Published September 8, 2016

Analysis of Genotype × Environment Interaction (G×E) Using SAS Programming

Mahendra Dia, Todd C. Wehner,* and Consuelo Arellano

ABSTRACT

Genotype × environment interaction (G×E) can lead to differences in the performance of genotypes across environments. A G×E analysis can be used to analyze the stability of genotypes and the value of test locations. We developed a SAS program (SASG×E) that calculates univariate stability statistics, descriptive statistics, pooled and yearly ANOVA, genotypic and location variation, cluster analysis for location, and correlations among stability parameters. Univariate stability statistics calculated are Wricke's ecovalence (W_i^2), Shukla's variance (σ_i^2), Lin and Binns cultivar superiority measure (P_i), Francis and Kannenberg coefficient of variation (CV_i), Kang's yield stability statistic (YS_i), Perkins and Jinks β (β_i), regression slope (b_i), and deviation from regression (S_d^2). Other output includes input files for analyzing stability in R software using AMMI and GGEBiplotGUI packages. SASG×E uses SAS programming language features (macro and structured query language [SQL]) for repetitive tasks, making it efficient and flexible for the simultaneous analysis of multiple dependent variables. SASG×E is free and intended for use by scientists studying the performance of polygenic or quantitative traits in multiple environments. The SASG×E program is presented here and is also available at http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html.

Core Ideas

- Genotype × environment interaction can lead to differences in genotype performance.
- G×E analysis can analyze genotype stability and the value of test locations.
- SASG×E uses SAS and R programming to compute uni- and multi-variate stability statistics.
- SASG×E output includes univariate stability statistics, ready to go input files, and R code for AMMI and GGE biplot analysis, ANOVA, descriptive statistics, cluster analysis of location, rank correlation among stability parameters, and Pearson correlation of location with average location performance.

Published in Agron. J. 108:1838–1852 (2016)
doi:10.2134/agronj2016.02.0085
Received 8 Feb. 2016
Accepted 12 Apr. 2016
Supplemental material available online

Copyright © 2016 by the American Society of Agronomy
5585 Guilford Road, Madison, WI 53711 USA
All rights reserved

Genotype × environment interaction (G×E) refers to the modification of genetic factors by environmental factors and to the role of genetic factors in determining the performance of genotypes in different environments. A G×E can occur for quantitative traits of economic importance and is often studied in plant and animal breeding, genetic epidemiology, pharmacogenomics, and conservation biology research. The traits include reproductive fitness, longevity, height, weight, yield, and disease resistance.

Selection of superior genotypes in target environments is an important objective of plant breeding programs. A target environment is a production environment used by growers. To identify superior genotypes across multiple environments, plant breeders conduct trials across locations and years, especially during the final stages of cultivar development. A G×E is said to exist when differences in genotype performance across environments result in a scale shift or a rank shift. Performance of genotypes can vary greatly across environments because of the effect of the environment on trait expression. Therefore, cultivars with high and stable performance are of value.

Because it is impossible to test genotypes in all target environments, plant breeders do indirect selection using their own multiple-environment trials or test environments. Genotype × environment interaction reduces the predictability of the performance of genotypes in target environments based on genotype performance in test environments. An important factor in plant breeding is the selection of suitable test locations because it accounts for G×E and maximizes gain from selection (Yan et al., 2011). An efficient test location is discriminating and is representative of the target environments for the cultivars to be released. Discriminating locations can detect differences among genotypes with only a few replications. Representative locations will make it likely that the genotypes selected will perform well in the target environments (Yan et al., 2011).

Analysis of variance (ANOVA) is useful in determining the existence, size, and significance of G×E. To determine G×E

M. Dia and T.C. Wehner, Dep. of Horticultural Science, North Carolina State Univ., Raleigh, NC 27695; and C. Arellano, Statistics Dep., North Carolina State Univ., Raleigh, NC 27695. *Corresponding author (tcwehner@gmail.com).

Abbreviations: ANOVA, analysis of variance; AMMI, additive main effects and multiplicative interaction; CL, cultigen or genotype; CLWT, cull fruit weight; CSV, comma-separated value; CV, coefficient of variation; EN, environment (location–year combination); GGE, genotype main effects plus genotype × environment interaction model; G×E, genotype × environment interaction; LC, location; LSD, least significant difference; MKWT, marketable fruit weight; MKMGHA, marketable megagrams per hectare; PC, principal component; RP, replication; SQL, structured query language; YR, year.

for a group of elite cultivars, genotypes are often considered to be fixed effects and environments random. However, for the purpose of estimating breeding values using best linear unbiased prediction, genotypes are considered to be random and environments fixed. Some researchers prefer that genotypes be considered random regardless of the stage of selection, provided that the objective is to select the best ones (Smith et al., 2005). If G×E is significant, additional stability statistics may be calculated.

Several statistical methods have been proposed for stability analysis. These methods are based on univariate and multivariate models. We have focused on the analysis of stability measured using SAS and R programming, so a brief description of each stability measure is provided below.

The most widely used methods are univariate stability models based on regression and variance estimates. The original concept of regression statistics as a stability measure was proposed by Yates and Cochran (1938) and later improved by Eberhart and Russell (1966). According to the regression model, stability is expressed as trait means (M), the slope of the regression line (b_i), and the sum of squares deviation from the regression (S_d^2). A high mean of a genotype is a precondition of stability. The slope (b_i) of regression indicates the response of the genotype to the environmental index, which is derived from the average performance of all genotypes in each environment. If b_i is not significantly different from 1, the genotype is adapted in all environments. A b_i >1 describes genotypes with higher sensitivity to environmental change (below-average stability) and greater specificity of adaptability to high-yielding environments. A b_i <1 provides a measure of greater resistance to environmental change (above-average stability) and therefore increasing specificity of adaptability to low-yielding environments.

A regression coefficient similar to that of Eberhart and Russell (1966) was proposed by Perkins and Jinks (1968). The Perkins and Jinks β (β_i) uses regression of G×E effects on environmental effects. Genotypes with β_i values not significantly different from 0.0 are judged to be stable, whereas those with significant β_i values are unstable. According to Becker and Leon (1988), the two regression statistics are equivalent ($\beta_i = b_i - 1$).

The variance parameters that measure stability statistics include ecovalence (W_i^2) proposed by Wricke (1962), stability variance (σ_i^2) proposed by Shukla (1972), the coefficient of variation (CV_i) suggested by Francis and Kannenberg (1978), the mean sum of squares (P_i) proposed by Lin and Binns (1988), and yield stability (YS_i) proposed by Kang (1993).

The ecovalence stability index (W_i^2) of a genotype is its contribution to the G×E sum of squares across environments. Shukla (1972) proposed an unbiased estimate (σ_i^2) of the variance of the G×E plus an error term associated with genotype. Shukla's stability variance (σ_i^2) is a linear combination of Wricke's ecovalence (W_i^2). Shukla's stability statistic measures the contribution of a genotype to the G×E and error term; therefore a genotype with low σ_i^2 is regarded as stable. According to Kang et al. (1987), W_i^2 and σ_i^2 are equivalent (rank correlation = 1.0) in ranking genotypes for stability.

The Kang (1993) stability statistic (YS_i) is a nonparametric stability procedure in which both the mean (M) and Shukla (1972) stability variance (σ_i^2) for a trait are used as selection criteria. This method gives equal weight to M and σ_i^2. According to this method, genotypes with YS_i greater than the mean YS_i are considered stable (Kang, 1993; Mekbib, 2003; Fan et al., 2007). We computed YS_i based on the procedure outlined by Mekbib (2003). The stability concept (P_i) of Lin and Binns (1988) is derived from the mean sums of squares for year nested in location. The ANOVA is computed for genotypes, including just two factors, i.e., location and year within location (FAO, 2015). High stability is indicated by a low P_i value, i.e., low temporal variation of genotype trait values. Similarly, according to Francis and Kannenberg (1978), a genotype with low coefficient of variation (CV_i) is regarded as stable.

For multivariate stability models, the additive main effects and multiplicative interaction (AMMI) model and genotype main effects plus G×E (GGE) model with a graphical display has gained in popularity for analyzing multiple-environment trial data (Casanoves et al., 2005; Dehghani et al., 2006). Proponents of the AMMI and GGE biplot methods disagree on the best method for analyzing multi-environment trial data (Gauch, 2006; Yan et al., 2007), although the two methods provide similar results (Gauch, 2006).

Yan et al. (2000) referred to biplots based on singular value decomposition of environment-centered or within-environment standardized two-way (genotype × environment) data matrix as GGE biplots. A GGE biplot was constructed from the first two principal components (PC1 and PC2), which explained the maximum variability in the data, derived by singular value decomposition of a two-way (genotype × environment) data matrix (Yan et al., 2000). The GGE biplot graphically displays the two-way (genotype × environment) data matrix and allows visualization of the interrelationship among environments and genotypes and their interactions (Yan and Kang, 2003). In a GGE biplot, the genotype effect and G×E effect are the two sources of variation that are relevant to genotype evaluation and mega-environment identification (Gauch and Zobel, 1996; Yan and Kang, 2003).

The AMMI model combines the analysis of variance (ANOVA, an additive model) to characterize genotype and environment main effects with principal component analysis (a multiplicative model) to characterize interactions (interaction principal components, IPCs) (Crossa et al., 1990; Gauch, 1992). An AMMI biplot scatters genotypes according to their IPC scores. Therefore, it is easy to qualitatively assess the differences in genotype stability and adaptability to the environments in a graphical representation. The closer the IPC is to zero, the more stable the genotypes are across the testing environments (Carbonell et al., 2004).

A genotype F ratio for each test location and correlation of the test location with the average location are important measures of location value. The genotype F statistic is the ratio of genotype variance to error variance (assuming all genotypes have equal variance). When the means of all genotypes are equal, then the F ratio will be close to 1. If ANOVA is computed by location, then a high genotype F ratio indicates high discriminating ability for that location. A high and significant value of the Pearson correlation of each location with the mean of all locations indicates a strong representation of the mean location performance.

Our objective was to develop a SAS program (SASG×E) that gives an output for genotype stability and location value. Hussein et al. (2000) provided a SAS program (SASG×ESTAB) that computes univariate stability statistics and correlations among them. However, the SAS program is not available at the listed server nor from the corresponding author. Others have found it unavailable as well (K. Lamkey, personal communication, 2015). Similarly, Piepho (1999) published a SAS mixed model procedure to compute univariate stability statistics.

Our program (SASG×E) is flexible, automates the computation of statistics on multiple traits simultaneously, and has flexible output. For example, SASG×E gives an output for univariate stability statistics including Wricke's ecovalence (W_i^2), Shukla's variance (σ_i^2), Lin and Binns P_i, Francis and Kannenberg coefficient of variation (CV_i), Kang's statistic (YS_i), Perkins and Jinks β (β_i), regression slope (b_i), and deviation from regression (S_d^2), along with their test of significance (Table 1). SASG×E also provides output files and preprocessing of R code that are ready to use for analyzing stability in RStudio using R software (RStudio, 2014) and the *agricolae* [AMMI( )function] (de Mendiburu, 2015) and GGEBiplotGUI (Frutos et al., 2014) packages. Finally, SASG×E provides the user with pooled and yearly ANOVA, descriptive statistics (mean, coefficient of variation, and standard deviation), cluster analysis of location, rank correlation among stability parameters, and Pearson correlation of location with average location performance. To understand the interrelationship of stability methods, the rank correlation is recommended over the Pearson correlation because the stability parameters cannot be assumed to be normally distributed (Becker, 1981).

SASG×E uses SAS/Macro for repetitive tasks and SAS/SQL for complex joins of SAS software (Version 9.3 and higher) (SAS Institute). SASG×E is freely available, annotated, and intended for scientists studying the performance of polygenic or quantitative traits under different environmental conditions. Here we provide the general features of the SASG×E program along with the functionality each of its modules and a supplemental material file with the SASG×E program, example input data, graphical illustrations for uploading an input data file into SASG×E and installing R packages in RStudio, outputs from example input data, and interpretation of univariate and multivariate statistical analysis. The log output, input data file template, and a full set of outputs from example input data along with the SASG×E program features are also available at http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html.

General Features of the SASG×E Program

SASG×E is a user friendly, annotated, flexible and efficient SAS program that allows users to analyze stability statistics for multiple dependent variables, simultaneously, of balanced multilocation trial data. SASG×E checks for missing records because it is not suitable for unbalanced data. For unbalanced data, model parameters can be estimated by restricted maximum likelihood for mixed models. Piepho (1999) provided a SAS program to compute different stability parameters independently using mixed model theory. Similarly, Dia et al. (2016a) provided an R program (RG×E) to compute stability measures using mixed models. SASG×E generates output into the same folder from where it read the input dataset and is intended for SAS-PC (Version 9.3 or higher) under the Microsoft Windows operating system. A schematic representation of SASG×E is presented in Fig. 1. Interpretation of

Table 1. Summary of the features of the SASG×E program.

Statistic	Method	Parameter	Stability decision
	Genotype stability		
Regression	Eberhart and Russell (1966)	b_i	did not differ significantly from 1
		S_d^2	did not differ significantly from 0
	Perkins and Jinks (1968)	β_i	did not differ significantly from 0
Variances	Wricke (1962)	W_i	choose low values
	Shukla (1972)	σ_i^2	choose low values
	Francis and Kannenberg (1978)	CV_i	choose low values
	Lin and Binns (1988)	P_i	choose low values
	Kang (1993)	YS_i	more than its mean
Multivariate	GGE biplot		see supplement file
	AMMI		see supplement file
Correlation	Spearman rank correlation among regression and variance parameters		
	Location value		
Variances	*F* test		high genotype *F* ratio indicates high discriminating ability of location
	SD		choose low values
Multivariate	GGE biplot		see supplement file
Correlation	cluster analysis		
	Pearson correlation of location with average location performance		
	Descriptive statistics		
	mean, SD, and CV		

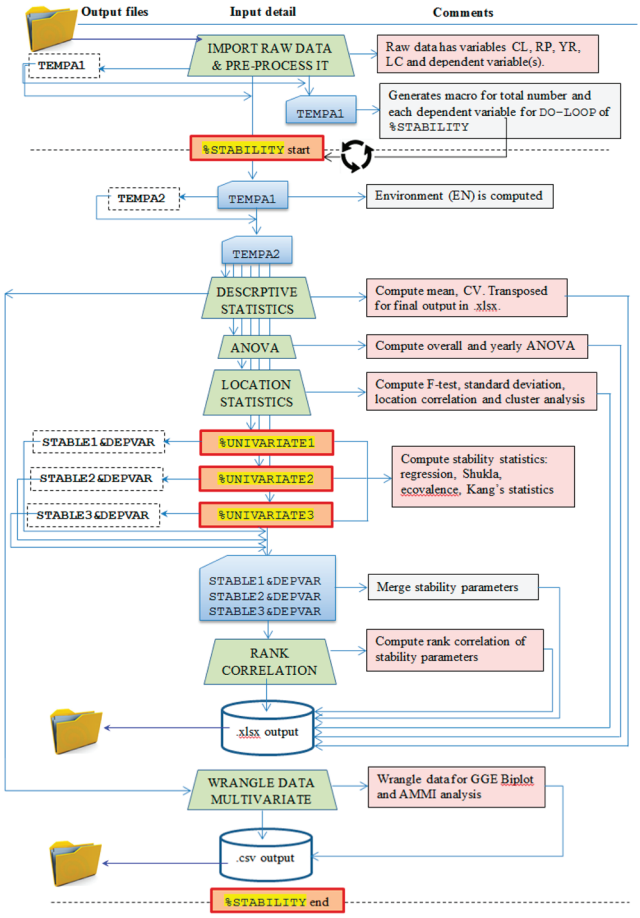


Fig. 1. Schematic representation of overall process of SASG×E program and macro `STABILITY`.

univariate stability statistics and output and interpretation of multivariate statistics of the example dataset are presented in the supplemental material and can also be obtained from http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html. Additionally, the sample input data used here were part of studies published on watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai] genotype stability (Dia et al., 2016b) and location value (Dia et al., 2016c). Dia et al. (2016b, 2016c) used SASG×E to analyze their data, and the results presented in these studies serve as an example.

Options Statement

We ran SASG×E with an options set for the program to run efficiently and make it easy to debug. Options such as MPRINT, SYMBOLGEN, and MLOGIC are helpful for debugging. However, MLOGIC and SYMBOLGEN are turned off in the production macro to improve program efficiency (SAS Institute, 2009). Similarly, the option LABEL was turned off so that auto-labeling can be prevented (e.g., in PROC MEANS) and, hence, the user will not be confused. Because SASG×E is not intended to print output in the SAS Output window, the options DATE and NUMBER are turned off. The option MPRINT is turned on so the user can view the text generated by macro execution in the SAS Log window. The statement DM LOG "CLEAR" is called in to clear the log from the previous session. The user has the option to uncomment and turn on the ODS HTML TURN OFF statement to suppress the default html output from certain procedures like PROC GLM and PROC CORR.

Import Input Data

SASG×E starts with user-entered fields. The user is required to feed the input data file location, name, and sheet name at the `%LET IPATH`, `%LET INAME`, and `%LET ISHEETNAME1` statements, respectively. SASG×E requires an input data file in Microsoft Excel (.xlsx type only). Highlighted fields are user entered in the code shown below. The user input records are converted into macro variables using the `%LET` statement:

```
%LET IPATH=...; /*INPUT FILE PATH*/
%LET INAME=...; /*INPUT FILE NAME*/
%LET ISHEETNAME1=...; /*INPUT FILE SHEET
  NAME*/
```

The input data file name, type, and location can be retrieved by right-clicking on the input data file and selecting Properties. Once in the Properties window, the user can see input data file details under the General tab. These details include file name, file type, and file location (Supplemental Fig. S1A). Similarly, the input data sheet name can be found on the left side of the bottom bar on the Excel file (Supplemental Fig. S1B). The input file location can also be viewed by clicking on the file path or address bar of the folder where the input data are stored (Supplemental Fig. S1C).

In SAS, the input data file is required to have missing records represented by a dot (.) and not to have a blank first row. We have also provided an input data file template, which is available at http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html. The input data file template is comprised of column names, including YR (year), LC (location), RP (replication), CL (cultigen or genotype), and dependent variables 1 to *n* (Supplemental Fig. S1D). An example input dataset is presented here. Hereafter, *genotype* is used to indicate cultigen, cultivar, or genotype. SASG×E is not sensitive to column position. However, SASG×E requires the user to not change the column names that are indicated in bold and uppercase. Dependent variables are indicated in lowercase and not bold, and the user is allowed to store multiple dependent variables. SASG×E is capable of analyzing multiple dependent variables simultaneously. Because the SAS programing language requires dataset and variable names with fewer than 32 characters, storing of short names for dependent variables and the input data set (<20 character is desired) is recommended.

SASG×E imports input data using PROC IMPORT. The sample input data set (`SASG×E_PROG_INPUT_DATA.XLSX`) consists of 3 yr, 5 locations, 2 replications, 10 genotypes, and 2 dependent variables (Fig. 2). The two dependent variables are marketable fruit weight (MKWT) and cull fruit weight (CLWT). Dependent variables are recorded at four different times for the same individuals within a plot (MKWT1–4 and CLWT1–4), and the unit is pounds per plot. A dataset or variable is longitudinal if it records the same type of information on the same individuals during a time period. Missing values are represented by a dot (.).

	A	B	C	D	E	F	G	H	I	J	K	L
1	YR	LC	RP	CL	MKWT1	CLWT1	MKWT2	CLWT2	MKWT3	CLWT3	MKWT4	CLWT4
2	2009	KN	1	8	0	0	79	0	27	0	20	0
3	2009	KN	1	10	0	0	136	13	10	0	19	3
4	2009	KN	1	3	0	0	72	0	31	0	16	0
5	2009	KN	1	5	31	0	60	0	12	15	8	4
6	2009	KN	1	7	0	0	69	7	22	0	18	0
7	2009	KN	1	4	0	0	138	0	18	0	0	0
8	2009	KN	1	9	0	0	121	0	17	0	0	0
9	2009	KN	1	2	62	0	70	0	19	0	29	0
10	2009	KN	1	6	0	0	116	0	66	0	0	0
11	2009	KN	1	1	0	0	37	0	19	0	39	0
12	2009	KN	2	3	27	0	83	0	50	0	21	4
13	2009	KN	2	10	0	0	54	22	62	16	4	6
14	2009	KN	2	8	0	0	59	10	95	0	17	0
15	2009	KN	2	9	4	0	13	0	18	8	15	0
16	2009	KN	2	2	0	10	171	0	26	0	0	28
17	2009	KN	2	4	5	0	5	124	0	4	11	6
18	2009	KN	2	7	42	10	27	0	0	0	11	0
19	2009	KN	2	1	0	0	118	11	13	0	15	5
20	2009	KN	2	5	38	0	62	0	66	0	16	0
21	2009	KN	2	6	0	0	138	0	18	0	0	0
22	2009	CI	1	7	47	7	55	0	84	0	0	0
23	2009	CI	1	3	72	0	17	7	9	5	0	0
24	2009	CI	1	1	54	0	87	0	20	0	0	0
25	2009	CI	1	4	33	0	60	0	0	0	0	0

Fig. 2. Screenshot of input sample data template consisting of year (YR), location (LC), replication (RP), genotype (CL), and dependent variables marketable fruit weight (MKWT1–4) and cull fruit weight (CLWT1–4) columns and top 25 rows.

Preprocess Input Data

Compute Sum of Longitudinal Variables

After importing data, SASG×E computes the sum of the longitudinal data (SUM function), renames genotype names (IF–ELSE–THEN statement), and drops those dependent variables for which stability statistics are not required (DROP statement). In the program below, dropped variables are highlighted and dependent variable values are converted from pounds per plot (lbs. plot^{-1}) to megagrams per hectare (Mg ha^{-1}). The two new dependent variables presented here are MKMGHA and CLMGHA (in bold below).

```
DATA TEMPA1 (RENAME=(CLT=CL));
  SET TEMPA1;

  MKWT=SUM(MKWT1,MKWT2,MKWT3,MKWT4);
    /*sum across the dependent variables*/
  CLWT=SUM(CLWT1,CLWT2,CLWT3,CLWT4);
    /*sum across the dependent variables*/

  MKMGHA=MKWT*0.40751; /*calculate yield Mg/ha
    for 12 ft plot size*/
  CLMGHA=CLWT*0.40751; /*factor 0.40751 converts
    lbs/plot to Mg/ha*/

        IF CL=01 THEN CLT='Mountain Hoosier';
  ELSE IF CL=02 THEN CLT='Hopi Red Flesh     ';
  ELSE IF CL=03 THEN CLT='Early Arizona      ';
  ELSE IF CL=04 THEN CLT='Starbrite F1       ';
  ELSE IF CL=05 THEN CLT='Stone Mountain     ';
  ELSE IF CL=06 THEN CLT='Stars-N-Stripes F1 ';
  ELSE IF CL=07 THEN CLT='AU-Jubilant        ';
  ELSE IF CL=08 THEN CLT='Calhoun Gray       ';
  ELSE IF CL=09 THEN CLT='Big Crimson        ';
  ELSE IF CL=10 THEN CLT='Legacy F1';

  DROP MKWT1 MKWT2 MKWT3 MKWT4 CLWT1
    CLWT2 CLWT3 CLWT4 MKWT CLWT CL;
RUN;
```

Except for YR, LC, RP, and CL, SASG×E treats other columns as dependent variables and computes stability statistics on each of them. Therefore, we suggest dropping dependent variables in this step if a stability statistic is not needed. This dataset step is optional. The user can skip this step if dependent variables in the input data do not need to be recreated. The SASG×E code will not be affected because dataset file names are the same (TEMPA1) as the file names in the previous step.

