UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

GILVANI MATEI

GENOME-WIDE SELECTION IN SOYBEANS AND OPTIMIZATION OF PHENOTYPING FOR GRAIN YIELD

TESE

PATO BRANCO
2017

GILVANI MATEI

GENOME-WIDE SELECTION IN SOYBEANS AND OPTIMIZATION OF PHENOTYPING FOR GRAIN YIELD

Tese apresentada ao Programa de Pós-Graduação em Agronomia da Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco, como requisito parcial à obtenção do título de Doutor em Agronomia - Área de Concentração: Produção Vegetal.

Orientador: Prof. Dr. Giovani Benin

PATO BRANCO

2017

M425g Matei, Gilvani.
Genome-Wide selection in soybeans and optimization of phenotyping for grain yield / Gilvani Matei. -- 2017.
99 f. : il. ; 30 cm.

Orientador: Prof. Dr. Giovani Benin
Tese (Doutorado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Agronomia. Pato Branco, PR, 201 7.
Bibliografia: f. 75 - 85.

1. Plantas - Melhoramento genético. 2. Marcadores genéticos. 3. Produtividade agrícola. I. Benin, Giovani, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Agronomia. III. Título.

CDD: (22.ed.) 630

Ficha Catalográfica elaborada por
Suélem Belmudes Cardoso CRB9/1630
Biblioteca da UTFPR Campus Pato Branco

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Câmpus Pato Branco
Diretoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Agronomia

TERMO DE APROVAÇÃO

Título da Tese n° XXX

GENOME-WIDE SELECTION IN SOYBEANS AND OPTIMIZATION OF PHENOTYPING FOR GRAIN YIELD

por

GILVANI MATEI

Tese apresentada às 14 horas 00 min. do dia 12 de 12 de 2017 como requisito parcial para obtenção do título de DOUTOR EM AGRONOMIA, Linha de Pesquisa – Produção Vegetal, Programa de Pós-Graduação em Agronomia (Área de Concentração: Produção vegetal) da Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco. O candidato foi arguido pela Banca Examinadora composta pelos membros abaixo designados. Após deliberação, a Banca Examinadora considerou o trabalho APROVADO.

Banca examinadora:

Dra. GRACIELA SALLAS
Nidera Semillas - Argentina

Dr. CLAUDIO ROBERTO CARDOSO DE GODOI
Nidera Sementes – Rio Verde, GO.

Dr. JOSE FRANCISCO FERRAZ DE TOLEDO
Pesquisador

Dr. GROVER SHANNON
University of Missouri - US

Prof. Dr. GIOVANI BENIN
Orientador

Prof. Dr. MOESES ANDRIGO DANNER
Coordenador do PPGA

"O Termo de Aprovação, devidamente assinado, encontra-se arquivado na Coordenação do Programa"

You are the meaning of conquest, unconditional support at all times, the direction of my life. I dedicate this thesis to my daughters Maria Eduarda and Geovanna.

I Dedicate.

ACKNOWLEDGEMENTSG

The completion of this doctoral thesis is the result of a period of great dedication and learning in my life. First of all, I thank Dr. Giovani Benin for his confidence and willingness in the work that produced this thesis, and to whom I owe a great part of the professional maturity that I have obtained during this period.

I give sincere thanks to my professional colleagues Rodolfo Rossi, Jose F. F. Toledo, Claudio R. C. Godoi, Marcio A. Capelin, and Rafael Fiabane for their support and contributions during very challenging but productive periods of work, and in shaping this thesis. On their behalf, I would also like to thank our other co-workers and university colleagues for their helpful contributions during this period.

I thank CAPES and CNPQ for granting the scholarship for my internship period in the United States of America. Thanks are due also to the University of Missouri, Columbia, USA, for the valuable internship, and particularly to Professors Grover Shannon and Henry Nguyen.

I have had the good fortune of being guided in life by my loved ones and special people. I thank my parents, Inêz Maria Cagnini Matei and Mario Gesser Matei, for their unconditional love, and for allowing me to develop my full potential, especially in the most difficult moments of this phase of my life. My two daughters, Maria Eduarda and Geovanna Matei, to whom this work is dedicated, thank you for the joy you have brought to my days, and especially for showing strength, intelligence, and simplicity with which you have overcome this difficult phase of our lives. To my siblings, Michelli and Rodrigo Matei, I am grateful for your affection, companionship, support, and for embodying the competence inherited from our parents. I am also grateful for the care and support of my Nona, Iolanda Risson Cagnini (who represents all my relatives), my future brothers-in-law, Michele Dresch and Alan Beal.

Finally, as in all stages of my life, I would like to express my gratitude to God, especially for placing each one of my friends and well-wishers on my path, and for having guaranteed all the conditions for me to complete my thesis successfully. Thank you very much!

“I believe luck is preparation meeting opportunity.”

Oprah Winfrey

ABSTRACT

MATEI, Gilvani. Genome-Wide Selection in soybeans and optimization of phenotyping for grain yield. 100 pp. Thesis (Ph.D in Agronomy) - Graduate Program in Agronomy (Area of Focus: Plant Production), Federal Technological University of Paraná. Pato Branco, 2017.

In a breeding program, several factors influence the selection of cultivars, mainly due to the high number of genotypes under evaluation and the reduced experimental capacity in the initial phases of the program. In this context, the present study was divided into four parts. The first one aimed to identify the core locations for evaluation and selection of soybean genotypes in the macro-regions 1 (M1) and 2 (M2), in generations with low seed availability. The data set consisted of 22 soybean genotypes grown in 23 sites for three years. The GGL + GGE and G analyses versus the GE analysis were used. The locations Chapada-RS and Maracaju-MS were the most representative sites and discriminant macro-regions 1 and 2, respectively. Identification of the core location is fundamental to evaluation, since it is where the number of test sites can be summarized to a single site by soybean growing macro-region. The second study aimed to evaluate the experimental accuracy of different statistical methods used to analyze the assays with large numbers of soybean genotypes. The grain yield data from 324 soybean genotypes, evaluated in six replicates, were used. The data were analyzed by using the randomized block design, triple lattice design, and Papadakis method. The experimental accuracy indicators of the Papadakis method were more favorable when compared to those of the randomized block and triple lattice designs. Two replicates could be used when analyzing the data without reducing experimental accuracy: a randomized complete block design or the Papadakis method. In the third study, the productive performance, adaptability, and stability of modern soybean cultivars were evaluated in multi-environment assays. A total of 46 cultivars were evaluated in eight environments, in the adaptation micro-regions 102, 201, and 202, during the 2014/2015 harvest. Genotype × complex environment interactions occurred with changes in the ranking of cultivars between the sites. Among the genotypes evaluated, the cultivar NA 5909 RG, parental to the RILs in the genome-wide selection (GWS) assay, was considered to be among the genotypes with higher mean productivities, and it also showed high adaptability and stability. The fourth study had three objectives: to evaluate the accuracy of genomic selection in soybean, to identify the effect of intra-population structure on the accuracy of genomic selection, and to compare the efficiencies of the phenotypic and genomic selections in soybean. The BayesB model with cross validation was used for analyzing the phenotype data from the 324 soybean genotypes. The accuracy of GS for phenotypic characters with genotypic data of 5403 SNP molecular markers was also evaluated. The results indicated that the genotypic accuracy was similar, irrespective of consideration of the population structure. It was observed that the population structure did not significantly affect the accuracy of the models for the traits evaluated. It was verified that with this methodology it is possible to halve the selection time and increase the selection efficiency by 123% for grain yield.

Keywords: Agronomy. Plants - Genetic Improvement. Genetic Markers.

RESUMO

MATEI, Gilvani. Seleção genômica ampla em soja e otimização da fenotipagem para produtividade de grãos. 100 f. Tese (Doutorado em Agronomia) – Programa de Pós-Graduação em Agronomia (Área de Concentração: Produção vegetal), Universidade Tecnológica Federal do Paraná. Pato Branco, 2017.

Em um programa de melhoramento genético varios fatores influenciam na seleção de cultivares, basicamente pelo elevado número de genótipos em avaliação e pela reduzida capacidade experimental em fases iniciais do programa. Nesse contexto, o presente trabalho foi dividido em quatro partes. O primeiro estudo objetivou identificar locais chaves para avaliação e seleção de genótipos de soja nas nas macrorregiões 1 (M1) e 2 (M2), em gerações com pouca disponibilidade de semente. O conjunto de dados consistiu em 22 genótipos de soja cultivados em 23 locais por 3 anos. As análises GGL + GGE e G vs. GE foram usadas. As localidades Chapada-RS e Maracaju-MS foram os locais mais representativos e discriminantes macrorregiões 1 e 2, respectivamente. A identificação das localidades chave é fundamental para a avaliação, onde o número de locais de ensaio pode se resumir a um único local por macrorregião sojícola. O segundo estudo teve como objetivo avaliar a precisão experimental de diferentes métodos de análise estatística para ensaios com elevado número de genótipos de soja. Foram usados dados de produtividade de grãos de 324 genótipos de soja, avaliados em 6 repetições. Os dados foram analisados considerando os delineamentos de blocos ao acaso, látice triplo e uso do método de Papadakis. Os indicadores de precisão experimental do método de Papadakis são mais favoráveis, quando comparados com os delineamentos de blocos ao acaso e látice triplo. Pode-se usar duas repetições e analisar os dados, usando o delineamento de blocos completamente casualizados ou método Papadakis, sem redução da precisão experimental. No terceiro estudo foi avaliado o desempenho produtivo, a adaptabilidade e a estabilidade de cultivares modernas de soja, em ensaios multiambientes. Foram avaliados 46 cultivares em oito ambientes, nas microrregiões de adaptação 102, 201 e 202, na safra 2014/2015. Ocorreu interação genótipo x ambiente complexa, com alterações do ranqueamento de cultivares entre os locais. Dentre os genótipos avaliados a cultivar NA 5909 RG, parental das RILs no ensaio GWS, esteve presente entre genótipos de maiores médias produtivas, apresentando também elevada adaptabilidade e estabilidade. O quarto estudo teve três objetivos: avaliar a precisão da SG na soja; identificar o efeito da estrutura intrapopulação na precisão da seleção genômica; e, comparar a eficiência da seleção fenotípica e genômica na soja. Foi utilizado o modelo BayesB com validação cruzada para dados fenótipicos e genótipicos de 324 genótipos de soja. Avaliou-se a precisão do GS para caracteres fenotípicos com dados genotípicos de 5403 marcadores SNPs. Os resultados indicaram que a precisão genotípica foi semelhante, considerando, ou não, a estrutura da população. Se observou que a estrutura da população não afetou significativamente a precisão dos modelos para os caracteres avaliados. Constatou-se que com esta metodologia torna-se possível reduzir pela metade o tempo de seleção e aumentar a eficiência de seleção em 123% para produtividade de grãos.

Palavras-chave: Agronomia. Plantas - Melhoramento Genético. Marcadores Genéticos.

LIST OF ILLUSTRATIONS

Figure 2.1 – The GGL + GGE ((genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)) biplot for the macrorregion 1 based on 2012/13–2015/16 crop seasons (Data for the VCU trials). The environments are displayed as the relevant location code, and the genotypes are represented by "+". The environments are: Abelardo Luz (ABL), Chapada (CHA), Campos Novos (CPN), Erechim (ERC), Guarapuava (GUA), Palma Sola (PSO), Realeza (RLZ), São Francisco de Assis (SFA), Santa Cruz do Sul (STC), Santo Augusto (STO), Vacaria (VAC). 26

Figure 2.2 – The biplot form to display the environment's ability to select for G (Genotypic effect) *vs*. GE (genotype x environment effects) (A) and the environment vector view of the GGE biplot with linear map (B), based on 2012/13–2015/16 crop seasons. Data for the VCU trials in soybean macrorregion 1. The environment is composed of a location and a year (For example: STC_13 = Santa Cruz do Sul in the year 2013). The environments are: Abelardo Luz (ABL), Chapada (CHA), Campos Novos (CPN), Erechim (ERC), Guarapuava (GUA), Palma Sola (PSO), Realeza (RLZ), São Francisco de Assis (SFA), Santa Cruz do Sul (STC), Santo Augusto (STO), Vacaria (VAC). 26

Figure 2.3 – The GGL + GGE ((genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)) biplot for the macrorregion 2 based on 2012/13–2015/16 crop seasons (Data for the VCU trials). The environments are displayed as the relevant location code, and the genotypes are represented by "+". The environments are: Brasilândia do Sul (BRA), Cascavel (CCV), Cândido Mota (CMO), Dourados (DOU), Londrina (LON), Maracaju (MCJ), Palotina (PLT), Ponta Porã (PPO), Perobal (PRB), Sidrolândia (SID), São Jorge do Ivaí (SJI), Ubiratã (UBI)....27

Figure 2.4 – The biplot form to display the environment's ability to select for G (Genotypic effect) *vs*. GE (genotype x environment effects) (A) and the environment vector view of the GGE biplot with linear map (B), based on 2012/13–2015/16 crop seasons. Data for the VCU trials in soybean macroregion 2. The environment is composed of a location and a year (For example: CCV_13 = Cascavel, in the year 2013). The environments are: Brasilândia do Sul (BRA), Cascavel (CCV), Cândido Mota (CMO), Dourados (DOU), Londrina (LON), Maracaju (MCJ), Palotina (PLT), Ponta Porã (PPO), Perobal (PRB), Sidrolândia (SID), São Jorge do Ivaí (SJI), Ubiratã (UBI). 28

Figure 3.1 - Selective accuracy (SA), least significant difference (LSD) by the Tukey test (5%) in percentage of the mean, and variation in the Fasoulas differentiation index (FDI); mean, lower limit (LL), and upper limit (UL) of the confidence interval, by resampling (1-α=0.95) for different number of replicates, and using the randomized complete block design (RCBD) and Papadakis method. 39

Figure 4.1 – Mean and stability for the 46 soybean cultivars (A), and for the cultivars division in early (B), medium (C) and late cycle (D), assessed in eight locations (seven of which in the State of Paraná (Cambé, Corbélia, Mamborê, Palotina, Realeza, São Jorge do Ivaí e São Miguel do Iguaçu) and one in State of São Paulo (Cândido Mota)), in 2014/2015 crop season. PC: main component. Cultivars: BMX Potência RR (1), DMario 58i (2), NK 7059 RR (3), A 6411RG (4), BMX Ativa RR (5), BMX Energia RR (6), NA 5909 RG (7), NS 4823 (8), BMX Turbo RR (9), NS 5858 (10), NS 6262 (11), SYN1059 RR (12), NS 6767 (13), TMG 7262RR (14), NS 4901 (15), NS 5258 (16), NS 5290 (17), NS 5401 RR (18), NS 6209 (19), NS6121RR (20), NS6823RR (21), M6210IPRO (22), M6410IPRO (23), 5958RSF IPRO (24), 6458RSF IPRO (25), 6563RSF IPRO (26), AS 3570IPRO (27), AS 3610IPRO (28), M5917IPRO (29), NS 5000 IPRO (30), NS 5106 IPRO (31), NS 5151 IPRO (32), NS 5445 IPRO (33), NS 5959 IPRO (34), NS 6909 IPRO (35), NS 7000 IPRO (36), NS 7209 IPRO (37), NS 7237 IPRO (38), NS 7300 IPRO (39), NS 7338 IPRO (40), NS 5727 IPRO (41), NS 6006 IPRO (42), NS6060IPRO (43), NS6700IPRO (44), NS6906IPRO (45) e TMG2158IPRO (46). 51

Figure 4.2. Ideal genotype for the 46 soybean cultivars (A), and for the cultivars division in early (B),

medium (C) and late cycle (D), assessed in eight locations (seven of which in the State of Paraná (Cambé, Corbélia, Mamborê, Palotina, Realeza, São Jorge do Ivaí e São Miguel do Iguaçu) and one in State of São Paulo (Cândido Mota)), in 2014/2015 crop season. PC: main component. Cultivars: BMX Potência RR (1), DMario 58i (2), NK 7059 RR (3), A 6411RG (4), BMX Ativa RR (5), BMX Energia RR (6), NA 5909 RG (7), NS 4823 (8), BMX Turbo RR (9), NS 5858 (10), NS 6262 (11), SYN1059 RR (12), NS 6767 (13), TMG 7262RR (14), NS 4901 (15), NS 5258 (16), NS 5290 (17), NS 5401 RR (18), NS 6209 (19), NS6121RR (20), NS6823RR (21), M6210IPRO (22), M6410IPRO (23), 5958RSF IPRO (24), 6458RSF IPRO (25), 6563RSF IPRO (26), AS 3570IPRO (27), AS 3610IPRO (28), M5917IPRO (29), NS 5000 IPRO (30), NS 5106 IPRO (31), NS 5151 IPRO (32), NS 5445 IPRO (33), NS 5959 IPRO (34), NS 6909 IPRO (35), NS 7000 IPRO (36), NS 7209 IPRO (37), NS 7237 IPRO (38), NS 7300 IPRO (39), NS 7338 IPRO (40), NS 5727 IPRO (41), NS 6006 IPRO (42), NS6060IPRO (43), NS6700IPRO (44), NS6906IPRO (45) e TMG2158IPRO (46). 53

Figure 5.1 – The eight geographical locations in this study. CM, Cândido Mota; COR, Corbélia; MAB, Mamborê; PAL, Palotina; REA, Realeza; CAM, Cambé; SJI, São Jorge do Ivaí; SMI, São Miguel do Iguaçu. PR, Paraná State; SP, São Paulo State. 58

Figure 5.2 – Decay of LD (r2) with physical map distances between markers in the examined genotypes, as determined by locally weighted polynomial regression. 64

Figure 5.3 – Scatterplot of the two first eigenvalues in a principal coordinates analysis of 4947 SNPs in 324 Brazilian soybean genotypes clustered into RIL (G1, in green) and elite genotype (G2, in red) subpopulations. 64

Figure 5.4 – Dendrogram obtained by cluster analysis of data from 4947 SNP markers in 324 individuals (divided into two subgroups): G1 (243 RILs, in green) and G2 (81 elite genotypes, in red) included in the GS. 65

Figure 5.5 – eBLUPs predicted by a mixed linear model (lme4) for the analyzed traits. Datasets corresponding to RILs and elite genotypes are shown in green and red, respectively. 66

LIST OF TABLES

Table 2.1 – Locations of conduction of value of cultivation and use (VCU) trials in the macroregions of soybean 1 (M1) and 2 (M2). 20

Table 2.2 - Soybean genotypes and lineages tested in the Value of Cultivation and Use (VCU) trials in 2012/13 to 2015/16 crops seasons in the M1 and M2 macroregions. 21

Table 2.3 – Numerical values of the analysis G vs. GE for locations with the ability to select superior genotypes (AEC_X), stable (G) and unstable (GE) (AEC_Y), Vector length and representativeness of the environment (Correlation with AEA) for Macroregion 1. 23

Table 2.4 – Numerical values of the analysis G vs. GE for locations with the ability to select superior genotypes (AEC_X), stable (G) and unstable (GE) (AEC_Y), Vector length and representativeness of the environment (Correlation with AEA) for Macroregion 2. 24

Table 3.1 – Variability indicators obtained using the randomized complete block design (RCBD), Papadakis (Papa) method, and triple lattice for nine environments (E) in Paraná (PR) and São Paulo (SP), Brazil, in the 2014/15 harvest. 35

Table 3.2 - Variability indicators obtained using the randomized complete block design (RCBD; six blocks), Papadakis method (Papa; six blocks), and duplicate triple lattice design in soybean trials conducted in the 2014/15 harvest. 38

Table 4.1 – Description of 46 soybean cultivars, maturity group, cycle, year of release, technology and releaser. 42

Table 4.2 – Identification of test locations used to evaluate 46 soybean cultivars, in 2014/2015 crop season. 44

Table 4.3 – Estimation of genetic parameters for each of the eight locations and for the set of locations, for the trait grain yield (GY) of 46 soybean cultivars. 46

Table 4.4 – Grain yield (kg ha-1) of soybean cultivars, grouping means by the Scott-Knott test, and mean of cultivars in eight. sites GY (X G), mean of each location (X L), and mean of 46 soybean cultivars classifid according to their cycle, in each site, in the 2014/2015 crop season. 48

Table 4.5 – Genetic effects (g), predicted genotypic values (u + g), gain, new mean of the genotype, rank, average genotypic value in the environments (u + g + gem) and methods of adaptability and stability using mixed models. 49

Table 5.1 – Geographical and climatic information on the eight locations selected for evaluation of 324 soybean genotypes in the 2014/15 crop season. 58

Table 5.2 – Coefficients of genetic and environmental variances and heritability determined from restricted maximum likelihood model estimates. 66

Table 5.3 – Accuracy of GS analysis for five traits in soybean based on the BayesB model and eBLUP considering homogeneous or stratified population structures. 67

Table 5.4 – Deviance information criterion values obtained for GS analysis* 67

Table 5.5 – Accuracy of GS vs. phenotypic selection 68

LIST OF ACRONYMS AND ABBREVIATIONS

ABL	Abelardo Luz
alt	Altitude
AEA	Average environment axis
α	Bootstrap confidence interval estimate
BRA	Brasilândia do Sul
CPN	Campos Novos
CMO	Cândido Mota
CCV	Cascavel
CHA	Chapada
CV%	Coefficient of variation
CRD	Completely randomized design
HA-GGE	Data scaled by standard deviation and adjusted heritability
DM	Days to maturity
DOU	Dourados
eBLUPs	Empirical best linear unbiased prediction
h	Environment vector length
ERC	Erechim
VCe	Error variation coefficient
FDI	Fasoulas differentiation index
r_g	Genetic correlation
g	Genetic effects
VCg	Genetic variation coefficient
GB	Genomic breeding
GEBVs	Genomic breeding values
GS	Genomic selection
GGE	Genotype main effects plus genotype × environment interaction
GA	Genotypic accuracy
CVgi%	Genotypic coefficient of variation
G	Genotypic effect
GE	Genotypic interaction
V_G	Genotypic variance
GY	Grain yield
GUA	Guarapuava
HMGV	Harmonic mean of genotypic values
R8	Harvest maturity
ha	Hectare

h^2_g	Heritability in the broad sense
IPF	Insertion of the first pod
lat	Latitude
LSD	Least significant difference
LD	Linkage disequilibrium
GGL	Location interaction
GL	Location interaction
LON	Londrina
long	Longitude
M1	Macroregions 1
M2	Macroregions 2
MCJ	Maracaju
MAS	Marker-assisted selection
MG	Maturity groups
m	Metros
NPK	Nitrogen:phosphorus:potassium
PSO	Palma Sola
PLT	Palotina
Papa	Papadakis
PR	Paraná State
PRB	Perobal
PA	Phenotypic accuracy
PB	Phenotypic breeding
V_F	Phenotypic variance
PH	Plant height
PH	Plant height
PPO	Ponta Porã
RPGV	Predicted genotypic values
u + g	Predicted genotypic values
PCs	Principal components
i.e.	Progenies from a single plant
QTLs	Quantitative trait loci
kg	Quilograma
RCBD	Randomized complete block design
RLZ	Realeza
RILs	Recombinant inbreed lines
RKHS	Reproducing kernel Hilbert space
CVe%	Residual coefficient of variation
STC	Santa Cruz do Sul

