Goal

This vignette introduces the CERIS-JGRA R package developed by Li *et al.* 2021 to model, explain, and predict trait dynamics from a multi-environment trial (MET). The Joint Regression Analysis (JRA) is one of the popular frameworks for MET. However, JRA's key limitation is that it is based on environmental mean, the average performance of the whole population at each environment. CERIS (Critical Environmental Regressor through Informed Search) identifies an explicit environmental index to replace environmental mean. The identified environmental index can be further integrated with genomic information to predict performance. The process of leveraging genomics and JGA is termed as Joint Genomic Regression Analysis (JGRA). The entire analysis framework is terms as CERIS-JGRA.

Outline

Here an environment is defined by the combination of year, location, and planting date. If the same population is planted in the same year at the same location but different planting dates, each combination of year, location, and planting date is counted as an environment. From each environment, the average performance of the entire population is calculated for that environment. For a MET with *m* environments, the vector for environmental mean has *m* values.

From a candidate pool, the current CERIS algorithm uses Pearson correlation to identify the environmental index, which has the strongest correlation with the vector of environmental mean. The candidate pool is built based on the combinations of environmental parameters and growing periods. A growing period is defined as *x* to *y* days-after-planting (DAP). Because of the varied planting dates among environments, the same growing period can vary among environments. For example, the same 1 to 10 DAP may correspond to May 1st to May 10th in one environment, but May 16th to May 25th in another environment. Environmental parameters can be common individual environmental factors that can be directly measured for any environment, such as daily maximum temperature, daily minimum temperature, day length, and precipitation. Environmental parameters can also be calculated based on two or more environmental factors. Here we provide some environmental parameters derived from temperature and day length (DL).

$$GDD = \frac{1}{2}(T_{max} + T_{min}) - T_{base}$$

$$PTT = GDD \times DL$$

$$PTR = GDD \div DL$$

$$PTS = (T_{max}^2 - T_{min}^2) \times DL^2$$

DL is daytime in hours, which is calculated based on the daylength function in the *geosphere* package based on latitude coordinate. T_{max} is the maximum temperature, T_{min} is the minimum temperature, while T_{base} is adjusted based on species. A typical daily temperature source is NOAA GHCN (https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn). You can build the environmental index pool from data recorded in your own in-field weather stations.

The CERIS-JGRA R package has two scripts, "CERIS_JGRA.r" for the main function and "Sub_functions.r" for subfunctions. After your input files are formatted and stored in the corresponding directory, modify a few lines (see next section) in CERIS_JGRA.r based on your settings.

The directory structure.
######## Top_dir
######## |----0Maize
######## | |--11Envs_envParas_DAP100.txt
######## | |--Env_meta_table.txt
######## | |--Trait_records.txt
######## | |--Genotype.txt
######## |----CERIS_JGRA.r
######## |----Sub_functions.r

CERIS for identifying environmental index

Three essential input files are required for CERIS. 1) Trait records; 2) Meta-information for environments; 3). Candidate pool for the environmental index. The corresponding files for the maize example (under 0Maize) are Trait_records.txt, Env_meta_table.txt, and 11Envs_envParas_DAP100. Follow the format to compile your own input files (Format the date as YYYY-MM-DD).

In line 22, modify the Tbase based on your species.
In line 25, modify the top_dir based on your setting.
In line 32, modify the experiment based on your setting.
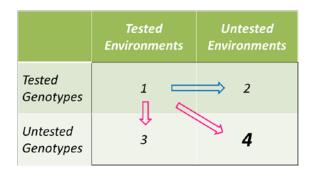In line 34, modify the trait that you want to analyze.

Run the CERIS_JGRA.r script to line 81. All the output files will be stored under the corresponding trait folder. For instance, /0Maize/FT/ has all output files for the FT trait in maize.

Open FT_9Envs_PTTPTR_0LOO_cor.txt and MaxR_DTA_9Envs_0LOO.png to select the environmental index. FT_9Envs_PTTPTR_0LOO_cor.txt stores all the correlations. Each row is for one growing period. The fifth to ninth columns are the original correlations (corresponding to the top row of MaxR_DTA_8Envs_0LOO.png), while the tenth to fourteenth columns are the negative values (corresponding to the bottom row of MaxR_DTA_8Envs_0LOO.png). In this FT example, it is PTR from 22 to 37 DAP was selected as the environmental index because of the strong correlation and that this growing period is in the early season before the earliest observation of FT trait expression.

Modify the values from line 85 to 87 based on the identified environmental index. Run the script to line 102.

CERIS-JGRA for performance prediction

Li *et al.* 2021 (not published yet) documented the extensive analysis with CERIS-JGRA for three diversity panels in three crop species and multiple traits. At this time, please refer to Li *et al.* 2018 and Guo *et al.* 2020 for the prediction scenarios (1to2, 1to3, and 1to4) and prediction strategies (reaction norm parameter and marker effect). In these two earlier studies, a single biparental mapping population and a single trait were analyzed.

Genotype information is only required for 1to3 and 1to4 prediction scenarios. (Note: Genotype.txt only contains a small subset of the total SNPs for this demo).

Run line 120 to 132. The corresponding excel files store the prediction results for each line under each environment.

********Note: This document will be updated as we conduct additional workshops.********

References

1. Xianran Li, Tingting Guo, Jinyu Wang, Wubishet Bekele, Sivakumar Sukumaran, Adam E. Vanous, James P. McNellie, Laura Tibbs Cortes, Marta Lopes, Kendall Lamkey, Mark E. Westgate, John McKay, Sotirios V. Archontoulis, Matthew P. Reynolds, Nicholas Tinker, Patrick S. Schnable, and Jianming Yu. 2021. Reinstating the environmental dimension for genome-wide association studies and genomic selection in crops. *Molecular Plant* (*In Revision*).

2. Tingting Guo, Qi Mu, Jinyu Wang, Adam Vanous, Xianran Li, and Jianming Yu. 2020. Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. *Genome Research* 30:673-683.

3. Xin Li, Tingting Guo, Qi Mu, Xianran Li, and Jianming Yu. 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *PNAS* 115:6679-6684.