Compute Environment and Quality Check

A new additional variable environment (EN) is created and a quality check of missing values is performed in dataset TEMPA2. Environment is a location–year combination, which is in bold below. Dataset TEMPA2 serves as input variables for descriptive statistics, ANOVA, location statistics, and stability parameters (Fig. 1).

```
DATA TEMPA2;
  SET TEMPA1;
  EN=TRIM(LC)||'-'||TRIM(LEFT(YR)); /*
  ENV=LOC*YEAR */
  IF LC=' ' OR YR=. OR RP=. OR &DEPVAR = . THEN
  DELETE;
RUN;
```

Defining Macro

The macro UNIVARIATE1 computes the regression slope (b_i), standard error of the slope, deviation from regression (S_d^2), *t*-test and *F*-test on the regression slope (H_0: $b_i = 1$) and deviation from regression (H_0: $S_d^2 = 0$), and level of significance using PROC GLM. The results of these statistics are captured in the UNIVARIATE1&DEPVAR.SAS file, where &DEPVAR is a dependent variable name (Supplemental Fig. S2A). The .SAS file can be located in the current SAS session from the following path: SAS program window → Explorer panel → Explorer tab → Libraries → Work library. The level of significance of the *t*-test and *F*-test at 0.05, 0.01, and 0.001 is represented by *, **, and ***, respectively. The user can independently compute the regression slope, standard error of the slope, deviation from regression, and their level of significance from the macro UNIVARIATE1 while feeding input fields (in bold) in the code below. The dataset name TEMPA2 in the code below is the input dataset, which is comprised of column names, including YR (year), LC (location), RP (replication), CL (cultigen or genotype), EN (environment = location–year combination), and dependent variables. A schematic diagram of the macro UNIVARIATE1 to show its process is presented in Fig. 3.

```
*%Macro UNIVARIATE1 (INDPVAR=ENV&DEPVAR);
*Comment: Compute environmental index;
%LET DEPVAR=DEPENDENT_VARIABLE_NAME;
  /*user needs to enter dependent variable name*/

PROC SQL NOPRINT;
  CREATE TABLE DSTERM AS SELECT EN, RP, YR,
    LC, MEAN(&DEPVAR) AS ENV&DEPVAR
  FROM TEMPA2 GROUP BY EN, RP, YR, LC ORDER
    BY EN, RP, YR, LC;
  CREATE TABLE DST02 AS SELECT A.*,
    B.ENV&DEPVAR
  FROM TEMPA2 AS A
```

```
  LEFT JOIN DSTERM AS B ON (A.EN=B.EN AND
    A.RP=B.RP)
  ORDER BY CL;
QUIT;

%LET INDPVAR = ENV&DEPVAR;
PROC GLM DATA = DST02 OUTSTAT=OUTMSEDS2
    PLOTS=NONE;
  BY CL;
  CLASS CL LC RP EN;
  MODEL &DEPVAR = &INDPVAR  EN RP
    /SOLUTION SS1;
  ODS OUTPUT OVERALLANOVA=ANOVADS2
    PARAMETERESTIMATES=PARMGLMDS2;
RUN;

DATA OUTMSEDS3 (RENAME=
    (_SOURCE_=SOURCE));
  SET OUTMSEDS2 (WHERE=(_SOURCE_ NE "RP")
    KEEP=CL _NAME_ _SOURCE_  DF SS);
  MS=SS/DF;
RUN;

PROC TRANSPOSE DATA=OUTMSEDS3
    (RENAME=(_NAME_=DEPENDENT))
    OUT=MSDS;
  BY CL DEPENDENT;
  ID SOURCE;
  VAR MS;
RUN;

PROC TRANSPOSE DATA=OUTMSEDS3
    (RENAME=(_NAME_=DEPENDENT))
    PREFIX=DF_OUT=FDS3(DROP=_NAME_);
  BY CL DEPENDENT;
  ID SOURCE;
  VAR DF;
RUN;

DATA REGCOEFDS;
  SET PARMGLMDS2 (WHERE =
    (PARAMETER="&INDPVAR") KEEP=CL
    PARAMETER DEPENDENT ESTIMATE STDERR);
RUN;

PROC SORT DATA= MSDS; BY CL DEPENDENT;
  RUN;
PROC SORT DATA= REGCOEFDS; BY CL
  DEPENDENT; RUN;

DATA SLOPE (RENAME= (BI=SLOPE
    DEVLMS=DEVREG));
  MERGE MSDS (IN=A DROP=_NAME_
    RENAME=( ERROR=MSE  &INDPVAR=LREGMS
    EN=DEVLMS))
  REGCOEFDS (RENAME= (ESTIMATE=BI))
  FDS3;
  BY CL DEPENDENT;
    T_HO1=(BI-1)/STDERR;
   /*null hypothesis: Slope=1*/
    PT_HO1=2*(1-PROBT(ABS(T_HO1),
      DF_ERROR));
    IF  PT_HO1 LE 0.001 THEN SIG_SLOPE="***";
    ELSE IF  PT_HO1 LE 0.01 THEN
      SIG_SLOPE="**";
    ELSE IF  PT_HO1 LE 0.05 THEN SIG_SLOPE="*";

    F_DEVREG=DEVLMS/MSE;
     /*null hypothesis: predicted–actual=0*/
    PF_HO0=1-PROBF(F_DEVREG, DF_EN,
      DF_ERROR);
    IF  PF_HO0 LE 0.001 THEN SIG_DEVREG="***";
    ELSE IF  PF_HO0 LE 0.01 THEN
      SIG_DEVREG="**";
    ELSE IF  PF_HO0 LE 0.05 THEN
      SIG_DEVREG="*";
  /*Concatenate level of significance with slope
    and deviation*/
  SLOPE1 = PUT(BI, z5.3);
  SLOPE2 = SLOPE1||LEFT(TRIM(SIG_SLOPE));
  DEVREG1 = PUT(DEVLMS, z12.3);
  DEVREG2 = DEVREG1||LEFT(TRIM(SIG_
    DEVREG));
  DROP SLOPE1 DEVREG1;
RUN;

*Post process the data;
*Output for slope and dev from reg;
PROC SQL NOPRINT;
  /*Prepare dataset for correlation*/
  CREATE TABLE STABLE1&DEPVAR
  AS SELECT CL, DEPENDENT AS TRAIT, SLOPE,
    DEVREG
  FROM SLOPE;
  /*Prepare dataset for stability output*/
  CREATE TABLE UNIVARIATE1&DEPVAR
  AS SELECT CL, DEPENDENT AS TRAIT, SLOPE2
    AS SLOPE, STDERR AS STDERR_SLOPE, T_HO1
    AS TTEST_SLOPE, PT_HO1 AS PROB_SLOPE,
    DEVREG2 AS DEVREG, F_DEVREG AS
    FTEST_DEVREG, F_HO0 AS PROB_DEVREG
  FROM SLOPE;
QUIT;
*%MEND UNIVARIATE1;
```

Macro `UNIVARIATE2` computes Wricke's ecovalence (W_i^2), Shukla's stability variance (σ_i^2), and Perkins and Jinks β (β_i) using PROC IML and Lin and Binns P_i and Francis and Kannenberg coefficient of variation (CV_i) using PROC GLM. The results of these statistics are captured in the `UNIVARIATE2&DEPVAR.SAS` file (SAS window → Explorer panel → Explorer tab → Libraries → Work library), where `&DEPVAR` is a dependent variable name (Supplemental Fig. S2B). As for the macro `UNIVARIATE1`, the user can independently compute Wricke's ecovalence (W_i^2), Shukla's stability variance (σ_i^2), Perkins and Jinks β (β_i), Lin and Binns cultivar superiority measure (P_i), and Francis and Kannenberg coefficient of variation (CV_i) from the macro `UNIVARIATE2`. User-defined, required fields (data set `TEMPA2` and the name of a dependent variable) for `UNIVARIATE2` are in bold in the code below. A schematic diagram of `UNIVARIATE2` to illustrate its process is presented in Fig. 4.

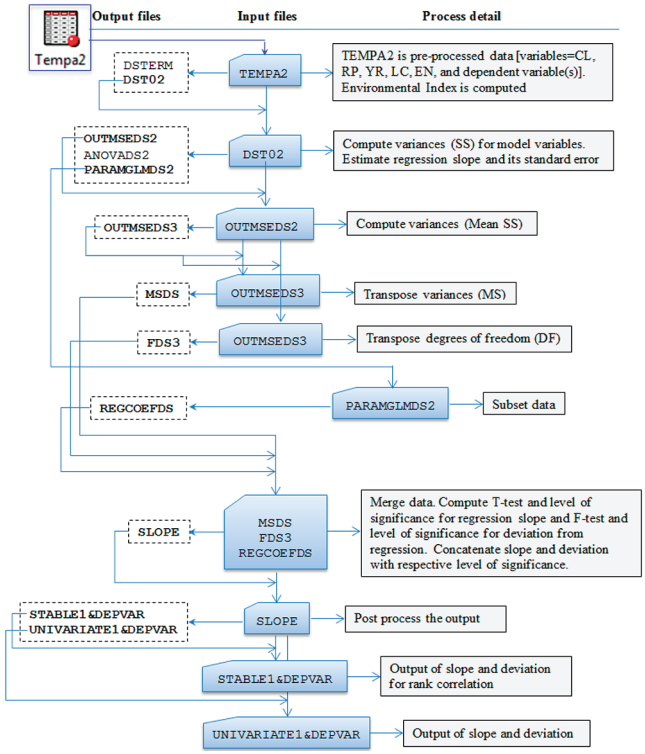


Fig. 3. Simplified flow diagram of the macro UNIVARIATE1 process.

```
*%MACRO UNIVARIATE2 (DEPVAR2=);
%LET DEPVAR=DEPENDENT_VARIABLE_NAME;
  /*User needs to enter dependent variable name*/

PROC SQL NOPRINT;
  ***Sum across ENV, Cultigen, Year, Location***;
  CREATE TABLE DSTECS AS SELECT EN, CL, YR,
    LC, SUM(&DEPVAR) AS SUM&DEPVAR
  FROM TEMPA2 GROUP BY EN, CL, YR, LC ORDER
    BY EN;

  ***List of distinct genotypes***;
  CREATE TABLE TEMP_CL
  AS SELECT DISTINCT (CL) AS DISTINCT_CL
  FROM TEMPA2 ORDER BY CL;

  ***Macro for total number of replications***;
  SELECT COUNT (DISTINCT(RP))INTO:
    TOTAL_RP TRIMMED
  FROM TEMPA2;

  ***Macro for total number of environments***;
  SELECT COUNT (DISTINCT(EN))INTO:
    TOTAL_EN TRIMMED
  FROM TEMPA2;
QUIT;
%PUT &TOTAL_RP = &TOTAL_EN =;

DATA DST01; SET DSTECS;
  BY EN;
  IF FIRST.EN THEN ET+1;
RUN;
PROC SORT DATA=DST01; BY CL; RUN;

%LET DEPVAR2=SUM&DEPVAR;
DATA DST01B;
  SET DST01;
  BY CL;
  ARRAY E(ET) E1-E&TOTAL_EN;
  RETAIN E1-E&TOTAL_EN;
  E=&DEPVAR2;
  IF LAST.CL THEN DO;
  OUTPUT;
    DO OVER E;
    E=.;
    END;
  END;
  KEEP E1-E&TOTAL_EN CL;
RUN;

PROC IML;
  *Reset AUTONAME;
  *Start MAIN;
    USE DST01B;
    READ ALL  INTO X;
    P=NROW(X); /*No of cultivars*/
    Q=NCOL(X); /*No of environments*/
    CMEAN=X[+,]/P; ** Column grand mean;
    CULT=J(P,Q);
    DO I={1} TO P;
      CULT[I,]=CMEAN[{1},{1}:Q];
       ***Generate matrix of column means (P,Q);
      END;
    U=X- CULT; **Residuals from overall mean;
    UM=U/Q; *** Get residual over number of
     col (responses);
    ENV=J(P,Q);
    DO K={1} TO Q;
      ENV[,K]=UM[,+];
      END;
    DIFF=U-ENV; /*Matrix of GXE residuals*/
    SSDIFF=(DIFF#DIFF)[,+];
    SUMSS=SUM(SSDIFF); /*Total SS resid*/

    N={&TOTAL_RP}; /*No of reps*/
    ECOV=SSDIFF/N; /*Wrickes ecovalence */
    L=P*(P-{1});
    E=(Q-{1})*(P-{1})*(P-{2});
    LSSDIFF=(SSDIFF*L)/N;
    F=J(P,{1},(SUMSS/N));
    SIG=LSSDIFF-F;

    SIGMA=SIG/E; /*Shuklas sigma*/
    TOT=SUM(X);
    GM=TOT/(P*Q);
    Z=J({1},Q,GM);
    ZJ=CMEAN-Z;
    SUMSQZJ=SUM(ZJ#ZJ);
    RAT=J(P,Q);
    DO R={1} TO P;
      RAT[R,]=ZJ[{1},{1}:Q];
      END;
    NEW=DIFF#RAT;

    BETA=(NEW/SUMSQZJ)[,+]; /*Regression of GEN
   on ENV means-using method of Perkins and Jinks*/
    GP=J(P,Q);
```

```
    DO C={1} TO Q;
      GP[,C]=BETA[{1}:P,{1}];
      END;
    BIZJ=RAT#GP;
    NEWDIFF=(DIFF-BIZJ);
    SI=(NEWDIFF#NEWDIFF)[,+];
    TS=P/((P-{2})*(Q-{2}));
    TOTSI=SUM(SI)/L;
    SP=((SI-TOTSI)*TS)/N; /*Shukla S squared*/
      CREATE IML_OUT VAR {BETA SIGMA
       ECOV};
      APPEND; CLOSE IML_OUT;
QUIT;

DATA TEMP_CL1 (RENAME=(DISTINCT_CL=CL));
  SET TEMP_CL;
  ID=_N_;
  TRAIT="&DEPVAR";
RUN;

DATA STAT2;
  SET IML_OUT;
  ID=_N_;
  RENAME BETA=BETA_PERKINS_AND_JINKS
   /*OUTPUT BETA_PERKINS AND JINKS*/
    SIGMA=SHUKLA /*OUTPUT
     SIGMA_SHUKLA*/
    ECOV=ECOVALENCE; /*OUTPUT WRICKE'S
     ECOVALENCE*/
RUN;

*Merge genotype name with stability parameters;
PROC SQL NOPRINT;
  CREATE TABLE TEMP_STABLE2
  AS SELECT A.CL, A.TRAIT, BETA_PERKINS_
   AND_JINKS, SHUKLA, ECOVALENCE
  FROM TEMP_CL1 AS A
  INNER JOIN STAT2 AS B ON A.ID=B.ID;
QUIT;

*Comment: Compute location*year variance (Pi)
  Lin and Binns (1988);
PROC SORT DATA = TEMPA2; BY CL; RUN;
*ODS TRACE ON/LISTING;
PROC GLM DATA = TEMPA2 PLOTS=NONE;
  BY CL;
  CLASS CL LC YR;
  MODEL &DEPVAR =LC YR LC*YR;
  RANDOM LC YR LC*YR/TEST;
    ODS OUTPUT
      OVERALLANOVA=OUTLANDB1
      RANDOMMODELANOVA=OUTLANDB2;
RUN;
*ODS TRACE OFF;

DATA OUTLANDB3_&DEPVAR;
  SET OUTLANDB2 (WHERE=
  (SOURCE="LC*YR"));
  MS1= PUT(MS, 10.5);
    IF  PROBF LE 0.001 THEN
      PI=TRIM(MS1)||TRIM(LEFT("***"));
    ELSE IF  PROBF LE 0.01 THEN
      PI=TRIM(MS1)||TRIM(LEFT("**"));
    ELSE IF  PROBF LE 0.05 THEN
      PI=TRIM(MS1)||TRIM(LEFT("*"));
    ELSE PI=MS1;
    KEEP CL DEPENDENT MS PROBF PI;
RUN;

*Comment: Compute genotypes coefficient of variation
  Francis & Kannenberg (1978);
PROC SQL NOPRINT;
  CREATE TABLE PLAISTEDCV_&DEPVAR AS
  SELECT  CL, CV(&DEPVAR) AS CVI
  FROM TEMPA2
  GROUP BY CL;
QUIT;

*Post process the data;
*Output for Shukla, ecovalence, beta, Pi and CV;
PROC SQL NOPRINT;
  /*Prepare dataset for correlation*/
  CREATE TABLE STABLE2&DEPVAR
  AS SELECT A.CL, A.TRAIT, A.BETA_PERKINS_
   AND_JINKS, A.SHUKLA, A.ECOVALENCE, B.MS
   AS PI, C.CVI
  FROM TEMP_STABLE2 AS A
  INNER JOIN OUTLANDB3_&DEPVAR AS B ON
   A.CL=B.CL
  INNER JOIN PLAISTEDCV_&DEPVAR AS C ON
   B.CL=C.CL;
  /*Prepare dataset for stability output*/
  CREATE TABLE UNIVARIATE2&DEPVAR
  AS SELECT A.CL, A.TRAIT, A.BETA_PERKINS_
   AND_JINKS, A.SHUKLA, A.ECOVALENCE,
   B.PI, C.CVI
  FROM TEMP_STABLE2 AS A
  INNER JOIN OUTLANDB3_&DEPVAR AS B ON
   A.CL=B.CL
  INNER JOIN PLAISTEDCV_&DEPVAR AS C ON
   B.CL=C.CL;
QUIT;
*%MEND UNIVARIATE2;
```

The macro `UNIVARIATE3` computes least square (LS) means, standard error of LS means, least significant difference (LSD) of the mean, and Kang's yield-stability statistic (YS_i). Here Kang's YS_i is computed based on the steps outlined by Mekbib (2003) in the code below. The results of macro `UNIVARIATE3` are captured in the `UNIVARIATE3&DEPVAR.SAS` file (SAS window → Explorer panel → Results tab → Libraries → Work library), where `&DEPVAR` is a dependent variable name (Supplemental Fig. S2C). The LSD is used to compare trait means across genotypes. As for other stability parameters, the user can compute the LS means, standard error of the LS means, LSD of the mean, and YS_i independently. However, the user is required to run the macros `UNIVARIATE2` and `UNIVARIATE3` in sequential order. This is because `UNIVARIATE3` reads the dataset `TEMP_STABLE2` from `UNIVARIATE2` (Fig. 5). The user-defined, required field (dataset `TEMPA2`) for the macro

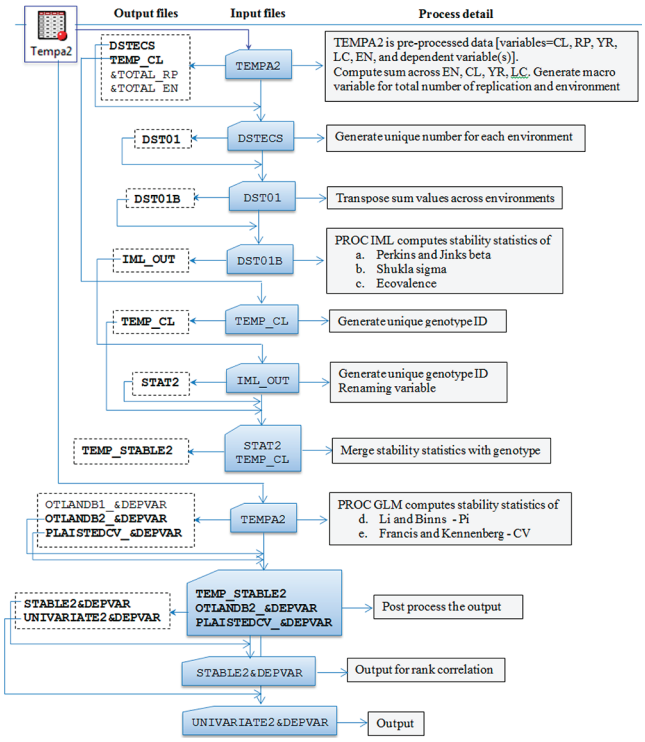


Fig. 4. Simplified flow diagram of the macro UNIVARIATE2 process.

UNIVARIATE3 is in bold in the code below. A schematic diagram of UNIVARIATE3 to illustrate its process is presented in Fig. 5.

```
*%Macro UNIVARIATE3;
%LET DEPVAR = DEPENDENT_VARIABLE_NAME;
  /*User needs to enter dependent variable name*/

*Comment: Macro for total no of REP and ENV;
*Comment: Compute environmental index;
PROC SQL NOPRINT;
  SELECT COUNT (DISTINCT(RP))INTO:
    TOTAL_RP TRIMMED FROM TEMPA2;
  SELECT COUNT (DISTINCT(EN))INTO:
    TOTAL_EN TRIMMED FROM TEMPA2;

  CREATE TABLE DSTERM AS SELECT EN, RP, YR,
    LC, MEAN(&DEPVAR) AS ENV&DEPVAR
    /*Environmental index*/
  FROM TEMPA2 GROUP BY EN, RP, YR, LC
    ORDER BY EN, RP, YR, LC;

  CREATE TABLE DST02 AS SELECT A.*,
    B.ENV&DEPVAR
  FROM TEMPA2 AS A
  LEFT JOIN DSTERM AS B ON (A.EN=B.EN
    AND A.RP=B.RP)
  ORDER BY CL;
QUIT;

PROC GLM
  DATA=DST02 OUTSTAT=OUTMSDS
    PLOTS=NONE;
  CLASS CL LC RP EN;
  MODEL &DEPVAR= EN RP(EN) CL (EN);
  LSMEANS CL(EN)/STDERR
    OUT=CLTLSMNDS1 SLICE=(EN CL);
  ODS OUTPUT OVERALLANOVA=ANOVADS
  FITSTATISTICS=DEPMEANDS;
RUN;

PROC SQL NOPRINT;
  CREATE TABLE CLTLSMNDS2
  AS SELECT CL, MEAN(LSMEAN) AS LSMEAN,
    MEAN(STDERR) AS STDERR
  FROM CLTLSMNDS1 GROUP BY CL
    ORDER BY CL;
QUIT;

DATA CLTLSMNDS;
  SET CLTLSMNDS2;
  _NAME_="&DEPVAR";
RUN;
PROC SORT DATA=CLTLSMNDS;BY CL;RUN;

DATA SEE1;
  IF _N_=1 THEN MERGE ANOVADS(IN=
    A WHERE=(SOURCE = 'Error') KEEP= SOURCE
    MS DF) DEPMEANDS(IN=B KEEP= DEPMEAN);
  ELSE SET CLTLSMNDS;
    SE_DIFF=SQRT(MS*(2*(1/(&TOTAL_EN*
    &TOTAL_RP))));
    T_DFE=TINV(0.975, DF); /*alpha=0.975*/
    LSD=T_DFE*SE_DIFF;
  IF LSMEAN LE (DEPMEAN-2*LSD)THEN
    SCORE_LSD=-3;
  ELSE IF LSMEAN LE (DEPMEAN-LSD) THEN
    SCORE_LSD=-2;
  ELSE  IF  LSMEAN LE DEPMEAN THEN
    SCORE_LSD=-1;

  IF LSMEAN GE (DEPMEAN+2*LSD) THEN
    SCORE_LSD=3;
  ELSE IF LSMEAN GE (DEPMEAN+LSD) THEN
    SCORE_LSD=2;
  ELSE  IF  LSMEAN GE DEPMEAN THEN
    SCORE_LSD=1;
RUN;

DATA SEE1;
  SET SEE1;
  IF _N_ GT 1;
RUN;
PROC SORT DATA=SEE1;BY CL;RUN;
PROC SORT DATA=TEMP_STABLE2;BY CL;RUN;
  /*Dataset TEMP_STABLE2 is generated from macro
  %UNIVARIATE2*/

DATA SEE2 (DROP = SHUKLA_);
  MERGE SEE1 TEMP_STABLE2 (KEEP =
    CL SHUKLA);
  BY CL;
  F_CALC=SHUKLA/MS;
  PF_SHUKLA=1-PROBF(F_CALC,
    (&TOTAL_EN-1),DF);

  IF PF_SHUKLA  LE 0.01 THEN SIG_SHUKLA=-8;
```

```
   ELSE IF  PF_SHUKLA LE 0.05 THEN
     SIG_SHUKLA=-4;
   ELSE IF  PF_SHUKLA LE 0.10 THEN
     SIG_SHUKLA=-2;
   ELSE SIG_SHUKLA=0;

   SHUKLA_= PUT(SHUKLA, 10.5);
   IF PF_SHUKLA LE 0.001 THEN SHUKLA_TEST=
     TRIM(SHUKLA_)||TRIM(LEFT("***"));
   ELSE IF PF_SHUKLA LE 0.01 THEN SHUKLA_
     TEST=TRIM(SHUKLA_)||TRIM(LEFT("**"));
   ELSE IF PF_SHUKLA LE 0.05 THEN SHUKLA_
     TEST=TRIM(SHUKLA_)||TRIM(LEFT("*"));
   ELSE SHUKLA_TEST=SHUKLA_;
RUN;

PROC RANK DATA=SEE2  OUT=RNK&DEPVAR;
   VAR LSMEAN;
   RANKS YRANK;
RUN;

PROC SORT DATA= RNK&DEPVAR;
   BY  DESCENDING YRANK;RUN;

DATA  RNK&DEPVAR;
   SET RNK&DEPVAR;
   SUMMED=YRANK +SCORE_LSD;
   YS=SUMMED +SIG_SHUKLA;
RUN;

*Comment mark (@) stable genotype based on
   YS > mean(YS);
PROC SQL NOPRINT;
   SELECT MEAN(YS)INTO: MEAN TRIMMED
     FROM RNK&DEPVAR;
QUIT;
%PUT MEAN =&MEAN;

DATA RNK&DEPVAR;
   SET RNK&DEPVAR;
   YS_=PUT(YS,8.0);
   IF YS GT &MEAN THEN YS_TEST=
     TRIM(YS_)||TRIM(LEFT("@"));
   ELSE YS_TEST=YS_;
   DROP YS_;
RUN;

*Post process the data;
*Output for LSMEAN, LSD, YS, test of SIG. for SHUKLA;
PROC SQL NOPRINT;
   /*Prepare dataset for correlation*/
   CREATE TABLE STABLE3&DEPVAR
   AS SELECT CL, _NAME_ AS TRAIT, LSMEAN,
     LSD, YS
   FROM RNK&DEPVAR;

   /*Prepare dataset for stability output*/
   CREATE TABLE UNIVARIATE3&DEPVAR
   AS SELECT CL, _NAME_ AS TRAIT, LSMEAN, LSD,
     SHUKLA_TEST AS SHUKLA, YS_TEST AS YS
   FROM RNK&DEPVAR;
QUIT;
*%MEND UNIVARIATE3;
```