STO	Santo Augusto
SFA	São Fco. de Assis
SJI	São Jorge do Ivaí
SP	São Paulo State
SA	Selective accuracy
SID	Sidrolândia
SNPs	Single Nucleotide Polymorphisms
SVP	Singular value partition
TGW	Thousand grain weight
UBI	Ubiratã
VAC	Vacaria
VCU	Value of cultivation and use
$V_{G\times A}$	Value of interaction variance
VC	Variation coefficient

SUMMARY

1 INTRODUCTION 17

2 IDENTIFICATION OF CORE LOCATIONS FOR SOYBEAN BREEDING IN SOUTHERN BRAZIL 19

2.1 Literature Review of Core Locations for Soybean Breeding in Southern Brazil ... 19

2.2 Materials and Methods Utilized to Identification of Core Locations for Soybean Breeding in Southern Brazil 20

2.3 Results about Identification of Core Locations for Soybean Breeding in Southern Brazil 23

2.4 Discussion about Identification of Core Locations for Soybean Breeding in Southern Brazil 29

2.5 Conclusion about Identification of Core Locations for Soybean Breeding in Southern Brazil 31

3 METHODS OF ANALYSIS AND NUMBER OF REPLICATES FOR TRIALS WITH LARGE NUMBERS OF SOYBEAN GENOTYPES 33

3.1 Literature Review of Methods of analysis and number of replicates for trials with large numbers of soybean genotypes 33

3.2 Materials and Methods of analysis and number of replicates for trials with large numbers of soybean genotypes 34

3.3 Results and Discussion about Methods of analysis and number of replicates for trials with large numbers of soybean genotypes.....36

3.4 Conclusion about Methods of analysis and number of replicates for trials with large numbers of soybean genotypes 41

4 AGRONOMIC PERFORMANCE OF MODERN SOYBEAN CULTIVARS IN MULTI-ENVIRONMENT TRIALS.....42

4.1 Literature Review of Agronomic Performance of Modern Moybean Cultivars in Multi-environment Trials 42

4.2 Materials and Methods of Agronomic Performance of Modern Soybean Cultivars in Multi-environment Trials 43

4.3 Results and Discussion of Agronomic Performance of Modern Soybean Cultivars in Multi-environment Trials 46

4.4 Conclusion of Agronomic Performance of Modern Soybean Cultivars in Multi-Environment Trials 55

5 GENOMIC SELECTION IN SOYBEAN: ACCURACY AND TIME GAIN IN RELATION TO PHENOTYPIC SELECTION 56

5.1 Introduction of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection.....56

5.2 Materials and Methods of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection 58

5.3 Results of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection.....64

5.4 Discussion of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection.....70

5.4 Conclusions of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection.....73

6 CONCLUSIONS.....74

REFERENCES.....76

GLOSSÁRIO.....87

1 INTRODUCTION

The constantly growing global demand for soybeans necessitates increasing the country's productivity. Selective plant breeding based on genetics has made significant advances in this direction possible, more specifically in the last few decades. However, breeding selection is practiced based on phenotypic features, often obtained through estimation methods that consider only the phenotype, which results in the narrowing of the genetic base and consequently, in the limitation of the genetic gains in successive breeding cycles.

In a genetic breeding program, several factors influence the selection of the ideal phenotype, mainly due to the high number of genotypes and the reduced experimental capacity in the initial phases of the program. In addition, to selecting a superior phenotype, it is necessary to consider the representativeness of the target selection environment, the efficiency of planning and experimentation used in the genotype evaluation assays, and the suitability of the selected target genotype group and its performance in multi-environments.

Choosing the best environment for the selection of superior genotypes has direct implications on the effectiveness of the breeding program, mainly due to the limitations of early seed availability and the proportion of lineages in these stages. As such, it is essential to conduct tests in a location representative of the target environment to maximize the genetic gains in the next generations.

Selection of the best genotype is associated with obtaining phenotypic information from the replicates sown in a specific place, which, once statistically analyzed, allows to improve the tests and the efficiency in identifying the most promising genotypes. However, the accuracy of the assays and the number of replicates suitable for trials with large numbers of soybean genotypes are still unknown.

The multi-environmental performance is what determines if a group of selected genotypes shows adaptability and production stability. That is, a genotype’s ability to respond predictably to environmental stimuli and the predictability in its performance in different environments determine its suitability and superiority, and based on these characteristics the selection of phenotypes with superior genotypes is

guaranteed.

However, the pace and proportion with which the breeding programs have been developing in the recent decades, has made it increasingly difficult to identify suitable genotypes early, simultaneously with increasing the selection gains and reducing the gap between segregating generations. In this context, the use of molecular markers associated with the selection and prediction of genetic values for grain yields stands out as an important tool in the selection process.

Genome-wide selection (GWS) has become an important tool in the early selection of superior genotypes, and it basically involves the simultaneous prediction of genetic effects based on a large number of genetic markers scattered throughout the genome. In this way, it seeks to capture the effects, both small and large, of all loci, and to explain much of the genetic variations of a quantitative character based on specific molecular markers.

In this context, the objective of this work was to develop a model for broad genomic selection through the use of SNP molecular markers, aiming to contribute to the early selection of superior lineages in soybean breeding programs. For this purpose, the main selection environments for the initial stages of the crop breeding program were selected. In addition, the number of ideal replicates and the efficiency of statistical methodologies applied for the selection of a large number of genotypes were evaluated. Finally, the soybean genotypes ideal for cultivation in eight growing environments in Brazil were selected through multi-environmental trials, aiming for new studies and validation of the developed model.

2 IDENTIFICATION OF CORE LOCATIONS FOR SOYBEAN BREEDING IN SOUTHERN BRAZIL

2.1 Literature Review of Core Locations for Soybean Breeding in Southern Brazil

Hybridization among homozygous soybean genotypes is performed with the aim of recombining alleles for increasing genetic variability. Subsequently, several succeeding generations of progeny are evaluated and selected until a new cultivar is developed and released. However, in generations with low seed availability, such as in segregating or advanced populations, and recently selected lines (i.e., progenies from a single plant), soybean breeders are often faced with a dilemma while choosing the best environment for the selection of superior genotypes. This is an important issue owing to limitations in seed availability in early generations. Moreover, these trials are usually conducted in few locations, and in many cases, at a single location. This has direct implications on the effectiveness of the breeding program.

It is essential to conduct trials at a location that is representative of the target environment of selection. Moreover, this representativeness must be consistent over the years (DIA, 2016b; YAN, 2016). Furthermore, location and selection must be efficient in differentiating superior genotypes (QIN et al., 2015; KRISHNAMURTHY et al., 2017). When the location combines both characteristics it is named as a core location (YAN, 2014).

The GGE (genotype main effects plus genotype × environment interaction) biplot analysis has been widely used for evaluating the suitability of test locations for numerous crops, e.g., watermelon (DIA et al., 2016a), spring durum wheat (KARIMIZADEH et al., 2016), rice (KRISHNAMURTHY et al., 2017), soybean (QIN et al., 2015), and oat (YAN, 2010). However, traditional methods use individual year-to-year analysis. This method produces results that are hard to interpret; thus, it may not be efficient if a pattern observed over years is not clear-cut. A new method designated as GGL + GGE (YAN, 2014, 2015) overcame this obstacle. In this biplot, data are summarized in a single biplot point, which is calculated from the mean of the two first principal components (PCs). This enables the establishment of a pattern among years. Thus, the GGL + GGE biplot is the most appropriate for comparing test

locations based on multi-year data (YAN, 2014). This analysis enables decision-making with more certainty regarding the representativeness of a location as compared with that of the target environment and its potential for the selection of superior genotypes. Therefore, this study aimed to identify core locations in soybean M1 and M2 in southern Brazil.

2.2 Materials and Methods Utilized to Identification of Core Locations for Soybean Breeding in Southern Brazil

Data from Value of Cultivation and Use (VCU) trials from the 2012–13, 2013–14, 2014–15, and 2015–16 crop seasons were used. Trials were conducted in 23 locations, 11 of which are in M1 and 12 in M2 macroregions (Table 2.1). Twenty-two genotypes were tested, including eight cultivars and 14 lineages (Table 2.2).

Experiments were conducted using a completely randomized block design with three replications. Plots consisted of four rows with a length of 5 m, and row spacing and plot spacing of 0.5 m. Sowing density was standardized for all genotypes at 30 seeds m^{-2}. Basis fertilization consisted of N-P-K mineral fertilizer with 7 kg N ha^{-1}, 70 kg P_2O_5 ha^{-1}, and 70 kg K_2O ha^{-1}. Harvest was performed using a plot combine when plants reached harvest maturity (R8). Both central rows of each plot (5 m^2) were harvested, and seed moisture content was routinely adjusted to 13%.

Statistical analysis for identification of core locations were performed using the GGEbiplot software (YAN, 2001). GGL + GGE analysis was used to identify core location in each of the two previously defined macroregions. In this analysis, location in the biplot was defined by the mean of both PC1 and PC2 over the tested years. When the data were scaled using HA-GGE (data scaled by standard deviation and adjusted heritability), the cosine of the angle between the vector of the location and the Average environment axis (AEA) line represents a good estimate of the genetic correlation (r_g) between the two (YAN and holland, 2010; YAN, 2014). Thus, the smaller the angle between the location and the mean environment, the more representative is the location. The vector length indicates the consistency of the results recorded over the years, i.e., its representativeness. When the biplot explanation is high, the vector length is proportional to the squared root of the

heritability (*h*) (YAN, 2014).

Table 2.1 – Locations of conduction of value of cultivation and use (VCU) trials in the macroregions of soybean 1 (M1) and 2 (M2).

Location	Code	Region	Crop season			
			2012	2013	2014	2015
Abelardo Luz	ABL	M1	X	X	X	X
Brasilândia do Sul	BRA	M2			X	X
Cascavel	CCV	M2	X	X	X	X
Chapada	CHA	M1	X	X	X	X
Cândido Mota	CMO	M2	X	X	X	
Campos Novos	CPN	M1	X		X	X
Dourados	DOU	M2	X	X	X	X
Erechim	ERC	M1	X	X		
Guarapuava	GUA	M1		X	X	X
Londrina	LON	M2	X	X	X	X
Maracaju	MCJ	M2	X	X		X
Palotina	PLT	M2	X	X	X	X
Ponta Porã	PPO	M2	X		X	X
Perobal	PRB	M2	X		X	X
Palma Sola	PSO	M1	X	X	X	
Realeza	RLZ	M1	X	X	X	X
São Fco. de Assis	SFA	M1			X	X
Sidrolândia	SID	M2	X		X	X
São Jorge do Ivaí	SJI	M2			X	X
Santa Cruz do Sul	STC	M1	X	X	X	X
Santo Augusto	STO	M1	X	X		X
Ubiratã	UBI	M2		X	X	X
Vacaria	VAC	M1	X	X		X

The G vs. GE analysis allows for inferences about the environmental capability (environment = location + year) for the selection of superior genotypes. In this analysis, the more farther to the right of the biplot the environment is, the higher the environmental potential for selection of superior genotypes. The vector length towards the double-arrowed line, average environment coordination (AEC), permits identification of the environmental potential to select genotypes through the genotypic effect (G) or interaction (GE genotype x environment effects, i.e., instability effect). In this case, the shorter the vector, the higher the potential of the environment to select genotypes through G, and the longer the vector (more distant from the AEA line in order to the AEC line) the higher its potential for selection of genotypes through GE, indicating genotype instability. Therefore, environments to the right side of the biplot

with short vectors are desirable, whereas environments to the left side with long vectors are not suitable. Environments to the left of the AEC line are inefficient for any type of selection and environments to the right of the AEC line, even with long vectors towards or close to the AEC line, are somewhat useful, e.g., for selecting unstable but not superior genotypes (YAN, 2014).

Table 2.2 - Soybean genotypes and lineages tested in the Value of Cultivation and Use (VCU) trials in 2012/13 to 2015/16 crops seasons in the M1 and M2 macroregions.

Genotype	Specification	Crop season			
		2012/13	2013/14	2014/15	2015/16
A 4724RG	RG	X	X		
BMX ENERGIA RR	RG	X	X		
Dmario 58i	RG	X	X	X	X
NA 5909 RG	RG	X	X	X	X
BMX TURBO RR	RG	X	X	X	X
NS 6262	RG	X	X	X	X
BMX Potência RR	RG	X	X	X	X
NK 7059 RR	RG	X	X	X	X
NS L01	L	X	X		
NS L02	L	X	X		
NS L06	L	X	X		
NS L07	L	X	X	X	X
NS L08	L	X	X		
NS L11	L	X	X		
NS L12	L	X	X	X	X
NS L13	L	X	X		
NS L14	L	X	X	X	X
NS L23	L			X	X
NS L24	L			X	X
NS L25	L			X	X
NSL26	L			X	X
NSL27	L			X	X

L = Line; RG = Released Genotype.

Vector analysis of the environments allows the construction of the environmental linear map (right side of the biplot). This bar indicates the distance between environments, where proximity among environments indicates positive genetic correlation among them (YAN, 2014). Furthermore, the position of environments on the linear map makes it possible to identify patterns in genotype and location interaction (GL) and GE data. Thus, if environments are placed on the linear map mainly by location and not year, dominance will be established by GL over GE.

However, if environments are mainly placed by year instead of location, dominance will be established by GE over GL (YAN, 2014).

For all analysis, the following parameters were used: data transformation (Transform = 0, no transformation); data scaling [Scaling = 2, data scaled by standard deviation (SD-scaled), and adjusted heritability (*h*-weighted)]; data centering [Data centering = 2, genotype + genotype environment interaction (G + GE), and singular value partition (SVP) = 2, focus on environment].

2.3 Results about Identification of Core Locations for Soybean Breeding in Southern Brazil

The GGL + GGE analysis revealed that the most representative test location (of the target environment, i.e., macroregion) is highly correlated with the mean environment, and that the results must be consistent over years (long vector). Furthermore, it is highly desirable that it allows genotype differentiation in the G vs. GE analysis and selection of superior genotypes mainly through the genotypic effect (G).

In the G vs. GE analysis based on various environments, the interpretation of the biplot may be limited by location overlapping. However, results that are more intelligible are presented in Tables 2.3 and 2.4, for M1 and M2, respectively. Data in these tables are $r_g h$ values (AEC_X), which is an index for evaluating test environments that is defined by the ratio between r_g (close to the cosine of the angle between environment and AEA) and *h* (environment vector length). This index is integrated for environmental evaluations and is useful even with a low explanation of the biplot (Yan, 2014). Therefore, the higher the AEC_X value, the higher is the environmental potential for selection of superior genotypes and the desirability of the environment. The AEC_Y column refers to the potential of a location for selection of genotypes through G or GE. The higher the modular AEC_Y value, the lower the potential of the location for selecting genotypes by G, and higher is the possibility of selecting based on the GE interactions. Thus, environments characterized by AEC_Y values between 0.30 and −0.30 were considered adequate for selecting genotypes through G, whereas environments with AEC_Y above these

values were considered adequate for selecting genotypes based on GE. In summary, a most desirable environment will be one characterized by a high AEC_X value and a close-to-zero AEC_Y value.

Table 2.3 – Numerical values of the analysis G vs. GE for locations with the ability to select superior genotypes (AEC_X), stable (G) and unstable (GE) (AEC_Y), Vector length and representativeness of the environment (Correlation with AEA) for Macroregion 1.

Tester	AEC_X	AEC_Y	Vector Length	Correlation with AEA
CPN_14	1.449	0.373	1.496	0.968
GUA_14	1.327	0.644	1.475	0.900
PSO_14	1.289	-0.370	1.341	0.961
ABL_14	1.252	-0.459	1.333	0.939
VAC_12	1.240	0.105	1.244	0.996
STC_13	1.160	0.703	1.356	0.855
VAC_15	1.155	0.112	1.160	0.995
CHA_15	1.145	0.602	1.294	0.885
ERC_13	1.143	0.816	1.404	0.814
RLZ_14	1.049	-0.352	1.106	0.948
CPN_15	1.047	-0.106	1.052	0.995
CPN_12	1.033	1.029	1.458	0.708
STO_13	0.995	0.673	1.201	0.828
GUA_15	0.981	-0.74	1.229	0.798
ERC_12	0.980	0.501	1.101	0.890
CHA_13	0.919	-0.264	0.956	0.961
ABL_15	0.917	-0.387	0.995	0.921
STO_12	0.830	1.147	1.416	0.586
CHA_12	0.783	0.788	1.111	0.705
PSO_12	0.766	0.940	1.212	0.632
STO_15	0.756	-0.201	0.782	0.966
STC_14	0.734	-0.415	0.843	0.870
ABL_13	0.684	-1.037	1.243	0.551
STC_12	0.641	1.305	1.454	0.441
RLZ_12	0.570	0.050	0.573	0.996
PSO_13	0.545	-0.935	1.082	0.503
ABL_12	0.507	0.877	1.013	0.501
RLZ_15	0.437	-0.642	0.776	0.563
CHA_14	0.435	-1.258	1.331	0.327
SFA_14	0.422	-1.112	1.189	0.355
VAC_13	0.317	-1.186	1.227	0.258
RLZ_13	0.308	0.109	0.327	0.943
SFA_15	0.134	-0.194	0.236	0.568
GUA_13	-0.184	-0.969	0.986	-0.187
STC_15	-1.040	-0.149	1.051	-0.990

A Tester is composed of a location and a year (For example: CPN_14 = Campos Novos, in the year

2014). The environments are: Abelardo Luz (ABL), Chapada (CHA), Campos Novos (CPN), Erechim (ERC), Guarapuava (GUA), Palma Sola (PSO), Realeza (RLZ), São Francisco de Assis (SFA), Santa Cruz do Sul (STC), Santo Augusto (STO), Vacaria (VAC).

Table 2.4 – Numerical values of the analysis G vs. GE for locations with the ability to select superior genotypes (AEC_X), stable (G) and unstable (GE) (AEC_Y), Vector length and representativeness of the environment (Correlation with AEA) for Macroregion 2.

Tester	AEC_X	AEC_Y	Vector Length	Correlation with AEA
MCJ_13	1.779	0.263	1.799	0.989
MCJ_12	1.643	0.517	1.722	0.954
LON_12	1.611	-0.234	1.628	0.990
PRB_12	1.373	0.006	1.373	1.000
DOU_12	1.360	0.359	1.407	0.967
CMO_13	1.316	-0.151	1.325	0.993
LON_13	1.307	-0.335	1.349	0.969
MCJ_15	1.133	-0.326	1.179	0.961
PPO_15	1.133	-0.326	1.179	0.961
CCV_12	1.091	-0.924	1.429	0.763
DOU_13	0.993	1.090	1.474	0.674
SID_15	0.973	-0.076	0.976	0.997
SID_12	0.937	0.790	1.225	0.765
PPO_12	0.918	-0.542	1.066	0.861
CMO_12	0.835	-1.197	1.460	0.572
UBI_15	0.793	0.752	1.093	0.726
PLT_15	0.762	0.469	0.895	0.852
DOU_15	0.678	0.349	0.763	0.889
PLT_12	0.655	-1.466	1.606	0.408
PLT_13	0.655	-1.466	1.606	0.408
CCV_14	0.642	1.070	1.248	0.514
SJI_15	0.560	0.360	0.665	0.841
CCV_15	0.446	0.290	0.532	0.838
DOU_14	0.420	-0.014	0.421	0.999
BRA_15	0.417	0.996	1.079	0.386
LON_15	0.272	0.198	0.336	0.809
LON_14	0.237	-0.936	0.966	0.246
PRB_15	0.224	-0.313	0.384	0.582
UBI_13	0.186	-0.487	0.522	0.356
CMO_14	-0.098	1.358	1.361	-0.072
BRA_14	-0.108	0.609	0.619	-0.174
SJI_14	-0.234	-0.624	0.666	-0.351
PPO_14	-0.253	0.163	0.301	-0.840
UBI_14	-0.274	-0.816	0.861	-0.319
SID_14	-0.335	-0.495	0.598	-0.56
PLT_14	-0.379	1.065	1.130	-0.335
PRB_14	-0.542	1.061	1.191	-0.455
CCV_13	-0.707	-1.036	1.254	-0.564

A Tester is composed of a location and a year (For example: MCJ_13 = Maracaju, in the year 2013). The environments are: Brasilândia do Sul (BRA), Cascavel (CCV), Cândido Mota (CMO), Dourados (DOU), Londrina (LON), Maracaju (MCJ), Palotina (PLT), Ponta Porã (PPO), Perobal (PRB), Sidrolândia (SID), São Jorge do Ivaí (SJI), Ubiratã (UBI).

The GGL + GGE analysis indicated that CHA and PSO were the most representative locations in M1. Both locations had high r_g with elevated mean environment and genetic correlation between them (Figure 2.1). The G vs. GE analysis revealed that environment CPN_14 presented the highest potential for selection of superior genotypes (highest $r_g h$) (Figure 2.2A, Table 2.3). However, CPN is less representative of M1, as compared with CHA and PSO (Figure 2.1). The CHA environments were placed to the right of the AEC line across years, which indicate their effectiveness as sites for selecting superior genotypes. The best performance among these locations occurred in the environment CHA_15. However, CHA_14 exceeded optimal vector length, which indicates that it would not be an effective site for selecting genotypes through G. Regarding PSO environments, PSO_14 showed the third highest potential for selecting superior genotypes. Linear correlation analysis (linear mapping), showed that environments CHA_12 and CHA_15 were highly correlated and close on the linear map, indicating that these two environments were similar over the years (Figure 2.2B). However, most environments were mainly grouped by year and not by location, which indicates that GE dominates over GL.

