untested Years and Environments

BaseG, BaseGxE, FullG and FullGxE models were evaluated in untested years and environments prediction scenarios based on their predictive ability. “Leave one year out” and “leave one environment out” cross-validations were performed by leaving one year and environment out at a time, respectively, its data as the validation set, and the remaining data as the training set. “Leave one environment out per ME” cross-validation was performed by splitting the dataset by MEs and performing a “Leave one environment out” cross validation within each ME. In all these scenarios, predictive ability was computed as the mean Pearson’s correlation between the predicted values and their corresponding $\widehat{G_{ij}}$. In order to compare the predictive ability of models trained with the complete dataset and by ME, the results of the “leave one environment out” were also averaged by ME. For the RRMs, covariate values in the validation set were used for predictions. Only years with more than 30 evaluated genotypes and environments with more than 15 evaluated genotypes were considered for computing predictive abilities.

2.5 RESULTS

2.5.1 Genotype by Environment Interaction Characterization

2.5.1.1 Variance Components

Total GEI (the sum of genotype by location, genotype by year, and genotype by sowing period) contributed a large proportion of the total variance, adding up to a total of 12.84% in *indica* and 10.14% in *japonica* (Table 2). The largest genotype by environment component was genotype by location interaction in *indica* and genotype by year interaction in *japonica*. Genotype by sowing period was the smallest component of GEI in both populations. GEI was much larger than the genotypic variance for both populations, with a ratio of 1:8 in *indica* and 1:2.5 in *japonica*. The genotypic variance was larger in *japonica* than in *indica* in both absolute and relative terms. Total environmental variance (the sum of location, year, and sowing period) was the main source of variance in both populations and contributed 45.68% in *indica* and 42.33% in *japonica*. The year component was large in both populations but larger in *japonica*, while the location and sowing period components were relatively more important in indica than in *japonica*. Block and residual variance components were large for both populations.

Table 2. Grain yield (GY) variance component estimates and proportion of the total variance for 3,796 indica genotypes evaluated in 107 YLSs (year, location, and sowing period combinations) and for 9,817 *japonica* genotypes evaluated in 108 YLSs.

Variance component	***indica***		***japonica***	
	GY variance	**Proportion**	**GY variance**	**Proportion**
	kg^2 ha^{-2}	%	kg^2 ha^{-2}	%
Location	743,193	13.9	201,558	5.0
Year	1,074,357	20.2	1,387,187	34.5
Sowing period	616,838	11.6	114,002	2.8
Genotype	82,428	1.5	160,088	4.0
Genotype by location	353,443	6.6	155,079	3.8
Genotype by year	258,791	4.9	193,071	4.8
Genotype by sowing period	71,910	1.4	59,541	1.5
Block	1,023,564	19.2	939,893	23.4
Residual	1,104,357	20.7	811,660	20.2
Total	5,328,880	100	4,022,076	100

2.5.1.2 Mega-environment Definition

Five MEs were found in *indica* (Figure 2a) and three in *japonica* (Figure 2b). The first two components of the GGE model explained 91.3 and 96.9% of the total variance for *indica* and *japonica*, respectively. The number of YLS per ME ranged from 4 to 14 in *indica* and from 13 to 16 in *japonica*. Overall GY per ME in *indica* was 9,818, 9,521, 8,971, 8,791, and 8,342 kg ha^{-1}, for ME3, ME5, ME2, ME1, and ME4, respectively, while in *japonica* it was 8,219, 8,098, and 7,835 kg ha^{-1} for ME2, ME3, and ME1, respectively. A pattern regarding sowing periods was observed in MEs for *indica*. Some MEs that had none or a minority of S1 sowing periods were observed (ME1 and ME5), while others had exclusively or a majority of S1 sowing periods (ME2 and ME3), and ME4 had most intermediate sowing periods and no S4. No pattern of location grouping in the ME was observed in *indica*, as all locations are present in every ME except for the C location in ME3, but this is a very small ME and the C location is also very infrequent. No clear pattern of grouping of locations and sowing periods in the MEs was observed for *japonica* as all sowing periods were observed in all three MEs in *japonica* and ME2 has YLS exclusively from the E location, but this is also the main and most common location.

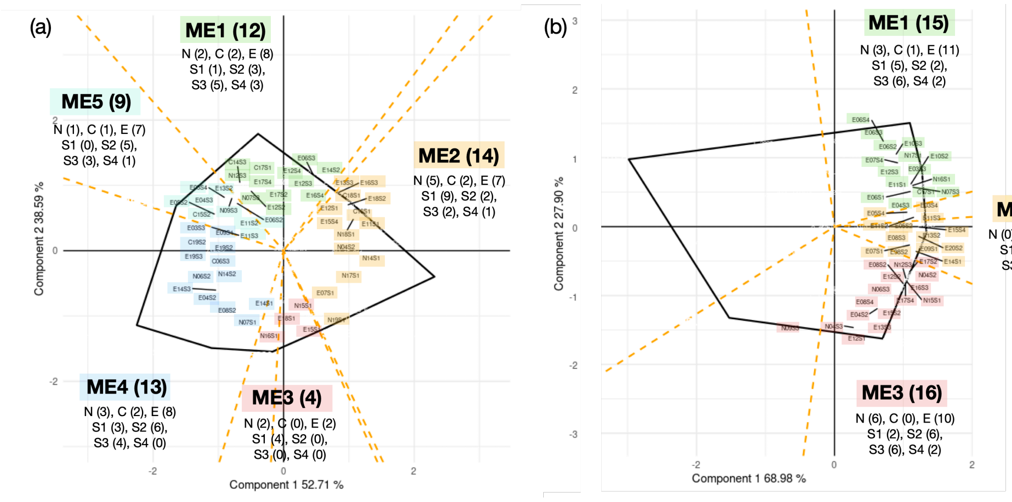
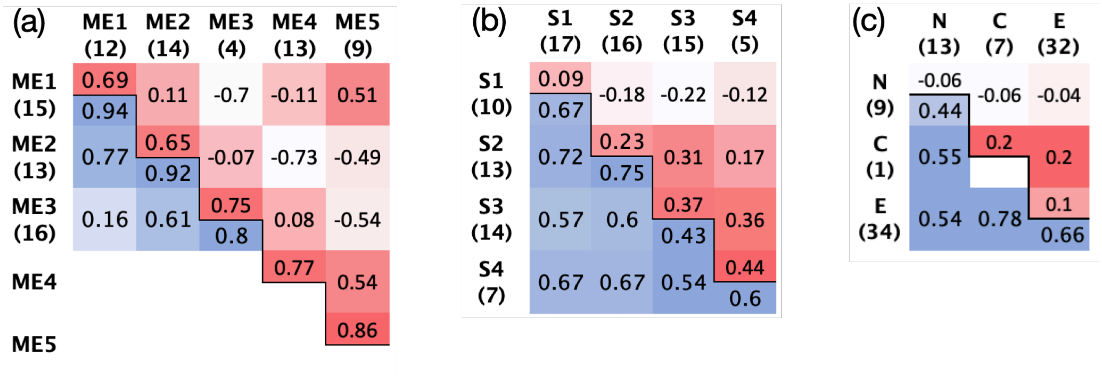


Figure 2. GGE model for grain yield (kg ha^{-1}) graphical augmented biplot representation considering (a) 52 *indica* genotypes evaluated in 52 YLSs (year, location, and sowing period combinations) and (b) 67 *japonica* genotypes in 44 YLSs. The number of YLSs for each mega-environment (ME) are presented inside parenthesis outside the polygon, as well as the number of YLSs for each location (N: the North, C: the Center, E: the East) and sowing periods (S1: early, trials sown until the 15th of October; S2, intermediate-early, trials sown from the 16th of October to the 31st of October; S3, intermediate-late, trials sown from the 1st to the 15th of November; S4, late, trials sown from the 16th of November on).

2.5.1.3 Correlations

The average (ungrouped) pairwise correlation between YLSs for each population was 0.06 for *indica* and 0.62 for *japonica*. Correlations within MEs were high in both populations but lower in *indica* than in *japonica* (Figure 3a). Mean correlation within each sowing period was lower in *indica* than in *japonica* (Figure 3b). In *indica*, the mean correlation between sowing period S1 and all other sowing periods were negative, otherwise, positive correlations were observed. Tfhe mean correlation within and between locations was low in *indica* and intermediate or high in *japonica* (Figure 3c). Mean pairwise correlation within and between combinations of locations and sowing periods (Figure 4) reinforced previous results of correlations between locations and sowing periods alone for both populations: in *indica*, S1 from N and E negatively correlate with the rest of the location and sowing periods, while no groups can be drawn in *japonica*.