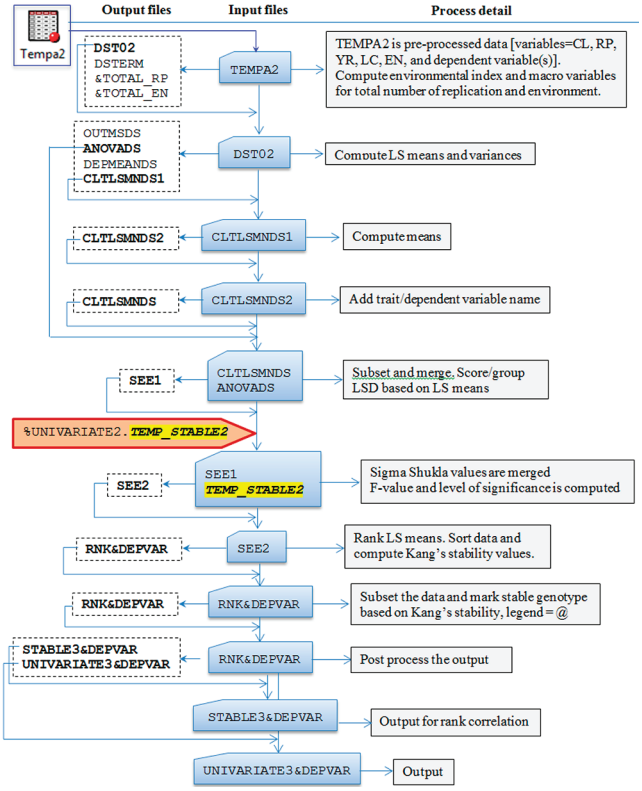


Fig. 5. Simplified flow diagram of the macro UNIVARIATE3 process.

The macro LEVELOFSIG concatenates the Spearman correlation value with the level of significance. The level of significance at 0.05, 0.01, and 0.001 is represented by *, **, and ***, respectively.

The macro OUTPUTEXCEL exports output files in Excel (.xlsx only) and to the same folder or location where the input data file is placed. Similarly, the macro OUTPUTCSV exports output files as comma-separated value (.CSV) files and to same folder or location where the input data file is placed. These .CSV files are loaded into RStudio to analyze multivariate stability statistics using the AMMI and GGE biplot models. The macros GENOTYPE, ENVIRONMENT, and LOCATION generate shorter names for genotypes or cultivars, environment, and location, respectively, so that visualization of the AMMI and GGE biplot output is legible.

Creating Macro Variables for Dependent Variables during DATA Step Execution

SASG×E creates macro variables for the total number of (&LAST_DEPVARIABLE) and each dependent variable (&DEPVARX) using SYMPUT routine in a DATA step so that it knows the number of iterations the program needs to run. To improve program efficiency, the DATA _NULL_ statement was used:

```
DATA _NULL_;
   SET START1 END=END_OF_DATASET;
   CALL SYMPUT ('DEPVARX'||TRIM(LEFT(_N_)),
     NAME); /*macro for dependent variable*/
   IF END_OF_DATASET THEN CALL SYMPUT
     ('LAST_DEPVARIABLE', COMPRESS(_N_));
RUN;
```

Compute Stability Statistics of All Dependent Variables Simultaneously, Using Macro STABILITY

The macro STABILITY computes different stability statistics for multiple dependent variables using iterative %DO and %END statements. During each iteration, one dependent variable is analyzed. The %DO loop stops processing after the stop value is equal to &LAST_DEPVARIABLE.

Input data are quality checked for missing records and the environment is defined in the DATA TEMPA2 statement. SASG×E removes rows having missing records for location, year, replication, or dependent variable. The environment (EN) is a combination of year and location. Descriptive statistics, including means, sum, and CV, are computed using PROC MEANS and PROC SQL. Using PROC TRANSPOSE, the results of the descriptive statistics are transposed in a user-friendly layout so that researchers can interpret them easily. SASG×E generates the following descriptive statistics:

- Trait mean across genotypes and environments (MEAN&DEPVAR.SAS)
- Trait sum across genotypes and environments (SUM&DEPVAR.SAS)
- Trait mean across genotypes, environments, and replications (ENV&DEPVAR.SAS)
- Trait mean across genotypes, years, locations, and replications (M_&DEPVAR_CYLR.XLSX)
- Trait mean across genotypes, years, and locations (M_&DEPVAR_CYL.XLSX)
- Trait mean across genotypes and years (M_&DEPVAR_CY.XLSX)
- Trait mean across genotypes and locations (M_&DEPVAR_CL.XLSX)
- Trait CV across genotypes and locations (CV_&DEPVAR_CL.XLSX)
- Trait mean across genotypes (M_&DEPVAR_C.XLSX)
- Trait mean across locations (M_&DEPVAR_L.XLSX)
- Trait mean across locations and years (M_&DEPVAR_LY.XLSX)
- Trait mean across genotypes and environments (M_&DEPVAR_CE.XLSX)
- Trait mean across genotypes, locations, and replications (M_&DEPVAR_CLR.XLSX)
- Trait mean across genotypes, environments. and replications (M_&DEPVAR_CER.XLSX)

The macro STABILITY automatically exports .xlsx output files of descriptive statistics to the same location or folder where the input data file is placed. The user can view .SAS output files of descriptive statistics in the current SAS session at the following location: SAS program window → Explorer panel → Explorer tab → Libraries → Work library.

An ANOVA is computed using PROC GLM to determine the size and significance of the G×E of the dependent variable. SASG×E considers genotypes, years, locations, and replications as random effects. An *F*-test is used to test the significance of each factor. The level of significance of the *F*-test at 0.05, 0.01 and 0.001 is represented by *, **, and ***, respectively. When two or more factors (out of three or more factors) are random, then an exact error term for the *F*-test is not found for all the factors. Therefore, PROC GLM estimates the correct variance component to test each factor. The output file OVERALL_ANOVA_&DEPVAR.XLSX reports pooled ANOVA results along with the test term for each factor (Supplemental Fig. S2D).

Similarly, SASG×E computes statistics for identifying a location's discriminating ability and representativeness. Location statistics include yearly ANOVA, overall genotype *F*-ratio across each location, standard deviation, correlation of average location with individual location, and cluster analysis. Results of the location statistics are captured in YEAR_ANOVA_&DEPVAR.XLSX, LOCATION_VALUE_&DEPVAR.XLSX, and LOCTREE_&DEPVAR.PDF. Cluster analysis output (.pdf) is exported to the same folder or location where the input data (.xlsx) file is placed (Supplemental Fig. S1C).

The macro STABILITY calls the macros UNIVARIATE1, UNIVARIATE2, and UNIVARIATE3 to compute univariate stability statistics. These univariate statistics include regression slope (b_i), standard error of the slope, deviation from regression (S_d^2), *t*-test on regression slope (H_0: $b_i = 1$), *F*-test deviation from regression (H_0: $S_d^2 = 0$), Wricke's ecovalence (W_i^2), Shukla's stability variance (σ_i^2), Perkins and Jinks β (β_i), Lin and Binns cultivar superiority measure (P_i), Francis and Kannenberg coefficient of variation (CV_i), LS means, standard error of LS means, LSD, and Kang's yield stability statistic (YS_i). The major reasons for defining the macros UNIVARIATE1, UNIVARIATE2, and UNIVARIATE3 outside of the macro STABILITY are to add keyword parameters and to avoid nested macros so that the program is efficient, flexible, and easy to debug.

SASG×E assigns ranks to genotypes for each stability parameter. Spearman's rank correlation is computed using PROC CORR on the ranks to measure the relationship between stability parameters. Genotypes are ranked in increasing order for decreased values of a dependent variable. However, for certain dependent variables such as disease, cull fruits percentage, etc., where a lower trait value is considered to be good, the user is required to assign a higher ordinal value to a lower trait value of such dependent variables. Otherwise, Spearman's rank correlation will give the wrong (or opposite) output and mislead the user. Genotypes are ranked in increasing order for increased values of S_d^2, W_i^2, σ_i^2, P_i, CV_i, and YS_i. A regression slope (b_i) approximating unity is considered to be stable; therefore genotypes are ranked in increasing order when $b_i > 1$ and decreasing order when $b_i < 1$. Similarly, β_i is stable near zero; therefore genotypes are ranked in increasing order when $\beta_i > 0$ and decreasing order when $\beta_i < 0$. The level of significance of correlation at 0.05, 0.01, and 0.001 is represented by *, **, and ***, respectively.

The macro STABILITY invoked the macros OUTPUTEXCEL and OUTPUTCSV to generate output files in .xlsx and .csv formats, respectively. These output files are automatically sent to the same location or folder where the input data file is placed (Supplemental Fig. S1A and S1C). The following are the output files generated by the macro OUTPUTEXCEL:

- Trait mean across genotype, year, location, and replication (M_&DEPVAR_CYLR.XLSX)
- Trait mean across genotype, year, and location (M_&DEPVAR_CYL.XLSX)
- Trait mean across genotype and year (M_&DEPVAR_CY.XLSX)
- Trait mean across genotype and location (M_&DEPVAR_CL.XLSX)
- Trait CV across genotype and location (CV_&DEPVAR_CL.XLSX)
- Trait mean across genotype (M_&DEPVAR_C.XLSX)
- Trait mean across location (M_&DEPVAR_L.XLSX)
- Trait mean across location and year (M_&DEPVAR_LY.XLSX)
- Trait mean across genotype and environment (M_&DEPVAR_CE.XLSX)
- Trait mean across genotype, environment, and replication (M_&DEPVAR_CER.XLSX)
- Trait mean across genotype, location, and replication (M_&DEPVAR_CLR.XLSX)
- Pooled analysis of variance (OVERALL_ANOVA_&DEPVAR.XLSX)
- Yearly analysis of variance (YEAR_ANOVA_&DEPVAR.XLSX)
- Location value (LOCATION_VALUE_&DEPVAR.XLSX)
- Univariate stability statistics (STAB_&DEPVAR.XLSX)
- Spearman's rank correlation (ANOVA_&DEPVAR.XLSX)
- Legend for location used in AMMI (LOC_LEGEND_&DEPVAR.XLSX)
- Legend for genotype used in AMMI and GGE biplot (GEN_LEGEND_&DEPVAR.XLSX)
- Legend for environment used in AMMI and GGE biplot (ENV_LEGEND_&DEPVAR.XLSX)

The following are the output files generated by the macro OUTPUTCSV:

- Input file for RStudio for GGE biplot (genotype × environment) analysis (BIPLOT_&DEPVAR_.CSV)
- Input file for RStudio for GGE biplot (genotype × location) analysis (BIPLOT2_&DEPVAR_.CSV)
- Input file for RStudio for AMMI (genotype × environment) analysis (AMMI1_&DEPVAR_.CSV)
- Input file for RStudio for AMMI (genotype × location) analysis (AMMI2_&DEPVAR_.CSV)

Multivariate Statistics

SASG×E does not compute multivariate statistics (AMMI and GGE biplot) for stability analysis per se. However, files (BIPLOT_&DEPVAR_.CSV, BIPLOT2_&DEPVAR_.CSV, AMMI1_&DEPVAR_.CSV, and AMMI2_&DEPVAR_.CSV) generated by SASG×E can be loaded into RStudio to compute multivariate stability statistics (RStudio, 2014). These files are compatible with AMMI (*agricolae*) and GGE biplot (GGEBiplotGUI) packages available at the CRAN repository and hence ready to go in RStudio. However, the user needs to be cautious with the case sensitivity of the R computing language. RStudio is an integrated tool designed to help the user be more productive with R computing software, and it requires R Version 2.11.1 or higher. To improve the visuals of AMMI and GGE biplot analysis, genotypes, locations, and environments are abbreviated as G1 to G*n*, LOC1 to LOC*n*, and ENV1 to ENV*n*, respectively, where *n* is the total number of entities. The user can view the respective abbreviations for a corresponding genotype, location, and environment in GEN_LEGEND_&DEPVAR.XLSX, LOC_LEGEND_&DEPVAR.XLSX, and ENV_LEGEND_&DEPVAR.XLSX files, respectively.

To analyze stability using the AMMI model, the user needs to select the AMMI( ) function of the *agricolae* package in the system library window of RStudio. If the *agricolae* package is not found in the system library, then the user can install it from the CRAN repository (Comprehensive R Archive Network, 2014) (Supplemental Fig. S3A). Additionally, we have provided *install.packages* and *library* commands in R code that installs the library from CRAN and calls the library, respectively. Then reference the path of the folder where the input data are located from Session in the window tool bar (Session in window tool bar → select Set Work Directory → select Choose Work Directory → select folder where data are kept) (Supplemental Fig. S3B). The user can also reference the path in code in the Console or R Script window. However, the user needs to be cautious about the requirement of a forward slash in the reference path in R computing language. The user can make use of the code below in the Console or R Script window to analyze the AMMI model. The output files AMMI1_&DEPVAR_.CSV and AMMI2_&DEPVAR_.CSV generate genotype × environment and genotype × location analysis results, respectively. The AMMI outputs from the example input dataset (Supplemental Fig. S4A and S4B) and their interpretation can be found in the supplemental material or at http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html. In AMMI analysis, the principal components (PCs) are often called interaction PCs. However, the AMMI( ) function of the *agricolae* package does not have the option to pass user-defined arguments to label the axis of AMMI biplots. Therefore, the default label PC represents the axis of AMMI biplots.

```
#Comment: User needs to replace input data file path
setwd("E:/PhD Research Work/PhD Articles/Articles for
   Publication/G×E SAS Prog/Sample Data")

#Comment: User needs to replace file name (AMMI2_MK-
   MGHA). It is case sensitive
Data = read.csv(file="AMMI1_MKMGHA.csv",
   header=TRUE)

#Comment: View top 6 rows of data
head(Data)

attach(Data)

#install library agricolae
#user can comment out install.packages command to
```

```
   turn it off
install.packages("agricolae")

#Call agricolae library
library (agricolae)

#Comment: User needs to replace dependent variable name
   (MKMGHA). It is case sensitive
model <- AMMI(Locality, Genotype, Rep, MKMGHA,
   console=FALSE)
model$ANOVA

#Comment: see help(plot.AMMI)
detach(Data)

#Comment: biplot
plot(model)

#Comment: triplot PC 1,2,3
plot(model, type=2, number=TRUE)

#Comment: biplot PC1 vs. DEPENDENT VARIABLE
plot(model, first=0,second=1, number=TRUE)
```

Similarly, the user needs to select the GGEBiplotGUI package in the system library window of RStudio to compute the GGE biplot model. If the GGEBiplotGUI package is not found in the system library, then the user should install it from the CRAN repository (Comprehensive R Archive Network, 2014) (Supplemental Fig. S3C). Additionally, we have provided *install.packages* and *library* commands in the R code shown below that will install the library from CRAN and call the library, respectively. Then, reference the path of the folder where the input data are located from Session in the window tool bar (Session in window tool bar → select Set Work Directory → select Choose Work Directory → select folder where data are kept) (Supplemental Fig. S3B). The user can also reference the path in code in the Console or R Script window. However, the user needs to be cautious about the requirement for a forward slash in the reference path in the R computing language, which is the opposite of SAS. The GGEBiplotGUI package accepts input data where rows are labeled and no blank [NA] records. Therefore, the system-defined function *rownames* was used to label the rows. Similarly, the user-defined function *na_check* was used to replace blank records with the trait mean for the genotype across locations [gge=na_check(gge,"Mean")] or by zero [gge=na_check(gge,"Zero")]. The user has the option to choose either the mean or zero to replace blank [NA] records. If the input data do not have missing records, then the program will process the data per se. The user can use the code below in the Console or R Script window to run GGEBiplot.

```
#Comment: User needs to replace input data file path
setwd("E:/PhD Research Work/PhD Articles/Articles for
   Publication/G×E SAS Prog/Sample Data")

#install library agricolae
#user can comment out install.packages command
   to turn it off
install.packages("GGEBiplotGUI")

#Call GGE Biplot library
library (GGEBiplotGUI)

#Comment: User needs to replace file name. It is
   case sensitive
gge=read.csv(file="BIPLOT2_MKMGHA.csv",
   header=TRUE)

#Comment: View top 6 rows of data
head(gge)

#Comment: colnames( ) gives column labels
#Comment: rownames( ) gives row labels
rownames(gge)=gge[,1]
gge=gge[,-1]

#Comment: View top 6 rows of data
head(gge)

#Make a function to find all the NA (blank) values and
   replace with either row_mean or zero
na_check=function(dat,check)
{
 for(i in 1:nrow(dat))
 {
  for(h in 1:ncol(dat))
  {
   if (is.na(dat[i,h])==T)
   {
    if (check=="Mean")
    {
     dat[i,h]=mean(na.omit(as.numeric(dat[i,])))
     {
      if(check=="Zero")
      {
       dat[i,h]=0
      }
     }
    }
   }
  }
 }
 return(dat)
}

#Comment: Replace blank record with mean or zero using
   user defined function na_check
gge=na_check(gge,"Mean")

#Comment: View top 6 rows of data
head(gge)

#Comment: GGE biplot analysis
GGEBiplot(Data=gge)
```

Upon executing the code, the Model Selection window opens and the user is required to populate the dropdowns for SVP, Centered By, and Scaled (Divided) By to generate appropriate biplots as described by Yan et al. (2007) (Supplemental Fig. S3D). The singular value partition (SVP) attribute of the Model Selection window has dropdown values of 1 and 2 (Supplemental Fig. S3D), where a value of 1 indicates that singular values are partitioned into genotype scores, enhancing the biplot for comparing genotypes, and a value of 2 indicates

that singular values are partitioned into environment scores, enhancing the biplot for visualizing the relationships among environments (Yan et al., 2007).

Similarly, the Centered By attribute in the Model Selection window has dropdown values of 0, 1, 2, or 3 (Supplemental Fig. S3D). A value of 0 means original data are used for visualization—effective for a dataset whose grand mean is close to 0. It is useful in quantitative trait loci × environment interaction and genotype covariate × environment interaction studies. A value of 1 means the grand mean is centered and useful when both row (genotype) and column (environment) main effects are of interest, while a value of 2 means the data are environment centered and a value of 3 means the data are double-centered—desirable if G×E is of sole interest and for gene expression studies.

The dropdown values of the Scaled By attribute in the Model Selection window include 0, 1, 2, and 3 (Supplemental Fig. S3D). The purpose of scaling is to put the variables in comparable ranges by dividing by different parameters. Scaling is necessary if the variables are of different units. A value of 0 means the data are not divided by anything. A value of 1 means the data are rescaled by the within-environment standard deviation and assumes all environments to be equally important. A value of 2 means the data are rescaled by within-environment standard errors. It removes any heterogeneity among environments with regard to their experimental errors while retaining the information about the environment's discriminating ability (Yan et al., 2007). A value of 3 means the data are rescaled by the environment means. It removes the differences in units and data ranges among the variables while retaining the discriminating ability of the environments (Yan et al., 2007).

The SASG×E output files `BIPLOT_&DEPVAR_.CSV` and `BIPLOT2_&DEPVAR_.CSV` generate genotype × environment and genotype × location biplot results, respectively. A detailed description and the differences between the AMMI and GGE biplot models was presented by Yan et al. (2007). The GGE biplot outputs of the example dataset (Supplemental Fig. S5A, S5B, and S5C) and their interpretation are presented in the supplemental material or can be found at http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html.

SUPPLEMENTAL MATERIAL

The supplemental material available online includes the SASG×E program, example input data, graphical illustrations for uploading an input data file in SASG×E and installing R packages in RStudio, outputs from the example input data, and interpretation of univariate and multivariate statistical analysis.

ACKNOWLEDGMENTS

We wish to thank Dr. David. A. Dickey and Ms. Joy Smith of North Carolina State University for their assistance with the statistical analysis. We are grateful to the anonymous reviewers for their comments.

REFERENCES

Becker, H.C. 1981. Correlation among some stability measures of phenotypic stability. Euphytica 30:835–840. doi:10.1007/BF00038812

Becker, H.C., and J. Leon. 1988. Stability analysis in plant breeding. Plant Breed. 101:1–23. doi:10.1111/j.1439-0523.1988.tb00261.x

Carbonell, S.A.M., J.A. de Azevedo Filho, L.A.S. Dias, A.F.F. Garcia, and L.K. de Morais. 2004. Common bean cultivars and lines interactions with environments. Sci. Agric. 61(2):169–177. doi:10.1590/S0103-90162004000200008

Casanoves, F., J. Baldessari, and M. Balzarini. 2005. Evaluation of multienvironment trials of peanut cultivars. Crop Sci. 45:18–26. doi:10.2135/cropsci2005.0018

Comprehensive R Archive Network. 2014. Available CRAN packages by name. R Found. Stat. Comput., Vienna. http://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-A (accessed 19 Mar. 2016).

Crossa, J., H.G. Gauch, and R.W. Zobel. 1990. Additive main effects and multiplicative interaction analysis of two maize cultivar trials. Crop Sci. 30:493–500. doi:10.2135/cropsci1990.0011183X003000030003x

Dehghani, H., A. Ebadi, and A. Yousefi. 2006. Biplot analysis of genotype by environment interaction for barley yield in Iran. Agron. J. 98:388–393. doi:10.2134/agronj2004.0310

de Mendiburu, F. 2015. agricolae: Statistical procedures for agricultural research. R package Version 1.2-3. R Found. Stat. Comput., Vienna. http://CRAN.R-project.org/package=agricolae (accessed 19 Mar. 2016).

Dia, M., T.C. Wehner, and C. Arellano. 2016a. Cucurbit breeding project: rg×e 1.1. Dep. of Hortic. Sci., North Carolina State Univ., Raleigh. http://cuke.hort.ncsu.edu/cucurbit/wehner/software.html (accessed 15 Mar. 2016).