Environments SFA and STC are the less representative of M1, with lower genetic association with the mean environment and short vectors in the GGL + GGE analysis (Figure 2.1). Furthermore, STC_15 and GUA_13 showed negative AEC_X values, which indicate negligible potential for selecting superior genotypes. Therefore, CHA, PSO, and CNP are the best candidates for a core location within M1, but CHA is more representative of M1 than the other locations are. Therefore, CHA can be considered the core location for this macroregion.

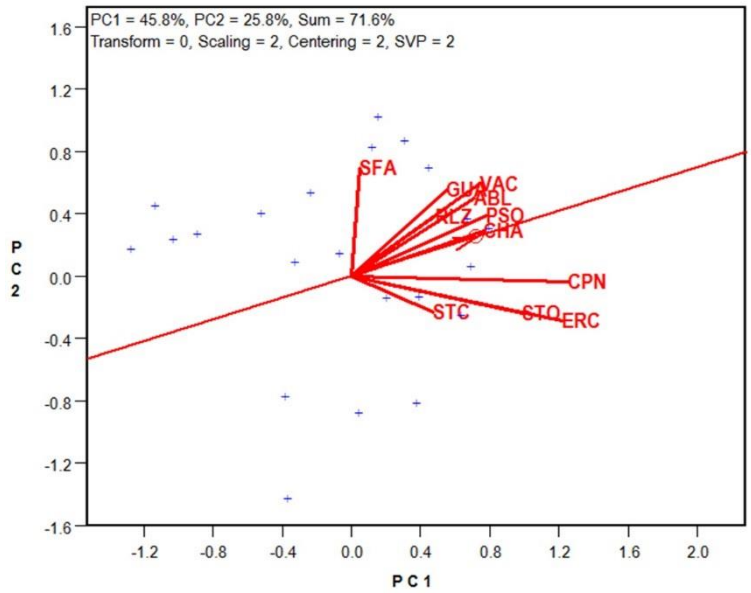


Figure 2.1 – The GGL + GGE ((genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)) biplot for the macrorregion 1 based on 2012/13–2015/16 crop seasons (Data for the VCU trials). The environments are displayed as the relevant location code, and the genotypes are represented by "+". The environments are: Abelardo Luz (ABL), Chapada (CHA), Campos Novos (CPN), Erechim (ERC), Guarapuava (GUA), Palma Sola (PSO), Realeza (RLZ), São Francisco de Assis (SFA), Santa Cruz do Sul (STC), Santo Augusto (STO), Vacaria (VAC).

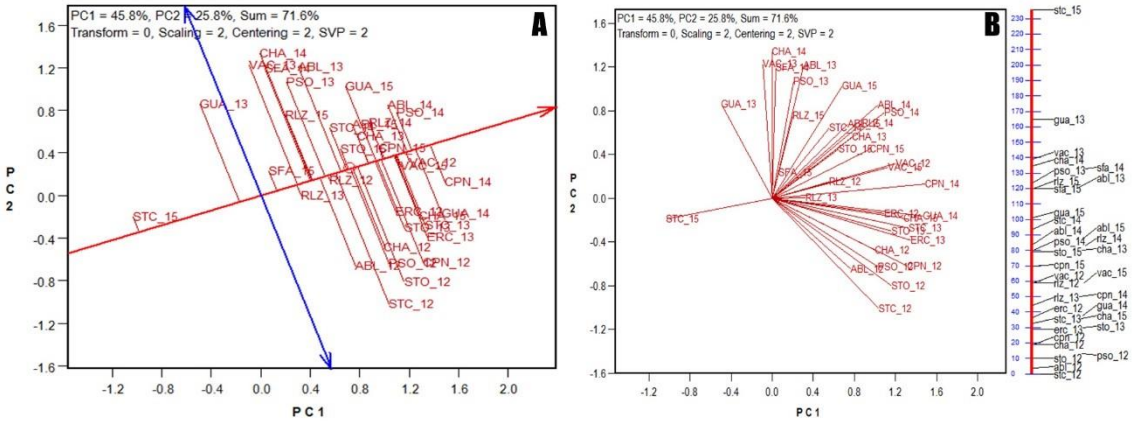


Figure 2.2 – The biplot form to display the environment's ability to select for G (Genotypic effect) *vs*. GE (genotype x environment effects) (A) and the environment vector view of the GGE biplot with linear map (B), based on 2012/13–2015/16 crop seasons. Data for the VCU trials in soybean macrorregion 1. The environment is composed of a location and a year (For example: STC_13 = Santa Cruz do Sul in the year 2013). The environments are: Abelardo Luz (ABL), Chapada (CHA), Campos Novos (CPN), Erechim (ERC), Guarapuava (GUA), Palma Sola (PSO), Realeza (RLZ), São Francisco de Assis (SFA), Santa Cruz do Sul (STC), Santo Augusto (STO), Vacaria (VAC).

In M2, the GGL + GGE analysis indicated that MCJ is the most representative location (highest r_g) (Figure 2.3). Furthermore, MCJ_13 and MCJ_12 are the environments with the highest potential for identifying superior genotypes,

since they are located to the right of the AEA line (higher $r_g h$). In addition, selection in MCJ_13 seems to be mostly dependent on G, because the environmental vector is shorter towards AEC than towards AEA (Figure 2.4A). Furthermore, the linear map indicated that MCJ_12, MCJ_13, and MCJ_15 are closely related, which showed high r_g among environments. Moreover, the GL effect partially dominates over the GE effect at these environments (Figure 2.4B). The G vs. GE analysis allowed the identification of CCV_13, PRB_14, PLT_14, SID_14, UBI_14, PPO_14, SJI_14, BRA_14, and SJI_14 as environments showing negative AEC_X values and located on the left of the AEC line (Figure 2.4A).

Furthermore, PLT, UBI, CCV, SJI, SDI, and PRB showed short vectors in the GGL + GGE analysis, which indicates inconsistency of results over the years and low heritability. In addition, it was observed that environments BRA and PLT are scarcely representative, whereas PLT, UBI, CCV, SJI, SID, and PRB present short vectors with inconsistent results over the tested years. Therefore, MCJ is the location most suitable for designation as a core location in M2 for evaluating segregating populations. In addition, MCJ is representative (high r_g and mean environment) and characterized by a long vector, which indicates consistent results over years. Furthermore, environments MCJ_12 and MCJ_13 exhibited the highest potential as sites where selection of superior genotypes can be conducted most efficiently.

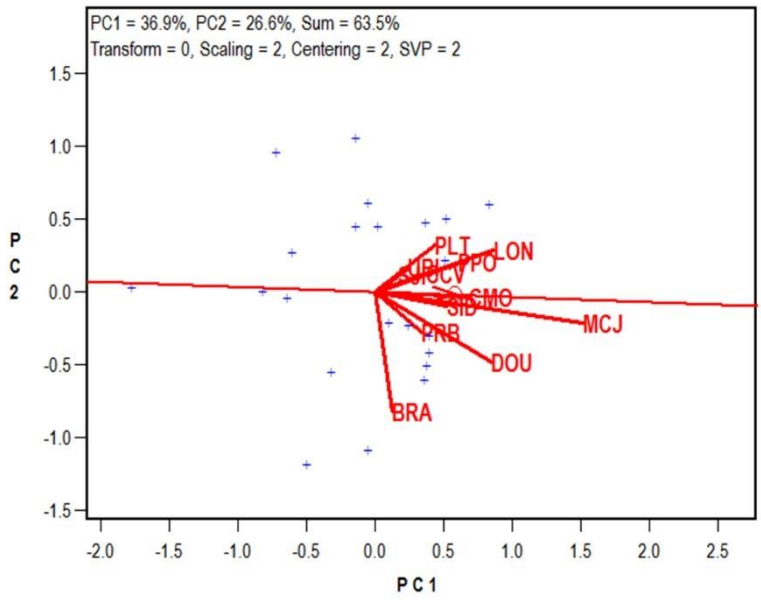


Figure 2.3 – The GGL + GGE ((genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)) biplot for the

macrorregion 2 based on 2012/13–2015/16 crop seasons (Data for the VCU trials). The environments are displayed as the relevant location code, and the genotypes are represented by "+". The environments are: Brasilândia do Sul (BRA), Cascavel (CCV), Cândido Mota (CMO), Dourados (DOU), Londrina (LON), Maracaju (MCJ), Palotina (PLT), Ponta Porã (PPO), Perobal (PRB), Sidrolândia (SID), São Jorge do Ivaí (SJI), Ubiratã (UBI).

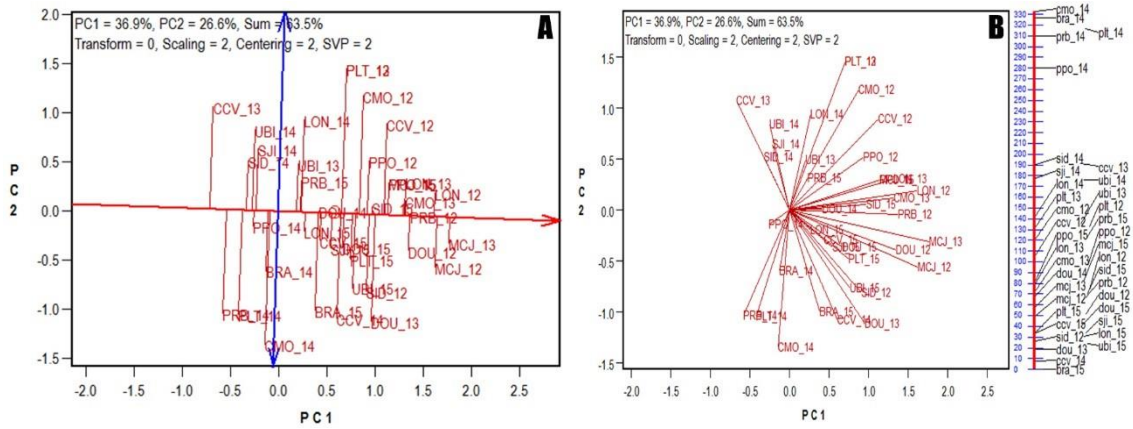


Figure 2.4 – The biplot form to display the environment's ability to select for G (Genotypic effect) *vs*. GE (genotype x environment effects) (A) and the environment vector view of the GGE biplot with linear map (B), based on 2012/13–2015/16 crop seasons. Data for the VCU trials in soybean macroregion 2. The environment is composed of a location and a year (For example: CCV_13 = Cascavel, in the year 2013). The environments are: Brasilândia do Sul (BRA), Cascavel (CCV), Cândido Mota (CMO), Dourados (DOU), Londrina (LON), Maracaju (MCJ), Palotina (PLT), Ponta Porã (PPO), Perobal (PRB), Sidrolândia (SID), São Jorge do Ivaí (SJI), Ubiratã (UBI).

2.4 Discussion about Identification of Core Locations for Soybean Breeding in Southern Brazil

The South of Brazil, specifically in M1, is characterized by higher altitude and mesothermic climate without dry seasons, as are Cfa or Cfb (temperatures in the coldest month range between −3 and 18 ºC). M1 consists of the Rio Grande do Sul and Santa Catarina states, in addition to the mid-southern and southeastern regions of the Paraná state, and the southern São Paulo state in Brazil. In contrast, M2, which consists of the western and northern Paraná, southern and western São Paulo, and southern Mato Grosso do Sul states, is considered a transitional region between Cfa and Cwa climates, with higher predominant temperatures in summer and dry winters (KASTER and FARIAS, 2012).

Segregating populations and recently selected lines had low seed availability. In this sense, it is necessary to identify the best test location, one that shows consistency in results across years and a sufficiently high genetic correlation

(r_g) with the target environment to make it representative (DIA et al., 2016b; YAN, 2016). In summary, locations CHA and MCJ were found to be the closest to a core location for the soybean-producing macroregions M1 and M2, respectively.

In addition to a high representativeness and consistent results across years, a core location must exhibit high potential for genotype differentiation (QIN et al., 2015). Moreover, a core location must allow selection of genotypes mainly by the genotypic effect (G) and not by instability (GE). Therefore, selection will be performed considering inherent characteristics in each genotype subject to as low as possible environmental effects. However, selection based on GE is interesting when focusing on discarding unstable genotypes. However, G selection is more attractive for breeding purposes, because superior genotypes can be selected with lower environmental interaction. Thus, the results obtained in core locations will be more reproducible at other locations, because the selection is mainly performed based on G. Accordingly, CHA and MCJ are appropriate locations; moreover, MCJ showed results that particularly enable superior genotype selection, mainly through genotypic effect in M2. Therefore, core locations CHA and MCJ are desirable environments in most years, according to the G vs. GE analysis. Environments with short vectors towards AEC are desirable for selecting superior genotypes through G (YAN, 2014), as observed in environments CHA_15, CHA_13, MCJ_12, and MCJ_13. However, locations to the left side of the AEC line present negative $r_g h$ and are inadequate for selecting superior genotypes (YAN, 2014).

In addition to the identification the best test locations, the possibility of identifying inappropriate locations for trials is another important characteristic of the GGE biplot (KRISHNAMURTHY et al., 2017). In this respect, environments on the left side of the AEC line on G vs. GE analysis may be replaced by other environments with higher potential for selecting superior genotypes, because they are not effective for selection, and therefore undesirable (YAN, 2014). Furthermore, non-representative locations with short vectors (inconsistent results) in the GGL + GGE analysis may be replaced. Our observations showed that SFA and STC in M1, and BRA, PLT, UBI, CCV, SJI, SID, and PRB in M2, had short vectors and an elevated angle (low r_g with the mean environment); thus, they do not represent the target environment (GGL + GGE analysis). Moreover, results at all these locations were inconsistent across

years. Consequently, they cannot be designated as core locations (YAN, 2015).

In the relationship among testers analysis, a linear map is built to facilitate identification of similarity among environments. Based on this analysis, it is possible to identify the patterns of, where GL represents genotype × location interaction, GY represents genotype × year interaction, and GLY represents genotype × location × year interaction (YAN, 2016). Thus, the GE interaction increases when the year effect is high.

In both macroregions, the environments effects are mainly ranked by year and not by location. Thus, a higher correlation is observed because of GE owing to the ranking of environments by year and not by location. This occurs because the year effect is naturally randomized and unpredictable; thus, GY and GLY are not reproducible (YAN, 2016). Minimizing environmental effects is essential for plant breeding, because environmental control decreases the effect of GE over GL in addition to enabling higher heritability and experimental precision. This was observed for MCJ: the environments at this location (especially, MCJ_12, MCJ_13, and MCJ_15) remained close on the linear map. A portion of GL may repeat with a decreased year effect, which is related to location effects such as sowing date, plant density, irrigation, soil fertility, and other crop management factors. These characteristics minimize the effects of interactions involving the year factor (YAN, 2016). Thus, in addition to choosing the most representative location with the highest potential for selecting superior genotypes, experimental area homogeneity and an efficient crop management practice are essential requisites of any core location, because the success of any breeding effort greatly depends on them.

Identifying the best location for selecting segregating populations and recently selected lines in a macroregion are essential, because limited seed availability is a common feature in these cases. Seed availability limitation usually limits the number of trial locations to a single one. Therefore, conducting experiments in the location that combines the highest representativeness and the highest potential for selecting superior genotypes — mainly through the genotypic effects — is essential for the success in evaluating and selecting new soybean cultivars.

2.5 Conclusion about Identification of Core Locations for Soybean Breeding in

Southern Brazil

Environments Chapada - RS and Maracaju - MS best approximate the definition of a core location in macroregion 1 and 2, respectively.

Locations Santa Cruz do Sul - RS and São Francisco de Assis - RS, in macroregion 1 and Perobal - PR, Brasilândia do Sul - PR, Ubiratã - PR, Palotina - PR, São Jorge do Ivaí - PR, Cascavel - PR and Sidrolândia - MS, in macroregion 2, are inadequate and should not be considered while designating core locations.

The identification of core locations is crucial when the availability of seeds is low because trials are normally conducted in one or few locations simultaneously in each macroregion.

3 METHODS OF ANALYSIS AND NUMBER OF REPLICATES FOR TRIALS WITH LARGE NUMBERS OF SOYBEAN GENOTYPES

3.1 Literature Review of Methods of analysis and number of replicates for trials with large numbers of soybean genotypes

Planning for genotype evaluation trials is a very important activity for qualification of results in genetic breeding programs. This activity is even more important when the number of entries (genotypes) is large, as usually occurs in breeding programs. For these cases, most appropriate designs and analysis methods are reported, e.g., single and triple lattice designs (RAMALHO et al., 2000) and the Papadakis spatial analysis method. This was applied in various cases with a smaller number of soybean genotypes (VOLLMAN et al., 2000; STORCK et al., 2008; BENIN et al., 2013), showing significant accuracy gains when compared to the random complete block design. Appropriate measures to assess the experimental precision in bean and soybean genotypes competition assays were studied (CARGNELUTTI FILHO et al., 2009), and they can be used to identify the best planning and data analysis.

Bean yield data with 25 to 400 genotypes and lattice design were used, and it was observed that the Papadakis method contributes to improve the local control efficiency (COSTA et al., 2005). In wheat, it was observed that the indices of experimental precision measurements improved with use of the Papadakis method, and the number of replicates necessary to predict genotype performance was reduced (STORCK et al., 2014). Similar results were obtained in soybean trials with a small number of genotypes, and it was observed that the Papadakis method allowed to reduce the number of replicates (STORCK et al., 2009).

In breeding programs, obtaining information about the best analysis method and the most appropriate number of replicates is extremely important to conduct competition assays with a large number of genotypes. This information allows to improve network testing and efficiency in the identification of most promising genotypes. Precision of trials, with a large number of soybean genotypes and analyzed by different methods, is still unknown. Number of replicates suitable for trials with a large number of genotypes is also not known. The aim of this study was to

evaluate both the experimental precision of different statistical analysis methods, with a large number of soybean genotypes, and their relationship with the number of replicates.

3.2 Materials and Methods of analysis and number of replicates for trials with large numbers of soybean genotypes

Data on soybean yield of nine trials were used (324 genotypes obtained in the 2014/15 harvest). Among the genotypes, lines from bi-parental crosses (generation $F_{7:8}$ of the *Nidera Sementes Ltda* breeding program) (240), advanced lines of the program (38), and commercial cultivars recommended for the soybean cultivation macroregions 1 and 2 (46) were evaluated. The triple lattice experimental design with three replicates was used. Each repetition (complete block) was arranged in six rows of three incomplete blocks (18 incomplete blocks) and each incomplete block contained 18 genotypes. Each plot comprised four 5-m length lines, with 0.50-m spacing between lines. The two central lines of each plot were used to obtain the grain yield data. A seeding density of 30 seeds m^{-2} was used, and the basic fertilization consisted of NPK (02:20:20; 350 kg ha^{-1}). Procedures for the control of weeds, pests, and diseases met the technical recommendations for the culture.

The trials were conducted in the state of Paraná (PR), in the cities of Cambé (lat: 23°11’ S; long: 51°17’ W; alt: 520m), Corbélia (lat: 24°32’ S; long: 53°18’ W; alt: 650m), Mamborê (lat: 24°13’ S; long: 52°32’ W; alt: 715m), Palotina (lat: 24°18’ S; long: 53°50’ W; alt: 330 m), Realeza (lat: 25°42’ S; long: 53°32’ W; alt: 520 m), São Jorge do Ivaí (lat: 23°25’ S; long: 52°18’ W; alt: 560m), and São Miguel do Iguaçu (lat: 25°15’ S; long: 54°14’ W; alt: 290m), and in the state of São Paulo (SP), in the city of Cândido Mota (lat: 22°46’ S; long: 50°23’ W; alt: 440m). In Corbélia city (PR), two side-by-side trials were performed, also forming an experiment with six complete blocks or a duplicate triple lattice design with 1944 plots.

Analysis of variance was performed according to the triple lattice design for each trial. Genes (CRUZ, 2013) software was used in these analyzis. The same trials were also analyzed using the randomized complete block design (RCBD) and spatial analysis or Papadakis method, using DelPapa (STORCK et al., 2015) software.

From these analyzes, precision measurements were obtained as follows: relative efficiency for the use of complete block in relation to completely randomized design (CRD); relative efficiency of the use of triple lattice in relation to RCBD; error variation coefficient (VCe); genetic variation coefficient (VCg); ratio VCe/VCg; and selective accuracy (SA), which was estimated as: $SA = (1-1/F)^{1/2}$, for $F>1$, and $SA = 0$, for $F<1$, in which F is the F test value for the genotype. The least significant difference (LSD) between genotypes was also obtained by the Tukey test. The Fasoulas differentiation index (FDI) was calculated using the expression , and m_i is the number of averages that the i^{th} genotype statistically exceeds, after the Tukey test ($\alpha=0.05$) was applied. The FDI is the percentage of the number of significant differences between means, which the method of multiple comparisons of means (Tukey) could detect in relation to the total number of pairs of means.

Mathematical model additivity was evaluated by the non-additivity test (STEEL et al., 1997). Regarding error estimates, homogeneity between genotype variances was verified using the Bartlett test (STEEL et al., 1997) with $\alpha=0.05$. Normality of distribution and error randomness was assessed by the Shapiro-Wilk test ($\alpha=0.05$) (STORCK et al., 2015).

Averages estimated by analysis with the random block design, and those obtained using the triple lattice analysis and Papadakis method, were used to estimate the Pearson linear correlation coefficients and Spearman non-parametric method.

Considering that a six-replicate experiment (six complete blocks) was also conducted in the city of Corbélia (PR), with plots arranged according to a 36-row and 54-column matrix, block resampling (replicate) was possible using the R (R Development Core Team, 2015) program. First, taking into account the six blocks, 2.000 replicas were resampled, with replacement, for analysis by the randomized complete block design (RCBD) and Papadakis (Papa) methods. In each method and analysis, the precision measurements (SA, VCe, VCg, VCg/VCe, LSD, and FDI) were obtained with the additivity, randomness, normality, and variance homogeneity assumptions. The 2000 results of each precision measurement were used to determine the mean value and the 0.025 (LL) and 0.975 (UL) percentiles as being a bootstrap confidence interval estimate ($\alpha=0.95$). The 2.000 resampling procedure was also performed in the cases of five, four, three, and two blocks per trial, and the mean,

LL, and UL values were obtained for each precision measurement and number of replicates.