(a)

	ME1 (12)	ME2 (14)	ME3 (4)	ME4 (13)	ME5 (9)
ME1 (15)	0.69 0.94	0.11	-0.7	-0.11	0.51
ME2 (13)	0.77	0.65 0.92	-0.07	-0.73	-0.49
ME3 (16)	0.16	0.61	0.75 0.8	0.08	-0.54
ME4				0.77	0.54
ME5					0.86

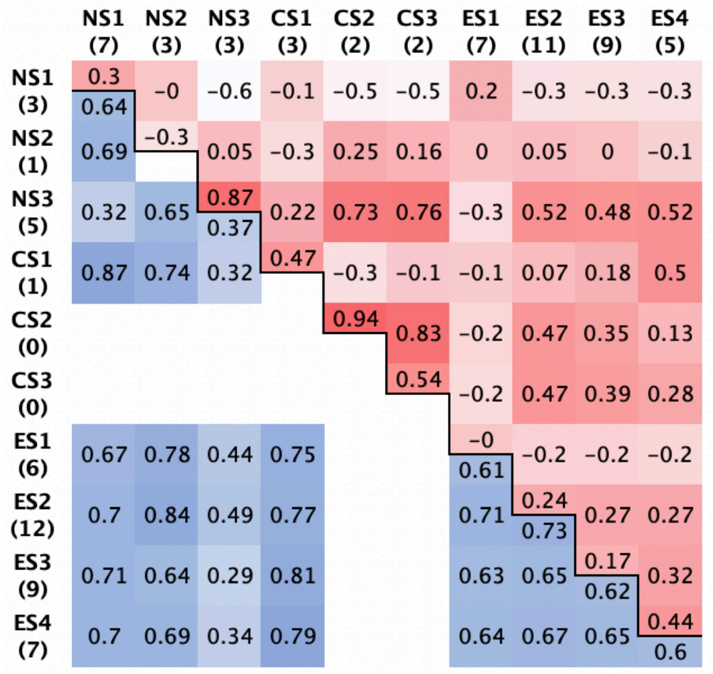
(b)

	S1 (17)	S2 (16)	S3 (15)	S4 (5)
S1 (10)	0.09 0.67	-0.18	-0.22	-0.12
S2 (13)	0.72	0.23 0.75	0.31	0.17
S3 (14)	0.57	0.6	0.37 0.43	0.36
S4 (7)	0.67	0.67	0.54	0.44 0.6

(c)

	N (13)	C (7)	E (32)
N (9)	-0.06 0.44	-0.06	-0.04
C (1)	0.55	0.2	0.2
E (34)	0.54	0.78	0.1 0.66

Figure 3. Mean Pearson's correlation within (diagonal) and between (off-diagonal) pairs of MEs (a), sowing periods (b), and locations (c) for *indica* (above diagonal, colored with red gradient) and *japonica* (below diagonal, colored with blue gradient). The number of YLSs for each category are presented within parenthesis. N: the North location, C: the Center location, E: the East location. S1: early trials, sown until the 15th of October; S2: intermediate-early trials, sown from the 16th of October to the 31st of October; S3: intermediate-late trials, sown from the 1st to the 15th of November; S4: late trials, sown from the 16th of November on.



	NS1 (7)	NS2 (3)	NS3 (3)	CS1 (3)	CS2 (2)	CS3 (2)	ES1 (7)	ES2 (11)	ES3 (9)	ES4 (5)
NS1 (3)	0.3 0.64	−0	−0.6	−0.1	−0.5	−0.5	0.2	−0.3	−0.3	−0.3
NS2 (1)	0.69	−0.3	0.05	−0.3	0.25	0.16	0	0.05	0	−0.1
NS3 (5)	0.32	0.65	0.87 0.37	0.22	0.73	0.76	−0.3	0.52	0.48	0.52
CS1 (1)	0.87	0.74	0.32	0.47	−0.3	−0.1	−0.1	0.07	0.18	0.5
CS2 (0)					0.94	0.83	−0.2	0.47	0.35	0.13
CS3 (0)						0.54	−0.2	0.47	0.39	0.28
ES1 (6)	0.67	0.78	0.44	0.75			−0 0.61	−0.2	−0.2	−0.2
ES2 (12)	0.7	0.84	0.49	0.77			0.71	0.24 0.73	0.27	0.27
ES3 (9)	0.71	0.64	0.29	0.81			0.63	0.65	0.17 0.62	0.32
ES4 (7)	0.7	0.69	0.34	0.79			0.64	0.67	0.65	0.44 0.6

Figure 4. Mean Pearson's correlation within (diagonal) and between (off-diagonal) pairs of combinations of locations and sowing periods for *indica* (above diagonal, colored with red gradient) and *japonica* (below diagonal, colored with blue gradient). The number of YLSs for

each combination of sowing date and location are presented within parenthesis. N: the North location, C: the Center location, E: the East location. S1: early trials, sown until the 15th of October; S2: intermediate-early trials, sown from the 16th of October to the 31st of October; S3: intermediate-late trials, sown from the 1st to the 15th of November; S4: late trials, sown from the 16th of November on.

2.5.2 Climatic Variable Selection

Selected CTW for each response variable are presented in Table 3. In *indica*, evapotranspiration and maximum temperature was found significant for the three effects, and wind was a shared selected variable between $\widehat{G_{ij}}$ and $\widehat{g_{ij}}$ effects. In *japonica*, mean temperature and wind speed were both shared between two of the three effects. Three CTW overlapping between effects was found in *indica* (MaxTemp 20-20, Evap 20-10, Wind 5-35) while none in *japonica*. A pattern regarding time windows was observed for *japonica* in the $\widehat{G_{ij}}$ effect, where all selected CTW correspond to the 35-5 time window, suggesting an earlier critical period is important. In *indica*, the 20-20 time window was important for the $\widehat{gxe_{ij}}$ effect, suggesting a more balanced around flowering critical period.

Table 3. Climatic variables and time windows (days before-after flowering date) selected through partial least squares (PLS) for each response variable ($\widehat{G}_{ij}$: adjusted genotype means in each environment, $\widehat{g}_{ij}$: genotype effect in each environment, $\widehat{gxe}_{ij}$: genotype-by-environment effect in each environment), ordered alphabetically, and variables (underscored) and interactions selected through the stepwise forward selection method for the *indica* and *japonica* breeding population. FullG model: selected random regression for genotype effect and climatic variables for the $\widehat{G}_{ij}$ effect, FullGxE model: selected random regression for genotype effect and CTWs for the $\widehat{gxe}_{ij}$ effect.

Population	***indica***			***japonica***		
Response variable	$\widehat{G}_{ij}$	$\widehat{g}_{ij}$	$\widehat{gxe}_{ij}$	$\widehat{G}_{ij}$	$\widehat{g}_{ij}$	$\widehat{gxe}_{ij}$
Variables selected through PLS	Evap 10-30 MaxTemp 20-20 Rad 20-10 Wind 5-35	Evap 20-10 MaxTemp 20-20 Wind 5-35	Evap 20-10 HeatU 20-20 MaxTemp 20-20 MeanTemp 20-20	Ampl 35-5 HeatU 35-5 MaxTemp 35-5 MeanTemp 35-5	Evap 35-5 PpDay 10-10 SunHs 35-5 Wind 10-30	MeanTemp 10-20 Precip 10-10 Rad 5-35 RelH 5-35 Wind 10-10
Interactions in FullG models	Evap 10-30 x MaxTemp 20-20 Evap 10-30 x Rad 20-10			Ampl 35-5 x MaxTemp 35-5 Ampl 35-5 x MeanTemp 35-5		
Interactions in FullGxE models				MeanTemp 10-20 x Rad 5-35 Rad 5-35 x Precip 10-10 Precip 10-10 x Wind 10-10		

2.5.3 Genomic Predictions

2.5.3.1 Model Selection

All RRMs including climatic variables improved predictive ability compared to the base models both in *indica* (Figure 5a) and in *japonica* (Figure 5b). The main and interaction terms selected for the FullG and FullGxE models in both populations are presented in Table 3. In *indica*, the FullG model had a mean predictive ability of 0.65, 28.3% higher than the BaseG model, while the FullGxE model had a mean predictive ability of 0.76, which was 1.9% higher than the BaseGxE model. In *japonica*, the FullG model reached a mean predictive ability of 0.75, 11.4% higher than the BaseG model, while the FullGxE model had a mean predictive ability of 0.86, 1.6% higher than the BaseGxE model. Wilcoxon rank-sum tests were statistically significant (p-value < 0.01) for every comparison that was made. Models with additional variables and interactions than FullG and FullGxE failed to converge, except for the last forward selection step in *indica* that lowered the predictive ability.

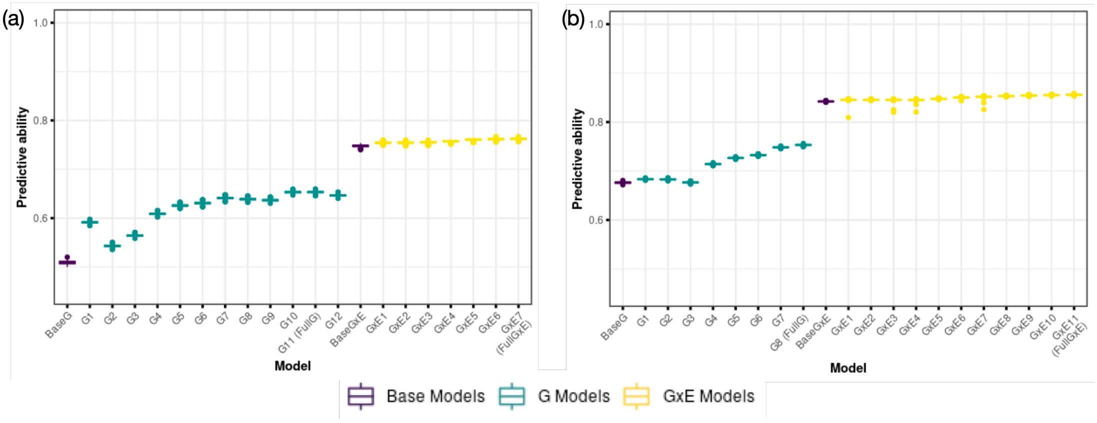


Figure 5. Boxplot of predictive ability measured as the correlation between predicted vs. observed grain yield (kg ha^{-1}) from the 1000 iterations of the 7-fold cross-validation for the *indica* (a) and *japonica* (b) rice breeding populations with the models compared in this study BaseG: genotypic main effect, G1 to G12: random regression for genotype effect and climatic covariates for the $\widehat{G_{ij}}$ effect in each step of the forward procedure, BaseGxE: genotypic and environmental main effects, GxE1 to GxE11: random regression for genotype effect and climatic covariates for the $\widehat{gxe}_{ij}$ effect and environmental main effect for each step of the forward procedure.