Dia, M., T.C. Wehner, R. Hassell, D.S. Price, G.E. Boyhan, S. Olson, et al. 2016b. Genotype × environment interaction and stability analysis for watermelon fruit yield in the United States. Crop Sci. 56. doi:10.2135/cropsci2015.10.0625

Dia, M., T.C. Wehner, R. Hassell, D.S. Price, G.E. Boyhan, S. Olson, et al. 2016c. Values of locations for representing mega-environments and for discriminating yield of watermelon in the U.S. Crop Sci. 56. doi:10.2135/cropsci2015.11.0698

Eberhart, S.A., and W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6:36–40. doi:10.2135/cropsci1966.0011183X000600010011x

Fan, X.M., M.S. Kang, H. Chen, Y. Zhang, J. Tan, and C. Xu. 2007. Yield stability of maize hybrids evaluated in multi-environment trials in Yunnan, China. Agron. J. 99:220–228. doi:10.2134/agronj2006.0144

FAO. 2015. Measures of yield stability and reliability. FAO, Rome. http://www.fao.org/docrep/005/y4391e/y4391e0a.htm#bm10.1 (accessed 19 Mar. 2016).

Francis, T.R., and L.W. Kannenberg. 1978. Yield stability studies in short-season maize: A descriptive method for grouping genotypes. Can. J. Plant Sci. 58:1029–1034. doi:10.4141/cjps78-157

Frutos, E., M.P. Galindo, and V. Leiva. 2014. An interactive biplot implementation in R for modeling genotype-by-environment interaction. Stochastic Environ. Res. Risk Assess. 28:1629–1641. doi:10.1007/s00477-013-0821-z

Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gauch, H.G. 2006. Statistical analysis of yield trials by AMMI and GGE. Crop Sci. 46:1488–1500. doi:10.2135/cropsci2005.07-0193

Gauch, H.G., and R.W. Zobel. 1996. AMMI analysis of yield trials. In: M.S. Kang and H.G. Gauch, editors, Genotype by environment interaction. CRC Press, Boca Raton, FL. p. 85–122. doi:10.1201/9781420049374.ch4

Hussein, M.A., A. Bjornstad, and A.H. Aastveit. 2000. SASG×ESTAB: A SAS program for computing genotype × environment stability statistics. Agron. J. 92:454–459. doi:10.2134/agronj2000.923454x

Kang, M.S. 1993. Simultaneous selection for yield and stability in crop performance trials: Consequences for growers. Agron. J. 85:754–757. doi:10.2134/agronj1993.00021962008500030042x

Kang, M.S., J.D. Miller, and L.L. Darrah. 1987. A note on relationship between stability variance and ecovalence. J. Hered. 78:107.

Lin, C.S., and M.R. Binns. 1988. A method for analyzing cultivar × location × year experiments: A new stability parameter. Theor. Appl. Genet. 76:425–430. doi:10.1007/BF00265344

Mekbib, F. 2003. Yield stability in common bean (*Phaseolus vulgaris* L.) genotypes. Euphytica 130:147–153. doi:10.1023/A:1022878015943

Perkins, J.M., and J.L. Jinks. 1968. Environmental and genotype–environment components of variability: III. Multiple lines and crosses. Heredity 23:339–356. doi:10.1038/hdy.1968.48

Piepho, H.S. 1999. Stability analysis using the SAS system. Agron. J. 91:154–160. doi:10.2134/agronj1999.00021962009100010024x

RStudio. 2014. RStudio: Integrated development environment for R. Version 0.98.1074. RStudio, Boston, MA. http://www.rstudio.com/ (accessed 19 Mar. 2016).

SAS Institute. 2009. SAS 9.2 macro language: Reference. SAS Inst., Cary, NC. http://support.sas.com/documentation/cdl/en/mcrolref/61885/PDF/default/mcrolref.pdf (accessed 19 Mar. 2016).

Shukla, G.K. 1972. Genotype stability analysis and its application to potato regional trials. Crop Sci. 11:184–190. doi:10.2135/cropsci1971.0011183X001100020006x

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143:449–462. doi:10.1017/S0021859605005587

Wricke, G. 1962. Evaluation method for recording ecological differences in field trials. Z. Pflanzenzuecht. 47:92–96.

Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar evaluation and mega-environment investigation based on GGE biplot. Crop Sci. 40:597–605. doi:10.2135/cropsci2000.403597x

Yan, W., and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

Yan, W., M.S. Kang, B.-L. Ma, S. Woods, and P.L. Cornelius. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Sci. 47:643–653. doi:10.2135/cropsci2006.06.0374

Yan, W., D. Pageau, J. Frégeau-Reid, and J. Durand. 2011. Assessing the representativeness and repeatability of test locations for genotype evaluation. Crop Sci. 51:1603–1610. doi:10.2135/cropsci2011.01.0016

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci. 28:556–580. doi:10.1017/S0021859600050978

SUPPLEMENT
(For Publication)

Analysis of Genotype x Environment Interaction (GxE) Using SAS Programming

Description

The supplemental material provides SASGxE program, example input data, graphical illustrations for uploading input data file in SASGxE and installing R packages in 'R studio', outputs from example input data, and interpretation of univariate and multivariate statistical analysis.

Output and Interpretation of univariate stability statistics

Linear regression coefficient (b_i)

Regression coefficient (b_i) of genotypes approximating unity (P<0.01) along with high trait mean is considered to stable across wide range of environment. When this is associated with low trait mean performance, genotypes are poorly adapted to all environments. A b_i greater than unity describes genotypes with higher sensitivity to environmental change (below average stability), and greater specificity of adaptability to high yielding environments. Output of b_i for trait MKMGHA from example input dataset is presented in Figure S2 Panel A.

Deviation from regression (S^2_d)

Genotype is considered to be stable when deviation from regression (S^2_d) is not significantly different from zero. Output of S^2_d for trait MKMGHA from example input dataset is presented in Figure S2 Panel A.

Perkins and Jinks (β_i)

Genotypes with slope β_i values not significantly different from 0.0 were judged to be stable, whereas those with significant β_i values were unstable. Output of β_i for trait MKMGHA from example input dataset is presented in Figure S2 Panel B.

Shukla stability variance (σ_i^2), Wricke's ecovalence (Wi), Lin and Binns Pi (P_i), Francis and Kannenberg coefficient of variation (CV_i)

Genotype with low σ_i^2, *Wi*, P_i and CV_i is regarded as stable. Output of σ_i^2, *Wi*, P_i, and CV_i for trait MKMGHA from example input dataset is presented in Figure S2 Panel B.

Kang stability statistic (YS_i)

According to this method, genotypes with YS_i greater than the mean YS_i are considered stable. Output of YS_i for trait MKMGHA from example input dataset is presented in Figure S2 Panel C and stable genotypes are marked with symbol '@'.

Output and interpretation of multivariate stability statistics

AMMI

Biplot (Trait vs. PC1)

The biplot abscissa and ordinate shows the trait main effect and first principal component (PC1) term, respectively. The horizontal and vertical lines that divide the graph into four quadrants indicate the interaction score of zero and trait gran mean, respectively. Displacement along the vertical axis indicates the interaction difference between genotypes and environments. Similarly, displacement along horizontal axis indicates the difference in genotype and environment main effect. The genotypes with PC1 scores close to zero indicate general adaptation across the environments, whereas with larger PC1 scores indicate specific adaptation of genotypes to environment which has same PC1 scores and sign. The relative magnitude and direction of genotypes along the abscissa and ordinate axis in biplot explain the response pattern of genotypes across the environments. The genotype with PC1 score close to zero and high trait mean is considered to be stable. The environment with PC1 scores close to zero indicates smaller variation in interaction and relative ranking of genotypes are stable at these locations. This biplot is also known as AMMI1 as it considers one PC into account. It is commonly used for genotypic evaluation. Output of trait (MKMGHA) vs PC1 biplot of example input dataset is presented in Figure S4 Panel A.

Biplot (PC1 vs. PC2)

The biplot abscissa and ordinate showed the first and second multiplicative axis term (PC1 and PC2), respectively. Horizontal and vertical lines passing from the origin (0, 0) of biplot divide the biplot into four quadrants. The distances from the origin of biplot indicate the amount of interaction exhibited by genotypes either over environment or environment over genotypes. The angle between genotype and environment vectors determined the nature of interaction; that is, it is positive for acute angles, negligible for right angles, and negative for obtuse angle. The angle formed by the vectors two environments provided an estimate of their correlation. Similarly, the small angles between genotype vectors inside same quadrant are similar in genetic performance. Connecting extreme genotype on biplot forms a polygon and the perpendicular to the side of polygon form sectors of genotype and environment. The genotypes at vertex of polygon are the winners in the environments included in that sector. When location falls in same sector across years then location is considered as separate mega-environment for genotype evaluation and recommendation. Location close to biplot origin is less interactive location and considered to be location good for selection of genotypes with average adaptation. This biplot is also known as AMMI2 as it considers two PC into account and it is used for mega-environment evaluation. Output of PC1 vs PC2 biplot for trait MKMGHA of example input dataset is presented in Figure S4 Panel B.

GGE Biplot

'Which-won-where' or polygon view

The 'which-won-where' or polygon view of the GGE biplot is an effective visual tool in mega-environment analysis. The perpendicular lines to the polygon sides divide the biplot into sector. If environments fall into different sectors, this suggests that different genotype won in different sector and thus genotype x environment interaction or crossover pattern exist. The winning genotype for a sector is the vertex genotype. Conversely, if all environments fall into a single sector, this indicates that a single genotype had the highest yield in all environments. Dividing target environment or

location into mega-environment is recommended if crossover patterns are repeatable across year. Locations fall within one sector is considered as one mega-environment. Mega-environment is set of location that consistently share the best set of genotypes across years. Output of 'which-won-where' view of the biplot for trait MKMGHA of example input dataset is presented in Figure S5 Panel A.

Mean vs. stability view

The 'mean vs. stability' view of GGE biplot facilitates genotype comparisons based on mean performance and stability across environments within a mega-environment. An 'ideal' genotype should have both high mean and high stability performance within a mega-environment. An ideal genotype is represented by the head of arrow on AEC abscissa (horizontal axis). The arrow shown on the AEC abscissa points in the direction of higher mean performance of genotypes and, consequently ranks the genotypes with respect to mean performance. The most stable genotype is located almost on the AEC abscissa (horizontal axis) and had a near-zero projections onto the AEC ordinate (vertical axis). It means, the rank of this genotype was highly consistent across the environment. Output of 'mean vs. stability' view of the biplot for trait MKMGHA of example input dataset is presented in Figure S5 Panel B.

Discriminative vs. representativeness view

The 'discriminating power vs. representative' view of GGE biplot identifies test environments that effectively identify superior genotypes for mega-environment. An 'ideal' test environment is a virtual environment that has the most ability to discriminate the genotype and represent the mega-environment. An ideal test environment is represented by the head of arrow on AEC abscissa (horizontal axis). Test environment with longest vectors from biplot origin are more discriminating of the genotypes. If the test environment has a very short vector or is close to biplot origin, it means genotypes performed similar in it. A short vector could also mean that environment is not represented well by PC1 and PC2 if the biplot does not explain most of the data (Yan et al., 2007). Similarly, the test environment that has small angle on AEC abscissa is more representative of the mega-environment than those that have larger angles with it. Output of 'discriminative vs. representative' view of the biplot for trait MKMGHA of example input dataset is presented in Figure S5 Panel C.