3.3 Results and Discussion about Methods of analysis and number of replicates for trials with large numbers of soybean genotypes

3.3.1 Methods of analysis

In all trials and analysis methods, genotype effects were significant ($\alpha<0.01$). According to the randomized complete block design (RCBD), analysis relative efficiency is low in all trials relative to the completely randomized design (CRD). Similarly, efficiency in using triple lattice is low relative to RCBD (Table 3.1). Block effect was significant ($\alpha<0.05$) in five of nine trials (55.0%) and significance did not result in greater relative efficiency. A study of soybean (226 trials) showed that the average relative efficiency in block use was equal to 135% (range: 102.7-215.0%) when the number of genotypes is low (10 to 20), and the block effect was significant (29.0% of the trials) (STORCK et al., 2008). The value for relative efficiency was equal to 127.0% in the evaluation of grain yield (100 soybean genotypes) using a triplicate triple lattice design (nine replicates) (BARONA et al., 2009). In a study with soybean yield data (25-60 genotypes) in lattice design (nine trials), the average relative efficiency in the lattice design was 186% (range: 101-402%), and the variation coefficient (VC) was 12%. The VC value was determined for the RCBD (17.5%) and Papadakis method (11.0%) with variation in the number of neighboring plots in the covariate estimate (VOLLMANN et al., 2000). Progenies of Eucalyptus (121) of two ages were analyzed in two different locations, according to the lattice design (11x11), with three replicates, and the Papadakis method efficiency was observed only when the lattice was also efficient (145%) (SOUZA et al., 2003). Thus, in this study, the precision gain due to the use of triple lattice was limited because high precision already existed in the analysis by use of the design in random complete blocks.

Table 3.1 – Variability indicators obtained using the randomized complete block design (RCBD), Papadakis (Papa) method, and triple lattice for nine environments (E) in Paraná (PR) and

São Paulo (SP), Brazil, in the 2014/15 harvest.

Variability indicators[(1)]	A1: Corbélia-PR			A2: Corbélia-PR			A3: Palotina-PR		
	RCBD	Papa	lattice	RCBD	Papa	lattice	RCBD	Papa	lattice
RE	100.3	-	101.7	101.3	-	101.8	99.8	-	104.4
SA	98.72	99.14	98.69	98.14	98.70	98.10	97.54	98.59	97.52
VCe	6.89	5.64	6.73	7.88	6.55	7.69	10.11	7.73	9.69
VCg	24.66	24.73	23.76	23.26	23.23	22.46	25.84	26.25	24.64
VCg/VCe	3.58	4.38	3.53	2.95	3.55	2.92	2.56	3.40	2.54
LSD	27.12	22.21	26.51	31.00	25.78	30.29	39.79	30.41	38.17
FDI	46.58	53.78	47.58	38.43	46.34	39.68	32.07	43.93	33.92
p-Random	0.222	0.000	-	0.116	0.000	-	0.000	0.000	-
p-Normal	0.000	0.000	-	0.357	0.732	-	0.003	0.098	-
p-Non-addict	0.087	-	-	0.701	-	-	0.344	-	-
p-Bartlet	0.001	0.013	-	0.894	0.996	-	0.816	0.030	-
	A4: Mamborê-PR			**A5: São Miguel do Iguaçu-PR**			**A6: São Jorge do Ivaí-PR**		
RE	104.9	-	100.9	118.6	-	99.0	100.9	-	106.9
SA	96.39	97.82	96.37	98.10	98.65	98.08	94.70	96.77	94.78
VCe	7.81	6.02	7.69	7.03	5.92	7.07	8.14	6.27	7.69
VCg	16.33	16.36	16.01	20.52	20.59	20.52	13.86	13.89	13.19
VCg/VCe	2.09	2.72	2.08	2.92	3.48	2.90	1.70	2.22	1.72
LSD	30.73	23.69	30.28	27.66	23.31	27.82	32.03	24.68	30.28
FDI	17.43	27.88	17.99	35.81	43.85	35.53	11.98	21.17	13.31
p-Random	0.090	0.000	-	0.302	0.000	-	0.004	0.000	-
p-Normal	0.420	0.746	-	0.000	0.000	-	0.009	0.676	-
p-Non-addict	0.789	-	-	0.055	-	-	0.106	-	-
p-Bartlet	0.575	0.583	-	0.000	0.001	-	0.687	0.020	-
	A7: Cambé-PR			**A8: Cândido Mota-SP**			**A9: Realeza-PR**		
RE	99.7	-	100.0	100.1	-	104.6	105.8	-	101.8
SA	91.33	94.83	91.28	97.10	98.29	97.16	96.59	98.07	96.57
VCe	8.65	6.70	8.63	8.96	6.81	8.58	5.56	4.12	5.43
VCg	11.21	11.56	11.14	20.99	21.00	20.35	11.99	11.95	11.66
VCg/VCe	1.30	1.73	1.29	2.34	3.08	2.37	2.16	2.90	2.15
LSD	34.06	26.37	34.00	35.25	26.78	33.79	21.89	16.23	21.39
FDI	5.07	12.83	5.08	25.30	37.90	26.99	18.94	31.33	19.63
p-Random	0.415	0.000	-	0.002	0.000	-	0.000	0.000	-
p-Normal	0.484	0.479	-	0.121	0.410	-	0.689	0.703	-
p-Non-addict	0.904	-	-	0.661	-	-	0.240	-	-
p-Bartlet	0.599	0.507	-	0.976	0.172	-	0.013	0.229	-

(1) RE: relative efficiency; SA: selective accuracy; VCe: error variation coefficient; VCg: genetic variation coefficient; LSD: least significant difference, by the Tukey test (5%) in percentage of the mean; FDI: Fasoulas differentiation index; and p-value for the randomness, normality, and non-additivity tests and error variance homogeneity (Bartlet); - : Variable not available.

Taking into account the precision measurements (SA, VCg/VCe, LSD, and FDI), use of the Papadakis method showed a higher efficiency (more accurate) in

all trials compared to the RCBD and triple lattice models. In the study of 226 trials by STORCK et al. (2008), LSD was reduced (from 41.5 to 30.6%) and FDI was increased (from 5.0 to 13.1) with use of the Papadakis method relative to RCBD. In this study, different values were obtained for LSD with the use of RCBD (31.1%) and Papadakis method (24.4%; 6.7% less). In addition, average values were also obtained for FDI using RCBD (25.7) and Papadakis method (35.4; 9.7 more units). Thus, despite the large number of genotypes assessed in this study, the results obtained for the two precision measures cited above were similar, indicating that the Papadakis method provides higher accuracy compared to RCBD and triple lattice.

Selective accuracy (SA) was very high (SA>90%; RESENDE, DUARTE, 2007) and similar in all trials and analysis methods, ranging from 91.3% (RCBD; Cambé-PR) to 99.1% (Papa; Corbélia-PR). It is likely that the methodologies did not show difference, due to the high precision (low experimental error) in these trials, with a large number of genotypes. In a study of 216 soybean trials, with a variable number of genotypes (10-20), the SA values were very high (90%>SA) in 15.5% of the trials and high (70%<SA<90%) in 59% of trials (CARGNELUTTI FILHO et al., 2009). In the present study, all (100%) trials (and methods) presented very high SA values, and thus accuracy differences between analysis methods are small, and all of them belong to the same (very high) accuracy class.

Regarding assumptions, the additivity model was not rejected in the nine trials. Regarding analysis in RCBD, randomness (5), variance homogeneity (6), and normality (5) were observed in the trials. Regarding use of the Papadakis method, results for normality and homogeneity were similar. In addition, lack of error randomness was observed due to adjustment of plot values as a function of mean errors in the neighboring plots.

Adjustment in the average values (by the Papadakis method or lattice design) did not change their ordination (via Spearman correlation) and relationship between means (via Pearson correlation) not adjusted in the analysis with the RCBD model (data not shown). Taking into account the high number of genotypes and the very high accuracy (SA>90%) in this study, it seems reasonable to assume that there is no bias in the adjustment of means when the triple lattice design and Papadakis method are used. However, spatial analysis in a trial of 115 soybean genotypes, led to

a different ordering of lines relative to the non-spatial analysis (DUARTE, VENCOVSKY, 2005).

On average, the precision measurements (SA, LSD, and FDI), as obtained by analysis of the duplicate triple lattice design (six replicates), are higher than those obtained by the RCBD or Papadakis method in six replicates (Table 3.2). However, the differences are small and the method can be chosen by precaution, in which efficiency of the analysis method depends on the number of entries and spatial variation of the plots in the experimental area.

Table 3.2 - Variability indicators obtained using the randomized complete block design (RCBD; six blocks), Papadakis method (Papa; six blocks), and duplicate triple lattice design in soybean trials conducted in the 2014/15 harvest.

Variability indicators(1)	RCBD	Papa	Lattice
SA (%)	98.48	98.97	99.18
VCe (%)	9.97	8.18	7.32
VCg (%)	23.04	23.06	22.47
VCg/VCe	2.31	2.82	3.07
LSD (%)	27.63	22.68	20.31
FDI (%)	43.58	51.26	55.67

(1) SA: selective accuracy; VCe: error variation coefficient; VCg: genetic variation coefficient; LSD: least significant difference, by the Tukey test (5%) in percentage of the mean; FDI: Fasoulas differentiation index.

3.3.2 Number of replicates

Influence of number of replicates on the experimental precision, as measured by SA, LSD and FDI is shown in figure 3.1. As can be seen, the accuracy gain is low from three replicates on, in the analyses of both randomized complete block design (RCBD) and Papadakis method. If amplitude of the confidence interval is considered in the precision measurements of this study, difference between accuracy values obtained for two or three replicates was not observed. A possible reason is that accuracy is already high, and little can be gained by increasing the number of replicates and varying the method of analysis. In RCBD, estimates of SA and FDI were lower than those obtained by the Papadakis method for any number of

replicates.

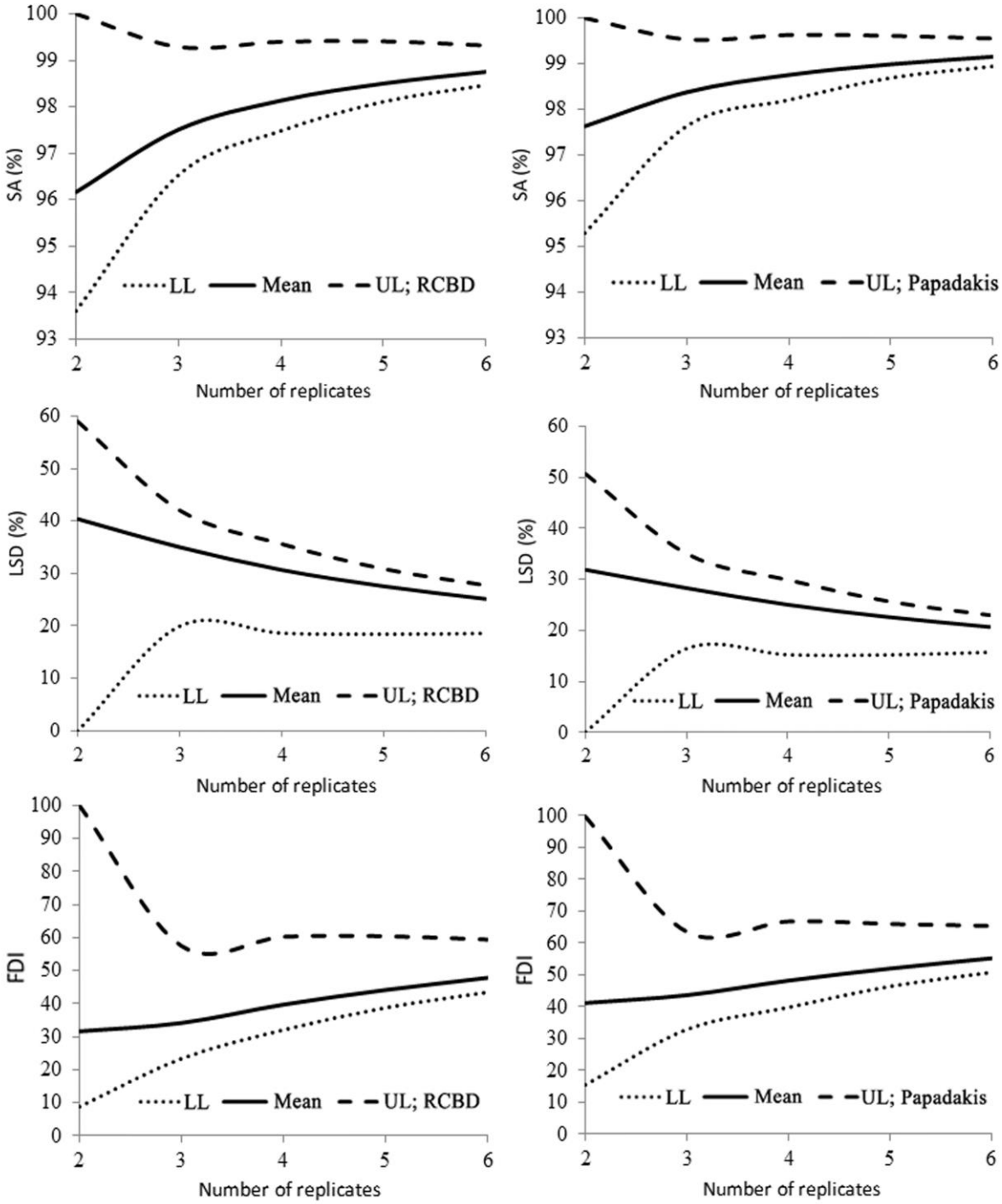


Figure 3.1 - Selective accuracy (SA), least significant difference (LSD) by the Tukey test (5%) in percentage of the mean, and variation in the Fasoulas differentiation index (FDI); mean, lower limit (LL), and upper limit (UL) of the confidence interval, by resampling (1-α=0.95) for different number of replicates, and using the randomized complete block design (RCBD) and Papadakis method.

In a study of 175 competition trials of irrigated rice in RCBD, with number of genotypes in the range of 5-36, SA was suitable to evaluate experimental precision in the trials and use of more than five replicates was of little contribution to precision gains (CARGNELUTTI FILHO et al., 2012). In another study with 101 maize yield trials, in RCBD, FDI was also suitable to classify the experimental precision (CARGNELUTTI FILHO, STORCK, 2007). In this study, it was observed that FDI values in RCBD are smaller than those obtained using the Papadakis method with any number of replicates.

A trial is considered very accurate if SA**>**90%. Such a high accuracy was obtained using two replicates, due to the LL values of the confidence interval for RCBD (93.6%) and Papadakis method (95.3%). Thus, use of only two repetitions can be recommended for trials containing 324 soybean genotypes, and the analysis can be performed using the RCBD or Papadakis method. Based on these results, however, it cannot be stated that the simple lattice method (two replicates) can be used, as this situation was not analyzed by resampling. Block resampling with replacement does not assures genotype distribution as recommended by the “simple lattice” method.

3.4 Conclusion about Methods of analysis and number of replicates for trials with large numbers of soybean genotypes

Papadakis method has more reliable experimental precision indicators when compared to the randomized complete block design and triple lattice method. For trials with 324 soybean genotypes, it is possible to use two replicates and analyze the data with the randomized complete block design or Papadakis method as a precaution to obtain selective accuracy above the range of high experimental precision.

4 AGRONOMIC PERFORMANCE OF MODERN SOYBEAN CULTIVARS IN MULTI-ENVIRONMENT TRIALS

4.1 Literature Review of Agronomic Performance of Modern Moybean Cultivars in Multi-environment Trials

Soybean (*Glycine max* (L.) Merr.) is one of the most important crops for the Brazilian economy. Its domestic production reached 96.2 million tons in the 2014/2015 crop season, with an average production of approximately 3,000 kg ha^{-1} (CONAB, 2016).

Genotype × environment interaction (GEI) is one of the main challenges of soybean breeding programs with respect to both cultivar selection and recommendation phases (BRANQUINHO et al., 2014). GEI is reflected in differing genotypic expressions in different growing environments, and reduces the association between phenotype and genotype, thereby reducing genetic progress in breeding programs (LOPES et al., 2012).

Data from multi-environment trials are necessary to assess the presence of GEI, yield, and genotype adaptability and stability. Adaptability is the ability of the genotype to respond predictably to environmental stimuli, and stability indicates the predictability of performance in different environments. Several methods for adaptability and stability analysis have been described in the literature, which differ according to the statistics—including analysis of variance, non-parametric regression, multivariate analysis, and mixed-model analysis—and parameters used. Methods based on mixed models enable the analysis of genotypes, including random effects analysis, and multivariate analysis has innovative solutions regarding the visualization of results.

The aforementioned mixed-model analysis methods, including restricted maximum likelihood/best linear unbiased prediction (REML/BLUP), enable estimation of variance components and prediction of genetic values free of environmental effects (PEIXOUTO et al., 2016). The following methods may be used: the harmonic mean of genotypic values (HMGV) to infer the mean and stability; relative performance of predicted genotypic values (RPGV) to analyze the genotypic adaptability and average production rate; and the harmonic mean of the relative performance of predicted

genotypic values (HMRPGV) to identify highly productive, adapted and stable genotypes (GOMEZ et al., 2014; SPINELLI et al., 2015; COSTA et al., 2015). Mixed models provide estimates of stability and genotypic adaptability because they classify the genotypic effects as random (RESENDE, 2004).

The use of multivariate statistics, involving tools, which include GGE biplots, enables the summarizing of data from a large dataset into a few principal components (YAN, 2015). Biplots assessing the mean, phenotypic stability, and ideal genotype enable the graphical representation of the performance of each cultivar, thereby facilitating the selection of superior genotypes (QIN et al., 2015).

The simultaneous use of mixed models, based on REML/BLUP and multivariate method, enables the exploration of different adaptability and stability concepts, thereby complementing the data collected, and thus increasing the efficacy of the selection of superior genotypes (ANDRADE et al., 2016).

This study differs from other published studies on soybean adaptability, stability, and crop performance, because it combines mixed-model methods and GGE biplots to assess cultivars widely grown in the Brazilian macroregions of adaptation 1 and 2.

This study aimed to assess the crop performance, adaptability, and stability of modern soybean cultivars, in multi-environment trials, and to identify the ideal cultivars for eight growing environments in Brazil.

4.2 Materials and Methods of Agronomic Performance of Modern Soybean Cultivars in Multi-environment Trials

Forty-six modern soybean cultivars, widely grown in the Brazilian soybean macroregions of adaptation 1 and 2, which were provided for cultivation from 2007 to 2013, were assessed (Table 4.1). The cultivars used were classified, based on their maturity groups (MG), into: early, MG = 4.8 to 5.7; medium, MG = 5.8 to 6.2; and late, MG = 6.3 to 7.3. The experiments were conducted in a randomized complete block design, with three replicates, in eight sites representative of the microregions of adaptation 102, 201, and 202, in the 2014/2015 crop season (Table 4.2). The sites were selected within microregion 201 and nearby regions, with similar sowing season

and climatic characteristics.

Table 4.1 – Description of 46 soybean cultivars, maturity group, cycle, year of release, technology and releaser.

Number	Cultivar	Maturity Group	Growth habit	Cycle	Year of release	Technology	Releaser
1	BMX Potência RR	6.7	I	Later	2007	RR	GDM Genética
2	DMario 58i	5.5	I	Early	2007	RR	GDM Genética
3	NK 7059 RR	6.2	I	Medium	2007	RR	Syngenta
4	A 6411RG	6.4	D	Later	2008	RR	Nidera
5	BMX Ativa RR	5.6	D	Early	2008	RR	GDM Genética
6	BMX Energia RR	5.3	I	Early	2008	RR	GDM Genética
7	NA 5909 RG	5.9	I	Medium	2008	RR	Nidera
8	NS 4823	4.8	I	Early	2008	RR	Nidera
9	BMX Turbo RR	5.8	I	Medium	2009	RR	GDM Genética
10	NS 5858	5.8	I	Medium	2010	RR	Nidera
11	NS 6262	6.2	I	Medium	2010	RR	Nidera
12	SYN1059 RR	5.9	I	Medium	2010	RR	Syngenta
13	NS 6767	6.7	I	Later	2011	RR	Nidera
14	TMG 7262RR	6.2	I	Medium	2011	RR	TMG
15	NS 4901	4.9	I	Early	2012	RR	Nidera
16	NS 5258	5.2	I	Early	2012	RR	Nidera
17	NS 5290	5.2	I	Early	2012	RR	Nidera
18	NS 5401 RR	5.4	I	Early	2012	RR	Nidera
19	NS 6209	6.2	I	Medium	2012	RR	Nidera
20	NS6121RR	6.1	I	Medium	2013	RR	Nidera
21	NS6823RR	6.8	I	Later	2013	RR	Nidera
22	M6210IPRO	6.2	I	Medium	2011	IPRO	Monsoy
23	M6410IPRO	6.4	I	Later	2011	IPRO	Monsoy
24	5958RSF IPRO	5.8	I	Medium	2012	IPRO	GDM Genética
25	6458RSF IPRO	6	I	Medium	2012	IPRO	GDM Genética
26	6563RSF IPRO	6.3	I	Later	2012	IPRO	GDM Genética
27	AS 3570IPRO	5.7	I	Early	2012	IPRO	Monsoy
28	AS 3610IPRO	6.1	I	Medium	2012	IPRO	Monsoy
29	M5917IPRO	5.9	I	Medium	2012	IPRO	Monsoy
30	NS 5000 IPRO	5	I	Early	2012	IPRO	Nidera
31	NS 5106 IPRO	5.1	I	Early	2012	IPRO	Nidera
32	NS 5151 IPRO	5.1	I	Early	2012	IPRO	Nidera
33	NS 5445 IPRO	5.4	I	Early	2012	IPRO	Nidera
34	NS 5959 IPRO	5.9	I	Medium	2012	IPRO	Nidera
35	NS 6909 IPRO	6.9	I	Later	2012	IPRO	Nidera
36	NS 7000 IPRO	7	I	Later	2012	IPRO	Nidera
37	NS 7209 IPRO	7.2	I	Later	2012	IPRO	Nidera
38	NS 7237 IPRO	7.2	I	Later	2012	IPRO	Nidera
39	NS 7300 IPRO	7.3	I	Later	2012	IPRO	Nidera
40	NS 7338 IPRO	7.3	I	Later	2012	IPRO	Nidera
41	NS 5727 IPRO	5.7	I	Early	2013	IPRO	Nidera
42	NS 6006 IPRO	6	I	Medium	2013	IPRO	Nidera
43	NS6060IPRO	6	D	Medium	2013	IPRO	Nidera
44	NS6700IPRO	6.7	I	Later	2013	IPRO	Nidera
45	NS6906IPRO	6.9	I	Later	2013	IPRO	Nidera
46	TMG2158IPRO	5.8	I	Medium	2013	IPRO	TMG

* I: Indetermined; D: Determined

This region has the highest soybean production in southern Brazil. The experimental units consisted of four 5 m rows, spaced 0.5 m between rows. The sowing density was 30 seeds m^{-2}, and base fertilization was performed using 350 kg ha^{-1} NPK (02:20:20). Mechanical methods were employed for sowing and harvesting. The character primarily studied was the grain yield (GY, kg ha^{-1}), assessed in the two central rows of each plot (5 m^2 useful area), with grain moisture corrected to 13% (wet basis). Crop treatments were conducted in accordance with the technical recommendations for soybean cultivation (OLIVEIRA, ROSA, 2014).