2.5.3.2 Prediction in Untested Years and Environments

RRMs (either FullG or FullGxE) outperformed models without climatic covariates for predicting untested years and environments in both populations, on average (Table 4). The best model for predicting untested years varied across years (Figure 6), but either RRM showed higher predictive ability than both baseline models in 13 out of 18 years (72.2%) in *indica* and in 12 out of 21 years (57.1%) in *japonica*. In *indica*, the overall predictive ability for untested years with both RRM was above 40% higher than either baseline model. Conversely, FullG was outperformed by all others when predicting untested environments, showing FullGxE the best mean predictive ability for this scenario. In *japonica*, the FullGxE model showed the highest overall predictive ability both in untested environments and years by 4-31%. Training with the complete dataset was better than training within each ME in all MEs except for ME5 in *indica* (Table 5). ME5 has absence of S1 sowing periods, and given the low correlation between S1 and the other sowing periods, we speculate that S1 from other MEs in the training population hinder in this case the predictive ability of using the complete dataset.

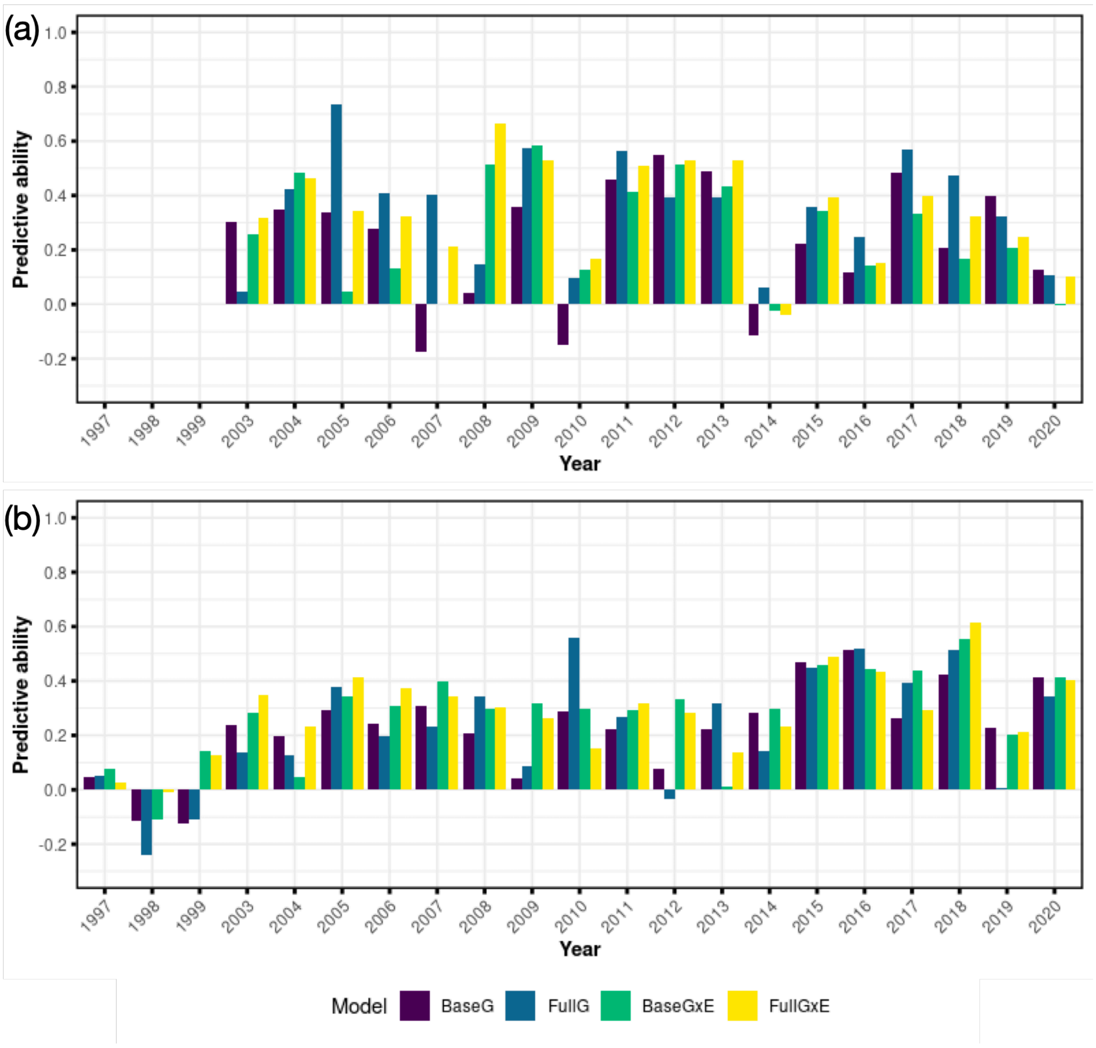


Figure 6. Predictive ability measured as the correlation between predicted vs. observed grain yield (kg ha^{-1}) values for the “leave one year out” prediction scenario for predicting untested years with the tested models (BaseG: genotypic main effect, FullG: selected random regression for genotype effect and climatic variables for the $\widehat{G_{ij}}$ effect, BaseGxE: genotypic and environmental main effects, FullGxE: selected random regression for genotype effect and CTWs for the $\widehat{gxe}_{ij}$ effect and environmental main effect) for the (a) *indica* and (b) *japonica* rice breeding populations.

Table 4. Overall predictive ability measured as the Pearson's correlation between predicted and observed grain yield values for the “leave one year out” and “leave one environment out” prediction scenarios for predicting untested environments prediction scenarios with the tested models (BaseG: genotypic main effect, FullG: selected random regression for genotype effect and climatic variables for the $\widehat{G_{ij}}$ effect, BaseGxE: genotypic and environmental main effects, FullGxE: selected random regression for genotype effect and CTWs for the $\widehat{gxe}_{ij}$ effect and environmental main effect) for the *indica* and *japonica* rice breeding populations. The best performing model for each scenario and population was underscored.

Population	Prediction scenario	BaseG	FullG	BaseGxE	FullGxE
indica	**“leave one year out”**	0.24	0.35	0.24	0.34
	“leave one environment out”	0.28	0.22	0.32	0.32
japonica	**“leave one year out”**	0.23	0.22	0.28	0.29
	“leave one environment out”	0.3	0.26	0.34	0.36

Table 5. Overall predictive ability by mega-environment (ME) measured as the Pearson's correlation between predicted and observed grain yield values for the prediction scenarios "leave one environment out" (trained with the complete dataset) and the "leave one environment out per mega-environment out" (trained by ME) for predicting untested environments prediction scenarios with the tested models (BaseG: genotypic main effect, FullG: selected random regression for genotype effect and climatic variables for the $\widehat{G_{ij}}$ effect, BaseGxE: genotypic and environmental main effects, FullGxE: selected random regression for genotype effect and CTWs for the $\widehat{gxe_{ij}}$effect and environmental main effect) for the *indica* and *japonica* rice breeding populations. The number of environments considered for averaging on each ME are presented within parenthesis. The best performing model for each ME and population was underscored.

Population	ME	Training population							
		Complete dataset				ME			
		Model				Model			
		BaseG	FullG	BaseGxE	FullGxE	BaseG	FullG	BaseGxE	FullGxE
indica	1 (14)	0.29	0.12	0.33	0.28	0.2	0.2	0.21	0.22
	2 (8)	0.28	0.21	0.35	0.32	0.19	0.19	0.27	0.26
	3 (2)	0.44	0.45	0.56	0.62	0.4	0.4	0.39	0.4
	4 (13)	0.24	0.15	0.28	0.31	0.14	0.14	0.23	0.21
	5 (10)	0.32	0.29	0.25	0.27	0.32	0.32	0.26	0.3
japonica	1 (24)	0.3	0.22	0.34	0.35	0.2	0.13	0.24	0.25
	2 (32)	0.39	0.33	0.42	0.42	0.29	0.28	0.38	0.39
	3 (23)	0.37	0.31	0.37	0.46	0.25	0.27	0.32	0.42

2.6 DISCUSSION

In this work we first characterized GY GEI with a set of tools: variance component estimation, ME definition and correlation between locations, sowing periods, and MEs. Then, we identified the most influential climatic covariates for GY and GEI. Finally, we demonstrated that GEI can be modelled using random regression single-step models based on climatic covariates and evaluated these models in tested and untested environments. These methods were applied to a large and highly unbalanced MET data of two different rice populations of subtropical rice (*indica* and *japonica*). Our results showed large GEI in both populations of the IRBP dataset, while it was larger in *indica* than in *japonica.* RRMs with selected climatic covariates improved the predictive ability for both tested and untested years and environments.

2.6.1 Genotype by Environment Interaction Characterization

By combining complementary ways of characterizing GEI, we were able to provide a comprehensive insight into the GEI structure in our dataset. Although this GEI characterization is useful to design more informed and efficient resource allocation strategies, the unbalanced and sparse nature of a breeding program dataset (which in our case had 71.6 and 74,0% missing and imputed datapoints in *indica* and *japonica,* respectively), demands caution on this regard. The genotypes that are tested on different sowing dates and locations are the more advanced (selected) ones and thus have less genetic variance in the selection environment, possibly affecting the estimated correlations between and within groups of environments.