SASGxE Program

```
/*****************************************************************/
/*************          USER INPUT FIELD START           **********/
/*****************************************************************/;

%LET IPATH = E:\PhD Research Work\PhD Articles\Articles for Publication\GxE SAS
Prog\Sample Data1; /*INPUT FILE PATH*/
%LET INAME = SASGxE_PROG_INPUT_DATA ; /*INPUT FILE NAME*/
%LET ISHEETNAME1 = SHEET2; /*INPUT FILE SHEET NAME*/

/*****************************************************************/
/*************           USER INPUT FIELD END            **********/
/*****************************************************************/;

OPTIONS NODATE NONUMBER NOLABEL NOMLOGIC MPRINT NOSYMBOLGEN;
TITLE;
```

```
FOOTNOTE;
DM LOG "CLEAR";
*ODS HTML CLOSE; /*OPTIONAL: USER CAN TURN IT ON*/
RUN;

*COMMENT: IMPORT INPUT DATA FROM EXCEL FILE(.XLSX ONLY);
PROC IMPORT
       OUT= WORK.TEMPA1
       DATAFILE= "&IPATH\&INAME..XLSX"
       DBMS=XLSX REPLACE;
       SHEET="&ISHEETNAME1";
RUN;

DATA TEMPA1 (RENAME=(CLT=CL));
       SET TEMPA1;

       MKWT=SUM(MKWT1,MKWT2,MKWT3,MKWT4); /*SUM ACROSS THE DEPENDENT VARIABLES*/
       CLWT=SUM(CLWT1,CLWT2,CLWT3,CLWT4); /*SUM ACROSS THE DEPENDENT VARIABLES*/

       MKMGHA=MKWT*0.40751; /*CALCULATE YIELD MG/HA FOR 12 FT PLOT SIZE*/
       CLMGHA=CLWT*0.40751; /*FACTOR 0.40751 CONVERTS LBS/PLOT TO MG/HA*/

        IF CL=01 THEN CLT='Mountain Hoosier  ';
       ELSE IF CL=02 THEN CLT='Hopi Red Flesh    ';
       ELSE IF CL=03 THEN CLT='Early Arizona     ';
       ELSE IF CL=04 THEN CLT='Starbrite F1      ';
       ELSE IF CL=05 THEN CLT='Stone Mountain    ';
       ELSE IF CL=06 THEN CLT='Stars-N-Stripes F1';
       ELSE IF CL=07 THEN CLT='AU-Jubilant       ';
       ELSE IF CL=08 THEN CLT='Calhoun Gray      ';
       ELSE IF CL=09 THEN CLT='Big Crimson       ';
       ELSE IF CL=10 THEN CLT='Legacy F1         ';

       DROP MKWT1 MKWT2 MKWT3 MKWT4 CLWT1 CLWT2 CLWT3 CLWT4 MKWT CLWT CL;
RUN;


*COMMENT: DEFINE MACRO FOR SLOPE AND DEVIATION FROM REGRESSION;
*COMMENT: DEVIATION FROM REG. = PREDICTED - ACTUAL;
*COMMENT: SLOPE IS TESTED FOR SIG. DIFFERENCE W/ VALUE ONE ;
*COMMENT: DEVIATION FROM REG. TESTED FOR SIG. DIFFERENCE FROM ZERO;

%MACRO UNIVARIATE1 ;
*COMMENT: COMPUTE ENVIRONMENTAL INDEX;
PROC SQL NOPRINT;
       CREATE TABLE DSTERM AS SELECT EN, RP, YR, LC, MEAN(&DEPVAR) AS ENV&DEPVAR
/*ENVIRNMENTAL INDEX*/
       FROM TEMPA2 GROUP BY EN, RP, YR, LC ORDER BY EN, RP, YR, LC;

       CREATE TABLE DST02 AS SELECT A.*, B.ENV&DEPVAR
       FROM TEMPA2 AS A
       LEFT JOIN DSTERM AS B ON (A.EN=B.EN AND A.RP=B.RP)
       ORDER BY CL;
QUIT;

%LET INDPVAR = ENV&DEPVAR;
*ODS TRACE ON/LISTING;
PROC GLM DATA= DST02 OUTSTAT=OUTMSEDS2 PLOTS=NONE;
       BY CL;
       CLASS CL LC RP EN;
       MODEL &DEPVAR =&INDPVAR   EN RP  /SOLUTION SS1;
       ODS OUTPUT OVERALLANOVA=ANOVADS2 PARAMETERESTIMATES=PARMGLMDS2;
RUN;
```

```
*ODS TRACE OFF;

DATA OUTMSEDS3(RENAME=(_SOURCE_=SOURCE));
       SET OUTMSEDS2(WHERE=(_SOURCE_ NE "RP") KEEP=CL _NAME_ _SOURCE_  DF SS);
       MS=SS/DF;
RUN;

PROC TRANSPOSE DATA=OUTMSEDS3 (RENAME=(_NAME_=DEPENDENT)) OUT=MSDS ;
       BY CL DEPENDENT;
       ID SOURCE ;
       VAR MS;
RUN;

PROC TRANSPOSE DATA=OUTMSEDS3 (RENAME=(_NAME_=DEPENDENT)) PREFIX=DF_
OUT=FDS3(DROP=_NAME_) ;
       BY CL DEPENDENT;
       ID SOURCE ;
       VAR DF;
RUN;

DATA REGCOEFDS;
       SET PARMGLMDS2(WHERE = (PARAMETER="&INDPVAR") KEEP=CL PARAMETER
DEPENDENT ESTIMATE STDERR);
RUN;

PROC SORT DATA= MSDS; BY CL DEPENDENT; RUN;
PROC SORT DATA= REGCOEFDS; BY CL DEPENDENT;RUN;

DATA SLOPE (RENAME=(BI=SLOPE DEVLMS=DEVREG));
       MERGE MSDS(IN=A DROP=_NAME_ RENAME=( ERROR=MSE  &INDPVAR=LREGMS
EN=DEVLMS))
       REGCOEFDS (RENAME=(ESTIMATE=BI))
       FDS3;
       BY CL DEPENDENT;
       T_HO1=(BI-1)/STDERR; /*NULL HYPOTHESIS: SLOPE=1 */
       PT_HO1=2*(1-PROBT(ABS(T_HO1), DF_ERROR));
       IF  PT_HO1 LE 0.001 THEN SIG_SLOPE="***";
       ELSE IF  PT_HO1 LE 0.01 THEN SIG_SLOPE="**";
       ELSE IF  PT_HO1 LE 0.05 THEN SIG_SLOPE="*";

       F_DEVREG=DEVLMS/MSE; /*NULL HYPOTHESIS: PREDICTED-ACTUAL = 0*/
       PF_HO0= 1-PROBF(F_DEVREG, DF_EN, DF_ERROR);
       IF  PF_HO0 LE 0.001 THEN SIG_DEVREG="***";
       ELSE IF  PF_HO0 LE 0.01 THEN SIG_DEVREG="**";
       ELSE IF  PF_HO0 LE 0.05 THEN SIG_DEVREG="*";

       /*CONCATENATE LEVEL OF SIGNIFICANCE WITH SLOPE AND DEVIATION*/
       SLOPE1 = PUT(BI, z5.3);
       SLOPE2 = SLOPE1||LEFT(TRIM(SIG_SLOPE));
       DEVREG1 = PUT(DEVLMS, z12.3);
       DEVREG2=DEVREG1||LEFT(TRIM(SIG_DEVREG));
       DROP SLOPE1 DEVREG1;
RUN;

*POST PROCESS THE DATA;
*OUTPUT FOR SLOPE AND DEV FROM REG;
PROC SQL NOPRINT;
       /*PREPARE DATASET FOR CORRELATION*/
       CREATE TABLE STABLE1&DEPVAR
       AS SELECT CL, DEPENDENT AS TRAIT, SLOPE, DEVREG
       FROM SLOPE;
       /*PREPARE DATASET FOR STABILITY OUTPUT*/
       CREATE TABLE UNIVARIATE1&DEPVAR
```

```
       AS SELECT CL, DEPENDENT AS TRAIT, SLOPE2 AS SLOPE,
                      STDERR AS STDERR_SLOPE, T_HO1 AS TTEST_SLOPE,
                      PT_HO1 AS PROB_SLOPE, DEVREG2 AS DEVREG,
                      F_DEVREG AS FTEST_DEVREG, PF_HO0 AS PROB_DEVREG
       FROM SLOPE;
QUIT;

%MEND UNIVARIATE1;


*COMMENT: DEFINING MACRO FOR ECOVALENCE, SHUKLAS SIGMA,
  PERKINS AND JINKS BETA, LIN AND BINNS PI, FRANCIS & KENNENBERG CV;
*COMMENT: COMPUTE WRICKES (1962) ECOVALENCE;
*COMMENT: COMPUTE PERKINS AND JINKS (1968) SLOPE ;
*COMMENT: COMPUTE SHUKLAS (1972) SIGMA;
*COMMENT: COMPUTE LOCATION*YEAR VARIANCE (Pi)-LIN AND BINNS (1988);
*COMMENT: COMPUTE GENOTYPES COEFFICIENT OF VARIATION- FRANCIS & KENNENBERG
(1968);
*COMMENT: COMPUTE REGRESSION OF GEN ON ENV MEANS-USING METHOD OF PERKINS AND
JINKS;

%MACRO UNIVARIATE2 ;
PROC SQL NOPRINT;
       ***SUM ACROSS ENV, CULTIGEN, YEAR, LOCATION***;
       CREATE TABLE DSTECS AS SELECT EN, CL, YR, LC, SUM(&DEPVAR) AS SUM&DEPVAR
       FROM TEMPA2 GROUP BY EN, CL, YR, LC ORDER BY EN;

       ***LIST OF DISTINCT GENOTYPES***;
       CREATE TABLE TEMP_CL
       AS SELECT DISTINCT (CL) AS DISTINCT_CL FROM TEMPA2 ORDER BY CL;

       ***MACRO FOR TOTAL NUMBER OF REPLICATION***;
       SELECT COUNT (DISTINCT(RP))INTO: TOTAL_RP TRIMMED FROM TEMPA2;

       ***MACRO FOR TOTAL NUMBER OF ENVIRONMENT***;
       SELECT COUNT (DISTINCT(EN))INTO: TOTAL_EN TRIMMED FROM TEMPA2;
QUIT;
%PUT &TOTAL_RP = &TOTAL_EN =;

DATA DST01 ; SET DSTECS;
       BY EN;
       IF FIRST.EN THEN ET+1;
RUN;
PROC SORT DATA=DST01; BY CL; RUN;

%LET DEPVAR2=SUM&DEPVAR;
DATA DST01B;
       SET DST01;
       BY CL;
       ARRAY E(ET) E1-E&TOTAL_EN;
       RETAIN E1-E&TOTAL_EN;
       E=&DEPVAR2;
       IF LAST.CL THEN DO;
       OUTPUT;
              DO OVER E;
              E=.;
              END;
       END;
       KEEP E1-E&TOTAL_EN CL;
RUN;

PROC IML;
*RESET AUTONAME;
```

```
*START MAIN;
USE DST01B;
       READ ALL  INTO X;
       P= NROW(X);  /*NO OF CULTIVAR*/
       Q= NCOL(X);  /*NO OF ENVIRONMENT*/
       CMEAN= X[+,]/P; ** COLUMN GRAND MEAN;
       CULT= J(P,Q);
       DO I={1} TO P;
        CULT[I,]= CMEAN[{1},{1}:Q];  ***GENEARTE MATRIX OF COLUMN MEANS (P,Q);
        END;
       U=X- CULT;  **RESIDUALS FROM OVERALL MEAN;
       UM=U/Q;      *** GET RESIDUAL OVER NUMBER OF COL (RESPONSES);
       ENV= J(P,Q);
       DO K={1} TO Q;
        ENV[,K]= UM[,+];
        END;
       DIFF=U-ENV; /*MATRIX OF GXE RESIDUALS*/
       SSDIFF=(DIFF#DIFF)[,+];
       SUMSS= SUM(SSDIFF); /*TOTAL SS RESID*/

       N={&TOTAL_RP}; /*NO OF REP*/
       ECOV=SSDIFF/N; /*WRICKES ECOVALENCE */
       L=P*(P-{1});
       E=(Q-{1})*(P-{1})*(P-{2});
       LSSDIFF=(SSDIFF*L)/N;
       F= J(P,{1},(SUMSS/N));
       SIG=LSSDIFF-F;

       SIGMA=SIG/E; /*SHUKLAS SIGMA*/
       TOT= SUM(X);
       GM=TOT/(P*Q);
       Z= J({1},Q,GM);
       ZJ=CMEAN-Z;
       SUMSQZJ= SUM(ZJ#ZJ);
       RAT= J(P,Q);
       DO R={1} TO P;
        RAT[R,]= ZJ[{1},{1}:Q];
        END;
       NEW=DIFF#RAT;

       BETA=(NEW/SUMSQZJ)[,+]; /*REGRESSION OF GEN ON ENV MEANS-USING METHOD OF
PERKINS AND JINKS*/
       GP= J(P,Q);
       DO C={1} TO Q;
        GP[,C]= BETA[{1}:P,{1}];
        END;
       BIZJ=RAT#GP;
       NEWDIFF=(DIFF-BIZJ);
       SI=(NEWDIFF#NEWDIFF)[,+];
       TS=P/((P-{2})*(Q-{2}));
       TOTSI= SUM(SI)/L;

       SP=((SI-TOTSI)*TS)/N; /*SHUKLA S SQUARED*/

              CREATE IML_OUT VAR {BETA SIGMA ECOV};
              APPEND ; CLOSE IML_OUT;
QUIT;

DATA TEMP_CL1 (RENAME=(DISTINCT_CL=CL));
       SET TEMP_CL;
       ID= _N_;
       TRAIT="&DEPVAR";
RUN;
```

```
DATA STAT2;
       SET IML_OUT;
       ID= _N_;
       RENAME BETA=BETA_PERKINS_AND_JINKS /*OUTPUT BETA_PERKINS AND JINKS*/
                 SIGMA=SHUKLA /*OUTPUT SIGMA_SHUKLA*/
                 ECOV=ECOVALENCE;/*OUTPUT WRICKE'S ECOVALENCE*/
RUN;

*MEREGE GENOTYPE NAME WITH STABILITY PARAMETERS;
PROC SQL NOPRINT;
       CREATE TABLE TEMP_STABLE2
       AS SELECT A.CL, A.TRAIT, BETA_PERKINS_AND_JINKS, SHUKLA,ECOVALENCE
       FROM TEMP_CL1 AS A
       INNER JOIN STAT2 AS B ON A.ID=B.ID;
QUIT;

*COMMENT: COMPUTE LOCATION*YEAR VARIANCE (Pi)-LIN AND BINNS (1988);

PROC SORT DATA = TEMPA2; BY CL; RUN;
*ODS TRACE ON/LISTING;
PROC GLM DATA= TEMPA2 PLOTS=NONE;
       BY CL;
       CLASS CL LC YR;
       MODEL &DEPVAR =LC YR LC*YR;
       RANDOM LC YR LC*YR/TEST;
              ODS OUTPUT
                     OVERALLANOVA=OUTLANDB1
                     RANDOMMODELANOVA = OUTLANDB2;
RUN;
*ODS TRACE OFF;

DATA OUTLANDB3_&DEPVAR;
       SET OUTLANDB2 (WHERE=(SOURCE="LC*YR"));
       MS1= PUT(MS, 10.5);
              IF  PROBF LE 0.001 THEN PI=TRIM(MS1)||TRIM(LEFT("***"));
       ELSE IF  PROBF LE 0.01 THEN PI=TRIM(MS1)||TRIM(LEFT("**"));
       ELSE IF  PROBF LE 0.05 THEN PI=TRIM(MS1)||TRIM(LEFT("*"));
       ELSE PI=MS1;
       KEEP CL DEPENDENT MS PROBF PI;
RUN;

*COMMENT: COMPUTE GENOTYPES COEFFICIENT OF VARIATION-FRANCIS & KANNENBERG
(1968);
PROC SQL NOPRINT;
       CREATE TABLE PLAISTEDCV_&DEPVAR AS
       SELECT  CL, CV(&DEPVAR) AS CVI
       FROM TEMPA2
       GROUP BY CL;
QUIT;

*POST PROCESS THE DATA;
*OUTPUT FOR SHUKLA, ECOVALENCE, BETA, PI AND CV;
PROC SQL NOPRINT;
       /*PREPARE DATASET FOR CORRELATION*/
       CREATE TABLE STABLE2&DEPVAR
       AS SELECT A.CL, A.TRAIT, A.BETA_PERKINS_AND_JINKS, A.SHUKLA,
                      A.ECOVALENCE, B.MS AS PI, C.CVI
       FROM TEMP_STABLE2 AS A
       INNER JOIN OUTLANDB3_&DEPVAR AS B ON A.CL=B.CL
       INNER JOIN PLAISTEDCV_&DEPVAR AS C ON B.CL=C.CL;
       /*PREPARE DATASET FOR STABILITY OUTPUT*/
       CREATE TABLE UNIVARIATE2&DEPVAR
```

```
       AS SELECT A.CL, A.TRAIT, A.BETA_PERKINS_AND_JINKS, A.SHUKLA,
                      A.ECOVALENCE, B.PI, C.CVI
       FROM TEMP_STABLE2 AS A
       INNER JOIN OUTLANDB3_&DEPVAR AS B ON A.CL=B.CL
       INNER JOIN PLAISTEDCV_&DEPVAR AS C ON B.CL=C.CL;
QUIT;

%MEND UNIVARIATE2;

*COMMENT: DEFINE MACRO FOR TRAIT LEAST SQUARE (LS) MEANS, LSD, KANGS (1995)
STABILITY PARAMETER-YS (MEKIB, 2003) ;
*COMMENT: STABILITY PARAMETER 'YS' IS CALCULATED BASED ON SHUKLA AND TRAIT
MEAN;
*COMMENT: STABILITY PARAMTER 'YS' IS CALCULATED AS PROCEDURE LISTED BY MEKIB,
F. EUPHYTICA, 2003;

%MACRO UNIVARIATE3;
*COMMENT: MACRO FOR TOTAL NO OF REP AND ENV;
*COMMENT: COMPUTE ENVIRONMENTAL INDEX ;
PROC SQL NOPRINT;
       SELECT COUNT (DISTINCT(RP))INTO: TOTAL_RP TRIMMED FROM TEMPA2;
       SELECT COUNT (DISTINCT(EN))INTO: TOTAL_EN TRIMMED FROM TEMPA2;

       CREATE TABLE DSTERM AS SELECT EN, RP, YR, LC, MEAN(&DEPVAR) AS ENV&DEPVAR
/*ENVIRNMENTAL INDEX*/
       FROM TEMPA2 GROUP BY EN, RP, YR, LC ORDER BY EN, RP, YR, LC;

       CREATE TABLE DST02 AS SELECT A.*, B.ENV&DEPVAR
       FROM TEMPA2 AS A
       LEFT JOIN DSTERM AS B ON (A.EN=B.EN AND A.RP=B.RP)
       ORDER BY CL;
QUIT;

PROC GLM
       DATA= DST02 OUTSTAT=OUTMSDS PLOTS=NONE;
       CLASS CL LC RP EN;
       MODEL &DEPVAR =  EN RP(EN) CL (EN);
       LSMEANS CL(EN)/STDERR  OUT=CLTLSMNDS1 SLICE=(EN CL);
       ODS OUTPUT OVERALLANOVA=ANOVADS FITSTATISTICS=DEPMEANDS;
RUN;

PROC SQL NOPRINT;
       CREATE TABLE CLTLSMNDS2
       AS SELECT CL, MEAN(LSMEAN) AS LSMEAN , MEAN(STDERR) AS STDERR
       FROM CLTLSMNDS1 GROUP BY CL ORDER BY CL;
QUIT;

DATA CLTLSMNDS;
       SET CLTLSMNDS2;
       _NAME_ = "&DEPVAR";
RUN;
PROC SORT DATA=CLTLSMNDS;BY CL;RUN;

DATA SEE1;
       IF _N_=1 THEN MERGE ANOVADS(IN=A  WHERE=(SOURCE = 'Error') KEEP= SOURCE
MS DF) DEPMEANDS(IN=B KEEP= DEPMEAN) ;
       ELSE SET CLTLSMNDS ;
              SE_DIFF=SQRT( MS*(2*(1/(&TOTAL_EN*&TOTAL_RP)))) ;
              T_DFE= TINV(0.975, DF); /*PROBABILITY=0.975 OR ALPHA=0.025*/
              LSD= T_DFE*SE_DIFF;
       IF LSMEAN LE (DEPMEAN-2*LSD)THEN SCORE_LSD=-3;
       ELSE IF LSMEAN LE (DEPMEAN-LSD) THEN SCORE_LSD=-2;
       ELSE  IF  LSMEAN LE DEPMEAN THEN SCORE_LSD=-1;
```

```
       IF LSMEAN GE (DEPMEAN+2*LSD)THEN SCORE_LSD= 3;
       ELSE IF LSMEAN GE (DEPMEAN+LSD) THEN SCORE_LSD= 2;
       ELSE  IF  LSMEAN GE DEPMEAN THEN SCORE_LSD= 1;
RUN;

DATA SEE1;
       SET SEE1;
       IF _N_ GT 1;
RUN;
PROC SORT DATA=SEE1;BY CL;RUN;
PROC SORT DATA=TEMP_STABLE2;BY CL;RUN;

DATA SEE2 (DROP = SHUKLA_);
       MERGE SEE1 TEMP_STABLE2 (KEEP = CL SHUKLA) ;
       BY CL;
       F_CALC=SHUKLA/MS;
       PF_SHUKLA=1-PROBF(F_CALC,(&TOTAL_EN-1),DF);

       IF PF_SHUKLA  LE 0.01 THEN SIG_SHUKLA=-8;
       ELSE IF  PF_SHUKLA LE 0.05 THEN SIG_SHUKLA=-4;
       ELSE IF  PF_SHUKLA LE 0.10 THEN SIG_SHUKLA=-2;
       ELSE   SIG_SHUKLA= 0;

       SHUKLA_= PUT(SHUKLA, 10.5);
        IF PF_SHUKLA LE 0.001 THEN SHUKLA_TEST=TRIM(SHUKLA_)||TRIM(LEFT("***"));
       ELSE IF PF_SHUKLA LE 0.01 THEN
SHUKLA_TEST=TRIM(SHUKLA_)||TRIM(LEFT("**"));
       ELSE IF PF_SHUKLA LE 0.05 THEN
SHUKLA_TEST=TRIM(SHUKLA_)||TRIM(LEFT("*"));
       ELSE SHUKLA_TEST=SHUKLA_;
RUN;

PROC RANK  DATA=SEE2  OUT=RNK&DEPVAR;
       VAR LSMEAN;
       RANKS YRANK;
RUN;

PROC SORT DATA=  RNK&DEPVAR;BY  DESCENDING YRANK;RUN;

DATA  RNK&DEPVAR;
       SET RNK&DEPVAR;
       SUMMED= YRANK +SCORE_LSD;
       YS= SUMMED +SIG_SHUKLA;
RUN;

*COMMENT MARK (@) STABLE GENOTYPE BASED ON YS > MEAN(YS);
PROC SQL NOPRINT;
       SELECT MEAN(YS)INTO: MEAN TRIMMED FROM RNK&DEPVAR;
QUIT;
%PUT MEAN =&MEAN;

DATA RNK&DEPVAR;
       SET RNK&DEPVAR;
       YS_ = PUT(YS,8.0);
       IF YS GT &MEAN THEN YS_TEST=TRIM(YS_)||TRIM(LEFT("@"));
       ELSE YS_TEST=YS_;
       DROP YS_;
RUN;

*POST PROCESS THE DATA;
*OUTPUT FOR LSMEAN, LSD, YS, TEST OF SIG. FOR SHUKLA;
PROC SQL NOPRINT;
```

```
        /*PREPARE DATASET FOR CORRELATION*/
        CREATE TABLE STABLE3&DEPVAR
        AS SELECT CL, _NAME_ AS TRAIT, LSMEAN, LSD, YS
        FROM RNK&DEPVAR;

        /*PREPARE DATASET FOR STABILITY OUTPUT*/
        CREATE TABLE UNIVARIATE3&DEPVAR
        AS SELECT CL, _NAME_ AS TRAIT, LSMEAN, LSD,
                        SHUKLA_TEST AS SHUKLA, YS_TEST AS YS
        FROM RNK&DEPVAR;
QUIT;

%MEND UNIVARIATE3;

*COMMENT: DEFINE MACRO FOR LEVEL OF SIGNIFICANCE;
*COMMENT: CONCETANATE CORRELATION VALUE W/ LEVEL OF SIGNIFICANCE;
%MACRO LEVELOFSIG (TEST=);
        &TEST.1= PUT(&TEST, 8.5);
        IF  P&TEST LE 0.001 THEN &TEST.2=&TEST.1||LEFT("***");
        ELSE IF  P&TEST LE 0.01 THEN &TEST.2=&TEST.1||LEFT("**");
        ELSE IF  P&TEST LE 0.05 THEN &TEST.2=&TEST.1||LEFT("*");
        ELSE &TEST.2=&TEST.1;
        DROP &TEST &TEST.1;
        RENAME &TEST.2=&TEST;
%MEND LEVELOFSIG;

*COMMENT: DEFINE MACRO FOR EXPORTING OUTPUTS OR RESULTS (.XLSX);
%MACRO OUPUTEXCEL (DATA=, NAME=);
        PROC EXPORT DATA= &DATA
        OUTFILE= "&IPATH\&NAME..xlsx"
        DBMS=xlsx REPLACE;
        SHEET="Sheet1";
        RUN;
%MEND OUPUTEXCEL;

*COMMENT: DEFINE MACRO FOR RENAMING GENOTYPE, ENVIRONMENT & LOCATION FOR
GGEBIPLOT & AMMI ANALYSIS;
%MACRO GENOTYPE;
        %DO j=1 %TO &TOTAL_CL;
        IF CUL = "&&CL&j" THEN GEN = "G&j";
        %END;
%MEND GENOTYPE;

%MACRO ENVIRONMENT;
        %DO k=1 %TO &TOTAL_EN;
        IF EN = "&&EN&k" THEN ENV = "ENV&k";
        %END;
%MEND ENVIRONMENT;

%MACRO LOCATION;
        %DO l=1 %TO &TOTAL_LC;
        IF LC = "&&LC&l" THEN LOC = "LOC&l";
        %END;
%MEND LOCATION;

*COMMENT: DEFINE MACRO FOR EXPORTING OUTPUTS OR RESULTS (.CSV);
%MACRO OUPUTCSV (DATA=, NAME=);
        PROC EXPORT DATA = &DATA
        OUTFILE = "&IPATH\&NAME..CSV"
        DBMS = CSV
        REPLACE;
        RUN;
%MEND OUPUTCSV;
```

```
/*********************************/

*COMMENT: DEFINE MACRO TO IDENTIFY NUMBER AND DIFFERENT
  TYPE OF DEPENDENT VARIABLES SO THAT STABILITY
  ANALSYS CAN BE PERFROMED SIMULTANEOUSLY;
PROC CONTENTS DATA=TEMPA1 OUT=START ORDER=VARNUM NOPRINT; RUN;

PROC SQL NOPRINT;
       CREATE TABLE START1
       AS SELECT *
       FROM START
       WHERE NAME NOT IN ('YR', 'LC', 'RP', 'CL');
QUIT;

DATA _NULL_;
       SET START1 END=END_OF_DATASET;
       CALL SYMPUT ('DEPVARX'||TRIM(LEFT(_N_)), NAME); /*MACRO FOR DEPENDENT
VARIABLE*/
       IF END_OF_DATASET THEN CALL SYMPUT ('LAST_DEPVARIABLE', TRIM(LEFT(_N_)));
RUN;
%PUT &LAST_DEPVARIABLE = ;

*COMMENT: DEFINING MACRO FOR SIMULTANEOUSLY ON DIFFERENT DEPENDENT VARIABLES;
*COMMENT: MACRO STABILITY INVOKES ABOVE LISTED MACRO'S IN IT;
%MACRO STABILITY (FINAL=&LAST_DEPVARIABLE);
%DO i=1 %TO &FINAL;
%LET DEPVAR = &&DEPVARX&i;

*COMMENT: CL = GENOTYPE , LC = LOCATION, YR = YEAR, EN = ENVIRONMENT, RP =
REPLICATION;
*COMMENT: DEFINE ENVIRONMENT;
*COMMENT: DATA QUALITY CHECK - REMOVE MISSING RECORDS;
DATA TEMPA2;
       SET TEMPA1;
       EN=TRIM(LC)||'-'||TRIM(LEFT(YR)); /* ENV = LOC*YEAR */
       IF LC=' ' OR YR =. OR RP=. OR &DEPVAR = . THEN DELETE;
RUN;

*COMMENT: LIST OF UNIQUE CL EN LC RP;
PROC SQL;
       CREATE TABLE TEMP_CL
       AS SELECT DISTINCT (CL) AS DISTINCT_CL FROM TEMPA2 ORDER BY CL;

    CREATE TABLE TEMP_EN
       AS SELECT DISTINCT (EN) AS DISTINCT_EN FROM TEMPA2 ORDER BY EN;

    CREATE TABLE TEMP_LC
       AS SELECT DISTINCT (LC) AS DISTINCT_LC FROM TEMPA2 ORDER BY LC;

    CREATE TABLE TEMP_RP
       AS SELECT DISTINCT (RP) AS DISTINCT_RP FROM TEMPA2 ORDER BY RP;
QUIT;

*COMMENT: MACRO FOR TOTAL # OF CL;
DATA _NULL_;
       SET TEMP_CL END=COUNT_CL;
       IF COUNT_CL THEN CALL SYMPUT('TOTAL_CL', TRIM(LEFT(_N_)));
RUN;

*COMMENT: MACRO FOR TOTAL # OF EN;
DATA _NULL_;
       SET TEMP_EN END=COUNT_EN;
```

```
      IF COUNT_EN THEN CALL SYMPUT('TOTAL_EN', TRIM(LEFT(_N_)));
RUN;

*COMMENT: MACRO FOR TOTAL # OF LC;
DATA _NULL_;
      SET TEMP_LC END=COUNT_LC;
      IF COUNT_LC THEN CALL SYMPUT('TOTAL_LC', TRIM(LEFT(_N_)));
RUN;

*COMMENT: MACRO FOR TOTAL # OF RP;
DATA _NULL_;
      SET TEMP_RP END=COUNT_RP;
      IF COUNT_RP THEN CALL SYMPUT('TOTAL_RP', TRIM(LEFT(_N_)));
RUN;

*COMMENT: MEAN AND CV COMPUTATION;