Table 4.2 – Identification of test locations used to evaluate 46 soybean cultivars, in 2014/2015 crop season.

Location	Macroregion	Microregion	Latitude / Longitude	Altitude (m)	Climate
Cambé, PR	2	201	23°16'S 51°16'W	520	Cfa
Cândido Mota, SP	2	201	22°44'S 50°23'W	440	Cwa
Corbélia, PR	2	201	24°47'S 53°18'W	650	Cfa
Mamborê, PR	2	201	24°19'S 52°31'W	715	Cfa
Palotina, PR	2	201	24°17'S 53°50'W	330	Cfa
Realeza, PR	1	102	25°46'S 53°31'W	520	Cfa
São Jorge do Ivaí, PR	2	202	23°25'S 52°17'W	560	Cfa
São Miguel do Iguaçu, PR	2	201	25°20'S 54°14'W	290	Cfa

Initially, variance components were assessed using restricted maximum likelihood (REML), and mean components were assessed using the best linear unbiased prediction (BLUP) method, employing the statistical package Selegen (RESENDE, 2002) with models 21 (for analysis of genetic parameters for each site) and 54 (for combined analysis of sites).

Analysis of variance was also performed to assess the presence of genotype × environment interactions. Subsequently, cluster analysis of means was performed using the Scott-Knott test at 5% probability and the statistical software

package Genes (CRUZ, 2013). The average production rates of each genotype, at each site, and for the set of sites were also indicated.

Data on genetic effects (g), predicted genotypic values (u + g), and the gain of each genotype would attain on removal of the environmental component, were also collected in the analysis performed using model 54 of the software package Selegen (RESENDE, 2002). The new genotype average was assessed with this gain, and rank analysis was performed using this new value. Furthermore, the mean genotypic value (u + g + gem) was assessed in the various environments; this indicated the average interaction with all study environments (RESENDE, 2002). The following parameters could also be assessed using this model: genotypic stability using HMGV; genotypic adaptability and crop performance, using RPGV multiplied by the overall mean (OM) of all sites (RPGV*OM); and genotypic stability and adaptability and crop performance, using HMRPGV*OM.

Stability was also assessed using the software GGEbiplot (Yan, 2001), which analyzes the stability of genotypes associated with their average production rates. For this purpose, the following parameters were used: data transformation (Transform = 0, without transformation), data scale (Scaling = 0, without scale), data centering (Data centering = 2, genotype plus genotype x environment interaction [G+GEI]), and singular-value partitioning (SVP = 1, focus on genotype). The concept of ideal genotype was also evaluated using the software GGEbiplot (YAN, 2001) using the same parameters as those used for the mean and stability analysis.

4.3 Results and Discussion of Agronomic Performance of Modern Soybean Cultivars in Multi-environment Trials

In the combined analysis, the estimation of heritability in the broad sense (h^2_g) for grain yield (GY) was 0.37 (± 0.05), which is lower than the estimate usually assessed for agronomic characters controlled by a few genes, albeit within the expected range for characters controlled by many genes having small effects, including GY (Table 4.3). Low h^2_g values indicate the need for a breakdown in the GEI, because they result from changes in the performance of genotypes at the study

sites (ROSADO et al., 2012). Interaction analysis enables maximizing selection gains, when testing homozygous clones or lines. Similar results were obtained by other authors (PINHEIRO et al., 2013; ROCHA et al., 2015; ANDRADE et al., 2016), who also found low heritability estimates for soybean GY. The value of interaction variance ($V_{G\times A}$), when higher than genotypic variance (V_G), also contributes to the low values of h^2_g estimates. In the individual analysis of sites, h^2_g was higher, ranging from 0.60 to 0.92, which indicates that a large part of phenotypic variance (V_F) resulted from V_G. The value of standard deviation, at each site, was higher than that assessed for the set of study environments, ranging from 0.19 to 0.23. However, these standard deviation values are within acceptable limits, thus indicating that the predictions are reliable for use in breeding (RESENDE, 2004).

Table 4.3 – Estimation of genetic parameters for each of the eight locations and for the set of locations, for the trait grain yield (GY Kg ha.$^{-1}$) of 46 soybean cultivars.

	Locations								**Mean of locations**	
State	**Paraná**							**São Paulo**		
Parameter	**Cambé**	**Corbélia**	**Mamborê**	**Palotina**	**Realeza**	**São Jorge do Ivaí**	**São Miguel do Iguaçu**	**Cândido Mota**	V_G	222102
V_G	198089	771929	903687	370107	327417	276953	683526	586966	V_{GxA}	292728
Ve	130764	105619	74431	69894	51372	117540	59123	87360	Ve	87014
V_F	328853	877548	978118	440001	378790	394494	742649	674326	V_F	601844
h^2_g	0,60 (±0,19)	0,88 (±0,22)	0,92 (±0,23)	0,84 (±0,22)	0,86 (±0,22)	0,70(±0,20)	0,92 (±0,23)	0,87 (±0,22)	h^2_g	0,37 (±0,05)
h^2_{mg}	0,82	0,96	0,97	0,94	0,95	0,88	0,97	0,95	h^2_{mg}	0,85
Acgen	0,91	0,98	0,99	0,97	0,97	0,94	0,99	0,98	Acgen	0,92
CVgi%	10,49	18,42	22,85	21,15	12,49	12,69	20,23	23,41	c^2_{int}	0,49
CVe%	8,53	6,81	6,56	9,19	4,95	8,26	5,95	9,03	rgloc	0,43
PEV	35727	33671	24147	21918	16273	34324	19156	27744	CVgi%	11,73
SEP	189,01	183,50	155,39	148,05	127,57	185,27	138,40	166,56	CVe%	7,34
Mean	**4242**	**4770**	**4160**	**2876**	**4582**	**4148**	**4087**	**3273**	**General mean**	**4017**

V_G, genotypic variance; V_{GxA}, genotype x environment interaction variance; Ve, residual variance; V_F, individual phenotypic variance. h^2_g, heritability of individual plots of the total genotypic effects (in the broad sense); c^2_{int}, coefficient of determination of the G x E interaction; h^2_{mg}, heritability of the mean of genotype; Acgen, accuracy in the selection of genotypes; rgloc, genotypic correlation between environments; CVgi%, coefficient of genotypic variation; CVe%, coefficient of residual variation; PEV, variance of the prediction error; SEP, standard deviation of the genotypic value predicted.

The genotypic coefficient of variation (CVgi%) was 11.73% in the combined analysis of sites, and ranged from 10.49%, in Cambé, Paraná State (PR), to 23.41%, in Cândido Mota, São Paulo (SP). Sites with higher CVgi% values favor the discrimination of genotypes; that is, they promote a wider performance range, thereby

favoring selection. The residual coefficient of variation (CVe%) ranged from 4.95% in Realeza, PR, to 9.19% in Palotina, PR. These values are considered low and indicate good experimental precision. Genotype selection accuracy (Acgen) for the set of sites was 0.92, and ranged from 0.91 in Cambé, PR, to 0.99 in Mamborê, PR and São Miguel do Iguaçu, PR, thereby, indicating the high experimental precision obtained in all study environments. This parameter involves correlating the true genotypic value of the genetic treatment with the genotypic value estimated or predicted from experimental data. These values may be classified within the very high accuracy class (Acgen > 0.90; RESENDE, DUARTE, 2007).

The genotypic correlation between performances in the various environments (rgloc) was 0.43. These values indicate the occurrence of complex interaction between genotypes and test sites, which entails different genotypic responses at the different sites where they are assessed, thus, changing the ranking between sites (COSTA et al., 2015). Furthermore, this also indicates that sites in the same soybean microregion of adaptation show considerable differences in cultivar performance. This is the case with microregion 201 (macroregion 2). The sites Realeza, PR, in microregion 102 (macroregion 1), and São Jorge do Ivaí, PR, in microregion 202 (macroregion 2), showed crop performance similar to that assessed in microregion 201. Furthermore, large variations in performance were observed, even in study sites with latitude variation smaller than 3°. Therefore, breeders must conduct several comparative trials of cultivars, within the same sub-region, to identify the specificity of each site where they intend to plant their cultivars.

The average GY of the trials was 4.017 kg ha^{-1} (Table 4.4), which is higher than the average of the Central-South region of Brazil (3.016 kg ha^{-1}), and the states of Paraná (3.294 kg ha^{-1}) and São Paulo (2.970 kg ha^{-1}), according to the Conab (2016). The average yields assessed in the trials, for each site, ranged from 2.876 kg ha^{-1} in Palotina, PR, to 4.770 kg ha^{-1} in Corbélia, PR.

In the set of study environments, the highest yields were observed in the NA 5909 RG, M6410IPRO, NS 5959 IPRO, NS6823RR, M5917IPRO, NS 6767, and 6563RSF IPRO cultivars with 4.851, 4.705, 4.670, 4.644, 4.590, 4.589, and 4.578 kg ha^{-1} GY, respectively.

Table 4.4 – Grain yield (kg ha-1) of soybean cultivars, grouping means by the Scott-Knott test, and mean of cultivars in eight. sites GY (X G), mean of each location (X L), and mean of 46 soybean cultivars classifid according to their cycle, in each site, in the 2014/2015 crop season.

Cultivar	Locations								XG
	Cambé, PR	Corbélia, PR	Mamborê, PR	Palotina, PR	Realeza, PR	São. J. do Ivaí, PR	São M. do Iguaçu, PR	Cândido Mota, SP	
BMX Potência RR	4151cB	6186 aA	4453bB	3072cD	4348cB	4706bB	3656dC	3859bC	4304c
DMario 58i	4241cB	4373dB	4804bA	2325eC	4623bA	3982cB	4580bA	2602dC	3941d
NK 7059 RR	4020cC	5813bA	3957cC	3579bD	4461bC	3955cC	5277aB	3364cD	4303c
A 6411RG	3495dB	4105eA	1174eD	2307eC	3997cA	2840eC	2286fC	2601dC	2851h
BMX Ativa RR	2770eB	3030fA	972eD	1743eC	3308dA	2525eB	2269fB	1653fC	2284i
BMX Energia RR	3624dB	4445dA	4850bA	2817dC	4953bA	4088cB	3580dB	3857bB	4027d
NA 5909 RG	4614bB	5557bA	5373aA	3240cC	5343aA	4808bB	5283aA	4588aB	4851a
NS 4823	3816cB	3129fC	3771cB	2251eD	5086aA	3635dB	2836eC	2463dD	3373g
BMX Turbo RR	4495bB	5108cB	5577aA	2445eE	5000bB	3928cC	4824bB	3202cD	4322c
NS 5858	3947cB	3548fB	3897cB	2242eC	4619bA	4301cA	4278cA	3437cB	3784e
NS 6262	4156cB	3938eB	4664bA	2230eD	4980bA	4195cB	3963cB	2726dC	3856e
SYN1059 RR	5099aA	4810cA	4475bB	2586dD	4749bA	4320cB	4278cB	3605cC	4240c
NS 6767	4999aB	5886bA	4686bB	3673bD	4104cD	4447cC	4909bB	4011aD	4589a
TMG 7262RR	4441bB	5123cA	4800bB	2158eC	5315aA	4634bB	4288cB	2502dC	4158c
NS 4901	3911cB	4182eB	3915cB	2255eD	5090aA	4067cB	3943cB	2832dC	3774e
NS 5258	4299bB	4052eB	4035cB	2175eC	4678bA	4670bA	4181cB	3677bB	3971d
NS 5290	4446bA	4228eA	4478bA	2424eB	4586bA	4441cA	4109cA	2317eB	3879e
NS 5401 RR	4093cA	3161fB	4008cA	2474dC	4083cA	4081cA	4033cA	2638dC	3571f
NS 6209	4199cB	5167cA	5253aA	3122cC	3984cB	3969cB	3607dB	3767bB	4134c
NS6121RR	3906cB	5345cA	3958cB	2886dC	4247cB	4189cB	3186eC	2976dC	3836e
NS6823RR	4586bB	6097aA	4277cC	3923bC	4699bB	4675bB	4458cB	4437aB	4644a
M6210IPRO	4328bC	5605bA	4257cC	3777bC	4117cC	4794bB	4705bB	3889bC	4434b
M6410IPRO	4613bB	5711bA	4184cC	2920dD	4815bB	5453aA	5296aA	4645aB	4705a
5958RSF IPRO	4686bB	5402cA	4657bB	3107cD	4188cC	3947cC	4951bB	3499cD	4305c
6458RSF IPRO	4056cC	4753dB	4668bB	3103cD	5130aA	4373cB	4158cC	3905bC	4268c
6563RSF IPRO	5342aA	5539bA	4970bB	3253cD	4763bB	4029cC	4373cC	4358aC	4578a
AS 3570IPRO	4200cA	3874eA	4063cA	2243eC	3834cA	4140cA	3084eB	3526cA	3621f
AS 3610IPRO	4329bB	5175cA	4625bA	3303cC	4776bA	4068cB	4870bA	4149aB	4412b
M5917IPRO	4215cC	5430cA	4793bB	3368cD	4797bB	4864bB	4870bB	4379aC	4590a
NS 5000 IPRO	4216cB	3360fC	3853cC	2785dD	4861bA	3762dC	3637dC	2217eE	3586f
NS 5106 IPRO	4131cC	4051eC	3961cC	2567dD	5416aA	4807bB	4370cC	2436dD	3967d
NS 5151 IPRO	4043cB	4238eB	3879cB	2656dD	5443aA	4093cB	3805cB	3259cC	3927d
NS 5445 IPRO	3328dC	4004eB	4355cB	2125eD	5056aA	3776dB	3172eC	2602dD	3552f
NS 5959 IPRO	4765bC	5120cB	5532aA	2725dE	5668aA	4490bC	5124aB	3936bD	4670a
NS 6909 IPRO	4579bA	4654dA	4600bA	2521dC	4665bA	4617bA	3796cB	2122eC	3944d
NS 7000 IPRO	4360bC	5730bA	3411dD	3866bD	4527bC	4923bB	4151cC	4152aC	4390b
NS 7209 IPRO	4178cB	6080aA	3026dD	4494aB	3489dC	4320cB	4613bB	3792bC	4249c
NS 7237 IPRO	4077cB	4895cA	3663cC	3437cC	3399dC	4326cB	2281fD	3019dC	3637f
NS 7300 IPRO	4114cC	6087aA	3975cC	3340cD	3091dD	3592dD	4852bB	3366cD	4052d

NS 7338 IPRO	4098cB	4628dA	3729cC	3647bC	3936cC	3629dC	4177cB	2712dD	3819e
NS 5727 IPRO	4722bA	3885eB	4274cB	2678dD	5182aA	3062eC	3178eC	2539dD	3690f
NS 6006 IPRO	4643bA	4420dB	5103aA	2754dC	4870bA	3961cB	4691bA	2505dC	4118c
NS6060IPRO	3224dB	3575fB	1343eD	1948eC	4898bA	2997eB	2150fC	2098eC	2779h
NS6700IPRO	4395bA	4691dA	4091cA	3369cB	4396cA	4376cA	4258cA	4251aA	4228c
NS6906IPRO	3940cC	6265aA	4474bB	3983bC	4279cC	3968cC	4945bB	3827bC	4460b
TMG2158PRO	5231aA	4975cA	4513bB	2388eD	4946bA	4005cC	4664bB	2275eD	4124c

*Means followed by the same uppercase in the row and lowercase letter in the column are not significantly different (P = 0.05) according to Scott-Knott test.

The highest absolute production (6.265 kg ha^{-1}) was obtained with the NS6906IPRO cultivar in Corbélia, PR, albeit with no significant differences from the BMX Potência RR, NS6823RR, NS 7300 IPRO, and NS 7209 IPRO cultivars in the same environment, which did not perform similarly at other sites. The BMX Ativa RR cultivar showed the worst average performance in the set of study sites, with 2.284 kg ha^{-1} GY.

The strongest, positive g values were obtained in the NA 5909 RG, M6410IPRO, NS 5959 IPRO, and NS6823RR cultivars, which therefore had the highest genetic values free of interaction (μ + g; Table 4.5). The strongest negative effects were obtained in the BMX Ativa RR, NS6060IPRO, and A 6411RG cultivars (all with growth habit determined), with genetic values far below the average. The new estimated means suggest that the genotype ranking remained similar to that obtained by comparing the fixed-model means, and changes occurred in phenotypes with intermediate ranking. Similarly, the predicted μ + g values and the average μ + g + gem values persisted in the same category when classified between genotypes, indicating that the same recommendation is made by both the parameters; thus, also enabling making recommendations for untested sites in the experimental network using μ + g values, because genotypic performance is free of interactions in this case. A similar result was also reported by Borges et al., (2012).

The NA 5909 RG, NS6823RR, M6410IPRO, and M5917IPRO cultivars were the most stable, and had the highest average production rates, based on the HMGV method; and the BMX Ativa RR, NS6060IPRO, and A 6411RG cultivars were the most unstable and least productive. The genotypic stability analysis using that method is related to the dynamic concept of stability by association with GY (Resende, 2004), and the lower the standard deviation of the genotypic performance between

sites is, the higher the HMGV will be. Thus, selection by HMGV simultaneously leads to both yield and stability selection (RESENDE, DUARTE, 2007).

Table 4.5 – Genetic effects (g), predicted genotypic values (u + g), gain, new mean of the genotype, rank, average genotypic value in the environments (u + g + gem) and methods of adaptability and stability using mixed models.

Cultivar	g	u+g	Gain	New mean	Rank	u+g+ gem	HMGV	RPGV *OM	HMRPGV *OM
BMX Potência RR	242	4260	424	4441	14	4300	4155	4300	4254
DMario 58i	-64	3953	247	4265	29	3942	3702	3904	3852
NK 7059 RR	242	4259	412	4429	15	4299	4178	4317	4260
A 6411RG	-988	3029	57	4075	44	2867	2559	2866	2605
BMX Ativa RR	-1468	2550	0	4017	46	2308	2056	2291	2111
BMX Energia RR	8	4025	295	4313	25	4027	3913	4042	3998
NA 5909 RG	706	4723	706	4723	1	4839	4718	4855	4829
NS 4823	-545	3472	81	4099	43	3382	3195	3363	3293
BMX Turbo RR	258	4276	454	4472	12	4318	4059	4278	4219
NS 5858	-198	3819	181	4198	35	3787	3631	3785	3736
NS 6262	-136	3881	214	4231	32	3859	3630	3823	3773
SYN1059 RR	189	4206	376	4394	18	4237	4073	4223	4203
NS 6767	484	4502	557	4574	6	4582	4497	4615	4572
TMG 7262RR	119	4136	354	4371	20	4156	3795	4081	3992
NS 4901	-206	3812	170	4187	36	3778	3582	3745	3716
NS 5258	-39	3978	282	4300	26	3972	3780	3960	3908
NS 5290	-118	3900	225	4243	31	3880	3634	3841	3782
NS 5401 RR	-378	3640	108	4126	41	3577	3445	3580	3533
NS 6209	98	4116	341	4359	21	4132	4028	4155	4104
NS6121RR	-153	3864	203	4220	33	3839	3701	3830	3800
NS6823RR	531	4548	593	4610	4	4635	4572	4683	4634
M6210IPRO	353	4370	505	4522	9	4428	4366	4473	4427
M6410IPRO	582	4599	644	4661	2	4695	4529	4704	4646
5958RSF IPRO	243	4261	438	4455	13	4301	4179	4306	4279
6458RSF IPRO	212	4230	399	4417	15	4265	4178	4283	4266
6563RSF IPRO	475	4492	545	4562	7	4570	4460	4590	4555
AS 3570IPRO	-336	3681	133	4150	39	3626	3493	3632	3586
AS 3610IPRO	334	4352	488	4505	10	4407	4332	4436	4412
M5917IPRO	484	4502	571	4588	5	4582	4501	4612	4585
NS 5000 IPRO	-365	3652	121	4138	40	3592	3422	3585	3520
NS 5106 IPRO	-42	3975	270	4288	27	3968	3721	3932	3862
NS 5151 IPRO	-76	3941	237	4254	30	3928	3789	3918	3896
NS 5445 IPRO	-394	3624	96	4114	42	3559	3337	3522	3466
NS 5959 IPRO	552	4570	613	4631	3	4661	4451	4639	4601
NS 6909 IPRO	-62	3955	259	4276	28	3945	3648	3895	3809
NS 7000 IPRO	315	4333	472	4489	11	4385	4301	4436	4353
NS 7209 IPRO	196	4213	387	4405	17	4246	4113	4328	4130
NS 7237 IPRO	-322	3695	145	4163	38	3642	3495	3681	3540
NS 7300 IPRO	30	4047	307	4325	24	4052	3895	4074	3956
NS 7338 IPRO	-168	3850	192	4209	34	3822	3744	3858	3807
NS 5727 IPRO	-277	3740	158	4175	37	3694	3492	3676	3597
NS 6006 IPRO	85	4103	319	4337	23	4117	3883	4087	4024
NS6060IPRO	-1048	2969	33	4050	45	2796	2473	2764	2546
NS6700IPRO	179	4196	366	4383	19	4225	4187	4276	4243
NS6906IPRO	375	4392	524	4541	8	4454	4356	4496	4416
TMG2158IPRO	91	4108	330	4347	22	4123	3778	4059	3960

*HMGV, harmonic mean of the genotypic values; RPGV*OM, relative performance of the genotypic values multiplied in all environments multiplied by the overall mean; HMRPGV*OM, harmonic mean of the relative performance of the genotypic values multiplied in all environments multiplied by the overall mean.