We observed important levels of GEI in both populations as GEI variance represented above 10% of the total variance and more than one ME was identified. Furthermore, part of the large observed residual variance in both populations may be due to the GEI corresponding to triple or higher order interactions between CTWs that were not accounted for in our models. Finding high levels of GEI was expected in a large dataset covering a span of 23 years, in three different locations. Moreover, GEI has a strong effect on quantitative traits (Bernardo, 2010) and rice GY is a well characterized quantitative trait (Xing & Zhang, 2010). Conversely, Monteverde et al. (2019) observed smaller GEI variance for a subset of our germplasm (5.5% for *indica* and 5.1% for *japonica*), but they analyzed three years of testing (2010-2012) and circumscribed to a single location. Similar or higher percentages of GEI were reported for other cereals in Uruguayan climatic conditions across many years and locations: Lado et al. (2016) found 7.6-16.9% of the variance for different yield testing stages due to GEI across five locations and five years in wheat, and Ceretta and van Eeuwijk (2008) found 14% of variance due to GEI in eight years and six testing locations in barley. The ratio between the genotypic and GEI variances was also within observed by these authors. This highlights the relevance of studying GEI, as it drives the larger relevant variation in a breeding program.

GEI was larger in *indica* than in *japonica*. We found GEI representing 12.84% of the total variance in the former and 10.14% in the latter. Moreover, we identified five MEs in *indica* and three in *japonica* and a higher contrast among MEs (i.e.: crossover GEI) was found in *indica*. Mean correlations between both ungrouped and grouped YLSs were lower in *indica* than in *japonica*. Furthermore, negative mean correlations were observed between ME in *indica* but not in *japonica*. These populations have different origins that determine their level of climatic adaptation (Khush, 1997). Low temperature is one of the most important climatic factors affecting rice in temperate regions (Kumar et al., 2021). Thus, the higher differential response to environmental changes in *indica* than in *japonica* in the Uruguay conditions can be explained by *indica* being more sensitive to cold than *japonica* (Shakiba et al., 2017).

Although genotype by sowing period interaction was the smallest GEI component in both populations, in *indica*, some MEs were dominated by S1 (i.e., early sowing period) and others had complete absence of it. Crossover GEI across sowing periods was observed for *indica* with negative correlations between S1 and every other sowing period. These results suggest that trials with different sowing dates are neccesary in the evaluation of *indica* germplasm in Uruguay. Early sowing period contributes to more favorable climatic conditions due to less incidence of lower temperature and higher radiation in the Uruguayan climatic conditions (Pérez de Vida, 2010). Given that *indica* materials are known to be negatively affected by low temperature (Yoshida, 1981), early sowing is recommended for *indica* cultivars in Uruguayan (Tseng et al., 2021). Moreover, the gains in productivity observed during the last years in the Uruguayan rice production have been attributed to the widespread adoption of early sowing of *indica* cultivars (Molina et al., 2021). On the other hand, *japonica* materials are more adapted to low temperatures than *indica* (Shakiba et al., 2017) and hence are less sensitive to the sowing period. Consequently, no grouping of sowing periods in the MEs nor negative correlations between sowing periods or their combination with locations were observed for *japonica*, suggesting that testing in different sowing periods might not be necessary for *japonica* evaluation.

GEI within and between locations was observed only in *indica*, with correlations within and between locations below 0.20 and genotype by location as the largest GEI variance component in this population. In *japonica,* where correlations within and between locations were above 0.40 and, in some cases, the correlation between was higher than the correlation within locations. This suggests that evaluation in different locations might not be necessary in *japonica*.

Some studies have been able to associate MEs with clear environmental differences like geographical regions (Yan et al., 2000) or environmental productivity (Lado et al., 2016), but this was not the case in our study. Although grouping by ME provided a useful characterization of GEI, it did not match with any repeatable pattern of location, sowing periods, nor its combinations, making this approach unsuitable for GP in untested environments in our study.

2.6.2 Climatic Variable Selection

The inclusion of climatic variables in GP models is a promising tool to predict performance in untested environments (Bustos-Korts et al., 2015). Some studies used a whole set of more than 50 available variables in the prediction model (Jarquín et al., 2014) but such a large number of covariates makes each explain a small amount of the total variance and them to be highly correlated with each other (Brancourt-Hulmel et al., 2000). This increases collinearity which has undesirable consequences on model inferences such as instability in parameter estimation (Dormann et al., 2013). Thus, variable selection has been reported as beneficial (Buntaran et al., 2021; Neyhart et al., 2022). Several methods for variable selection have been presented in the literature: R2 (Piepho, 2019), F-test (Buntaran et al., 2021), stepwise feature selection (Neyhart et al., 2021), and PLS (Vargas et al., 1998). Consensus on which method works best has not been reached and further research needed in that field has been suggested (Buntaran, 2021). We used PLS, a method frequently used in the chemometrics community that is especially useful for selecting highly correlated variables (Wold et al., 2001) like the ones we had.

Our results matched previously reported climatic variables causing abiotic stress and helped us gain further insight into them. Climatic variables related to temperature, wind, radiation, and precipitation from different time windows were found significant for the different analyzed response variables related to GY. Temperature and wind during reproduction negatively affect GY by causing spikelet sterility (Yoshida, 1981; Marchezan & da Silva Aude, 1993; Raju et al., 2013; Alvarado, 2002). Moreover, radiation is directly involved in grain filling capacity. Precipitation does not affect water availability in a flooded paddy system (Carracelas et al., 2019), but it was strongly negatively correlated with radiation in our dataset (Supplemental Fig. S1). Hence, precipitation can be seen here as an indirect and inverse measure of radiation. Monteverde et al. (2019) also found a different set of significant climatic variables and time windows for *indica* and *japonica* populations both for GY and milling quality traits. This highlights the importance of breeding and analyzing *indica* and *japonica* separately. Temperature, evapotranspiration, and wind variables for GY in *indica* and temperature and wind variables in *japonica* were also found significant elsewhere (Monteverde et al., 2019). Within each population, different sets of climatic variables were selected for each of the response variables we analyzed. The selected climatic variables for the $\widehat{G_{ij}}$ effect capture environmental effects that may be managed through agronomic practices. On the other hand, climatic variables related to $\widehat{g_{ij}}$ and $\widehat{gxe_{ij}}$ effects, which have a purely genetic component, reflect environmental factors that differentially affect some genotypes and their adaptation. Thus, dissecting $\widehat{g_{ij}}$ and $\widehat{gxe_{ij}}$ is of the special interest for the breeder.

2.6.3 Genomic Predictions

2.6.3.1 Model Selection

RRMs with climatic covariates are beneficial only if appropriate covariates are used. After initially selecting covariates on a biological consideration basis, model fit improvement with the candidate covariates is necessary (Buntaran et al., 2021). In this work, we implemented a two-stage covariate selection procedure: first CTWs affecting GY were identified by PLS, and second, incorporated to a RRM and selected by stepwise forward selection based on predictive ability. The model fit metric we used for CTWs selection (i.e., prediction ability based on 7-fold cross-validation), recreated a scenario with unbalanced data between environments, with tested or untested genotypes in tested environments (Malosetti et al., 2016). The progressive inclusion of CTWs and their interactions lead in some cases to convergence failure in the last steps of the forward procedure, which could have been caused by overparameterization and unsuitable variable centering and scaling (Buntaran et al., 2021). Nevertheless, marginal increments in predictive ability were observed during all the last steps of the stepwise forward selection.

Our procedure was successful and allowed us to select the model with the highest predictive ability. In both populations, BaseGxE was better than BaseG, which highlights the importance of separately modelling the environmental effect. All RRMs including PLS-selected climatic variables for both the $\widehat{G_{ij}}$ and $\widehat{gxe_{ij}}$ effect improved predictive ability compared to base models in both populations. This means that covariates capture, even partially, the environmental and GEI effects. The fact that FullG model had lower predictive ability than the BaseGxE model in both populations suggests that selected climatic variables cannot fully account for the main environmental effect. Jarquín et al. (2014) improved their models' fit by adding a residual term for the environment effect. This residual term models the variability that environmental variables failed to capture, but the main limitation of this is that it can only be done for tested environments. The relative increment in prediction ability of FullG over BaseG (28.3 and 11.4% for *indica* and *japonica,* respectively) were higher than those found in FullGxE over BaseGxE (1.9 and 1.6% for *indica* and *japonica,* respectively). The higher contribution of $\widehat{G_{ij}}$ related climatic variables than $\widehat{gxe_{ij}}$ related variables to the baseline models in terms of predictive ability could be attributable to the fact that environmental variance is larger than the GEI-related variance. We observed lower general predictive ability for *indica* than *japonica* which could be explained both by the larger *japonica* training data, and by the smaller GEI variance in *japonica*. Furthermore, the gain in the predictive ability of the RRMs over the baseline models was higher in *indica* than in *japonica*, showing that the modelling of the environment and GEI with climatic covariates is more beneficial when more GEI is present.

2.6.3.2 Prediction in Untested Years and Environments

Predicting untested years and environments is the most challenging issue in plant breeding when GEI is significant (Malosetti et al., 2016). RRMs with the best model fit based on the 7-fold cross-validation predictive ability metric (i.e., FullG and FullGxE models) were selected for predicting under such a scenario. The improvements we observed for RRMs over baseline models were higher than the ones obtained by Monteverde et al. (2019) comparing a reaction norm model including environmental covariates with GBLUP for the "leave one year out" prediction scenario in a subset of our dataset. Even though the environmental effect cannot be predicted for untested environments, BaseGxE models showed higher predictive abilities than BaseG models due to the benefit of modelling the environmental effect in the training population. Either RRM outperformed both baseline models for the majority of untested years in both populations. Even though, the high inter-annual variability of the Uruguayan climatic conditions (Tiscornia et al., 2016) and GEI determined that the best model for predicting unobserved years varied across years. This could be due to the occasional relative importance of other climatic variables during those specific years, because of a non-linear GEI response to modelled variables, or both (Bose et al., 2012). The highest overall predictive ability for untested years in *indica* was obtained with the FullG model, but it had only 0.01 higher predictive ability than the FullGxE model. In *japonica*, the latter was the best model for predicting untested years. For untested environments, the highest overall predictive ability was obtained with the FullGxE model for both populations. In sum, our RRMs (either FullG or FullGxE) showed higher predictive ability on average than models without covariates for predicting untested years and environments, in both populations. Based on the joint consideration of results across our different prediction scenarios, the FullGxE model, which includes the environmental effect and CTWs selected for the $\widehat{gxe_{ij}}$ effect, can be recommended.