PROC SQL NOPRINT;
      CREATE TABLE MEANCYLR AS /*OUTPUT USED FOR ANALYSIS*/
      SELECT YR, LC, RP, CL, EN, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY CL, YR, LC, RP
      ORDER BY CL, YR, LC, RP;

      CREATE TABLE MEANCYL AS
      SELECT YR, LC, CL, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY CL, YR, LC
      ORDER BY CL, YR, LC;

      CREATE TABLE MEANCY AS
      SELECT YR, CL, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY CL, YR
      ORDER BY CL, YR;

      CREATE TABLE MEANCL AS
      SELECT  LC, CL, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY CL, LC
      ORDER BY CL, LC;

      CREATE TABLE CVCL AS
      SELECT  LC, CL, CV(&DEPVAR) AS CV&DEPVAR
      FROM TEMPA2
      GROUP BY CL, LC
      ORDER BY CL, LC;

      CREATE TABLE MEANC AS /*OUTPUT -MEAN OF CUL*/
      SELECT CL, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY CL
      ORDER BY CL;

      CREATE TABLE MEANL AS /*OUTPUT -MEAN OF LOC*/
      SELECT LC, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY LC
      ORDER BY LC;

      CREATE TABLE MEANLY AS /*OUTPUT -MEAN OF LOC OVER YEAR*/
      SELECT YR, LC, MEAN(&DEPVAR) AS MEAN&DEPVAR
      FROM TEMPA2
      GROUP BY LC, YR
```

```
       ORDER BY LC, YR;

       CREATE TABLE MEANCE AS
       SELECT CL, EN, MEAN(&DEPVAR) AS MEAN&DEPVAR
       FROM TEMPA2
       GROUP BY CL, EN
       ORDER BY CL, EN;

       CREATE TABLE MEANCER AS /*OUTPUT USED FOR AMMI (GEN X ENV) ANALYSIS IN
R*/
       SELECT CL, EN, RP, MEAN(&DEPVAR) AS MEAN&DEPVAR
       FROM TEMPA2
       GROUP BY CL, EN, RP
       ORDER BY CL, EN, RP;

       CREATE TABLE MEANCLR AS /*OUTPUT USED FOR AMMI (GEN X LOC) ANALYSIS IN
R*/
       SELECT CL, LC, RP, MEAN(&DEPVAR) AS MEAN&DEPVAR
       FROM TEMPA2
       GROUP BY CL, LC, RP
       ORDER BY CL, LC, RP;

QUIT;

PROC TRANSPOSE DATA=MEANCE  OUT=DSTGGEBIPLOT (RENAME=(_NAME_ = TRAIT));
/*OUTPUT- MEAN CUL OVER ENV*/
       BY CL;
       ID EN;
       VAR MEAN&DEPVAR;
RUN;

PROC TRANSPOSE DATA=MEANLY  OUT=MEAN_LCYR (RENAME=(_NAME_ = TRAIT))PREFIX =
YR_; /*OUTPUT- MEAN LOC OVER YEAR*/
       BY LC;
       ID YR;
       VAR MEAN&DEPVAR;
RUN;

PROC TRANSPOSE DATA=MEANCY  OUT=MEAN_CLYR (RENAME=(_NAME_ = TRAIT))PREFIX =
YR_; /*OUTPUT- MEAN CUL OVER REP*/
       BY CL;
       ID YR;
       VAR MEAN&DEPVAR;
RUN;

PROC TRANSPOSE DATA=MEANCL  OUT=MEAN_CLLC (RENAME=(_NAME_ = TRAIT)); /*OUTPUT-
MEAN CUL OVER LOC*/
       BY CL;
       ID LC;
       VAR MEAN&DEPVAR;
RUN;

PROC TRANSPOSE DATA=CVCL  OUT=CV_CLLC (RENAME=(_NAME_ = TRAIT)); /*OUTPUT-
COEFF OF VAR (CV) CUL OVER LOC*/
       BY CL;
       ID LC;
       VAR CV&DEPVAR;
RUN;

*COMMENT: COMPUTE OVERALL ANOVA;
*COMMENT: ALL FACTORS (CL LC YR RP) ARE CONSIDERED AS RANDOM;
PROC GLM
       DATA= TEMPA2 OUTSTAT=TEMP_ANOVA_A1 PLOTS=NONE;
```

```
	CLASS CL LC RP EN  YR ;
	MODEL &DEPVAR = LC YR LC*YR  RP(LC*YR) CL CL*LC CL*YR CL*LC*YR;
	RANDOM   LC YR LC*YR  RP(LC*YR) CL*LC CL*YR CL*LC*YR/TEST;
	ODS OUTPUT
		OVERALLANOVA=TEMP_ANOVA_A2
		FITSTATISTICS=TEMP_ANOVA_A3
		RANDOMMODELANOVA = TEMP_ANOVA_A4
		EXPECTEDMEANSQUARES = TEMP_ANOVA_A5;
RUN;

PROC SQL NOPRINT;
	CREATE TABLE OVERALL_ANOVA1_&DEPVAR AS SELECT * FROM TEMP_ANOVA_A4
	WHERE SOURCE IN ("LC", "YR", "LC*YR", "RP(LC*YR)", "CL", "CL*LC",
"CL*YR", "CL*LC*YR", "Error: MS(Error)");
QUIT;

*COMMENT:ANOVA ACROSS YEARS;
PROC SORT DATA=TEMPA2; BY YR; RUN;

*ODS TRACE ON/LISTING;
PROC GLM DATA= TEMPA2 OUTSTAT=TEMP_ANOVA_Y1 PLOTS=NONE;
	BY YR;
	CLASS CL RP LC;
	MODEL &DEPVAR = CL LC  CL*LC RP(LC) ; /*MKMGHA MKTHHA ERPC PRCCW KGPF HH
BRX LYC*/
	RANDOM CL LC  CL*LC RP(LC)/TEST;
	ODS OUTPUT
			OVERALLANOVA = TEMP_ANOVA_Y2
			FITSTATISTICS=TEMP_ANOVA_Y3
			RANDOMMODELANOVA = TEMP_ANOVA_Y4
			EXPECTEDMEANSQUARES = TEMP_ANOVA_Y5;
RUN;
*ODS TRACE OFF;

*COMMENT: PREPARE OUTPUT FOR ANOVA ACROSS YEARS;
PROC SQL NOPRINT;
	CREATE TABLE YEAR_ANOVA1_&DEPVAR AS SELECT * FROM TEMP_ANOVA_Y4
	WHERE SOURCE IN ("CL", "LC",  "CL*LC", "RP(LC)", "Error: MS(Error)");
QUIT;

*COMMENT: F-TEST OF OVERALL GENOTYPE PERFORMANCE IN EACH LOCATION;
PROC SORT DATA=TEMPA2; BY LC; RUN;
*ODS TRACE ON/LISTING;
PROC GLM DATA= TEMPA2 OUTSTAT=TEMP_ANOVA_L1 PLOTS=NONE;
	BY LC;
	CLASS CL RP YR;
	MODEL &DEPVAR = CL YR CL*YR RP(YR) ;
	RANDOM CL YR CL*YR RP(YR)/TEST;
	ODS OUTPUT
			OVERALLANOVA = TEMP_ANOVA_L2
			FITSTATISTICS=TEMP_ANOVA_L3
			RANDOMMODELANOVA = TEMP_ANOVA_L4
			EXPECTEDMEANSQUARES = TEMP_ANOVA_L5;
RUN;
*ODS TRACE OFF;

PROC SQL NOPRINT;
	CREATE TABLE FTEST_&DEPVAR AS
	SELECT LC, DEPENDENT, FVALUE FROM TEMP_ANOVA_L4
	WHERE SOURCE = "CL";
QUIT;

*COMMENT: STANDARD DEVIATION (SD) OF LOCATION;
```

```
PROC SQL NOPRINT;
       CREATE TABLE SD_LOCATION_&DEPVAR
       AS SELECT LC, STD (&DEPVAR) AS SD&DEPVAR
       FROM TEMPA2
       GROUP BY LC;
QUIT;

*COMMENT: CORRELATION BETWEEN AVERAGE LOCATION VS. INDIVIDUAL LOCATION
PERFORMANCE ACROSS ALL GENOTYPES;
PROC SQL NOPRINT;
       CREATE TABLE LOC_MEAN
              AS SELECT LC,YR, RP, MEAN(&DEPVAR)AS &DEPVAR._LM
              FROM TEMPA2 GROUP BY LC, YR, RP;
       CREATE TABLE LOC_GRANDM
              AS SELECT YR, MEAN(&DEPVAR)AS &DEPVAR._GLM
           FROM TEMPA2 GROUP BY YR;
QUIT;

 DATA LOC_MEAN1;
       SET LOC_MEAN;
       TRAIT = "&DEPVAR";
       DROP RP;
 RUN;
 DATA LOC_GRANDM1;
       SET LOC_GRANDM;
       TRAIT = "&DEPVAR";
 RUN;

 PROC SQL NOPRINT;
       CREATE TABLE LOC_MEAN2
       AS SELECT A.*, B.&DEPVAR._GLM
       FROM LOC_MEAN1 AS A
       LEFT JOIN LOC_GRANDM1 AS B ON A.TRAIT=B.TRAIT AND A.YR = B.YR
       ORDER BY LC;
 QUIT;

*ODS TRACE ON/LISTING;
PROC CORR DATA = LOC_MEAN2 OUTP = CORRLC1;
       BY LC;
       VAR &DEPVAR._LM &DEPVAR._GLM;
       ODS OUTPUT PEARSONCORR = CORRLC2;
RUN;
*ODS TRACE OFF;

DATA CORRLC_&DEPVAR (RENAME=(TEST2 = CORRELATION));
       SET CORRLC2 (WHERE =(VARIABLE = "&DEPVAR._LM"));
       TRAIT = "&DEPVAR";
       TEST= PUT(&DEPVAR._GLM, 8.5);
                     IF  P&DEPVAR._GLM LE 0.001 THEN
TEST2=TEST||TRIM(LEFT("***"));
              ELSE IF  P&DEPVAR._GLM LE 0.01 THEN TEST2=TEST||TRIM(LEFT("**"));
              ELSE IF  P&DEPVAR._GLM LE 0.05 THEN TEST2=TEST||TRIM(LEFT("*"));
              ELSE TEST2=TEST;
       KEEP LC TRAIT TEST2;
RUN;

*MERGE LOCATION MEAN, F-RATIO, CORRELATION, STANDARD DEVIATION;
 PROC SQL NOPRINT;
       CREATE TABLE LOCATION_VALUE_&DEPVAR
       AS SELECT A.LC AS LOCATION,
                     A.TRAIT,
                     B.MEAN&DEPVAR AS MEAN,
                     C.FVALUE AS F_RATIO,
```

```
                    D.SD&DEPVAR AS STD_DEV,
                    A.CORRELATION
      FROM CORRLC_&DEPVAR AS A
             INNER JOIN MEANL AS B ON A.LC=B.LC
             INNER JOIN FTEST_&DEPVAR AS C ON B.LC=C.LC
             INNER JOIN SD_LOCATION_&DEPVAR AS D ON C.LC=D.LC;
 QUIT;

*CLUSTER ANALYSIS OF LOCATIONS;
PROC MEANS DATA=TEMPA2 NWAY NOPRINT;
      CLASS LC;
      VAR &DEPVAR;
      OUTPUT OUT=DSTCLUSTER MEAN =;
RUN;

ODS PDF FILE = "&IPATH\LOCTREE_&DEPVAR..PDF";
PROC CLUSTER DATA=DSTCLUSTER METHOD=WARD OUTTREE=LOCTREE
      PLOTS=(DENDOGRAM(VERTICAL HEIGHT=RSQ));
      VAR &DEPVAR;
      ID LC;
RUN;
ODS PDF CLOSE;


*COMMENT: INVOKE MACRO FOR SLOPE AND DEVIATION FROM REG;
*COMMENT: DEVIATION FROM REG = PREDICTED - ACTUAL;
*COMMENT: SLOPE IS TESTED FOR SIG DIFFERENCE W/ ONE ;
*COMMENT: DEV FROM REG TESTED FOR SIG DIFFERENCE FROM ZERO;
%UNIVARIATE1;

*COMMENT: INVOKE MACRO FOR WRICKES ECOVALENCE, SHUKLAS SIGMA,
               PERKINS AND JINKS BETA, LIN AND BINNS PI, FRANCIS & KENNENBERG
CV;
*COMMENT: COMPUTE WRICKES (1962) ECOVALENCE;
*COMMENT: COMPUTE PERKINS AND JINKS (1968) SLOPE ;
*COMMENT: COMPUTE SHUKLAS (1972) SIGMA;
*COMMENT: COMPUTE LOCATION*YEAR VARIANCE (Pi)-LIN AND BINNS (1988);
*COMMENT: COMPUTE GENOTYPES COEFFICIENT OF VARIATION-FRANCIS & KENNENBERG
(1968);
*COMMENT: COMPUTE REGRESSION OF GEN ON ENV MEANS-USING METHOD OF PERKINS AND
JINKS;
%UNIVARIATE2;

*COMMENT: INVOKE MACRO FOR TRAIT LS MEANS, LSD, KANGS (1995) STABILITY
PARAMETER-YS (MEKIB, 2003) ;
*COMMENT: STABILITY PARAMETER 'YS' IS CALCULATED BASED ON SHUKLA AND TRAIT
MEAN;
*COMMENT: STABILITY PARAMTER 'YS' IS CALCULATED AS PROCEDURE LISTED BY MEKIB,
F. EUPHYTICA, 2003;
%UNIVARIATE3;


*COMMENT: MERGE STABLITY RESULTS;
PROC SQL NOPRINT;
      CREATE TABLE STABLE&DEPVAR /*FINAL OUTPUT OF GENOTYPE STABILITY
PARAMETERS*/
      AS SELECT A.CL, A.TRAIT, C.LSMEAN, C.LSD, A.SLOPE, A.DEVREG,
                    B.BETA_PERKINS_AND_JINKS, C.SHUKLA, B.ECOVALENCE,
                    B.PI, B.CVI, C.YS
      FROM UNIVARIATE1&DEPVAR AS A
      INNER JOIN UNIVARIATE2&DEPVAR AS B ON A.CL=B.CL
      INNER JOIN UNIVARIATE3&DEPVAR AS C ON B.CL=C.CL;
QUIT;
```

```
*COMMENT: RANK GENOTYPES AND COMPUTE SPEARMAN RANK CORRELATION;
*COMMENT: GENOTYPE RANKING BASED ON MEAN YIELD, SLOPE, DEV FROM REG , SHUKLA,
AND YS AND SPEARMAN CORRELATION;
*COMMENT: SLOPE AND BETA ARE RANKED ASCENDING AND DESCENDING WHEN VALUE > 0 AND
< 0, RESPECTIVELY;
PROC SQL NOPRINT;
       CREATE TABLE TEMP_RANK1
       AS SELECT A.CL, C.LSMEAN, A.SLOPE, A.DEVREG,
                      B.BETA_PERKINS_AND_JINKS, B.SHUKLA, B.ECOVALENCE,
                      B.PI, B.CVI, C.YS
       FROM STABLE1&DEPVAR AS A
       INNER JOIN STABLE2&DEPVAR AS B ON A.CL=B.CL
       INNER JOIN STABLE3&DEPVAR AS C ON B.CL=C.CL
       ORDER BY CL;

       CREATE TABLE TEMP_RANK2
       AS SELECT CL, SLOPE
       FROM TEMP_RANK1 WHERE SLOPE GE 0 ORDER BY SLOPE;

       CREATE TABLE TEMP_RANK3
       AS SELECT CL, SLOPE
       FROM TEMP_RANK1 WHERE SLOPE LT 0 ORDER BY SLOPE DESC;

       CREATE TABLE TEMP_RANK4
       AS SELECT CL, BETA_PERKINS_AND_JINKS
       FROM TEMP_RANK1 WHERE BETA_PERKINS_AND_JINKS GE 0 ORDER BY
BETA_PERKINS_AND_JINKS;

       CREATE TABLE TEMP_RANK5
       AS SELECT CL, BETA_PERKINS_AND_JINKS
       FROM TEMP_RANK1 WHERE BETA_PERKINS_AND_JINKS LT 0 ORDER BY
BETA_PERKINS_AND_JINKS DESC;
QUIT;

DATA TEMP_RANK6;
       SET TEMP_RANK2 TEMP_RANK3;
       RNK_SLOPE = _N_;
RUN;

DATA TEMP_RANK7;
       SET TEMP_RANK4 TEMP_RANK5;
       RNK_BETA_PERKINS_AND_JINKS = _N_;
RUN;

PROC SQL NOPRINT;
       CREATE TABLE TEMP_RANK8
       AS SELECT A.CL, A.RNK_SLOPE, B.RNK_BETA_PERKINS_AND_JINKS
       FROM TEMP_RANK6 AS A
       INNER JOIN TEMP_RANK7 AS B ON A.CL=B.CL
       ORDER BY A.CL;
QUIT;

PROC RANK DATA = TEMP_RANK1 OUT=RANK1&DEPVAR;
       VAR LSMEAN SHUKLA ECOVALENCE YS DEVREG PI CVI;
       RANKS RNK_LSMEAN RNK_SHUKLA RNK_ECOVALENCE RNK_YS RNK_DEVREG RNK_PI
RNK_CVI;
RUN;

DATA RANK2&DEPVAR;
       SET RANK1&DEPVAR (KEEP = CL RNK_SHUKLA RNK_ECOVALENCE RNK_DEVREG RNK_PI
RNK_CVI);
```

```
      RNK1_SHUKLA = &TOTAL_CL - RNK_SHUKLA +1;
    RNK1_ECOVALENCE = &TOTAL_CL - RNK_ECOVALENCE+1;
    RNK1_DEVREG = &TOTAL_CL - RNK_DEVREG+1;
      RNK1_PI = &TOTAL_CL - RNK_PI+1;
      RNK1_CVI = &TOTAL_CL - RNK_CVI+1;
      DROP RNK_SHUKLA RNK_ECOVALENCE RNK_DEVREG RNK_PI RNK_CVI;
      RENAME
            RNK1_SHUKLA = RNK_SHUKLA
          RNK1_ECOVALENCE = RNK_ECOVALENCE
          RNK1_DEVREG = RNK_DEVREG
            RNK1_PI = RNK_PI
            RNK1_CVI = RNK_CVI;
RUN;

PROC SQL NOPRINT; /*OUTPUT FOR RANKS*/
      CREATE TABLE RANK3&DEPVAR
      AS SELECT A.CL,
                     B.RNK_LSMEAN AS MEAN,
                     A.RNK_SLOPE AS SLOPE_REG,
                     A.RNK_BETA_PERKINS_AND_JINKS AS BETA_PERKINS_JINKS,
                     C.RNK_SHUKLA AS SHUKLA,
                     C.RNK_ECOVALENCE AS ECOVALENCE_WRICKE,
                     C.RNK_DEVREG AS DEVIATION_REG,
                     C.RNK_PI AS LIN_BINNS_PI,
                     C.RNK_CVI AS CV,
                     B.RNK_YS AS KANG_YS
      FROM TEMP_RANK8 AS A
      INNER JOIN RANK1&DEPVAR AS B ON A.CL=B.CL
      INNER JOIN RANK2&DEPVAR AS C ON B.CL=C.CL
      ORDER BY CL;
QUIT;

PROC CORR DATA = RANK3&DEPVAR OUTS=CORRSPEAR1&DEPVAR;
      VAR MEAN SLOPE_REG BETA_PERKINS_JINKS SHUKLA
            ECOVALENCE_WRICKE DEVIATION_REG KANG_YS LIN_BINNS_PI CV;
      ODS OUTPUT SPEARMANCORR = CORRSPEAR2&DEPVAR;
RUN;

PROC SQL NOPRINT;
      CREATE TABLE CORRSPEAR&DEPVAR /*OUTPUT FOR RANK CORRELATION*/
      AS SELECT _NAME_ AS STABILITY_METHOD, MEAN, SLOPE_REG,
                                 BETA_PERKINS_JINKS, SHUKLA,
                                 ECOVALENCE_WRICKE, DEVIATION_REG, KANG_YS,
                                 LIN_BINNS_PI, CV
    FROM CORRSPEAR1&DEPVAR WHERE _NAME_ NE '';
QUIT;

DATA CORRSPEAR3&DEPVAR; /*OUTPUT FOR RANK CORRELATION W/ LEVEL OF SIG*/
      RETAIN VARIABLE MEAN SLOPE_REG DEVIATION_REG BETA_PERKINS_JINKS
               SHUKLA ECOVALENCE_WRICKE LIN_BINNS_PI CV KANG_YS
               PMEAN PSLOPE_REG PDEVIATION_REG PBETA_PERKINS_JINKS
               PSHUKLA PECOVALENCE_WRICKE PLIN_BINNS_PI PCV PKANG_YS;
      SET CORRSPEAR2&DEPVAR;
      /*COMMENT: INVOKE MACRO FOR LEVEL OF SIGNIFICANCE;*/
    /*COMMENT: USED FOR CONCETANATING CORR VALUE W/ LEVEL OF SIGNIFICANCE;*/
      %LEVELOFSIG (TEST=MEAN);
      %LEVELOFSIG (TEST=SLOPE_REG);
      %LEVELOFSIG (TEST=DEVIATION_REG);
      %LEVELOFSIG (TEST=BETA_PERKINS_JINKS);
      %LEVELOFSIG (TEST=SHUKLA);
      %LEVELOFSIG (TEST=ECOVALENCE_WRICKE);
      %LEVELOFSIG (TEST=LIN_BINNS_PI);
      %LEVELOFSIG (TEST=CV);
```

```
      %LEVELOFSIG (TEST=KANG_YS);
RUN;

*COMMENT: INVOKING MACRO FOR EXPORTING OUTPUT/RESULTS (.XLSX);
%OUPUTEXCEL (DATA= MEANCYLR, NAME=M_&DEPVAR._CYLR); /*OUTPUT - TRAIT MEAN OVER
CUL, YR, LC AND REP*/
%OUPUTEXCEL (DATA=MEANCYL, NAME=M_&DEPVAR._CYL); /*OUTPUT -MEAN OF CUL OVER
YEAR AND LOC*/
%OUPUTEXCEL (DATA=MEANC, NAME=M_&DEPVAR._C); /*OUTPUT -MEAN OF CUL*/
%OUPUTEXCEL (DATA=MEANL, NAME=M_&DEPVAR._L); /*OUTPUT -MEAN OF LOC*/
%OUPUTEXCEL (DATA=MEANCER, NAME=M_&DEPVAR._CER); /*OUTPUT -MEAN OF CUL OVER ENV
AND REP*/
%OUPUTEXCEL (DATA=MEANCLR, NAME=M_&DEPVAR._CLR); /*OUTPUT -MEAN OF CUL OVER LOC
AND REP*/
%OUPUTEXCEL (DATA=DSTGGEBIPLOT, NAME=M_&DEPVAR._CE); /*OUTPUT -MEAN OF CUL OVER
ENV*/
%OUPUTEXCEL (DATA= MEAN_LCYR, NAME= M_&DEPVAR._LY ); /*OUTPUT -MEAN LOC OVER
YEAR*/
%OUPUTEXCEL (DATA=MEAN_CLYR, NAME=M_&DEPVAR._CY); /*OUTPUT -MEAN CUL OVER REP*/
%OUPUTEXCEL (DATA=MEAN_CLLC , NAME=M_&DEPVAR._CL ); /*OUTPUT -MEAN CUL OVER
REP*/
%OUPUTEXCEL (DATA=CV_CLLC , NAME=CV_&DEPVAR._CL ); /*OUTPUT -COEFF OF VAR CUL
OVER REP*/
%OUPUTEXCEL (DATA=OVERALL_ANOVA1_&DEPVAR, NAME=OVERALL_ANOVA_&DEPVAR); /*OUTPUT
-ANOVA USING GLM*/
%OUPUTEXCEL (DATA=YEAR_ANOVA1_&DEPVAR, NAME= YEAR_ANOVA_&DEPVAR); /*OUTPUT -
YEARLY ANOVA USING GLM*/
%OUPUTEXCEL (DATA=LOCATION_VALUE_&DEPVAR, NAME= LOCATION_VALUE_&DEPVAR);
/*OUTPUT -LOCATION VALUES*/
%OUPUTEXCEL (DATA=STABLE&DEPVAR, NAME=STAB_&DEPVAR); /*OUTPUT -STABILITY
METHODS*/
%OUPUTEXCEL (DATA=CORRSPEAR3&DEPVAR, NAME=SPEAR_&DEPVAR); /*OUTPUT -RANK
CORRELATION W/ LEVEL OF SIG.*/

 *COMMENT: COMPUTING INPUT FILES FOR GGEBIPLOT ANALYSIS IN R SOFTWARE;
      DATA _NULL_;
            SET TEMP_CL;
            CUL = '_'||LEFT(DISTINCT_CL);
            CALL SYMPUT ('CL'||TRIM(LEFT(_N_)), CUL); /*MACRO FOR CL NAME*/
      RUN;

      DATA _NULL_;
            SET TEMP_EN;
            CALL SYMPUT ('EN'||TRIM(LEFT(_N_)), DISTINCT_EN); /*MACRO FOR EN
NAME*/
      RUN;

      DATA _NULL_;
            SET TEMP_LC;
            CALL SYMPUT ('LC'||TRIM(LEFT(_N_)), DISTINCT_LC); /*MACRO FOR LC
NAME*/
      RUN;

      DATA DSTGGEBIPLOT1;
            LENGTH GEN ENV $6.;
            SET MEANCE;
            CUL = '_'||LEFT(CL);
            /*COMMENT: INVOKING MACRO FOR RENAMING GENOTYPE AND ENVIRONMENT
FOR GGEBIPLOT ANALYSIS*/
            %GENOTYPE;
            %ENVIRONMENT;
      RUN;
```

```
	PROC SORT DATA =DSTGGEBIPLOT1;
		BY GEN;
	RUN;

	PROC TRANSPOSE DATA=DSTGGEBIPLOT1  OUT=GGEBIPLOT2&DEPVAR (DROP=_NAME_);
		BY GEN; /*OUTPUT- READY TO GO INPUT FILES FOR GGEBIPLOT ANALYSIS
USING R SOFTWARE*/
		ID ENV;
		VAR MEAN&DEPVAR;
	RUN;

	DATA GGEBIPLOTCXL;
		LENGTH GEN $6.;
		SET MEANCL;
		CUL = '_'||LEFT(CL);
		/*COMMENT: INVOKING MACRO FOR RENAMING GENOTYPE AND ENVIRONMENT
FOR GGEBIPLOT ANALYSIS*/
		%GENOTYPE;
	RUN;

	PROC SORT DATA =GGEBIPLOTCXL;
		BY GEN;
	RUN;

	PROC TRANSPOSE DATA=GGEBIPLOTCXL  OUT=GGEBIPLOTCXL&DEPVAR (DROP=_NAME_);
		BY GEN; /*OUTPUT- READY TO GO INPUT FILES FOR GGEBIPLOT ANALYSIS
USING R SOFTWARE*/
		ID LC;
		VAR MEAN&DEPVAR;
	RUN;

	*COMMENT: COMPUTE INPUT FILES FOR AMMI ANALYSIS IN R SOFTWARE;
	DATA AMMI1&DEPVAR; /*OUTPUT- READY TO GO INPUT FILES FOR AMMI (GEN x ENV)
ANALYSIS USING R SOFTWARE*/
		RETAIN ENV GEN RP MEAN&DEPVAR;
		LENGTH GEN ENV $6.;
		SET MEANCER;
		CUL = '_'||LEFT(CL);
		/*COMMENT: INVOKE MACRO FOR RENAMING GENOTYPE AND ENVIRONMENT FOR
AMMI ANALYSIS*/
		%GENOTYPE;
		%ENVIRONMENT;
		DROP  CL CUL EN;
		RENAME ENV=Locality GEN=Genotype RP=Rep MEAN&DEPVAR = &DEPVAR;
	RUN;

	DATA AMMI2&DEPVAR; /*OUTPUT- READY TO GO INPUT FILES FOR AMMI (GEN x LOC)
ANALYSIS USING R SOFTWARE*/
		RETAIN LOC GEN RP MEAN&DEPVAR;
		LENGTH GEN LOC $6.;
		SET MEANCLR;
		CUL = '_'||LEFT(CL);
		/*COMMENT: INVOKE MACRO FOR RENAMING GENOTYPE AND LOCATION FOR
AMMI ANALYSIS*/
		%GENOTYPE;
		%LOCATION;
		DROP  CL CUL LC;
		RENAME LOC=Locality GEN=Genotype RP=Rep MEAN&DEPVAR = &DEPVAR;
	RUN;

*COMMENT: LEGEND FOR GENOTYPE, ENVIRONMENT & LOCATION SIGN USED IN AMMI AND
GGEBIPLOT ANALYSIS;
	PROC SQL NOPRINT;
```

```
            CREATE TABLE LEGEND_GENO&DEPVAR AS SELECT DISTINCT GEN, CUL FROM
DSTGGEBIPLOT1 ORDER BY GEN;
            CREATE TABLE LEGEND_ENV&DEPVAR AS SELECT DISTINCT ENV, EN FROM
DSTGGEBIPLOT1 ORDER BY ENV;
      QUIT;

      DATA LEGEND_LOC&DEPVAR;
            LENGTH LOC $6.;
            SET TEMP_LC (RENAME=(DISTINCT_LC = LC));
            %LOCATION;
      RUN;

*COMMENT: INVOKE MACRO FOR EXPORTING OUTPUT/RESULTS (.XLSX);
      %OUPUTEXCEL (DATA=LEGEND_GENO&DEPVAR, NAME=GEN_LEGEND_&DEPVAR); /*OUTPUT
-GENOTYPE LEGEND*/
      %OUPUTEXCEL (DATA=LEGEND_ENV&DEPVAR, NAME=ENV_LEGEND_&DEPVAR); /*OUTPUT -
ENVIRONMENT LEGEND*/
      %OUPUTEXCEL (DATA=LEGEND_LOC&DEPVAR, NAME=LOC_LEGEND_&DEPVAR); /*OUTPUT -
LOCATION LEGEND*/

*COMMENT: INVOKE MACRO FOR EXPORTING OUTPUT/RESULTS (.CSV);
%OUPUTCSV (DATA=GGEBIPLOT2&DEPVAR, NAME=BIPLOT_&DEPVAR); /*OUTPUT USED FOR
GGEBIPLOT (GEN x ENV) ANALYSIS IN R SOFTWARE*/
%OUPUTCSV (DATA=GGEBIPLOTCXL&DEPVAR, NAME=BIPLOT2_&DEPVAR); /*OUTPUT USED FOR
GGEBIPLOT (GEN X LOC) ANALYSIS IN R SOFTWARE*/
%OUPUTCSV (DATA=AMMI1&DEPVAR, NAME=AMMI1_&DEPVAR); /*OUTPUT USED FOR AMMI (GEN
X ENV) ANALYSIS IN R SOFTWARE*/
%OUPUTCSV (DATA=AMMI2&DEPVAR, NAME=AMMI2_&DEPVAR); /*OUTPUT USED FOR AMMI (GEN
X LOC) ANALYSIS IN R SOFTWARE*/

%END;
%MEND STABILITY;

*COMMENT: INVOKE MACRO FOR GENOTYPE X ENVIRONMENTAL INTERACTION OF ALL
DEPENDENT VARIABLES SIMULTANEOUSLY;
%STABILITY (FINAL=&LAST_DEPVARIABLE);
Run;
```