The NA 5909 RG, NS6823RR, M6410IPRO, NS 6767, M5917IPRO, and NS 5959 IPRO cultivars had the highest RPGV*OM values. Selection using RPGV*OM enables the identification of the most adapted genotypes by capitalizing on the ability of each genotype to respond favorably to an improvement in the production environment. Furthermore, this parameter is associated with average production rate, which enables the identification of genotypes that are both well adapted and productive. This method can be compared to the method reported by Annicchiarico

(1992), because it uses relative performance. However, these two methods differ in their measurement of adaptability, which is performed genotypically in the RPGV*OM and phenotypically in the method by Annicchiarico (CARBONELL et al., 2007).

The NA 5909 RG, M6410IPRO, NS6823RR, and NS 5959 IPRO cultivars had the highest values, based on the HMRPGV*OM method, which indicates that they are, simultaneously, the most productive, stable, and adapted to the study sites. The BMX Ativa RR, NS6060IPRO, and A 6411RG cultivars had the worst crop performances, adaptability, and stability. This method has the advantage of assessing the relative performance of genotypes in the genotypic context, unlike other widely used methods, including the methods by Lin & Binns (1988) and Annichiarico (1992), which analyze the values in the phenotypic context (BORGES et al., 2010).

In the total set of cultivars, NA 5909 RG (7), NS 5959 IPRO (34), and M6410IPRO (23) had the highest average production rates, based on the GGE biplot method (Figure 4.1).

The classification is performed in relation to the single-arrow line indicating that the farther on the right it is, the higher the genotype average will be. The AS 3570IPRO, NS 6209, 6563RSF IPRO, and NA 5909 RG cultivars were the most stable, because they showed a small projection in relation to the two-arrow line. However, these genotypes are considered to respond poorly to environmental changes. The AS 3570IPRO cultivar failed to show either high stability or average production rate, thereby failing to meet the breeding objectives. However, the NA 5909 RG cultivar had adequate values of both characteristics.

Among the early cultivars, BMX Energia RR and DMario 58i had the highest average production rates, and the NS 4901 was the most stable cultivar. NA 5909 RG, NS 5959 IPRO, and M5917IPRO were the most productive, and 5958RSF IPRO the most stable medium-cycle cultivars. Among the late-cycle cultivars, M6410IPRO had the best crop performance associated with high stability. Similarly, the NS 6767 and NS6823RR cultivars were also productive and stable. The A 6411RG, NS 7237 IPRO, and NS 7338 IPRO cultivars had high stability.

However, they had the worst crop performances. Stability is measured biologically by the GGE biplot method; that is, the genotype has a consistent performance among all the environments, but fails to respond to environmental

improvements (JAMSHIDMOGHADDAM; POURDAD, 2013).

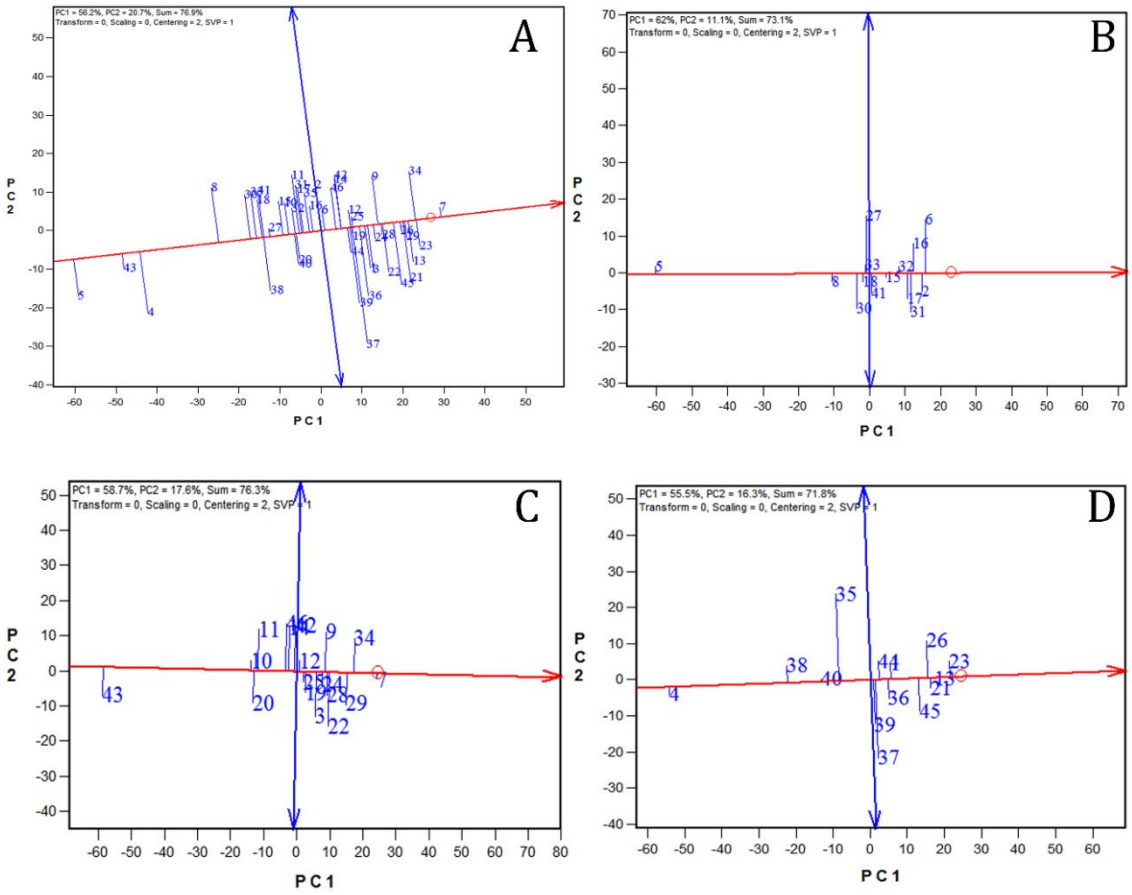


Figure 4.1 – Mean and stability for the 46 soybean cultivars (A), and for the cultivars division in early (B), medium (C) and late cycle (D), assessed in eight locations (seven of which in the State of Paraná (Cambé, Corbélia, Mamborê, Palotina, Realeza, São Jorge do Ivaí e São Miguel do Iguaçu) and one in State of São Paulo (Cândido Mota)), in 2014/2015 crop season. PC: main component. Cultivars: BMX Potência RR (1), DMario 58i (2), NK 7059 RR (3), A 6411RG (4), BMX Ativa RR (5), BMX Energia RR (6), NA 5909 RG (7), NS 4823 (8), BMX Turbo RR (9), NS 5858 (10), NS 6262 (11), SYN1059 RR (12), NS 6767 (13), TMG 7262RR (14), NS 4901 (15), NS 5258 (16), NS 5290 (17), NS 5401 RR (18), NS 6209 (19), NS6121RR (20), NS6823RR (21), M6210IPRO (22), M6410IPRO (23), 5958RSF IPRO (24), 6458RSF IPRO (25), 6563RSF IPRO (26), AS 3570IPRO (27), AS 3610IPRO (28), M5917IPRO (29), NS 5000 IPRO (30), NS 5106 IPRO (31), NS 5151 IPRO (32), NS 5445 IPRO (33), NS 5959 IPRO (34), NS 6909 IPRO (35), NS 7000 IPRO (36), NS 7209 IPRO (37), NS 7237 IPRO (38), NS 7300 IPRO (39), NS 7338 IPRO (40), NS 5727 IPRO (41), NS 6006 IPRO (42), NS6060IPRO (43), NS6700IPRO (44), NS6906IPRO (45) e TMG2158IPRO (46).

The ideal cultivar—the closest to the center of the concentric circles—is defined based on high yield and stability criteria (YAN, 2015). Thus, in the combined analysis, the NA 5909 RG and M6410IPRO cultivars may be considered ideal (Figure 4.2). BMX Energia RR and DMario 58i stood out among all the early cultivars, and the NA 5909 RG, which proved ideal, stood out among the medium-cycle cultivars.

Among the late cultivars, M6410IPRO was the closest to the ideal cultivar. Identifying adapted and stable genotypes for a wide region enables breeders to use this source of germplasm towards developing new cultivars for adaptation to a wide range of environments.

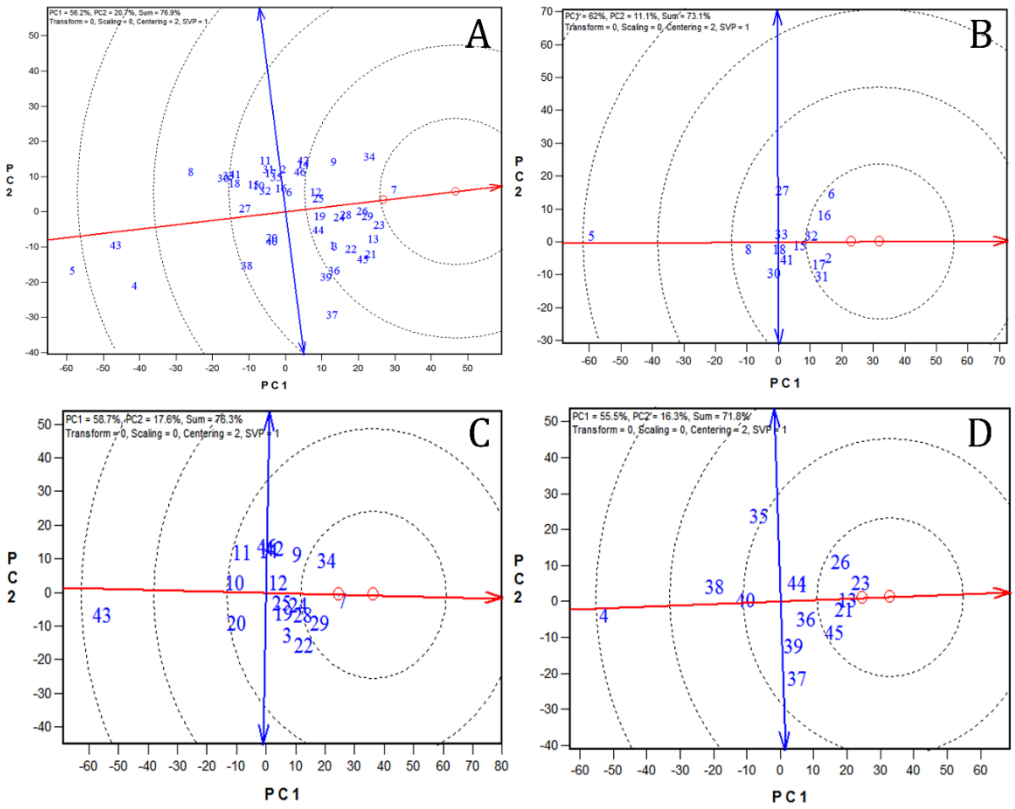


Figure 4.2. Ideal genotype for the 46 soybean cultivars (A), and for the cultivars division in early (B), medium (C) and late cycle (D), assessed in eight locations (seven of which in the State of Paraná (Cambé, Corbélia, Mamborê, Palotina, Realeza, São Jorge do Ivaí e São Miguel do Iguaçu) and one in State of São Paulo (Cândido Mota)), in 2014/2015 crop season. PC: main component. Cultivars: BMX Potência RR (1), DMario 58i (2), NK 7059 RR (3), A 6411RG (4), BMX Ativa RR (5), BMX Energia RR (6), NA 5909 RG (7), NS 4823 (8), BMX Turbo RR (9), NS 5858 (10), NS 6262 (11), SYN1059 RR (12), NS 6767 (13), TMG 7262RR (14), NS 4901 (15), NS 5258 (16), NS 5290 (17), NS 5401 RR (18), NS 6209 (19), NS6121RR (20), NS6823RR (21), M6210IPRO (22), M6410IPRO (23), 5958RSF IPRO (24), 6458RSF IPRO (25), 6563RSF IPRO (26), AS 3570IPRO (27), AS 3610IPRO (28), M5917IPRO (29), NS 5000 IPRO (30), NS 5106 IPRO (31), NS 5151 IPRO (32), NS 5445 IPRO (33), NS 5959 IPRO (34), NS 6909 IPRO (35), NS 7000 IPRO (36), NS 7209 IPRO (37), NS 7237 IPRO (38), NS 7300 IPRO (39), NS 7338 IPRO (40), NS 5727 IPRO (41), NS 6006 IPRO (42), NS6060IPRO (43), NS6700IPRO (44), NS6906IPRO (45) e TMG2158IPRO (46).

The methods of identifying ideal genotype via GGE, and stability via HMGV, consistently identified NA 5909 RG and M6410IPRO as superior cultivars. These two methods are not always consistent in identifying adapted and stable genotypes. Yang et al., (2009) indicated that their simultaneous use is advantageous,

because the assessment of those parameters analyzes the phenotype when using GGE, and the genotype when using mixed models. These methods also showed agreement regarding the cultivars with the worst performances, wherein BMX Ativa RR, NS6060IPRO, and A 6411RG were the least stable and productive.

4.4 Conclusion of Agronomic Performance of Modern Soybean Cultivars in Multi-Environment Trials

i. The NA 5909 RG, M6410IPRO, NS 5959 IPRO, NS6823RR, M5917IPRO, NS 6767, and 6563RSF IPRO cultivars are the most productive in the study environments, and the BMX Ativa RR cultivar has the worst crop performance.

ii. The NA 5909 RG, NS6823RR, M6410IPRO, and NS 5959 IPRO cultivars show high yield, adaptability and stability and may be considered ideal for cultivation in the study sites.

iii. There are modern soybean cultivars ideal for cultivation in the Brazilian soybean’s microregions of adaptation 102, 201, and 202 (all indetermined).

5 GENOMIC SELECTION IN SOYBEAN: ACCURACY AND TIME GAIN IN RELATION TO PHENOTYPIC SELECTION

5.1 Introduction of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection

The genetic gain for soybean cultivation (*Glycine max* L. Merrill) has varied between 0.5% and 1.8% per year in different countries (KARMAKAR & BHATNAGER, 1996; MORRISON et al., 1999; KOESTER et al., 2014; de FELIPE et al., 2016), with an average gain of 1.3% per year (RAY et al., 2013). The average annual genetic gain of the major producing countries (Brazil, United States, and Argentina) is 1.1% (DE FELIPE et al., 2016). To meet the demand for soy protein for both human consumption and animal feed that is projected for 2050 (9 billion people and increased purchasing power of developing countries), soybean production must increase by approximately 55% (RAY et al., 2013; HATFIELD and WALTHALL, 2015) and would need to reach 282 million tons (ALEXANDRATOS & BRUINSMA, 2012, HATFIELD and WALTHALL, 2015). However, the current genetic approaches are inadequate to meet this future demand, necessitating new breeding techniques such as genomic selection (GS).

GS consists of the application of statistical models to predict the breeding value of individuals in a population (SPINDEL et al., 2015). To this end, a large number of markers spread over the entire genome is required to encompass polymorphisms responsible for phenotypic variation. GS differs from traditional marker-assisted selection by the fact that it is not limited to previously selected markers; instead, breeding values can be predicted based on all available marker data (MEUWISSEN, 2007; SPINDEL et al., 2015). GS incorporates all marker effects or genetic loci in the whole genome to estimate genomic breeding values (GEBVs) (MEUWISSEN et al., 2001; SHU et al., 2013). One of the major advantages of GS in crop breeding is that it reduces the time required to complete breeding cycles (HICKEY et al. 2014), thereby accelerating genetic improvement per unit time. Soybean is particularly amenable to GS due to its moderate genome size and the rapid progress in soybean genome sequencing (SCHMUTZ et al. 2010; MA et al., 2016).

Several statistical models have been developed for GS analysis including the BayesB model, which uses a Bayesian approach and parametric regression (GIANOLA, 2013, DESTA and ORTIZ, 2014) and takes into account linkage effects between alleles of different loci, which may generate a spurious disequilibrium effect. In this manner, BayesB allows indirect correction of the effect of population structure; it also identifies quantitative trait loci in disequilibrium in an additive manner based on markers (XAVIER et al., 2016). This method also has the advantage that it allows markers to have large and/or null effects (HABIER et al. 2011). BayesB has already been used for GS in economically important species such as wheat (HEFFNER et al., 2011; THAVAMANIKUMAR et al., 2015) and soybean (XAVIER et al., 2016). In the latter, the BayesB model had the highest accuracy in genomic prediction for genetic architecture and heritability characteristics such as grain yield, days to maturity, plant height, pod number, node number, and pods per node (XAVIER et al., 2016). It was reported that the BayesB model enabled better prediction of grain yield in wheat in different environments as compared to the reproducing kernel Hilbert space (RKHS) kernel method (PÉREZ-RODRÍGUEZ et al., 2012) and was effective for selecting various traits in this species (HEFFNER et al., 2011).

Evaluating the influence of population structure on the accuracy of GEBV estimates and accuracy is important. The use of a more generalized set of unrelated genotypes makes GS more flexible and relevant in plant breeding programs. Results from several recent studies suggest that population structure should be considered when assessing the potential of GS (GUO et al., 2014; ISIDRO et al., 2015; MA et al., 2016), since it can influence the accuracy of the prediction depending of the evaluated trait(s); for instance, the accuracy within each subpopulation can be overestimated relative to the total population (Ma et al., 2016).

Estimating the gains from GS in relation to phenotypic selection is important for plant breeding programs (OLIVEIRA et al., 2012). The reduction in time spent on breeding programs can be considered as an estimate of the efficiency of GS. Current programs in Brazil run for several generations per year with the goal of releasing new cultivars more rapidly. About 6 years is required to obtain a new cultivar from initial crosses to lines for value for cultivation and use (VCU) trials. Therefore,

reducing the time to generate a new cultivar is desirable in breeding programs.

The objectives of this study were to apply the BayesB GS model to 243 recombinant inbred lines (RILs) and 81 selected varieties using 4947 genome-wide single nucleotide polymorphisms (SNPs) in order to (1) assess the prediction accuracy for grain yield (GY), 1000-grain weight (TGW), plant height (PH), days to maturity (DM), and insertion of the first pod (IFP); and; (2) evaluate the effect of intrapopulation structure on GS accuracy; and (3) compare the efficiencies of phenotypic selection and GS in soybean.

5.2 Materials and Methods of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection

5.2.1 Plant material, field experiments, and phenotyping

A total of 324 genotypes of 243 RILs obtained from crossing the NA 5909RG cultivar and the genetically dissimilar NS 5000 IPRO cultivar were used in field experiments. Both cultivars are high yielding and NA 5909RG is adapted to a waste region in South Central Brazil whereas NS 5000 IPRO is adapted to South Brazil environments. The other 81 genotypes consisted of elite material; 46 were already registered in Brazil while the others were pre-release cultivars.

Field experiments were conducted in a randomized complete block design with three replications, and were performed at eight locations during the 2014/15 crop season (Figure 5.1 and Table 5.1). Seeds were sown in September 2014 in four-row plots (5.0 m long), with 0.50-m spacing between rows, at a density of 30 plants m^{-2}. Base fertilization consisted of 350 kg ha^{-1} of nitrogen:phosphorus:potassium (02:20:20). Standard practices for soybean cultivation were used at all locations. The following traits were evaluated: GY (in kg ha−1) was assessed in the two central rows of each plot (5 m2 useful area), with grain moisture corrected to 13% (wet basis); TGW (in g) was estimated from seeds harvested in the useful area from three replicates of 200 grains; PH (in cm) was measured for 10 plants per location at R7 stage; DM was measured when plants reached R8 stage; and IFP (in cm) was measured in 10 plants at each location and took into consideration the

ground level before harvest. For PH, DM, and IFP only one replication was evaluated per location since all genotypes were homozygous.

Figure 5.1 – The eight geographical locations in this study. CM, Cândido Mota; COR, Corbélia; MAB, Mamborê; PAL, Palotina; REA, Realeza; CAM, Cambé; SJI, São Jorge do Ivaí; SMI, São Miguel do Iguaçu. PR, Paraná State; SP, São Paulo State.

Table 5.1 – Geographical and climatic information on the eight locations selected for evaluation of 324 soybean genotypes in the 2014/15 crop season.

Location	Latitude / Longitude	Altitude (m)	Climate
Cambé-PR	23°16' S 51°16' W	520	Cfa
Cândido Mota-SP	22°44' S 50°23' W	440	Cwa
Corbélia-PR	24°47' S 53°18' W	650	Cfa
Mamborê-PR	24°19' S 52°31' W	715	Cfa
Palotina-PR	24°17' S 53°50' W	330	Cfa
Realeza-PR	25°46' S 53°31' W	520	Cfa
São Jorge do Ivaí-PR	23°25' S 52°17' W	560	Cfa
São Miguel do Iguaçu-PR	25°20' S 54°14' W	290	Cfa

PR: Paraná State; SP: São Paulo State; Cfa: Humid subtropical climate; Cwa: Monsoon-influenced humid subtropical climate according to Köppen–Geiger climate classification (Peel et al., 2007)

5.2.2 Genotyping

The plants were sown in a greenhouse under controlled temperature (25°C ± 3°C), humidity (60% ± 10%), and natural photoperiod in the summer of 2014/15. Completely developed leaves of 20 plants of each genotype were collected at the V6 stage to obtain genetic material for genotyping. The collected leaves were immediately frozen in liquid nitrogen and stored at −80°C until use. DNA was extracted according to a previously published protocol (ALJANABI & MARTINEZ, 1997). DNA quantity and quality were analyzed on a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and DNA integrity was verified on a 0.8% agarose gel.