MEs represent repeatable patterns of GE because, by definition, GEI within ME is minimized (Braun et al., 2016). Thus, selecting genotypes within each ME goes in line with the strategy to handle GEI by reducing it. Given sufficient breeding resources are allocated to each ME, trait heritability and selection efficiency could be enhanced in this manner (Windhausen et al., 2012). Moreover, some studies have reported improvements in predictive ability by predicting within MEs (Burgueño et al., 2012; Dawson et al., 2013; Lado et al., 2016). Our results showed that predicting untested environments within ME was not consistently better than predicting untested environments with the complete dataset. Windhausen et al. (2012) had not observed improvements in predicting within subdivided environments either. A possible explanation for this is that any eventual interaction between genotype's response to each covariate across MEs is not strong enough to compensate the loss of predictive ability due to using a reduced dataset.

2.6.4 Final Remarks

We showed the usefulness of RRMs for GY prediction in a real plant breeding dataset and in both tested and untested years and environments prediction scenarios. This was achieved in a GS framework that accommodates phenotypic information of genotypes without genomic information (i.e.: single-step or HBLUP) accounting for selfing generations.

RRMs were first introduced in animal breeding by Henderson (1982) and Laird and Ware (1982). RRMs have been extensively used for the analysis of test day milk production records in dairy cattle breeding, as well as for growth traits, survival analysis, fertility (Schaeffer, 2004) and GEI in the reaction of cows to heat stress (Ravagnolo & Misztal, 2000). RRMs have later been used in plant breeding for time covariates in wheat (Sun et al., 2017) and in rice (Campbell et al., 2018, 2019; Baba et al., 2020) and more recently to accommodate per-trial environmental data for analyzing wheat MET (Buntaran et al., 2021; Tolhurst et al., 2021). We extended this by computing per-plot climatic covariates. RRMs in a large dataset like ours are useful because covariate information from a wide range of environmental conditions is covered and can be exploited to model GEI (Schaeffer, 2004). If covariate values in the new environment are available, the performance of genotypes in new environments can be predicted with RRMs (Buntaran et al., 2021). Several tools are available nowadays that make it easier the access to daily climatic data like the National Aeronautics and Space Administration (NASA) Prediction of Worldwide Energy Resource (POWER) project (https://power.larc.nasa.gov/). Jarquín et al. (2014) proposed a reaction norm model that has been used in several other studies (Lado et al., 2016; Saint Pierre et al., 2016; Monteverde et al., 2019; Jarquín et al., 2021; Neyhart et al., 2022). In that reaction norm model, an extensive number of environmental covariates are used to compute an environmental covariance matrix, possessing the unrealistic assumption that there is a single environmental variance component. Moreover, slopes for the different covariates have the same variance and there is no correlation between them. This represents a key difference from the RRMs we used, given that we allowed (co)variance between and within intercept and slope to be a free parameter, which is essential to ensure invariance with respect to translation and scale transformation of the covariates (Wolfinger, 1996). Heslot et al. (2014) reported not being able to implement a RRM with such a covariance structure in a similarly large wheat dataset due to extensive RAM usage with ASReml-R. We were able to easily accommodate it for more than one covariate by using blupf90, a software that is extensively used for animal breeding. Still, fitting a large number of covariates in RRMs does represent a challenge. Dimension reduction models like FA that include environmental covariates have been proposed and resulted in larger variance due to environmental known factors being explained than RRMs (Tolhurst et al., 2021). However, FA structures cannot be used for prediction in untested environments, so RRMs still are an appealing approach.

In all our RRMs, the response of genotypes to climatic covariates was modelled following a linear trend. Buntaran et al. (2021) observed averages of the standard error of prediction dropping considerably with the addition of a quadratic term in their RRM. Moreover, the use of orthogonal Legendre polynomials in RRM has been proven to benefit by offering numerical stability by reducing the correlation between random regression coefficients and computing error (Schaeffer, 2004). Despite RRMs being invariant to translation and scale transformation of the covariates, the scale of the covariate has been reported to affect model convergence and the estimates of the covariance parameters (Buntaran et al., 2021). Thus, additional research including the use of Legendre polynomials of different orders, as well as different scaling and centering of covariates adjustment, could allow going one step further on the understanding of the nature of GY GEI and our ability to predict it. Additionally, the intercept and slope estimates we obtained for each genotype with the RRMs could be used to perform genome-wide association studies that can help to understand the genetic factors underlying environmental responses (Li et al., 2021).

In sum, in this work we characterized GEI in two rice breeding populations, identified climatic variables affecting GY, and incorporated them in mixed RRM to predict GY in tested and untested environments. We were able to extract broad conclusions about the nature of GEI in the Uruguayan rice production conditions, showing large amounts of GEI over GY differently affecting the *indica* and *japonica* populations we analyzed. We demonstrated that RRMs can be used to improve GY predictive ability in a realistic plant breeding scenario. The use of these models will be especially beneficial for crop breeding programs that are focused on yield or other quantitative traits that are affected by environmental conditions and show strong GEI.

2.7 REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. *Journal of Dairy Science,* 93(2), 743–752. https://doi.org/10.3168/jds.2009-2730

Alvarado, J. R. (2002). Influence of air temperature on rice population, length of period from sowing to flowering and spikelet sterility. *Proceedings of the second temperate rice conference.* International Rice Research Institute.

Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., Tinker, N. A., & Jannink, J. L. (2013). Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. *Crop Science,* 53(5), 1894–1906. https://doi.org/10.2135/cropsci2012.09.0526

Baba, T., Momen, M., Campbell, M. T., Walia, H., & Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PLoS ONE,* 15(2), 1–17. https://doi.org/10.1371/journal.pone.0228118

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

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science,* 48(5), 1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo, R. (2010). *Breeding for quantitative traits in plants.* 2nd ed. Stemma Press, Woodbury, MN. ISBN 978-0-9720724-1-0

Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics,* 129(12), 2323-2332. https://doi.org/10.1007/s00122-016-2772-5

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., … Crossa, J. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science,* 55(1), 154–163. https://doi.org/10.2135/cropsci2014.07.0460

Bonnecarrère, V., Quero, G., Monteverde, E., Rosas, J., Pérez de Vida, F., Cruz, M., … Borsani, O. (2015). Candidate gene markers associated with cold tolerance in vegetative stage of rice (*Oryza sativa* L.). *Euphytica,* 203(2), 385–398. https://doi.org/10.1007/s10681-014-1290-2

Bose, L., Nagaraju, M., & Singh, O. (2012). Genotype x environment interaction and stability analysis of lowland rice genotypes. *Journal of Agricultural Sciences,* Belgrade, 57(1), 1–8. https://doi.org/10.2298/jas1201001b

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23, 2633-2635

Brancourt-Hulmel, M., Denis, J. B., & Lecomte, C. (2000). Determining environmental covariates which explain genotype environment interaction in winter wheat through probe genotypes and biadditive factorial regression. *Theoretical and Applied Genetics,* 100(2), 285–298. https://doi.org/10.1007/s001220050038

Braun, H.-J., Rajaram, S., & van Ginkel, M. (1996). CIMMYT’s approach to breeding for wide adaptation. *Euphytica,* 92(1–2), 175–183. https://doi.org/10.1007/BF00022843

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics,* 103(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Brzozowski, L. J., Campbell, M. T., Hu, H., Caffe, M., Gutiérrez, L., Smith, K. P., … Jannink, J. (2022). Generalizable approaches for genomic prediction of metabolites in plants. *The Plant Genome*, 15(2), 1–20. https://doi.org/10.1002/tpg2.20205

Buntaran, H., Forkman, J., & Piepho, H.-P. (2021). Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. *Theoretical and Applied Genetics,* 134(5), 1513–1530. https://doi.org/10.1007/s00122-021-03786-2

Burgueño, J., Crossa, J., Cornelius, P. L., & Yang, R. C. (2008). Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. *Crop Science,* 48(4), 1291-1305. doi:10.2135/cropsci2007.11.0632

Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Science,* 52(2), 707–719. https://doi.org/10.2135/cropsci2011.06.0299

Bustos-Korts, D., Malosetti, M., Chapman, S., & van Eeuwijk, F. (2015). Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. In X. Yin & P. C. Struik (Eds.), *Crop Systems Biology: Narrowing the gaps between crop modelling and genetics* (pp. 55–82). https://doi.org/10.1007/978-3-319-20562-5

Butler, D. G., Cullis, B.R., A. R. Gilmour, Gogel, B.G. & Thompson, R. (2017). *ASReml-R Reference Manual Version 4.* VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Campbell, M., Momen, M., Walia, H., & Morota, G. (2019). Leveraging Breeding Values Obtained from Random Regression Models for Genetic Inference of Longitudinal Traits. *The Plant Genome,* 12(2), 180075. https://doi.org/10.3835/plantgenome2018.10.0075

Campbell, M., Walia, H., & Morota, G. (2018). Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. *Plant Direct,* 2(9), 1–11. https://doi.org/10.1002/pld3.80

Carracelas, G., Hornbuckle, J., Rosas, J., & Roel, A. (2019). Irrigation management strategies to increase water productivity in *Oryza sativa* (rice) in Uruguay. *Agricultural Water Management,* 222, 161–172. https://doi.org/10.1016/j.agwat.2019.05.049

Ceretta, S., & van Eeuwijk, F. (2008). Grain yield variation in malting barley cultivars in Uruguay and its consequences for the design of a trials network. *Crop Science,* 48(1), 167–180. https://doi.org/10.2135/cropsci2006.06.0428

Crossa, J. (2012). From Genotype × Environment Interaction to Gene × Environment Interaction. *Current Genomics,* 13(3), 225–244. https://doi.org/10.2174/138920212800543066

Crossa, J., & Cornelius P. L. (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Science,* 37, 406–415. doi:10.2135/crop sci1997.0011183X003700020017x

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., … Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science,* 22(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., … Jannink, J. L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research,* 154, 12–22. https://doi.org/10.1016/j.fcr.2013.07.020

De los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics,* 193(2), 327–345. https://doi.org/10.1534/genetics.112.143313

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography,* 36(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dumble, S. (2022). *GGEBiplots: GGE Biplots with 'ggplot2'.* R package version 0.1.3. https://CRAN.R-project.org/package=GGEBiplots

Fehr, W. R. (1987). *Principles of Cultivars Development. Vol. 1.* Macmillan, New York.