Example data used in SASGxE program

YR	LC	RP	CL	MKWT1	CLWT1	MKWT2	CLWT2	MKWT3	CLWT3	MKWT4	CLWT4
2009	KN	1	8	0	0	79	0	27	0	20	0
2009	KN	1	10	0	0	136	13	10	0	19	3
2009	KN	1	3	0	0	72	0	31	0	16	0
2009	KN	1	5	31	0	60	0	12	15	8	4
2009	KN	1	7	0	0	69	7	22	0	18	0
2009	KN	1	4	0	0	138	0	18	0	0	0
2009	KN	1	9	0	0	121	0	17	0	0	0
2009	KN	1	2	62	0	70	0	19	0	29	0
2009	KN	1	6	0	0	116	0	66	0	0	0
2009	KN	1	1	0	0	37	0	19	0	39	0
2009	KN	2	3	27	0	83	0	50	0	21	4
2009	KN	2	10	0	0	54	22	62	16	4	6
2009	KN	2	8	0	0	59	10	95	0	17	0

2009	KN	2	9	4	0	13	0	18	8	15	0
2009	KN	2	2	0	10	171	0	26	0	0	28
2009	KN	2	4	5	0	5	124	0	4	11	6
2009	KN	2	7	42	10	27	0	0	0	11	0
2009	KN	2	1	0	0	118	11	13	0	15	5
2009	KN	2	5	38	0	62	0	66	0	16	0
2009	KN	2	6	0	0	138	0	18	0	0	0
2009	CI	1	7	47	7	55	0	84	0	0	0
2009	CI	1	3	72	0	17	7	9	5	0	0
2009	CI	1	1	54	0	87	0	20	0	0	0
2009	CI	1	4	33	0	60	0	0	0	0	0
2009	CI	1	6	99	0	28	6	0	0	0	0
2009	CI	1	10	122	0	11	6	0	0	0	0
2009	CI	1	5	62	0	108	0	0	0	0	0
2009	CI	1	8	0	0	0	0	0	0	0	0
2009	CI	1	9	106	41	21	0	0	0	0	0
2009	CI	1	2	57	0	36	12	22	15	0	0
2009	CI	2	3	0	0	68	0	141	11	0	0
2009	CI	2	10	32	0	143	5	25	7	0	0
2009	CI	2	8	80	0	46	0	0	9	0	0
2009	CI	2	9	76	0	33	0	0	0	0	0
2009	CI	2	2	34	5	97	0	0	0	0	0
2009	CI	2	4	31	0	104	0	12	8	0	0
2009	CI	2	7	42	0	69	0	21	9	0	0
2009	CI	2	1	29	0	75	3	8	0	0	0
2009	CI	2	5	43	0	62	45	5	0	0	0
2009	CI	2	6	119	0	28	0	29	12	0	0
2009	SC	1	9	108	0	208	0	157	0	0	0
2009	SC	1	1	91	0	85	0	120	0	0	0
2009	SC	1	3	147	0	0	0	0	0	0	0
2009	SC	1	**8**	165	0	154	0	124	0	0	0
2009	SC	1	6	25	0	131	0	46	0	0	0
2009	SC	1	10	160	0	51	0	0	0	0	0
2009	SC	1	4	91	0	101	0	20	0	0	0
2009	SC	1	2	78	0	155	0	60	0	0	0
2009	SC	1	7	108	0	97	0	0	0	0	0
2009	SC	1	5	44	0	169	0	54	0	0	0
2009	SC	2	7	0	0	111	0	43	0	0	0
2009	SC	2	5	119	0	151	0	0	0	0	0
2009	SC	2	6	37	0	115	0	82	0	0	0
2009	SC	2	8	128	0	57	0	26	0	0	0

2009	SC	2	1	66	0	129	0	24	0	0	0
2009	SC	2	4	26	0	82	0	105	0	0	0
2009	SC	2	10	59	0	107	0	54	0	0	0
2009	SC	2	9	60	0	42	0	0	0	0	0
2009	SC	2	2	70	0	120	0	62	0	0	0
2009	SC	2	3	97	0	105	0	18	0	0	0
2009	GA	1	7	0	0	144	11	0	0	0	0
2009	GA	1	4	0	0	202	0	0	0	0	0
2009	GA	1	1	0	0	117	0	0	0	0	0
2009	GA	1	9	14	0	105	0	0	0	0	0
2009	GA	1	2	51	0	94	0	0	0	0	0
2009	GA	1	10	113	0	33	0	0	0	0	0
2009	GA	1	**5**	37	0	62	0	0	0	0	0
2009	GA	1	6	28	0	75	0	0	0	0	0
2009	GA	1	3	36	0	159	27	0	0	0	0
2009	GA	1	8	0	0	253	0	0	0	0	0
2009	GA	2	9	29	0	195	0	0	0	0	0
2009	GA	2	1	40	0	262	0	0	0	0	0
2009	GA	2	7	47	0	187	0	0	0	0	0
2009	GA	2	10	54	0	96	0	0	0	0	0
2009	GA	2	3	76	0	88	0	0	0	0	0
2009	GA	2	4	60	0	151	11	0	0	0	0
2009	GA	2	6	32	0	135	0	0	0	0	0
2009	GA	2	8	83	0	116	0	0	0	0	0
2009	GA	2	2	58	0	277	0	0	0	0	0
2009	GA	2	5	21	0	96	0	0	0	0	0
2009	FL	1	8	47	7	55	0	84	0	0	0
2009	FL	1	10	72	0	17	7	9	5	0	0
2009	FL	1	3	54	0	87	0	20	0	0	0
2009	FL	1	5	33	0	60	0	0	0	0	0
2009	FL	1	7	99	0	28	6	0	0	0	0
2009	FL	1	4	10	0	9	0	32	6	0	0
2009	FL	1	9	7	0	11	0	10	0	0	0
2009	FL	1	2	89	0	13	0	0	29	0	0
2009	FL	1	6	0	0	25	0	50	0	0	0
2009	FL	1	1	26	0	82	0	105	0	0	0
2009	FL	2	3	59	0	107	0	54	0	0	0
2009	FL	2	10	60	0	42	0	0	0	0	0
2009	FL	2	8	70	0	120	0	62	0	0	0
2009	FL	2	9	97	0	105	0	18	0	0	0
2009	FL	2	2	0	0	144	11	0	0	0	0

2009	FL	2	4	12	0	15	8	0	0	15	4
2009	FL	2	7	0	0	53	0	15	0	10	0
2009	FL	2	1	8	0	8	0	0	0	3	2
2009	FL	2	5	25	0	29	0	12	0	17	13
2009	FL	2	6	0	0	0	0	16	7	17	8
2010	KN	1	9	121	0	133	0	0	0	0	0
2010	KN	1	1	21	0	191	0	0	0	0	0
2010	KN	1	3	94	0	87	0	0	0	0	0
2010	KN	1	**8**	127	0	113	0	0	0	0	0
2010	KN	1	6	114	0	107	0	0	0	0	0
2010	KN	1	10	147	0	47	0	0	0	0	0
2010	KN	1	4	103	0	170	0	0	0	0	0
2010	KN	1	2	90	0	0	0	0	0	0	0
2010	KN	1	7	136	0	149	0	0	0	0	0
2010	KN	1	5	117	0	159	0	0	0	0	0
2010	KN	2	7	78	0	179	0	0	0	0	0
2010	KN	2	5	193	0	101	0	0	0	0	0
2010	KN	2	6	124	0	139	0	0	0	0	0
2010	KN	2	8	97	0	212	0	0	0	0	0
2010	KN	2	1	55	0	196	0	0	0	0	0
2010	KN	2	4	30	0	80	0	0	0	0	0
2010	KN	2	10	44	0	129	0	0	0	0	0
2010	KN	2	9	65	0	105	0	0	0	0	0
2010	KN	2	2	195	0	22	0	0	0	0	0
2010	KN	2	3	79	0		0	0	0	0	0
2010	CI	1	7	68	0	0	0	53	12	0	0
2010	CI	1	4	80	0	43	0	0	0	0	0
2010	CI	1	1	48	0	90	0	66	30	0	0
2010	CI	1	9	63	0	38	0	40	37	0	0
2010	CI	1	2	53	0	108	0	30	4	0	0
2010	CI	1	10	41	0	22	0	45	9	0	0
2010	CI	1	**5**	0	0	127	0	11	0	0	0
2010	CI	1	6	0	0	0	0	18	21	0	0
2010	CI	1	3	10	0	9	0	32	6	0	0
2010	CI	1	8	7	0	11	0	10	0	0	0
2010	CI	2	9	89	0	13	0	0	29	0	0
2010	CI	2	1	0	0	25	0	50	0	0	0
2010	CI	2	7	0	0	66	0	0	0	0	0
2010	CI	2	10	0	0	0	0	26	14	0	0
2010	CI	2	3	0	0	0	0	94	6	0	0
2010	CI	2	4	0	0	67	8	104	27	0	0

2010	CI	2	6	47	0	0	0	0	0	0	0
2010	CI	2	8	28	0	39	0	0	17	0	0
2010	CI	2	2	0	0	17	0	30	0	0	0
2010	CI	2	5	0	0	25	2	13	2	0	0
2010	SC	1	8	0	0	155	31	0	0	0	0
2010	SC	1	10	0	0	77	26	0	0	0	0
2010	SC	1	3	0	0	118	40	0	0	0	0
2010	SC	1	5	0	0	123	14	0	0	0	0
2010	SC	1	7	0	0	103	0	0	0	0	0
2010	SC	1	4	0	0	100	26	0	0	0	0
2010	SC	1	9	0	0	116	18	0	0	0	0
2010	SC	1	2	0	0	109	0	0	0	0	0
2010	SC	1	6	0	0	72	9	0	0	0	0
2010	SC	1	1	0	0	110	18	0	0	0	0
2010	SC	2	3	0	0	141	33	0	0	0	0
2010	SC	2	10	0	0	109	12	0	0	0	0
2010	SC	2	8	0	0	83	8	0	0	0	0
2010	SC	2	9	0	0	110	10	0	0	0	0
2010	SC	2	2	0	0	94	11	0	0	0	0
2010	SC	2	4	0	0	150	14	0	0	0	0
2010	SC	2	7	0	0	190	20	0	0	0	0
2010	SC	2	1	0	0	122	12	0	0	0	0
2010	SC	2	5	0	0	173	97	0	0	0	0
2010	SC	2	6	0	0	128	14	0	0	0	0
2010	GA	1	5	107	0	95	0	0	0	0	0
2010	GA	1	10	62	0	98	0	0	0	0	0
2010	GA	1	6	163	0	83	0	0	0	0	0
2010	GA	1	8	62	0	44	0	0	0	0	0
2010	GA	1	7	0	0	28	0	81	0	0	0
2010	GA	1	4	0	0	41	0	72	0	35	0
2010	GA	1	9	4	0	37	0	16	12	19	0
2010	GA	1	2	31	0	104	0	12	8	0	0
2010	GA	1	3	42	0	69	0	21	9	0	0
2010	GA	1	1	29	0	75	3	8	0	0	0
2010	GA	2	3	43	0	62	45	5	0	0	0
2010	GA	2	10	119	0	28	0	29	12	0	0
2010	GA	2	8	108	0	208	0	157	0	0	0
2010	GA	2	9	91	0	85	0	120	0	0	0
2010	GA	2	2	147	0	0	0	0	0	0	0
2010	GA	2	4	165	0	154	0	124	0	0	0
2010	GA	2	7	28	0	24	0	17	0	0	0

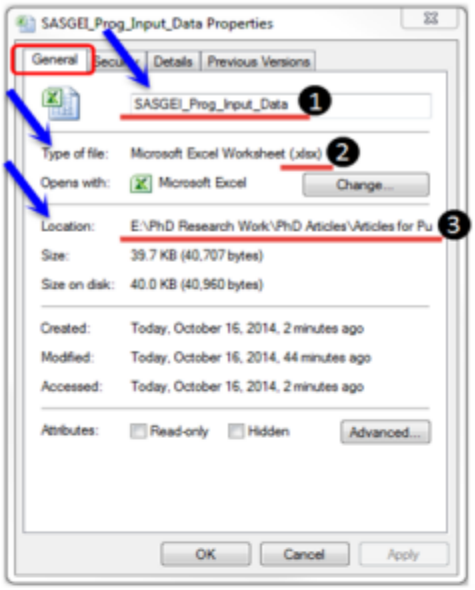
2010	GA	2	1	17	0	21	0	0	0	18	10
2010	GA	2	5	33	0	23	0	0	0	7	0
2010	GA	2	6	12	0	15	8	0	0	15	4
2010	FL	1	8	0	0	0	0	35	0	0	0
2010	FL	1	10	0	0	7	0	18	0	6	0
2010	FL	1	3	0	0	11	0	25	0	0	0
2010	FL	1	5	0	0	15	0	25	0	0	0
2010	FL	1	7	9	0	12	0	3	0	0	0
2010	FL	1	4	8	0	10	0	0	0	30	0
2010	FL	1	9	6	0	0	0	9	0	12	0
2010	FL	1	2	30	0	16	0	0	0	0	11
2010	FL	1	6	13	0	21	0	0	0	10	0
2010	FL	1	1	8	0	11	0	23	0	24	0
2010	FL	2	3	10	0	50	0	0	0	0	0
2010	FL	2	10	28	0	24	0	17	0	0	0
2010	FL	2	8	17	0	21	0	0	0	18	10
2010	FL	2	9	33	0	23	0	0	0	7	0
2010	FL	2	2	12	0	15	8	0	0	15	4
2010	FL	2	4	0	0	53	0	15	0	10	0
2010	FL	2	7	8	0	8	0	0	0	3	2
2010	FL	2	1	25	0	29	0	12	0	17	13
2010	FL	2	5	0	0	0	0	16	7	17	8
2010	FL	2	6	0	0	26	0	49	7	0	0
2011	KN	1	7	67	0	72	0	0	0	0	0
2011	KN	1	3	92	0	87	0	0	0	0	0
2011	KN	1	1	88	0	18	0	0	0	0	0
2011	KN	1	4	63	0	97	0	0	0	0	0
2011	KN	1	6	51	0	101	0	0	0	0	0
2011	KN	1	10	97	0	87	0	0	0	0	0
2011	KN	1	5	117	3	75	0	0	0	0	0
2011	KN	1	8	129	17	66	14	0	0	0	0
2011	KN	1	9	79	4	105	0	0	0	0	0
2011	KN	1	2	148	0	0	0	0	0	0	0
2011	KN	2	3	26	0	96	0	0	0	0	0
2011	KN	2	10	86	0	173	0	0	0	0	0
2011	KN	2	8	164	0	75	0	0	0	0	0
2011	KN	2	9	112	11	37	0	0	0	0	0
2011	KN	2	2	60	0	100	0	0	0	0	0
2011	KN	2	4	91	40	137	22	0	0	0	0
2011	KN	2	7	132	9	67	0	0	0	0	0
2011	KN	2	1	115	42	28	0	0	0	0	0

2011	KN	2	5	93	15	73	0	0	0	0	0
2011	KN	2	6	66	0	114	0	0	0	0	0
2011	CI	1	9	115	0	46	0	0	0	0	0
2011	CI	1	1	94	0	175	0	0	0	0	0
2011	CI	1	3	60	0	136	0	0	0	0	0
2011	CI	1	**8**	115	0	142	0	0	0	0	0
2011	CI	1	6	67	0	189	0	0	0	0	0
2011	CI	1	10	95	0	72	0	0	0	0	0
2011	CI	1	4	134	0	33	0	0	0	0	0
2011	CI	1	2	126	0	69	0	0	0	0	0
2011	CI	1	7	66	0	169	0	0	0	0	0
2011	CI	1	5	110	0	52	0	0	0	0	0
2011	CI	2	7	82	0	121	0	0	0	0	0
2011	CI	2	5	242	0	32	0	0	0	0	0
2011	CI	2	6	212	0	74	0	0	0	0	0
2011	CI	2	8	134	0	10	0	0	0	0	0
2011	CI	2	1	176	0	69	0	0	0	0	0
2011	CI	2	4	119	0	33	0	0	0	0	0
2011	CI	2	10	107	0	95	0	0	0	0	0
2011	CI	2	9	62	0	98	0	0	0	0	0
2011	CI	2	2	163	0	83	0	0	0	0	0
2011	CI	2	3	62	0	44	0	0	0	0	0
2011	SC	1	7	0	0	28	0	81	0	0	0
2011	SC	1	4	0	0	41	0	72	0	35	0
2011	SC	1	1	4	0	37	0	16	12	19	0
2011	SC	1	9	0	0	0	0	0	0	0	0
2011	SC	1	2	11	0	18	0	6	0	0	0
2011	SC	1	10	5	0	20	0	24	10	17	0
2011	SC	1	**5**	0	0	6	0	10	0	0	0
2011	SC	1	6	0	0	0	0	0	0	20	0
2011	SC	1	3	0	0	0	0	0	0	0	0
2011	SC	1	8	0	0	0	0	14	0	0	0
2011	SC	2	9	0	0	0	5	0	0	14	0
2011	SC	2	1	0	0	10	0	0	0	0	0
2011	SC	2	7	22	0	18	0	34	0	37	0
2011	SC	2	10	35	0	20	0	31	0	20	0
2011	SC	2	3	0	0	26	0	0	0	0	0
2011	SC	2	4	19	0	22	0	38	0	44	0
2011	SC	2	6	15	0	16	0	10	0	16	0
2011	SC	2	8	19	9	24	0	28	18	28	0
2011	SC	2	2	25	0	63	0	81	0	16	0

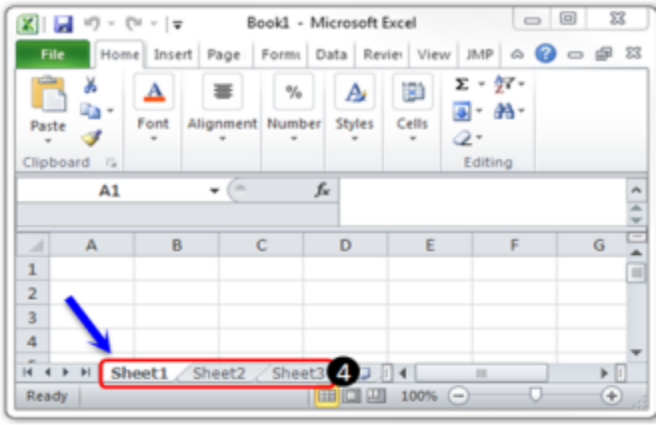
2011	SC	2	5	22	0	26	0	10	0	20	0
2011	GA	1	9	34	0	19	0	0	0	0	0
2011	GA	1	1	19	0	0	0	0	0	0	0
2011	GA	1	3	17	0	20	0	0	0	0	0
2011	GA	1	**8**	33	0	34	0	0	0	0	0
2011	GA	1	6	0	0	0	0	0	0	0	0
2011	GA	1	10	0	0	0	0	0	0	0	0
2011	GA	1	4	45	0	5	0	0	0	0	0
2011	GA	1	2	17	0	1	0	0	0	0	0
2011	GA	1	7	9	0	65	0	0	0	0	0
2011	GA	1	5	28	0	46	0	0	0	0	0
2011	GA	2	7	83	0	20	0	0	0	0	0
2011	GA	2	5	25	0	99	0	0	0	0	0
2011	GA	2	6	4	0	21	0	0	0	0	0
2011	GA	2	8	10	0	7	0	0	0	0	0
2011	GA	2	1	81	0	15	0	0	0	0	0
2011	GA	2	4	112	0	47	0	0	0	0	0
2011	GA	2	10	67	0	40	0	0	0	0	0
2011	GA	2	9	82	0	61	0	0	0	0	0
2011	GA	2	2	41	0	40	0	0	0	0	0
2011	GA	2	3	36	0	98	0	0	0	0	0
2011	FL	1	8	0	0	0	0	35	0	0	0
2011	FL	1	10	0	0	7	0	18	0	6	0
2011	FL	1	3	0	0	11	0	25	0	0	0
2011	FL	1	5	0	0	15	0	25	0	0	0
2011	FL	1	7	9	0	12	0	3	0	0	0
2011	FL	1	4	8	0	10	0	0	0	30	0
2011	FL	1	9	6	0	0	0	9	0	12	0
2011	FL	1	2	30	0	16	0	0	0	0	11
2011	FL	1	6	13	0	21	0	0	0	10	0
2011	FL	1	1	8	0	11	0	23	0	24	0
2011	FL	2	3	10	0	50	0	0	0	0	0
2011	FL	2	10	28	0	24	0	17	0	0	0
2011	FL	2	8	17	0	21	0	0	0	18	10
2011	FL	2	9	33	0	23	0	0	0	7	0
2011	FL	2	2	12	0	15	8	0	0	15	4
2011	FL	2	4	0	0	53	0	15	0	10	0
2011	FL	2	7	8	0	8	0	0	0	3	2
2011	FL	2	1	25	0	29	0	12	0	17	13
2011	FL	2	5	0	0	0	0	16	7	17	8
2011	FL	2	6	0	0	26	0	49	7	0	0

References:

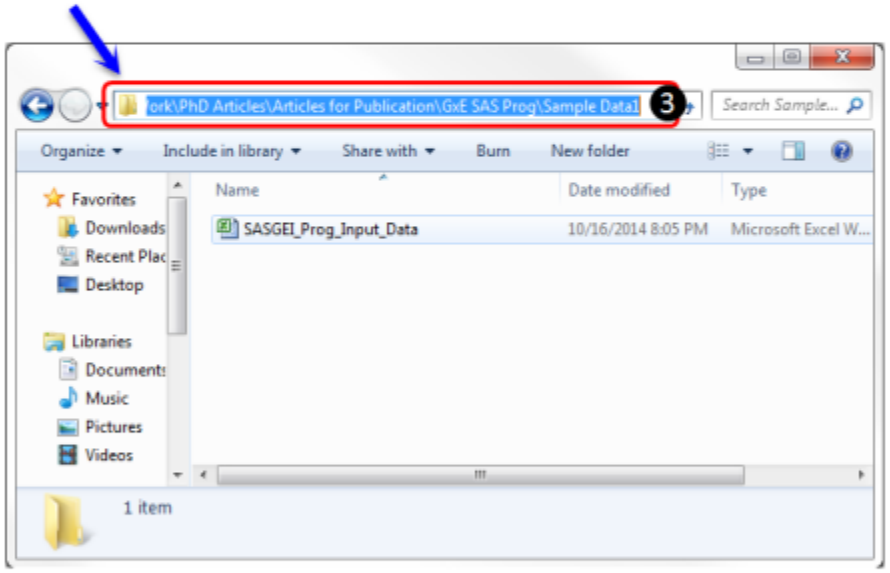
Yan W, Kang MS, Ma B, Woods S, Cornelius PL. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Sci. 47:643-653.



