



CONSOLIDATING 23 YEARS OF HISTORICAL DATA FROM A PUBLIC SUBTROPICAL IRRIGATED RICE BREEDING PROGRAM

S. Scheffel¹, I. Rebollo¹, P. Blanco¹, F. Molina¹, F. Pérez de Vida¹, J. E. Rosas^{1*}¹INIA, Uruguay. *jrosas@inia.org.uy

Introduction

Rice is one of the major staple foods worldwide. INIA's rice breeding program (IRBP) has been delivering Uruguay's most grown rice cultivars since 1987. The program is currently organized into four components based on germplasm and grain quality types: Tropical Japonica, Indica, Temperate Japonica and Clearfield. Up to date, field trial data was stored separately, hindering the joint analysis of multiple trials and years.

Objective

The aim of this work is to gather all available information of IRPB's field trials to facilitate joint analysis and thus make a more efficient use of the program's information.

Materials and methods

Original data consisted of records from 1997 to 2020 of field trials, disease, and grain quality evaluation, of inbred lines in all evaluation stages and its pedigree records, available in separate spreadsheets with a variety of formats.

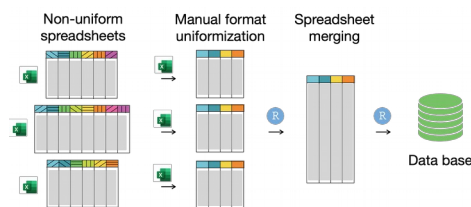


Figure 1. Data processing for gathering each trial's information into a database.

The data processing (Figure 1) included merging all available spreadsheets, standardization of missing data and levels of categorical variables, and application of data quality with statistical and agronomic criteria. The full dataset was analyzed with a joint model with all terms as random variables to study variance partition, and by each trial separately to compute trial's heritability (H^2) and coefficient of variation (CV)

Results and Discussion

The final unified database consisted of 1,526,415 data points corresponding to 1,007 trials, 12 locations, 23 years, 92,541 records, and 19,447 lines under evaluation, of which 965 had genotypic information (Figure 2). The most represented component is Tropical Japonica (49.8%), followed by Indica (38.8%), Clearfield (9%) and Temperate Japonica (2.5%). Most information corresponds to one location, UEPL, and advanced evaluation stages are present in three locations.

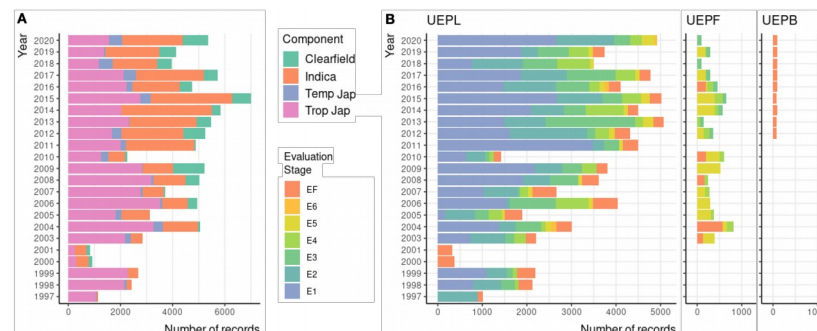


Figure 2. Number of records by year colored by component (A) and number of records by location colored by evaluation stage (B).

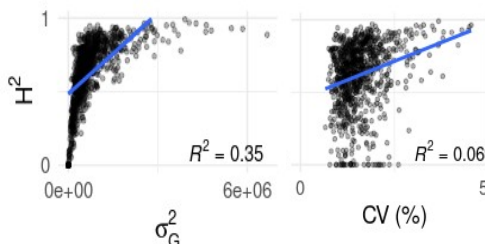


Figure 3. Trial's heritability (H^2) by genetic variance (A) and by coefficient of variation (B). The regression line is shown in blue.

Trial's H^2 ranged from 0 to 1 with a mean of 0.62 (Figure 3). CV ranged from 6.59 to 45.44, with a mean of 15.84. Trials with lower H^2 also had low genetic variance and CV.

Grain yield correlated significantly with all variables, and negatively with some disease and quality related variables (Figure 4).

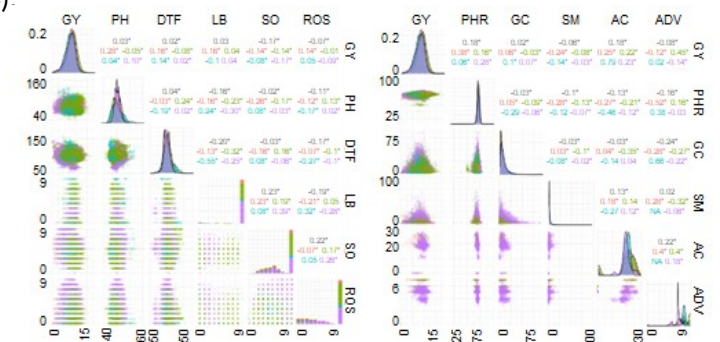
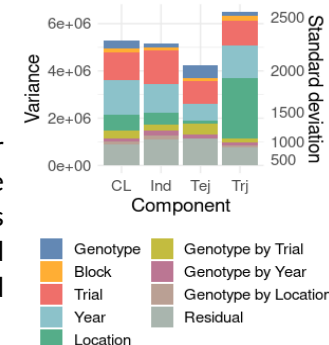


Figure 4. Distribution (lower triangle), correlations (upper triangle) and density (diagonal) of grain yield (GY), plant height (PH), days to flowering (DTF), leaf blast (LB), Sclerotium oryzae (SO), Rhizoctonia oryzae-sativae (ROS), proportion of head rice (PHR), grain chalkiness (GC), stained milled grain (SM), amilose content (AC) and alkali digestion value (ADV).

Figure 5. Variance and standard deviation analysis per component colored by source of variation.



All components showed a similar pattern of variation partition (Figure 5). Most phenotypic variance was explained by environment related effects (trial, year, location) and genotype by environment effects.

Conclusion

The consolidated dataset generated in this work is enabling joint analyses with higher genetic value prediction accuracy, contributing to improve the genetic gain and efficiency of IRBP.