Genotyping was performed using the Soy 6K Infinium chip SNP panel (Illumina, San Diego, CA, USA) supplied by Eurofins BioDiagnostics (River Falls, WI, USA). A total of 5403 SNPs were obtained. Filtering was applied and markers with 0.01< minor allele frequency < 0.99 and > 10% missing data were removed. The missing genotypic data were replaced with corresponding average values for each locus. After editing, 4947 SNPs were ultimately retained for analysis.

5.2.3 Statistical analysis

5.2.3.1 Phenotypic data

Phenotypic data were analyzed using linear mixed models to estimate genetic values (empirical best linear unbiased prediction, eBLUP) for each individual. This method allows estimation of variance components and heritability for each trait. The analysis was performed using the lme4 v.1.1-8 package (BATES et al., 2015), considering as fixed effects the replication within each location (for TGW and GY) and only location information for PH, IFP, and DM. Genotype effects were considered as random (containing information from individuals). The statistical model was as follows:

$$y = X\beta + Zu + \varepsilon$$

where y is the phenotypic value obtained for the analyzed treatment; β is the vector of fixed effects, u is the vector of random effects; X and Z are the incidence matrices of the fixed (containing replication information within the evaluation site) and random (containing information from individuals) effects, respectively; and ε is the vector of random errors associated with elements of vector y, with parameters and relationships described as follows:

$$V = ZGZ' + R;\ G = A\sigma_a^2;\ \text{and}\ R = I\sigma_a^2$$

where G is the matrix of genetic variances and covariances; A is the matrix of random effects correlations; I is an identity matrix; σ_a^2 is the estimation of variance of the random effects of individuals; and σ_e^2 is the variance estimate of the residuals. The structure of averages, variances, and covariances is as follows:

$$\begin{bmatrix} y \\ u \\ e \end{bmatrix} NMV \left\{ \begin{bmatrix} x\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V' & GZ & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \right\}$$

Heritability was calculated using estimates of variance components obtained from the restricted maximum likelihood model. Thus, heritability (h^2) was determined as the ratio between the additive variance of individuals (σ_a^2) and the total phenotypic variance (additive + environmental variances) (σ_a^2 + σ_e^2).

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

5.2.3.2 Linkage disequilibrium (LD) – methodology

The decay of LD for the analyzed population was calculated using TASSEL v.5 software (BRADBURY et al., 2007) taking into consideration loci with 0.01 < frequency < 0.99. Heterozygous loci were discarded and LD was computed in windows with intervals of 1000 markers. A graph illustrating the relationship between the LD coefficient (r2) and distance (bp) was generated using the ggplot2 package (WICKHAM, 2009) in R (R CORE TEAM, 2016).

5.2.3.3 Intra-population structural analysis

Correspondence was observed between data obtained for genotyping and phenotyping, with viable data for all 324 individuals of the original population. A comparison matrix was used to compute the variance at each locus, and these values were used for the main coordinate analysis using the ecodist package (GOSLEE et al., 2007), which was performed in order to verify the allocation of individuals given the intrapopulation structure.

In addition to principal coordinates analysis, the data were analyzed by hierarchical clustering. An array of Euclidean distances was generated; the diagonal matrix of genetic distances was subjected to cluster analysis based on the unweighted pair group method with arithmetic mean, and a dendrogram of genetic distances was constructed using the hclust command in R (R CORE TEAM, 2016).

5.2.3.4 GS analysis

Data from 4947 SNP markers were used to construct a genotype matrix for the 324 individuals of the original population. GS analysis was performed using the BayesB model (MEUWISSEN et al., 2001), which was applied using the Bayesian generalized linear regression package (PÉREZ & DE LOS CAMPOS, 2014) in R (R CORE TEAM, 2016). The analysis was performed by first considering the total set of individuals as a single population (3) and then using matrices related to the effects of markers in a stratified manner. This analysis considered matrices constructed from information unique to individuals in the identified groups (4). The stratification corrected for the effects of intrapopulation structure based on correction for the population effect in the principal coordinates analysis. Thus, the GS analysis used the following models:

Homogeneous regression (3):

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_0 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Stratified regression (4):

$$\begin{bmatrix} \acute{y}_1 \\ \acute{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ 0 \end{bmatrix} \beta_1 + \begin{bmatrix} 0 \\ X_2 \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

where the matrices containing the marker information (X_1 and X_2) interact either together (homogeneous regression) or separately (stratified regression) to predict the effects of the markers ($\beta_0, \beta_1, \beta_2$) in the analysis to obtain GEBVs ($\acute{y}_1, \acute{y}_2$).

5.2.3.5 Model accuracy

Accuracy values were determined based on the correlation between GEBV predicted by jackknife and eBLUPs estimated for each genotype with the linear mixed model. Cross-validation was used to evaluate the accuracy of the GS model. The jackknife procedure was based on the 10-fold-validation methodology. Thus, the dataset was randomly divided into 10 equal parts. GEBV prediction was performed for one of the groups from values generated with the training set composed of the other nine groups. GEBV was obtained through 1,000,000 iterations by random sampling of the training and validation models with a burn-in of 50,000 and a value of thinning default equal to 5 to avoid autocorrelation between sampled points. The best values for these parameters were verified based on the convergence of genetic and residual variance graphs.

5.2.3.6 Deviance analysis

The deviances for each trait and for homogeneous and stratified population structures were compared based on the average of five deviance information criterion values. The analysis was performed with the following parameters: 1,000,000 iterations, burn-in of 50,000, and value of thinning equal to five (the same values used for GEBV), and differed only by the omission of the jackknife technique.

5.2.3.7 Efficiency of genomic selection

The duration of a soybean selection cycle is approximately 6 years. In this study, we investigated the gain in selection efficiency using GS as compared to phenotypic selection. GS efficiency was measured for the same period as phenotypic selection and for a reduction in selection from 6 years to 1 year in a year-by-year manner. The efficiency of GS in relation to phenotypic selection was calculated according to a previous report (OLIVEIRA et al., 2012):

$$Efficiency = \left(\frac{GA * PB}{PA * GB}\right)$$

where GA is genotypic accuracy; GB is genomic breeding; PA is phenotypic accuracy; and PB is phenotypic breeding. Efficiency was calculated for homogeneous and stratified population structures.

5.3 Results of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection

5.3.1 Population structure analysis

For the evaluated genotypes, the LD between pairs of SNPs declined sharply to r^2 = 0.1 at around 7000 kb (Figure 5.2).A total of 4947 SNPs were used to verify the population structure of the 324 genotypes. The first two principal coordinates explained 52.25% of the molecular variation present in the genotypes (Figure 5.3), with principal coordinates 1 and 2 accounting for 38.66% and 13.59%, respectively, of the total variance. The scatterplot revealed the formation of two subgroups, G1 and G2; the former comprised RILs generated by crossing cultivars NA 5909RG and NS 5000 IPRO, and the latter comprised elite cultivars and lines from VCU trials. The same two subgroups were observed in the dendrogram (Figure 5.4).

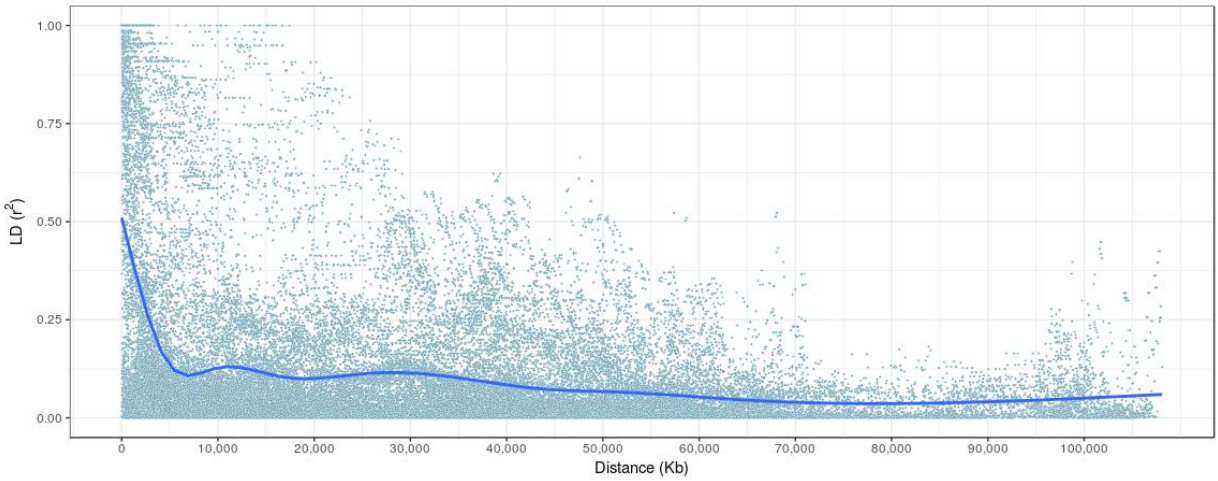


Figure 5.2 – Decay of LD (r2) with physical map distances between markers in the examined genotypes, as determined by locally weighted polynomial regression.

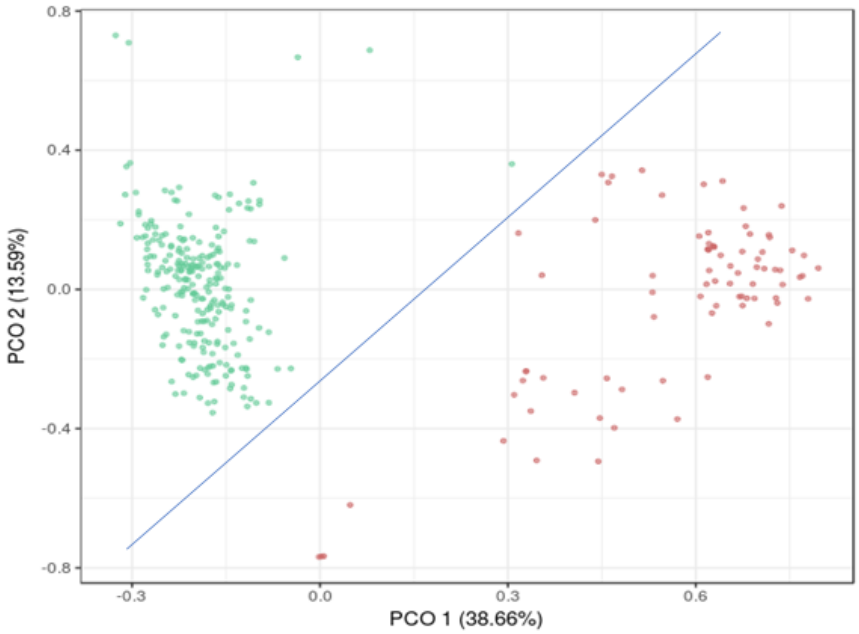


Figure 5.3 – Scatterplot of the two first eigenvalues in a principal coordinates analysis of 4947 SNPs in 324 Brazilian soybean genotypes clustered into RIL (G1, in green) and elite genotype (G2, in red) subpopulations.

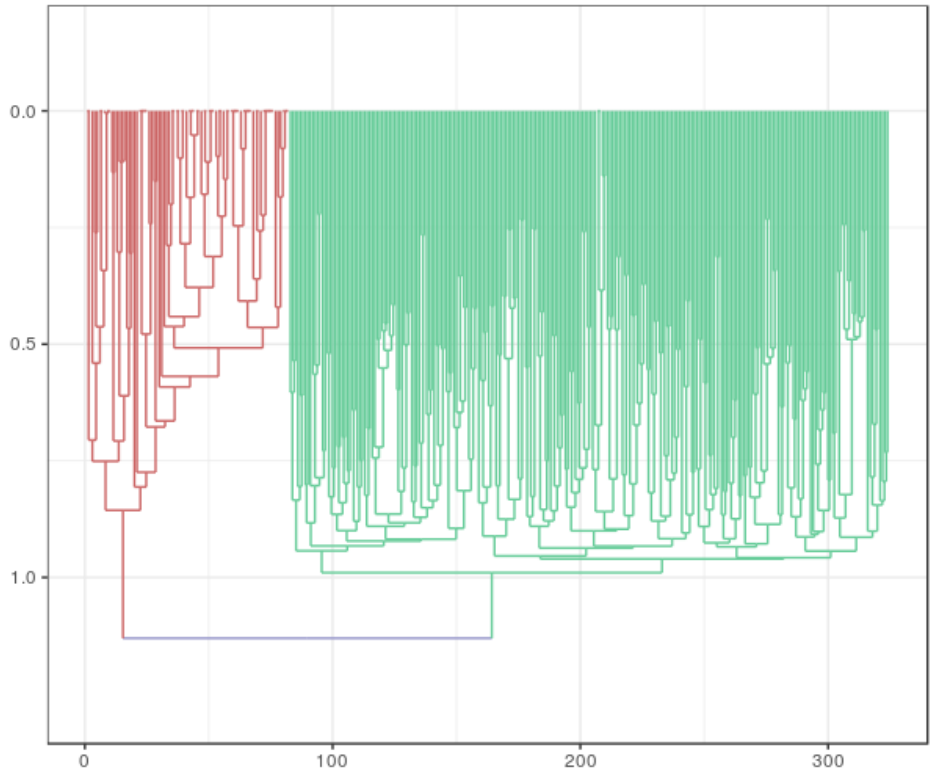


Figure 5.4 – Dendrogram obtained by cluster analysis of data from 4947 SNP markers in 324 individuals (divided into two subgroups): G1 (243 RILs, in green) and G2 (81 elite genotypes, in red) included in the GS.

5.3.2 eBLUP for evaluated traits

The eBLUP indicated that the distribution of traits for the G1 and G2 subpopulations was close to a Gaussian distribution. In addition, predicted values for GY, PH, DM, and IFP were higher for G1 than for G2 (Figure 5.5). GY is the most important trait in soybean and therefore, the focus of most breeding programs. The highest GY value for G1 was 4500 kg ha^{-1} as compared to 3750 kg ha^{-1} for G2. This indicates that the breeding program could identify the most productive genotypes in crosses. PH was higher in RILs than in cultivars (95 vs. 80 cm). The DM of G1 was 110–115 days as compared to 105 days for G2. IFP for RILs and cultivars was 12 and 10 cm, respectively. In contrast to the other traits, average TGW was lower for RILs than for cultivars (170 vs. 175 g).

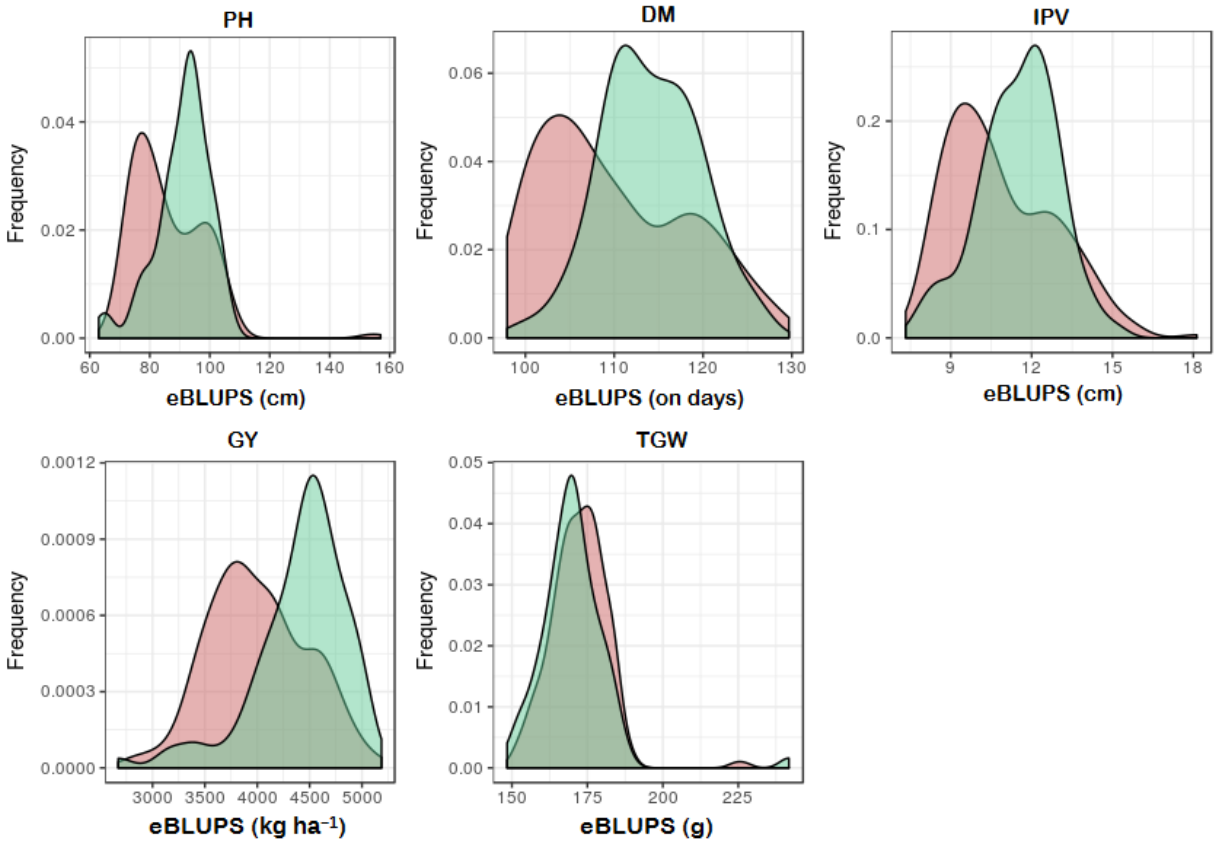


Figure 5.5 – eBLUPs predicted by a mixed linear model (lme4) for the analyzed traits. Datasets corresponding to RILs and elite genotypes are shown in green and red, respectively.

5.3.3 Variance and heritability

The highest variance, both additive and environmental, was obtained for GY (Table 5.2). Narrow-sense heritability (h^2) ranged from 0.73 for DM to 0.10 for TGW. The h^2 for GY was 0.42. PH, IFP, and TGW had h^2 values of 0.266, 0.280, and 0.102, respectively. The variation in magnitude can be explained by differences in the complexity of genetic control of the traits.

Table 5.2 – Coefficients of genetic and environmental variances and heritability determined from restricted maximum likelihood model estimates

Trait	GV	EV	h^2
Plant height	183.54	506.39	0.27
Days to maturity	63.29	22.91	0.73
Insertion of the first pod	4.58	11.77	0.28
Grain yield	261246.26	362821.72	0.42

Thousand grain weight	152.45	1334.29	0.10

5.3.4 Accuracy of genomic prediction

The accuracy of GS was similar regardless of whether population structure was considered (Table 5.3). DIC values were also similar in five independent replicates both with and without correction for population structure (Table 5.4).

Table 5.3 – Accuracy of GS analysis for five traits in soybean based on the BayesB model and eBLUP considering homogeneous or stratified population structures

Trait	**Accuracy**	
	Homogeneous	**Stratified**
Plant height	0.6830	0.6801
Days to maturity	0.8327	0.8327
Insertion of the first pod	0.7178	0.7286
Grain yield	0.7224	0.7291
Thousand grain weight	0.5010	0.4902

Table 5.4 – Deviance information criterion values obtained for GS analysis*

Rep	**Plant Height**		**Days to Maturity**		**Insertion of the First Pod**		**Grain Yield**		**Thousand Grain Weight**	
	Homog.	**Stratified**	**Homog.**	**Stratified**	**Homog.**	**Stratified**	**Homog.**	**Stratified**	**Homog.**	**Stratified**
1	2305.14	2308.68	1839.90	1828.69	1074.53	1076.96	4656.39	4655.92	2278.96	2279.49
2	2305.40	2306.02	1839.52	1830.69	1074.52	1078.04	4654.19	4650.35	2278.83	2279.47
3	2305.08	2305.82	1840.10	1828.81	1074.72	1077.76	4662.25	4651.21	2278.46	2277.03
4	2304.90	2305.11	1839.83	1828.44	1074.22	1076.91	4661.28	4652.66	2278.82	2276.98
5	2305.15	2306.85	1839.74	1829.40	1074.99	1076.71	4654.74	4649.77	2278.41	2279.51
Mean	**2305.13**	**2306.50**	**1839.82**	**1829.21**	**1074.60**	**1077.27**	**4657.77**	**4651.98**	**2278.70**	**2278.50**

*Homogeneous and stratified analyses were performed for traits of agronomic importance in soybean

5.3.5 Superiority of genomic selection

The maximum accuracy of GS was higher than that of phenotypic selection (Table 5.5). Even for a selection cycle of 6 years, GS was superior in terms of PH, IFP, GY, and TGW traits, whereas accuracy for DM was 2.8% higher for phenotypic selection than for GS.

Considering a homogeneous population structure, GS was superior to

phenotypic selection for PH, DM, IFP, GY, and TGW (164%, 94%, 171%, 123%, and 212%, respectively) when the length of the selection cycle was reduced from 6 to 3 years. When population structure was considered (stratified analysis), GS was superior to phenotypic selection for PH, DM, IFP, GY, and TGW (163%, 94%, 175%, 125%, and 206%, respectively).