Finlay, K., & Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research,* 14(6), 742. https://doi.org/10.1071/AR9630742

Henderson, C.R. (1982). Analysis of covariance in the mixed model: higher level, nonhomogeneous, and random regressions. *Biometrics,* 38, 623–640.

Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J. L. (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. https://doi.org/10.1007/s00122-013-2231-5

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., … de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics,* 127(3), 595–607. https://doi.org/10.1007/s00122-013-2243-1

Jarquín, D., de Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., … Lorenz, A. (2021). Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. *Frontiers in Genetics,* 11(March), 1–11. https://doi.org/10.3389/fgene.2020.592769

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

Khush G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology,* 35, 25–34.

Kucheryavskiy, S. (2020). mdatools – R package for chemometrics. *Chemometrics and Intelligent Laboratory Systems,* 198, 103937. 10.1016/j.chemolab.2020.103937

Kumar, M., Sandhya, S., Kumar, P., Singh, A. G., & Jukanti, A. K. (2021). Abiotic Stress Tolerance in Rice: Insight in Climate Change Scenario. In *Integrative Advances in Rice Research.* IntechOpen. https://doi.org/10.5772/intechopen.98909

Lado, B., Barrios, P. G., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype × Environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Science,* 56(5), 2165–2179. https://doi.org/10.2135/cropsci2015.04.0207

Laird, N.M., & Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics,* 38, 963–974.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, X., Guo, T., Wang, J., Bekele, W. A., Sukumaran, S., Vanous, A. E., … Yu, J. (2021). An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Molecular Plant,* 14(6), 874–887. https://doi.org/10.1016/j.molp.2021.03.010

Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters,* 2. doi:10.1088/1748-9326/2/1/014002

Malosetti, M., Bustos-Korts, D., Boer, M. P., & van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype x Environment interactions. *Crop Science,* 56(5), 2210–2222. https://doi.org/10.2135/cropsci2015.05.0311

Malosetti, M., Ribaut, J. M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, 4:44. https://doi.org/10.3389/fphys.2013.00044

Marchezan, E., & da Silva Aude, M. I. (1993). Adverse effect of the wind in irrigated rice. *Ciência Rural,* 23, 379–381. https://doi.org/10.1590/S0103-84781993000300025

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A., & Vitezica, Z. (2018). *Manual for BLUPF90 family of programs*. http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Molina, F., Terra, J., Roel, A., Oxley, A., Marella, M., Casterá, F., Platero, A., García, F., Rovira, G., & Escostegui, C. (2021). Indicadores tecnológicos-productivos zafra arrocera 2020-2021. In J. Terra, S. Martínez, H. Saravia, & B. Mesones (Eds.), *Arroz 2020* (pp. 1-5). Montevideo: INIA.

Monteverde, E., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J. E., Bonnecarrère, V., … McCouch, S. (2019). Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (*Oryza sativa* L.) Grown in Subtropical Areas. *G3 Genes|Genomes|Genetics,* 9(5), 1519–1531. https://doi.org/10.1534/g3.119.400064

Monteverde, E., Rosas, J. E., Blanco, P., Pérez de Vida, F., Bonnecarrère, V., Quero, G., … McCouch, S. (2018). Multienvironment Models Increase Prediction Accuracy of Complex Traits in Advanced Breeding Lines of Rice. *Crop Science,* 58(4), 1519–1530. https://doi.org/10.2135/cropsci2017.09.0564

Neyhart, J. L., Gutiérrez, L., & Smith, K. P. (2021). Optimizing the choice of test locations for multi-trait genotypic evaluation. *Crop Science,* 62, 192-202. https://doi.org/10.1002/csc2.20657

Neyhart, J. L., Silverstein, K. A. T., & Smith, K. P. (2022). Accurate Predictions of Barley Phenotypes Using Genomewide Markers and Environmental Covariates. *Crop Science,* 1–31. https://doi.org/https://doi.org/10.1002/csc2.20782

Pérez de Vida, F. (2010). Aspectos de la ecofisiología del cultivo de arroz en Uruguay: II. Importancia de la fecha de siembra en la productividad. In *Arroz: resultados experimentales 2009-2010* (pp. 8-12). Treinta y Tres: INIA*.*

Piepho, H-P. (2019) A coefficient of determination (R^2) for generalized linear mixed models. *Biometrical Journal,* 61(4), 860–872. https://doi.org/10.1002/bimj.201800270

Piepho, H. P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics,* 53, 761–766. doi:10.2307/2533976

Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, 97, 195–201. doi:10.1007/s001220050885

Quero, G., Gutiérrez, L., Monteverde, E., Blanco, P., Pérez de Vida, F., Rosas, J., … Bonnecarrère, V. (2018). Genome-Wide Association Study Using Historical Breeding Populations Discovers Genomic Regions Involved in High-Quality Rice. *The Plant Genome*, 11(3), 1–12. https://doi.org/10.3835/plantgenome2017.08.0076

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Raju, N. S., Senguttuve, P., Voleti, S. R., Prasad, A. S. H., Bhadana, V. P., Revathi, P., … Viraktamat, B. C. (2013). Stability Analysis of Flowering and Yield Traits to High Temperature Stress Adopting Different Planting Dates in

Rice (*O. sativa* L.). *International Journal of Agricultural Research,* 8(3), 137–148. https://doi.org/10.3923/ijar.2013.137.148

Ravagnolo, O., & Misztal, I. (2000). Genetic component of heat stress in dairy cattle, parameter estimation. *Journal of Dairy Science,* 83(9), 2126–2130. https://doi.org/10.3168/jds.S0022-0302(00)75095-8

Rebollo, I., Rosas, J. E., & Aguilar, I. (2020). Efficient computation of the additive relationship matrix and its inverse in self-breeding individuals. *VI International Congress of Quantitative Genetics*. https://icqg6.org/wp-content/uploads/2021/04/ines-rebollo-EfficientComputation.mp4

Rebollo, I., Scheffel, S., Blanco, P., Pérez de Vida, F., Molina, F., Martínez, S., … Rosas, J. E. (2022). *Consolidating 23 years of historical data from an irrigated subtropical rice breeding program* [Unpublished manuscript].

Rosas, J. E., Martínez, S., Blanco, P., Pérez de Vida, F., Bonnecarrère, V., Mosquera, G., … Gutiérrez, L. (2018). Resistance to Multiple Temperate and Tropical Stem and Sheath Diseases of Rice. *The Plant Genome,* 11(1), 0. https://doi.org/10.3835/plantgenome2017.03.0029

RStudio Team. (2020). *RStudio: Integrated Development for R.* RStudio. http://www.rstudio.com/

Saint Pierre, C., Burgueño, J., Crossa, J., Fuentes Dávila, G., Figueroa López, P., Solís Moya, E., … Singh, S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific Reports,* 6, 1–11. https://doi.org/10.1038/srep27312

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science,* 86(1–3), 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9

Shakiba, E., Edwards, J. D., Jodari, F., Duke, S. E., Baldo, A. M., Korniliev, P., … Eizenga, G. C. (2017). Genetic architecture of cold tolerance in rice (*Oryza sativa*) determined through high resolution genome-wide analysis. *PLoS ONE,* 12(3), 1–22. https://doi.org/10.1371/journal.pone.0172133

Smith, A. B., Cullis, B. R., & Thompson, R. (2001). Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics,* 57, 1138–1147. https ://doi.org/10.1111/j.0006-341X.2001.01138.x

Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J., & Sorrells, M. E. (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *The Plant Genome,* 10(2). https://doi.org/10.3835/plantgenome2016.11.0111

Tester, M., & Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science,* 327(5967), 818–822. https://doi.org/10.1126/science.1183700

Tiscornia, G., Cal, A., & Giménez, A. (2016). Análisis y caracterización de la variabilidad climática en algunas regiones de Uruguay. *Revista de Investigaciones Agropecuarias,* 42(1), 66–71.