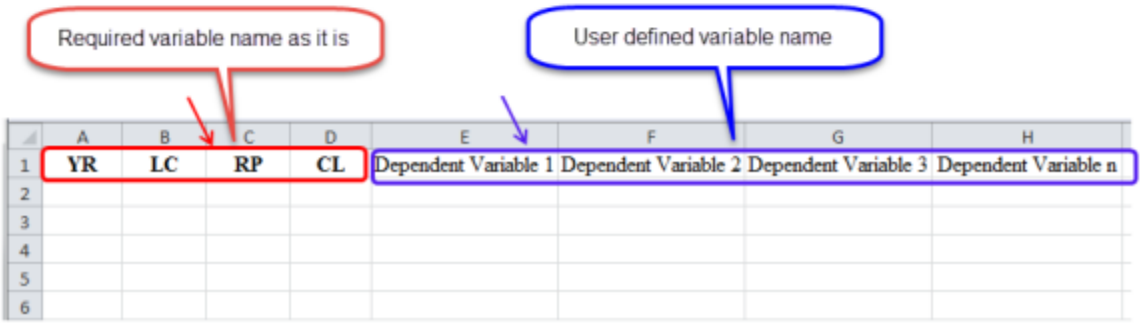
Panel A



Panel B

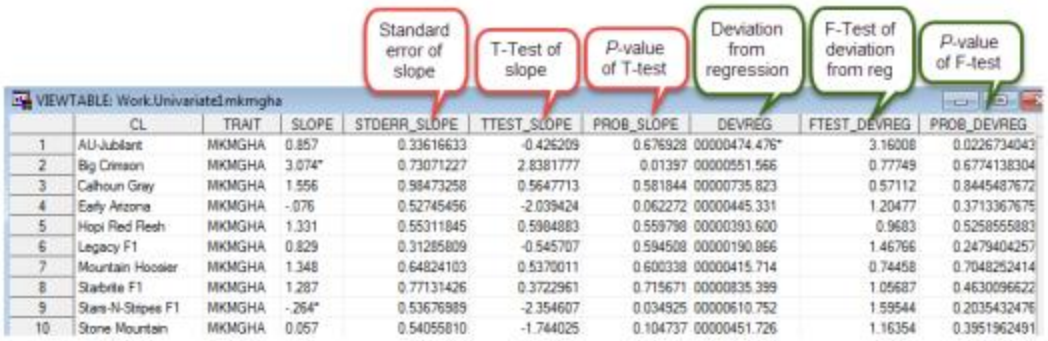


Panel C



Panel D

Figure S1. Screenshot of ‘Properties’ window (Panel A) of input data (.xlsx) file represents file name❶, file type❷ and file location❸; an example Excel file (Panel B) represents sheet name❹; folder (Panel C) represent the location❸ of input data file (.xlxs); and input data template (Panel D). The required variable year (**YR**), location (**LC**), replication (**RP**), and cultigen or genotype (**GN**) are represented in ‘bold’, ‘capital case’ and enclosed in a separate box (Panel D).



	CL	TRAIT	SLOPE	STDERR_SLOPE	TTEST_SLOPE	PROB_SLOPE	DEVREG	FTEST_DEVREG	PROB_DEVREG
1	AU-Jubilant	MKMGHA	0.857	0.33616633	-0.426209	0.676928	00000474.476*	3.16008	0.0226734043
2	Big Crimson	MKMGHA	3.074*	0.73071227	2.8381777	0.01397	00000551.566	0.77749	0.6774138304
3	Calhoun Gray	MKMGHA	1.556	0.98473258	0.5647713	0.581844	00000735.823	0.57112	0.8445487672
4	Early Arizona	MKMGHA	-.076	0.52745456	-2.039424	0.062272	00000445.331	1.20477	0.3713367675
5	Hopi Red Flesh	MKMGHA	1.331	0.55311845	0.5984883	0.559798	00000393.600	0.9683	0.5258555883
6	Legacy F1	MKMGHA	0.829	0.31285809	-0.545707	0.594508	00000190.866	1.46766	0.2479404257
7	Mountain Hoosier	MKMGHA	1.348	0.64824103	0.5370011	0.600338	00000415.714	0.74458	0.7048252414
8	Starbrite F1	MKMGHA	1.287	0.77131426	0.3722961	0.715671	00000835.399	1.05687	0.4630096622
9	Stars-N-Stripes F1	MKMGHA	-.264*	0.53676989	-2.354607	0.034925	00000610.752	1.59544	0.2035432476
10	Stone Mountain	MKMGHA	0.057	0.54055810	-1.744025	0.104737	00000451.726	1.16354	0.3951962491

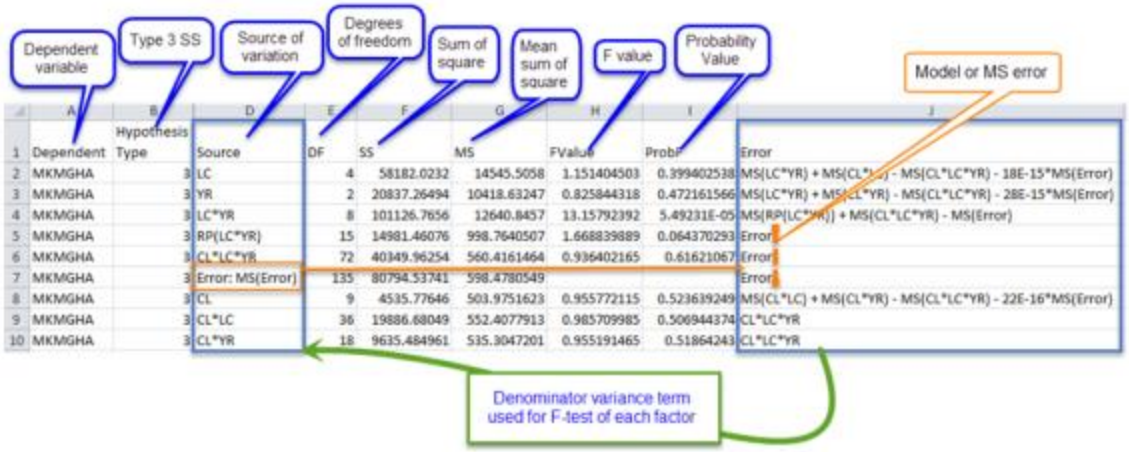
Panel A

	CL	TRAIT	BETA_PERKINS_AND_JINKS	SHUKLA	ECOVALENCE	PI	CVI
1	AU-Jubilant	MKMGHA	-0.06422444	523.9940333	6645.0901508	1473.63089***	55.7027
2	Big Crimson	MKMGHA	0.051701924	263.1007483	3723.0853587	1926.41082	74.776
3	Calhoun Gray	MKMGHA	0.448752791	1171.0621	13892.2525	4046.21425*	77.2177
4	Early Arizona	MKMGHA	-0.3281333	618.5130514	7703.703153	745.12845	52.2443
5	Hopi Red Flesh	MKMGHA	-0.02920956	392.536093	5172.7612191	1214.00342	57.5507
6	Legacy F1	MKMGHA	-0.10199315	185.7981163	2857.2958802	1067.86029**	50.569
7	Mountain Hoosier	MKMGHA	-0.01595743	449.2536724	5807.9981084	1951.60463*	60.878
8	Starbrite F1	MKMGHA	-0.2426858	1024.015605	12245.331758	1149.07738	59.1849
9	Stars-N-Stripes F1	MKMGHA	0.156099866	533.4552098	6751.055327	2324.50560***	69.6964
10	Stone Mountain	MKMGHA	0.125649101	383.6783533	5073.5545349	1786.15529**	61.159

Panel B

	CL	TRAIT	LSMEAN	LSD	SHUKLA	YS
1	Calhoun Gray	MKMGH	63.057919386	12.492150369	1171.06210*	7@
2	Starbrite F1	MKMGH	58.409127963	12.492150369	1024.01561	8@
3	Hopi Red Flesh	MKMGH	56.746470251	12.492150369	392.53609	9@
4	Mountain Hoosier	MKMGH	55.460157329	12.492150369	449.25367	8@
5	Stone Mountain	MKMGH	54.882322072	12.492150369	383.67835	7@
6	AU-Jubilant	MKMGH	53.034603407	12.492150369	523.99403	4
7	Big Crimson	MKMGH	52.910197871	12.492150369	263.10075	3
8	Legacy F1	MKMGH	52.377878425	12.492150369	185.79812	2
9	Stars-N-Stripes F1	MKMGH	50.30126961	12.492150369	533.45521	1
10	Early Arizona	MKMGH	49.032103351	12.492150369	618.51305	0

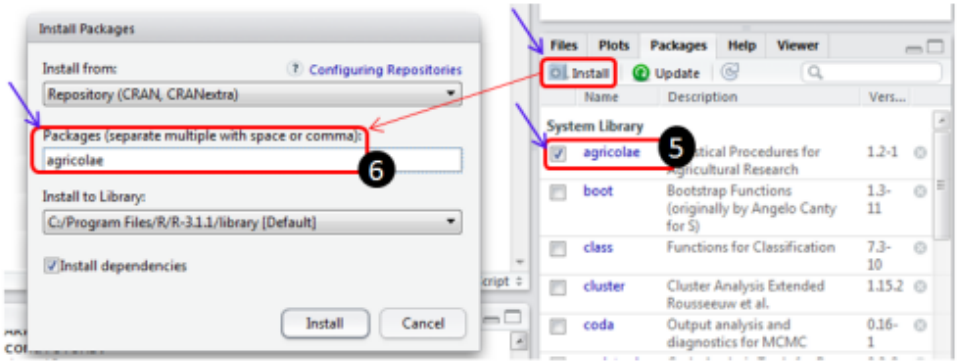
Panel C



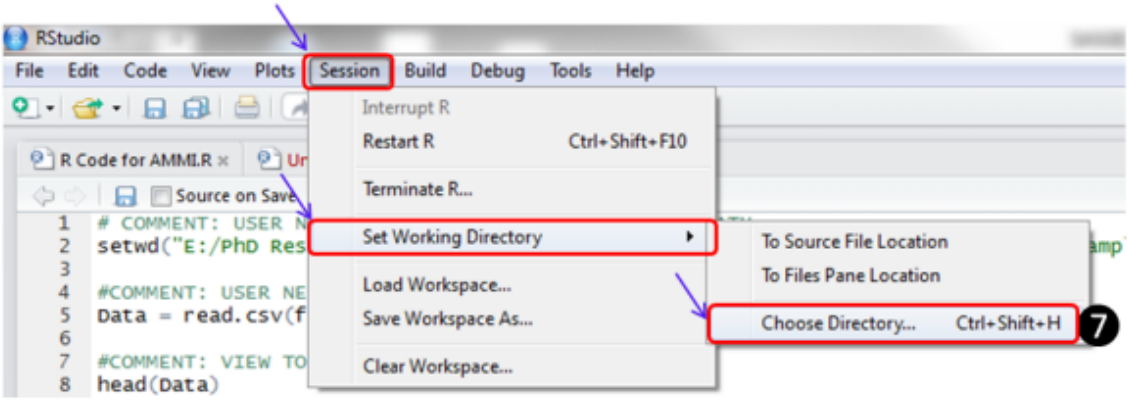
	A	B	D	E	F	G	H	I	J
1	Dependent	Hypothesis Type	Source	DF	SS	MS	FValue	Prob	Error
2	MKMGHA	3	LC	4	58182.0232	14545.5058	1.151404503	0.399402538	MS(LC*YR) + MS(CL*LC) - MS(CL*LC*YR) - 18E-15*MS(Error)
3	MKMGHA	3	YR	2	20837.26494	10418.63247	0.825844318	0.472161566	MS(LC*YR) + MS(CL*YR) - MS(CL*LC*YR) - 28E-15*MS(Error)
4	MKMGHA	3	LC*YR	8	101126.7656	12640.8457	13.15792392	5.49231E-05	MS(RP(LC*YR)) + MS(CL*LC*YR) - MS(Error)
5	MKMGHA	3	RP(LC*YR)	15	14981.46076	998.7640507	1.668839889	0.064370293	Error
6	MKMGHA	3	CL*LC*YR	72	40349.96254	560.4161464	0.936402165	0.61621067	Error
7	MKMGHA	3	Error: MS(Error)	135	80794.53741	598.4780549			Error
8	MKMGHA	3	CL	9	4535.77646	503.9751623	0.955772115	0.523639249	MS(CL*LC) + MS(CL*YR) - MS(CL*LC*YR) - 22E-16*MS(Error)
9	MKMGHA	3	CL*LC	36	19886.68049	552.4077913	0.985709985	0.506944374	CL*LC*YR
10	MKMGHA	3	CL*YR	18	9635.484961	535.3047201	0.955191465	0.51864243	CL*LC*YR

Panel D

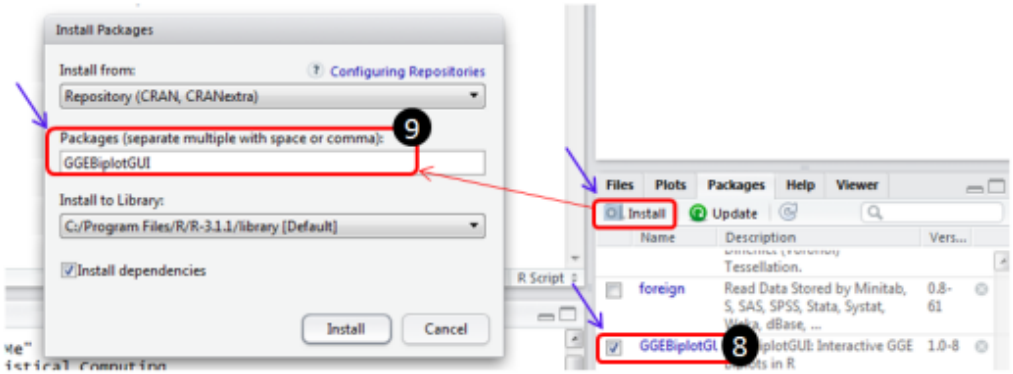
Figure S2. Screenshot of `SLOPE&DEPVAR.SAS` (Panel A), `UNIVARIATE2&DEPVAR.SAS` (Panel B), `UNIVARIATE3&DEPVAR.SAS` (Panel C), and OVERALL_ANOVA_`&DEPVAR`.XLSX (Panel D) files for dependent variable MKMGHA. Panel A represents analysis of regression slope (b_i), standard error of slope, deviation from regression (S^2_d), *T*-test of slope (H_0: $b_i = 1$), *F*-test of deviation from regression (H_0: $S^2_d = 0$) and level of significance of *T*-test and *F*-test. Panel B represents analysis of Wricke's ecovalence (W^2_i), Shukla's stability variance (σ_i^2), Perkins and Jinks beta (β_i), Lin and Binns Pi (P_i) and Francis and Kannenberg coefficient of variation (CV_i). Panel C represents analysis of least square (LS) means, standard error of LS means, Shukla's stability variance (σ_i^2) with test of significance, and Kang's Yield-Stability statistic (YS_i). Panel D represents analysis of variable with source of variation (Source), degrees of freedom (DF), sum of squares (SS), mean sum of squares (MS), *F*-test value (FValue), probability value (ProbF) and error term used as denominator value for computing *F*-test (Error).



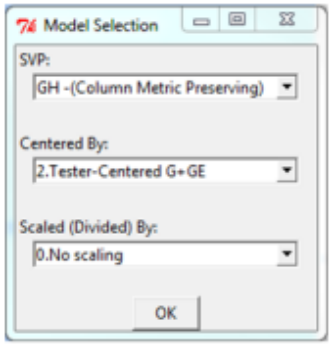
Panel A



Panel B



Panel C



Panel D

Figure S3. Screenshots of R software in 'RStudio'. Panel A represent AMMI package `agricolae`❺ in 'System Library' window and installing AMMI package `agricolae` ❺ from CRAN repository❻. Panel B represent the reference path of folder where input data is located from 'Session' in 'window tool bar' ('session' in window tool bar → select 'set work directory' → select 'choose work directory' → select folder where data is kept)❼. Panel C represent GGEBiplot package `GGEBiplotGUI`❽ in 'System Library' window and installing GGEBiplot package `GGEBiplotGUI`❽ from CRAN repository ❾. Panel D represent 'Model Selection' window, which pops-up after executing `GGEBiplotGUI` model.

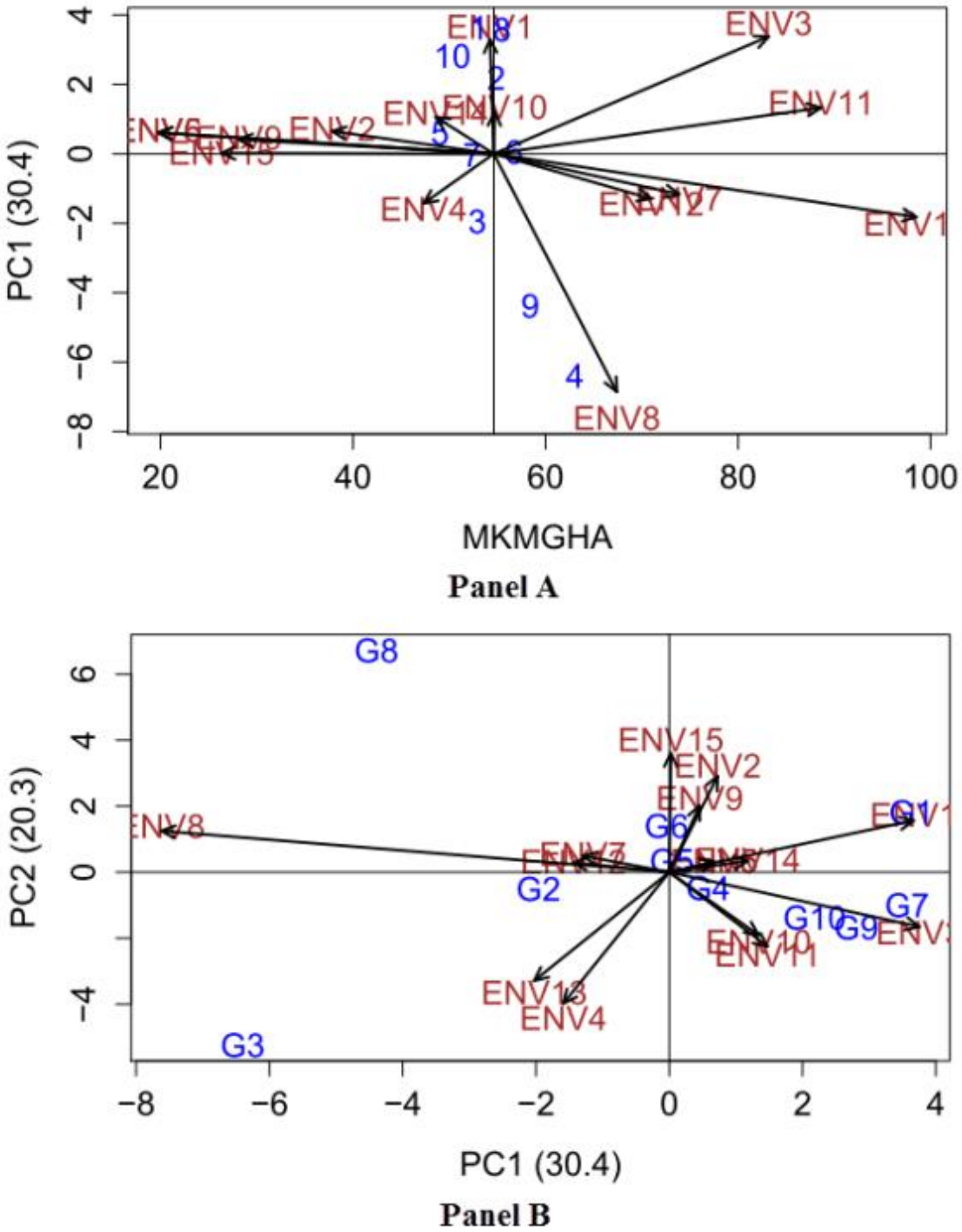


Figure S4. Trait vs. PC1 (Panel A) and PC1 vs. PC2 (Panel B) view of AMMI analysis for trait MKMGHA of example input data of SASGxE program.

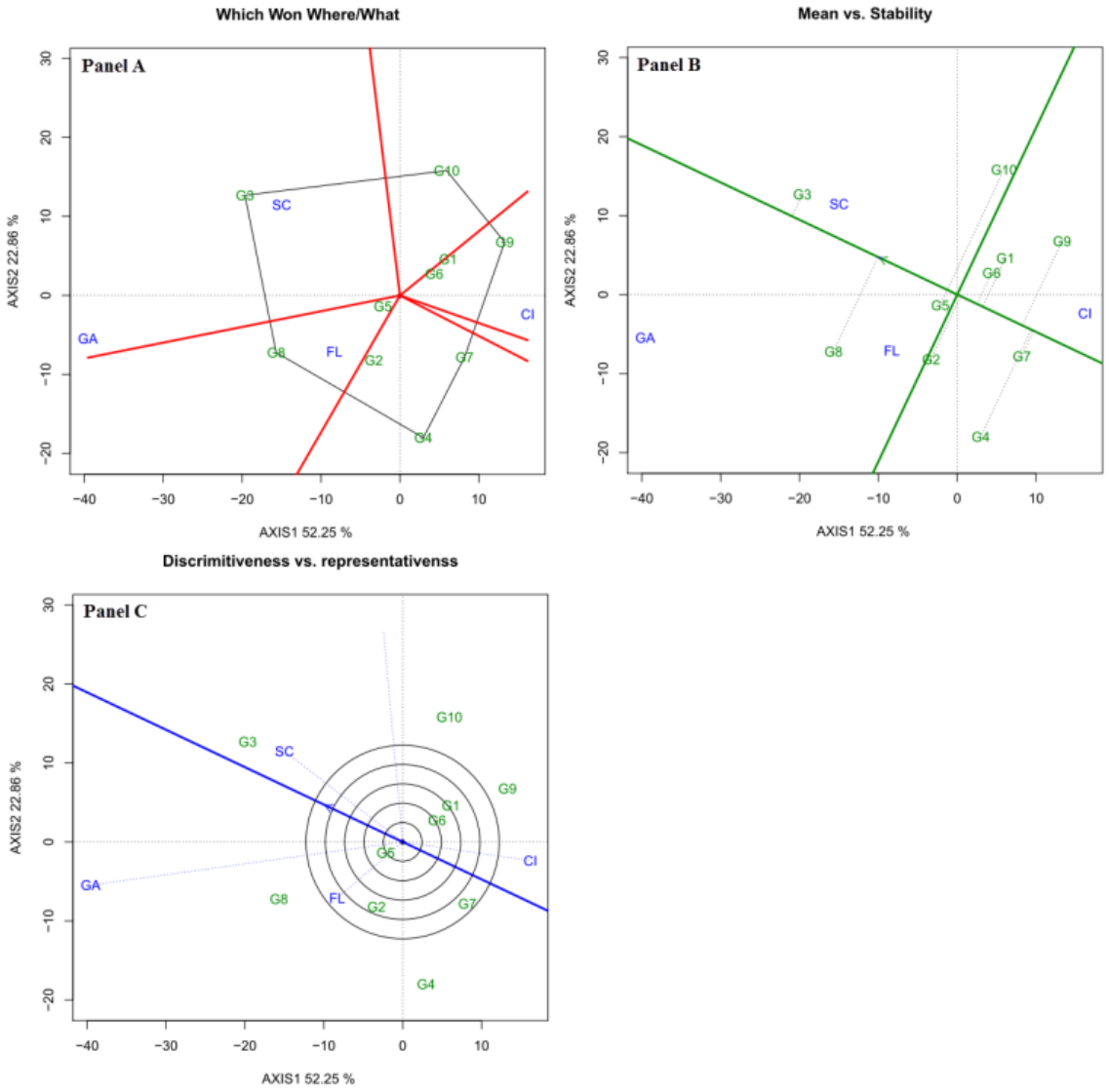


Figure S5. Which-won-where or polygon (Panel A), mean vs. stability (Panel B), and discriminative vs. representative (Panel C) view of GGE biplot for trait MKMGHA of example input data of SASGxE program.