Tolhurst, D. J., Gaynor, R. C., Gardunia, B., Hickey, J. M., & Gorjanc, G. (2021). Genomic Selection Using Environmental Covariates Within an Integrated Factor Analytic Linear Mixed Model. *Research Square,* 0–25. https://doi.org/10.21203/rs.3.rs-995887/v1

Tolhurst, D. J., Mathews, K. L., Smith, A. B., & Cullis, B. R. (2019). Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. *Journal of Animal Breeding and Genetics*, 136(4), 279–300. https://doi.org/10.1111/jbg.12404

Tseng, M.-C., Roel, Á., Macedo, I., Marella, M., Terra, J., Zorrilla, G., & Pittelkow, C. M. (2021). Field-level factors for closing yield gaps in high-yielding rice systems of Uruguay. *Field Crops Research*, 264(January), 108097. https://doi.org/10.1016/j.fcr.2021.108097

Van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype × Environment interactions? *Crop Science,* 56(5), 2119–2140. https://doi.org/10.2135/cropsci2015.06.0375

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

Vargas, M., Crossa, J., Sayre, K., Reynolds, M., Ramírez, M. E., & Talbot, M. (1998). Interpreting genotype x environment interaction in wheat by partial least squares regression. *Crop Science*, *38*(3), 679–689. https://doi.org/10.2135/cropsci1998.0011183X003800030010x

Windhausen, V. S., Wagener, S., Magorokosho, C., Makumbi, D., Vivek, B., Piepho, H. P., … Atlin, G. N. (2012). Strategies to subdivide a target population of environments: Results from the CIMMYT-led maize hybrid testing programs in Africa. *Crop Science,* 52(5), 2143–2152. https://doi.org/10.2135/cropsci2012.02.0125

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems,* 58(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics,* 1(2), 205–230. https://doi.org/10.2307/1400366

Xing, Y., & Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annual Review of Plant Biology,* 61, 421–442. https://doi.org/10.1146/annurev-arplant-042809-112209

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science,* 40(3), 597–605. https://doi.org/10.2135/cropsci2000.403597x

Yoshida, S. (1981). *Fundamentals of rice crop science.* International Rice Research Institute.

Zhao, K., Tung, C.W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., … McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications,* 2(1), 467. https://doi.org/10.1038/ncomms1467

3 DISCUSIÓN GENERAL

En este trabajo se utilizaron los datos históricos del PMGA que cubren 23 años de evaluación fenotípica con el fin de: 1) caracterizar la IGA en los ambientes objetivo del PMGA, 2) identificar variables climáticas que afectan el rendimiento y 3) evaluar el uso de modelos de regresión aleatoria como estrategia para modelar y predecir el rendimiento en ambientes y años conocidos y desconocidos. Los resultados mostraron que la IGA es grande en ambas poblaciones analizadas, pero lo es aún más en *indica* que en *japonica*. Si bien no se encontraron patrones reproducibles de IGA entre las combinaciones de fecha de siembra y localidades del programa, se observaron diferencias importantes entre las siembras tempranas de *indica* y los demás ambientes. Las variables climáticas que afectaron el rendimiento y la IGA estuvieron asociadas a temperatura, radiación, viento y precipitación, y fueron distintas para *indica* y *japonica*. Los modelos de regresión aleatoria que incorporaron covariables climáticas mejoraron la habilidad predictiva en ambientes y años evaluados y no evaluados. Dado que esta tesis se inscribe en las líneas de investigación del PMGA, resulta pertinente evaluar los aportes de este trabajo para el diseño de estrategias de mejoramiento en el PMGA.

Este trabajo contribuye al entendimiento de la IGA en las condiciones de evaluación del germoplasma de arroz en Uruguay del PMGA. Los resultados presentan bases sobre las cuales tomar decisiones para la evaluación y selección de materiales para los ambientes objetivo del PMGA. El exhaustivo análisis de IGA que se realizó permitió cuantificar su magnitud y la de los distintos factores que la componen en las dos principales poblaciones de mejoramiento que maneja el programa. Si bien la interacción entre el genotipo y la fecha de siembra fue el menor componente de la varianza en ambas poblaciones, en *indica* se observó importante IGA con cambio de ranking entre la fecha de siembra temprana y las demás. Estos resultados muestran que en *indica* es pertinente mantener la evaluación de rendimiento en al menos dos épocas de siembra contrastantes: una anterior y otra posterior al 15 de octubre. La respuesta favorable de los materiales *indica* frente a las fechas de siembra tempranas es ampliamente conocida (Pérez de Vida, 2010, Deambrosi et al., 2019). De hecho, la creciente adopción de fechas de siembra temprana, junto con el uso de cultivares *indica* adaptados a dicho manejo como INIA Merín, son algunos de los principales factores que explican los récords de rendimiento obtenidos en las últimas zafras (Molina et al., 2021). Sin embargo, existe también una gran IGA dentro de la fecha de siembra temprana. Profundizar en la caracterización de esta IGA será necesario en el futuro para mejorar el rendimiento en este tipo de ambiente tan importante para el PMGA. Por el contrario, en *japonica* el desempeño relativo de los cultivares es similar a lo largo de todas las fechas de siembra. Con base en estos resultados, se podría considerar la conveniencia de evaluar el rendimiento de materiales de tipo *japonica* en una única época de siembra.

Diversos autores han demostrado el potencial de la SG aplicada al mejoramiento genético de arroz (Guo et al., 2014, Grenier et al., 2015, Onogi et al., 2015, Spindel et al., 2015, 2016, Morais Júnior et al., 2017, 2018a, 2018b, Monteverde et al., 2018). Sin embargo, existen varios factores que hacen de su aplicación práctica un desafío: algunos son genéticos y otros, estadísticos (Crossa et al., 2017), económicos o de implementación (Rutkoski et al., 2015). Quedan aún muchos aspectos por cubrir para profundizar la comprensión y el alcance de la técnica y determinar cómo y cuándo podría aplicarse efectivamente para cumplir con las expectativas generadas. Rutkoski et al. (2015) advierten que sería recomendable que cada programa de mejoramiento pruebe si la SG es una estrategia adecuada dados sus recursos y objetivos específicos. Monteverde (2019) abordó varios aspectos de la implementación de SG en el PMGA, entre ellos, el impacto en la precisión de las predicciones del uso de información multiambiente, modelando su estructura de covarianza, así como también la incorporación de covariables ambientales a la predicción en modelos de mínimos cuadrados parciales y norma de reacción, resultando beneficiosas todas las incorporaciones. Sin embargo, aún quedaban por probar los modelos de regresión aleatoria para la incorporación de covariables ambientales así como el impacto que tendría el uso de información fenotípica de individuos de los que no se cuenta con información genotípica, sobre la precisión de las predicciones. Esto permitiría sacar provecho de la extensa información fenotípica histórica y climática disponible. Este trabajo demostró que la información fenotípica histórica y los modelos de regresión aleatoria con covariables climáticas mejoran las predicciones del rendimiento en distintos escenarios de predicción. La utilización de la extensa información fenotípica disponible en el programa permitió cubrir un amplio rango de condiciones climáticas. Además, se identificaron modelos que mejoraron la habilidad predictiva tanto para ambientes conocidos como para años y ambientes desconocidos. Estos modelos permiten estudiar la respuesta diferencial de cada genotipo frente a las condiciones climáticas relevantes (Jarquín et al., 2021). Adicionalmente, las estimaciones obtenidas podrían ser utilizadas en el marco del mapeo asociativo para ayudar a entender los factores genéticos que las determinan (Li et al., 2021).

Esta tesis genera nuevos insumos para evaluar la distribución de recursos y mejorar la eficiencia en el uso de recursos del PMGA. Por un lado la caracterización de la IGA del presente estudio permite identificar los grupos de ambientes que aportan la información clave para la selección por rendimiento y estabilidad del rendimiento. Por otro lado, la implementación rutinaria de los MRA aquí planteados permitiría un uso más eficiente de la información fenotípica y climática disponible. En este sentido, resulta importante mencionar que INIA cuenta con estaciones climáticas en las distintas estaciones experimentales, las que toman información diaria de un amplio conjunto de variables. El empleo de los modelos aquí evaluados sería una herramienta que sacaría provecho adicional a esa valiosa infraestructura disponible en el Instituto.

Finalmente, se destaca la generación de datos genómicos listos para ser usados, así como *pipelines* para su generación y uso que son también un valioso aporte de este trabajo. Dicha información queda como insumo para ser utilizado por el PMGA en futuros trabajos que impliquen la utilización de datos genómicos y modelos de SG.

4 CONCLUSIONES

La IGA es de importante magnitud en ambas poblaciones del PMGA. En *indica* esta es mayor que en *japonica*. En *indica* existe un importante patrón de IGA con cambio de ranking del rendimiento entre la fecha de siembra temprana y las demás. Sin embargo, existe también una gran IGA dentro de la fecha de siembra temprana. Profundizar en la caracterización de esta IGA será necesario en el futuro para mejorar el rendimiento en este tipo de ambiente. No hay un patrón claro de agrupamiento entre fechas de siembra en *japonica* ni por localidades en ninguna de las poblaciones de mejoramiento analizadas.

Las variables climáticas que se relacionan con el rendimiento en *indica* y *japonica* no son las mismas.

La predicción genómica del rendimiento tanto para ambientes conocidos como desconocidos mejora al modelar la respuesta diferencial de cada genotipo a covariables climáticas relacionadas con el rendimiento mediante MRA. La utilización de estos modelos permitiría hacer un uso más eficiente de los recursos del PMGA.

5 BIBLIOGRAFÍA

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science. 93 (2): 743-752. doi: 10.3168/jds.2009-2730.

Bernardo R. 2016. Bandwagons I, too, have known. Theoretical and Applied Genetics. 129 (12): 2323-2332. doi: https://doi.org/10.1007/s00122-016-2772-5.

Bernardo R. 2010. Breeding for quantitative traits in plants. 2nd ed. Woodbury: Stemma Press. 390 p.

Bonnecarrère V, Rosas J, Ferraro B. 2019. Economic impact of marker-assisted selection and rapid generation advance on breeding programs. Euphytica. 215 (12). doi: 10.1007/s10681-019-2529-8.

Braun H-J, Rajaram S, van Ginkel M. 1996. CIMMYT’s approach to breeding for wide adaptation. Euphytica. 92: 175–183. doi: 10.1007/BF00022843.

Buntaran H, Forkman J, Piepho H-P. 2021. Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. Theoretical and Applied Genetics. 134 (5): 1513-1530. doi: 10.1007/s00122-021-03786-2.

Burgueño J, de los Campos G, Weigel K, Crossa J. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Science. 52 (2): 707–719. doi: 10.2135/cropsci2011.06.0299.

Crossa J. 2012. From Genotype × Environment Interaction to Gene × Environment Interaction. Current Genomics. 13 (3): 225-244. doi: 10.2174/138920212800543066.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science. 22 (11): 961-975. doi: 10.1016/j.tplants.2017.08.011.

Deambrosi, E, Zorrilla De San Martín, G, Lauz, M, Blanco, P, Terra, J. 2019. Rompiendo el techo de rendimiento del cultivo de arroz. Montevideo: INIA. 104 p.

FAO (Food and Agriculture Organization). 2020. Crops and livestock products. [En línea] Último acceso 2 junio 2022. Disponible en: https://www.fao.org/faostat/en/#rankings/countries_by_commodity_exports

FAO. 2018. [En línea] Último acceso 9 diciembre 2018. Disponible en: http://www.fao.org/faostat/en/#rankings/commodities_by_regions_exports

Fehr WR. 1987. Principles of Cultivars Development. Vol. 1. New York: Macmillan. 536 p.

Gauch HG, Zobel RW. 1997. Identifying Mega-Environments and Targeting Genotypes. Crop Science. 37 (2): 311–326. doi: 10.2135/cropsci1997.0011183X003700020002x.

Gholami M, Wimmer V, Sansaloni C, Petroli C, Hearne SJ, Covarrubias-Pazaran G, Rensing S, Heise J, Pérez-Rodríguez P, Dreisigacker S, Crossa J, Martini JWR. 2021. A Comparison of the Adoption of Genomic Selection Across Different Breeding Institutions. Frontiers in Plant Science. 12 (November): 1–6. doi: 10.3389/fpls.2021.728567.

Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, Tohme J, Courtois B, Ahmadi N. 2015. Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS ONE. 10 (8): 1–25. doi: 10.1371/journal.pone.0136594.

Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G. 2014. The impact of population structure on genomic prediction in stratified populations. Theoretical and applied genetics. 127 (3): 749-762. doi: 10.1007/s00122-013-2255-x.

IndexMundi. 2018. [En línea] Último acceso 9 diciembre 2018. Disponible en: https://www.indexmundi.com/agriculture/?commodity=milled-rice&graph=yield

Jannink JL, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics. 9 (2): 166-177. doi: 10.1093/bfgp/elq001.

Jarquín D, de Leon N, Romay C, Bohn M, Buckler ES, Ciampitti I, Edwards J, Ertl D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Holland JB, Hooker D, Kaeppler SM, Knoll J, Lee EC, Lawrence-Dill CJ, Lynch JP, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Smith M, Springer N, Thomison P, Tuinstra M, Wisser RJ, Xu W, Yu J, Lorenz A. 2021. Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. Frontiers in Genetics. 11 (March): 1-11. doi: 10.3389/fgene.2020.592769.

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueño J, de los Campos G. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics. 127 (3): 595–607. doi: 10.1007/s00122-013-2243-1.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology. 35: 25-34.

Lado B, Barrios PG, Quincke M, Silva P, Gutiérrez L. 2016. Modeling genotype × Environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Science. 56 (5): 2165-2179. doi: 10.2135/cropsci2015.04.0207.

Lanfranco B, Saldias R, Penengo C, Ferraro B. 2018. Setting transformational pathways consistent with post-2015 SDGs: The case of Uruguay s rice sector. En: International Association of Agricultural Economists (IAAE) Conference. 37 p.

Lavecchia, A, Méndez J. 1999. Evaluación del comportamiento de cuatro cultivares, El Paso 144, INIA Cuaró, INIA Tacuarí e INIA Caraguatá, en tres épocas de siembra en dos localidades (Tacuarembó y Artigas). En: INIA Tacuarembó. Programa Nacional de Arroz. Arroz: resultados experimentales zafra 1998-1999. Tacuarembó. INIA. 52-57.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25 (14): 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li X, Guo T, Wang J, Bekele W, Sukumaran S, Vanous AE, McNellie JP, Cortes LT, Lopes M, Lamkey KR, Westgate ME, McKay J, Archontoulis S V., Reynolds MP, Tinker N, Schnable PS, Yu J. 2021. An Integrated Framework Reinstating the Environmental Dimension for GWAS and Genomic Selection in Crops. Molecular Plant. 135907. doi: 10.1016/j.molp.2021.03.010.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 157 (4): 1819–29. doi: 11290733.

Molina F, Terra J, Roel A, Oxley A, Marella M, Casterá F, Platero A, García F, Rovira G, Escostegui C. 2021. Indicadores tecnológicos-productivos zafra arrocera 2020-2021. En: Terra J, Martínez S, Saravia H, Mesones B. (Eds.). Arroz 2021. Montevideo: INIA. 1-5.

Monteverde E. 2019. Implementing Genomic Selection in the Uruguayan Rice Breeding Program. Tesis de doctorado. Ithaca, USA. Cornell University. 155 p.

Monteverde E, Gutiérrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G, McCouch S. 2019. Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (*Oryza sativa* L.) Grown in Subtropical Areas. G3 Genes|Genomes|Genetics. 9 (5): 1519-1531. doi: 10.1534/g3.119.400064.

Monteverde E, Rosas JE, Blanco P, Pérez de Vida F, Bonnecarrère V, Quero G, Gutiérrez L, McCouch S. 2018. Multienvironment Models Increase Prediction

Accuracy of Complex Traits in Advanced Breeding Lines of Rice. Crop Science. 58 (4): 1519-1530. doi: 10.2135/cropsci2017.09.0564.

Morais Júnior OP, Breseghello F, Duarte JB, Coelho ASG, Borba TCO, Aguiar JT, Neves PCF, Morais OP. 2018a. Assessing prediction models for different traits in a rice population derived from a recurrent selection program. Crop Science. 58 (6): 2347-2359. doi: 10.2135/cropsci2018.02.0087.

Morais Júnior OP, Duarte JB, Breseghello F, Coelho ASG, Magalhães Júnior AM. 2018b. Single-step reaction norm models for genomic prediction in multienvironment recurrent selection trials. Crop Science. 58 (2): 592-607. doi: 10.2135/cropsci2017.06.0366.

Morais Júnior OP, Duarte JB, Breseghello F, Coelho ASG, Borba TCO, Aguiar JT, Neves PCF, Morais OP. 2017. Relevance of additive and non-additive genetic relatedness for genomic prediction in rice population under recurrent selection breeding. Genetics and Molecular Research. 16 (4). doi: 10.4238/gmr16039849.

Mrode R. 2014. Linear Models for the Prediction of Animal Breeding Values. 3rd Edition. Boston: CABI. 343 p.

Onogi A, Ideta O, Inoshita Y. 2015. Exploring the areas of applicability of whole-genome prediction methods for Asian rice (*Oryza sativa* L.). Theoretical and applied genetics. 128: 41-53. doi: 10.1007/ s00122-014-2411-y.

Pérez de Vida F, Blanco P, Carracelas, G, Vargas J. 2015. Evaluación final de cultivares. En: ARROZ-SOJA, 2015, Treinta y Tres, UY. Resultados experimentales 2014-2015. Treinta y Tres. INIA. 4-6.

Pérez de Vida F, Macedo I. 2013. Aspectos de la ecofisiología del cultivo de arroz en Uruguay: Incidencia de factores climáticos en la productividad experimental de cultivares. En: Arroz-Soja. Resultados experimentales 2012-13. Treinta y Tres. INIA. 99-101.

Pérez de Vida F. 2010. Aspectos de la ecofisiología del cultivo de arroz en Uruguay. II. Importancia de la fecha de siembra en la productividad. En: Arroz: resultados experimentales 2009-2010. Treinta y Tres. INIA. 8-12.

Ravagnolo O, Misztal I. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. Journal of Dairy Science. 83 (9): 2126-2130. doi: 10.3168/jds.S0022-0302(00)75095-8.

Roel A, Baethgen WE. 2007. Towards the development of a spatial decision support system (SDSS) for the application of climate forecasts in Uruguayan rice production sector. En: Sivakumar JVK, Hansen J (Eds.). Climate prediction and agriculture: Advances and challenges. Berlin: Springer. 89-97.

Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME. 2015. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. The Plant Genome. 8 (1): 1-10. doi: 10.3835/plantgenome2014.09.0046.

Schmid M, Bennewitz J. 2017. Invited review: Genome-wide association analysis for quantitative traits in livestock - A selective review of statistical models and experimental designs. Archives Animal Breeding. 60 (3): 335-346. doi: 10.5194/aab-60-335-2017.

Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink JL, McCouch S. 2016. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity. 116 (4): 395-408. doi: 10.1038/hdy.2015.113.

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink JL, McCouch SR. 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 (2): 1-25. doi: 10.1371/journal.pgen.1004982.

Tolhurst DJ, Mathews KL, Smith AB, Cullis BR. 2019. Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. Journal of Animal Breeding and Genetics. 136 (4): 279-300. doi: 10.1111/jbg.12404.

Uruguay XXI. 2018. Oportunidades de inversión. Agronegocios. [En línea] Último acceso 15 octubre 2018. Disponible en: https://www.uruguayxxi.gub.uy/uploads/informacion/b5eb6050ca43af07330c66ca7fb7ba70f0cd8189.pdf

Yan W, Hunt LA, Sheng Q, Szlavnics Z. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science. 40 (3): 597-605. doi: 10.2135/cropsci2000.403597x