Table 5.5 – Accuracy of GS vs. phenotypic selection

PB*	GB	Homogeneous				Stratified			
		Plant Height							
		PA	GA	EFF	SUP	PA	GA	EFI	SUP
6	6	0.5157	0.6830	1.3244	32.4413	0.5157	0.6801	1.3188	31.8790
6	5	0.5157	0.6830	1.5893	58.9296	0.5157	0.6801	1.5825	58.2548
6	4	0.5157	0.6830	1.9866	98.6620	0.5157	0.6801	1.9782	97.8185
6	3	0.5157	0.6830	2.6488	164.8827	0.5157	0.6801	2.6376	163.7580
6	2	0.5157	0.6830	3.9732	297.3240	0.5157	0.6801	3.9564	295.6370
6	1	0.5157	0.6830	7.9465	694.6481	0.5157	0.6801	7.9127	691.2740
PB	**GB**	**Days to Maturity**							
		PA	GA	EFF	SUP	PA	GA	EFI	SUP
6	6	0.8567	0.8327	0.9719	-2.8067	0.8567	0.8327	0.9720	-2.8037
6	5	0.8567	0.8327	1.1663	16.6320	0.8567	0.8327	1.1664	16.6356
6	4	0.8567	0.8327	1.4579	45.7899	0.8567	0.8327	1.4579	45.7945
6	3	0.8567	0.8327	1.9439	94.3866	0.8567	0.8327	1.9439	94.3927
6	2	0.8567	0.8327	2.9158	191.5799	0.8567	0.8327	2.9159	191.5890
6	1	0.8567	0.8327	5.8316	483.1598	0.8567	0.8327	5.8318	483.1780
PB	**GB**	**Insertion of First Pod**							
		PA	GA	EFF	SUP	PA	GA	EFI	SUP
6	6	0.5291	0.7178	1.3566	35.6643	0.5291	0.7286	1.3771	37.7055
6	5	0.5291	0.7178	1.6280	62.7972	0.5291	0.7286	1.6525	65.2466
6	4	0.5291	0.7178	2.0350	103.4965	0.5291	0.7286	2.0656	106.5583
6	3	0.5291	0.7178	2.7133	171.3287	0.5291	0.7286	2.7541	175.4111
6	2	0.5291	0.7178	4.0699	306.9930	0.5291	0.7286	4.1312	313.1166
6	1	0.5291	0.7178	8.1399	713.9860	0.5291	0.7286	8.2623	726.2332
PB	**GB**	**Grain Yield**							
		PA	GA	EFF	SUP	PA	GA	EFI	SUP
6	6	0.6470	0.7224	1.1165	11.6538	0.6470	0.7291	1.1269	12.6893
6	5	0.6470	0.7224	1.3398	33.9845	0.6470	0.7291	1.3523	35.2272
6	4	0.6470	0.7224	1.6748	67.4807	0.6470	0.7291	1.6903	69.0340
6	3	0.6470	0.7224	2.2331	123.3076	0.6470	0.7291	2.2538	125.3787
6	2	0.6470	0.7224	3.3496	234.9614	0.6470	0.7291	3.3807	238.0680
6	1	0.6470	0.7224	6.6992	569.9227	0.6470	0.7291	6.7614	576.1360
PB	**GB**	**Thousand Grain Weight**							
		PA	GA	EFF	SUP	PA	GA	EFI	SUP
6	6	0.3202	0.5010	1.5646	56.4647	0.3202	0.4902	1.5309	53.0918
6	5	0.3202	0.5010	1.8776	87.7577	0.3202	0.4902	1.8371	83.7102
6	4	0.3202	0.5010	2.3470	134.6971	0.3202	0.4902	2.2964	129.6377

6	3	0.3202	0.5010	3.1293	212.9294	0.3202	0.4902	3.0618	206.1836
6	2	0.3202	0.5010	4.6939	369.3941	0.3202	0.4902	4.5928	359.2755
6	1	0.3202	0.5010	9.3879	838.7883	0.3202	0.4902	9.1855	818.5509

*DM, days to maturity; EFF, efficiency [(GA*PB)/(PA*GB)]; GA, genomic accuracy; GB, genomic breeding; GY, grain yield; IFP, insertion of the first pod (IFP); PA, phenotypic accuracy; PH, plant height; PB, phenotypic breeding; SUP, superiority of genomic selection in % (Oliveira et al., 2012); TGW, 1000-grain weight (OLIVEIRA et al., 2012).

5.4 Discussion of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection

5.4.1 Population structure

Population structure is an important consideration in GS analysis (ISIDRO et al. 2015; MA et al., 2016). In this study, principal coordinates analysis and the dendrogram revealed two subpopulations—one formed by RILs and the other by cultivars and advanced lines—among the examined genotypes. The population structure can generate spurious LD and can artificially inflate GS accuracy (MA et al., 2016). This can be avoided by correcting for population structure. We used two strategies to evaluate the effect of population structure; the first involved the use of the BayesB model, which considers the effects of LD associated with linkage and thereby rules out the effects of spurious LD, and the second was by stratified regression.

The prediction accuracy did not change substantially when population structure was considered, as evidenced by the similar DIC values for all evaluated traits. These results indicate that structural effects such as those observed for this population do not necessarily influence selection accuracy. The lack of differences among regression models with and without consideration of population structure can also be attributed to the use of the BayesB model for the analysis. Thus, this model can eliminate false positives and be used for GS in populations that exhibit structure effects comparable to those observed in the present analysis. Simpler models can also be used for selection based on the accuracy of GS. However, further analysis is required to validate the effectiveness of this model. Prediction accuracy was reduced to 5.27% for PH and 67.07% for yield per plant with the ridge regression (RR)-BLUP model (MA et al., 2016). An analysis using the RR-rrBLUP, Bayesian RR, RKHS, and

expectation–maximization models also highlighted the need to consider population structure to improve prediction accuracy (HUANG et al., 2016).

The set of cultivars used in this study occupies a large part of the area currently cultivated with soybean in the South-Central region of Brazil. Advanced lines, however, are materials that are currently being tested in VCU trials and may be released as new cultivars for the same region in the coming years. The RILs originated from a simple cross between two cultivars that are widely available in Brazil (NA 5909RG) and Argentina (NS 5000 IPRO). Thus, our attempt to develop GS in this population has real applications for the breeding programs of these populations. It should be noted that populations in breeding programs usually have a certain structure due to selection and drift as well as inbreeding. Thus, it is interesting to apply GS models to actual breeding populations although it is not the best strategy for developing GS models. Our results demonstrate that it is not necessary to develop populations in attempts to introduce GS to a soybean breeding program.

5.4.2 Prediction values

Selection for specific traits may benefit from GS. The selective accuracy was high for all traits examined here regardless of whether they were controlled by a large or small number of genes. GS can also be useful for altering the selection trends observed in a breeding program.

The use of descriptive statistics can demonstrate how a breeding program is being addressed, since the obtained results reflect the selection processes carried out in the program. There are several important traits in soybean breeding, including GY, PH, IFP, DM, and TGW. TGW is one of the main components of GY. IFP directly influences mechanical harvesting, with a larger value facilitating harvesting in steep areas, thereby reducing crop losses. DM defines the sowing and harvesting window so that the ideal period of development can be targeted and adverse periods avoided. This study is unprecedented because it evaluated the gain in efficiency by GS compared to phenotypic selection and identified the effects of population structure on GS efficiency using agronomic and adaptive traits.

The average yield of soybean in Brazil is 2,882 kg ha−1 (CONAB, 2017),

with productivity above 6000 kg ha−1 already being relatively common. The evaluated RILs have a productivity much higher than the national average and application of GS can help to further improve productivity, given the high selective accuracy obtained for GY. The reduction in TGW of RILs does not necessarily represent a problem, as this is only one of many factors that affect the GY of soybean crop (along with number of plants per area, number of pods per plant, and number of grains per pod) and can be compensated for by these other components. Ultimately, a lower TGW may lead to a reduction in the amount of seed needed for new planting, and can thus reduce costs for the farmer.

The BayesB model has already been used in several cultures and for traits with different genetic architecture and heritabilities (HEFFNER et al., 2011; THAVAMANIKUMAR et al., 2015; XAVIER et al., 2016). A prediction accuracy of rGS = 0.47 for GY was obtained using ridge RR-BLUP (MA et al., 2016). In our work, prediction accuracy was rGS = 0.72 for GY. BayesB was shown to yield better results than RR-BLUP in populations with low LD, high heritability, large sample size, and lower causal mutation relative to sample size (WIMMER et al., 2002; HABIER et al., 2007).

5.4.3 Heritability

Heritability is among the most important parameters in agriculture and plant breeding (VISSCHER et al., 2008) Traits with low heritability can only be improved slowly by traditional methods (FALCONER & MACKAY 1996). In this sense, GS accelerates genetic progress even for low-heritability traits (HEFFNER et al., 2009) since it allows better adjustment of models and has greater potential to detect genetic variance as compared to phenotypic selection.

5.4.4 Reduction in time required to complete a selection cycle

In soybean breeding programs it is desirable to reduce the time required between artificial hybridizations, the evaluation of the desired genotypes in VCU trials,

and the release of a new cultivar. GS is useful in this sense since it overcomes the limiting factors of time and financial cost in breeding programs. GS can replace the time-intensive phenotypic evaluation of complex traits with GEBVs and shorten breeding cycle length, thereby increasing gains per unit time (HEFFNER et al., 2009). In addition, gains may be achieved through elimination of intermediate steps between the identification of heterozygous genotypes and selection of homozygous lines (BASSI et al., 2016; ZHANG et al., 2017). In addition, costs can be reduced at each stage of the breeding program (ABELL et al., 2014; GORJANC et al., 2017) by eliminating intermediate steps such as driving segregating populations to appropriate regions.

5.4 Conclusions of Genomic Selection in Soybean: Accuracy and Time Gain in Relation to Phenotypic Selection

In this study, high selective accuracy was obtained by GS for all examined traits of soybean. We also showed that actual breeding populations can be used to determine GS models, without the need to develop specific populations for this purpose. Correcting for population structure by stratified regression combined with the use of the BayesB model ensured that selection accuracies did not vary significantly. Our results demonstrate that GS is superior to phenotypic selection, reducing the selection time by 50% increases the selection efficiency from 94% for Days to Maturity to 212% in Thousand Grain Weight. For GY, the increase in selection efficiency is shown at 123%.

6 CONCLUSIONS

Studies that consider the implementation of techniques aimed at the identification and selection of genotypes with high productive potential, both for use per se and hybridization purposes, are of great relevance in supporting the current breeding programs of plants. This demand is more incipient in much improved species, such as soybeans, where narrowing of the genetic base has limited genetic progress.

With this approach, considering the quantity/quality of data obtained during the experimentation process of this study, it was possible to obtain key answers for the selection processes used in soybean breeding programs: i. The results of this study allowed the identification of favorable sites for selection during the early generations in soybean breeding programs, when seed availability is generally low, and the number of genotypes is high. The environments of Chapada-RS and Maracaju-MS were considered the best environments for the macro-regions 1 and 2, respectively. ii. The results obtained in the present study allowed to infer that it is possible to use two replicates, and to obtain a selection accuracy above the range of high experimental accuracy in assays with 324 soybean genotypes. These results are important for formulating the GWS models, where the number of the phenotypically characterized genotypes is high. Moreover, the indicators of experimental accuracy provided by the Papadakis method are more favorable when compared to those of the randomized block and triple lattice designs. iii. The main cultivars with low and high productive performances were identified in the multi-environment assays, and it was possible to observe a wide variability in the characteristic yields of grains in a sample of the genotypes used in the GWS prediction assay. The cultivar NA 5909 RG, parental to the lineages used in the GWS prediction, presented high adaptability, stability, and grain yield. This cultivar can be considered as the ideal genotype for the analyzed environments. iv. High selection accuracy was obtained for all the traits studied, indicating the potential of using genomic selection in the breeding of the soybean crop. This suggests that it is possible to use real populations in whole genome predictions since the BayesB model is efficient in controlling population structuring in genomic selection.”

Additionally, under the conditions of this research, it was verified that genomic selection is superior to phenotypic selection for the evaluated traits. This methodology makes it possible to halve the selection time and increase the selection efficiency by 123% for grain yield. In this approach, the GWS methodology is recommended because of its high accuracy when used in soybean breeding programs, provided that phenotyping is highly accurate and the predictive models are formulated within the selection environment.

Finally, it is crucial to mention the importance of the present study for the development of future assays in soybean breeding programs, both for the prediction of genotypic values by the GWS methodology, and for improving the control of the experimental factors through the methodologies used.

REFERENCES

KARIMIZADEH, R.; ASGHARI, A.; CHINIPARDAZ, R.; SOFALIAN, O.; GHAFFARI, A. Application of GGE biplot analysis to evaluate grain yield stability of rainfed spring durum wheat genotypes and test locations by climatic factors in Iran. 2016. **Crop Breeding Journa**l, v. 4, p. 41–49. DOI:10.22092/cbj.2016.107106.

ABELL, C. E., DEKKERS, J. C., ROTHSCHILD, M. F., MABRY, J. W., STALDER, K. J. Total cost estimation for implementing genome-enabled selection in a multi-level swine production system. **Genetics Selection Evolution**, v. 46, n. 1, p. 32, 2014.

ALEXANDRATOS, N., AND J. BRUINSMA. World agriculture towards 2030/2050, the 2012 revision. ESA Working Paper 12–03, June 2012. Rome: Food and Agriculture Organization of the United Nations (FAO).

ALJANABI, S. M., MARTINEZ, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. **Nucleic acids research**, v.25, n. 22, p. 4692– 4693, 1997.

ANDRADE, A.C.B.; SILVA, A.J. da; FERRAUDO, A.S.; UNÊDA-TREVISOLI, S.H.; DI MAURO, A.O. Strategies for selecting soybean genotypes using mixed models and multivariate approach. **African Journal of Agricultural Research**, v. 11, p. 23–31, 2016. DOI: 10.5897/AJAR2015.9715.

ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Breeding**, v. 46, p. 269–278, 1992.

BARONA, M.A.A. et al. Epistasis for grain yield in soybeans. **Bragantia**, v. 68, n. 2, p. 313–318, 2009. Available from: <http://www.scielo.br/pdf/brag/v68n2/04.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0006-87052009000200004.

BASSI, F. M., BENTLEY, A. R., CHARMET, G., ORTIZ, R., CROSSA, J. Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). **Plant Science**, v. 242, p. 23–36, 2016.

BATES, D., MARTIN MAECHLER, BEN BOLKER, STEVE WALKEr. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n.1, p. 1–48, 2015.<doi:10.18637/jss.v067.i01>.

BATTENFIELD, S. D., C. GUZMÁN, R. C. GAYNOR, R. P. SINGH, R. J. PEÑA, S. DREISIGACKER, A. K. FRITZ, AND J. A. POLAND. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. **Plant Genome** , v. 9, 2016. doi:10.3835/plantgenome2016.01.0005

BENIN, G. et al. Improving the precision of genotype selection in wheat performance trials. **Crop Breeding and Applied Biotechnology**, v. 13, p. 234–240, 2013. Available from: <http://www.scielo.br/pdf/cbab/v13n4/a03v13n4.pdf>. Accessed: June 24, 2016. doi: 10.1590/S1984-70332013000400003.

BORGES, V.; SOARES, A.A.; REIS, M.S.; RESENDE, M.D.V.; CORNÉLIO, V.M.O.; LEITE, N.A.; VIEIRA, A.R. Desempenho genotípico de linhagens de arroz de terras altas utilizando metodologia de modelos mistos. **Bragantia**, v. 69, p. 833–841, 2010. DOI: 10.1590/S0006-87052010000400008.

BORGES, V.; SOARES, A.A.; RESENDE, M.D.V. de.; REIS, M. de S.; CORNELIO, V.M. de O.; LEITE, N.A.; SOARES, P.C.; COSTA JÚNIOR, G.T. Value for cultivation and use of upland rice cultivars tested in multi-environments. **Crop Breeding and Applied Biotechnology**, v. 12, p. 25–33, 2012. DOI: 10.1590/S1984-70332012000100004.

BrADBURY PJ, ZHANG Z, KROON DE, CASSTEVENS TM, RAMDOSS Y, BUCKLER ES. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, p. 2633-2635, 2007.

BRANQUINHO, R.G.; DUARTE, J.B.; SOUZA, P.I.M. de; SILVA NETO, S.P. da; PACHECO, R.M. Estratificação ambiental e otimização de rede de ensaios de genótipos de soja no Cerrado. **Pesquisa Agropecuária Brasileira**, v. 49, p. 783–795, 2014. DOI: 10.1590/S0100-204X2014001000005.

CARBONELL, S.A.M.; CHIORATO, A.F.; RESENDE, M.D.V. de; DIAS, L.A. dos S.; BERALDO, A.L.A.; PERINA, E.F. Estabilidade de cultivares e linhagens de feijoeiro em diferentes ambientes no Estado de São Paulo. **Bragantia**, v. 66, p. 193–201, 2007.

CARGNELUTTI FILHO, A. et al. Measurements of experimental precision and number of replicates in rice genotype trials. **Pesquisa Agropecuária Brasileira**, v. 47, n. 3, p. 336–343, 2012. Available from: <http://www.scielo.br/pdf/pab/v47n3/04.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2012000300004.

CARGNELUTTI FILHO, A. et al. Measures of experimental precision in common bean and soybean genotype trials. **Pesquisa Agropecuária Brasileira**, v. 44, n. 10, p. 1225–1231, 2009. Available from: <http://www.scielo.br/pdf/pab/v44n10/v44n10a03.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2009001000003.

CARGNELUTTI FILHO, A.; STORCK, L. Evaluation statistics of the experimental precision in corn cultivar trials. **Pesquisa Agropecuária Brasileira**, v. 42, n. 1, p. 17–24, 2007. Available from: <http://www.scielo.br/pdf/pab/v42n1/03.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2007000100003.

CHARMET, G., STORLIE, E., OURY, F. X., LAURENT, V., BEGHIN, D., CHEVARIN, L., ROBERT, O. Genome-wide prediction of three important traits in bread wheat. **Molecular Breeding**, v. 34, n. 4, p. 1843–1852, 2014. http://doi.org/10.1007/s11032-014-0143-y

CONAB. Companhia Nacional de Abastecimento. Séries históricas. Available at: <http://www.conab.gov.br/conteudos.php?a=1252&&Pagina_objcmsconteudos=3>. Accessed on: Sep. 1 2016.

COSTA, A.F.; LEAL, N.R.; VENTURA, J.A.; GONÇALVES, L.S.A.; AMARAL JÚNIOR, A.T. do; COSTA, H. Adaptability and stability of strawberry cultivars using a mixed model. Acta Scientiarum. **Agronomy**, v. 37, p. 435–440, 2015. DOI: 10.4025/actasciagron.v37i4.18251.

COSTA, J.R. et al. Neighborhood and spatial analysis in plant breeding. **Pesquisa Agropecuária Brasileira**, v. 40, n. 11, p. 1073–1079, 2005. Available from: <http://www.scielo.br/pdf/pab/v40n11/a04v4011.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2005001100004.

CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum – Agronomy**, v. 35, n. 3, p. 271–276, 2013. Available:<http://periodicos.uem.br/ojs/index.php/ActaSciAgron/article/view/21251/pdf>. Accessed: June 24, 2016. doi: 10.4025/actasciagron.v35i3.21251.

CRUZ, C.D. GENES: a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum. **Agronomy**, v. 35, p. 271–276, 2013. DOI: 10.4025/actasciagron.v35i3.21251.

De FELIPE, M., J. A. GERDE, AND J. L. ROTUNDO. 2016. Soybean Genetic Gain in Maturity Groups III to V in Argentina from 1980 to 2015. **Crop Sci**, v. 56, p. 3066–3077, 2016. doi:10.2135/cropsci2016.04.0214

De LOS CAMPOS, G., HICKEY, J. M., PONG-WONG, R., DAETWYLER, H. D., CALUS, M. P. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v.193, n. 2, p. 327–345, 2013.

DESTA, Z. A., ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in plant science**, v. 19, n. 9, p. 592–601, 2014.

DIA, M.; WEHNER, T.C.; HASSELL, R.; PRICE, D.S.; BOYHAN, G.E.; OLSON, S.; KING, S.; DAVIS, A.R.; TOLLA, G.E.; BERNIER, J.; JUAREZ, B. Value of Locations for Representing Mega-Environments and for Discriminating Yield of Watermelon in the US. 2016B. **Crop Science** v. 56, p. 1726–1735. DOI: 10.2135/cropsci2015.11.0698

DIA, M.; WEHNER, T.C.; PERKINS-VEAZIE, P.; HASSELL, R.; PRICE, D.S.; BOYHAN, G.E.; OLSON, S.M.; KING, S.R.; DAVIS, A.R.; TOLLA, G.E.; BERNIER, J. Stability of fruit quality traits in diverse watermelon cultivars tested in multiple environments. **Horticulture research**, v. 3, 2016A. DOI:10.1038/hortres.2016.66

DUARTE, J.B.; VENCOVSKY, R. Spatial statistical analysis and selection of genotypes in plant breeding. **Pesquisa Agropecuária Brasileira**, v. 40, n. 2, p. 107–114, 2005. Available from: <http://www.scielo.br/pdf/pab/v40n2/23816.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2005000200002.

GIANOLA, D. Priors in whole-genome regression: the Bayesian alphabet returns. **Genetics**, v. 194, n. 3, p. 573-596, 2013.

GIANOLA, D., DE LOS CAMPOS, G., HILL, W. G., MANFREDI, E., FERNANDO, R. Additive genetic variability and the Bayesian alphabet. **Genetics**, v. 183, n. 1, p. 347–363, 2009.

GOMEZ, G.M.; UNÊDA-TREVISOLI, S.H.; PINHEIRO, J.B.; DI MAURO, A.O. Adaptive and agronomic performances of soybean genotypes derived from different genealogies through the use of several analytical strategies. **African Journal of Agricultural Research**, v. 9, p. 2146–2157, 2014. DOI: 10.5897/AJAR2014.8700.

GORJANC, G., BATTAGIN, M., DUMASY, J. F., ANTOLIN, R., GAYNOR, R. C., HICKEY, J. M. Prospects for cost-effective genomic selection via accurate within-family imputation. **Crop Science**, v. 57, n. 1, p. 216–228, 2017.

GORJANC, G., JENKO, J., HEARNE, S. J., HICKEY, J. M. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. **BMC genomics**, v. 17, n. 1, p. 30, 2016.

GOSLEE, S.C. AND URBAN, D.L. The ecodist package for dissimilarity-based analysis of ecological data. **Journal of Statistical Software**, v. 22, n.7, p. 1–19, 2007.

HABIER, D., FERNANDO, R. L., DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n.4, p. 2389–2397, 2007.

HABIER, D., FERNANDO, R. L., KIZILKAYA, K., GARRICK, D. J. Extension of the Bayesian alphabet for genomic selection. **BMC bioinformatics**, v. 12, n.1, p.186, 2011.

HATFIELD, J. L., AND C. L. WALTHALL. Meeting Global Food Needs: Realizing the Potential via Genetics × Environment × Management Interactions. **Agron. J.**, v. 107, p. 1215–1226, 2015. doi:10.2134/agronj15.0076

HATFIELD, J. L., WALTHALL, C. L. Meeting global food needs: realizing the potential via genetics× environment × management interactions. **Agronomy Journal**, v. 107, n.4, p. 1215–1226, 2015.

HE, S., SCHULTHESS, A. W., MIRDITA, V., ZHAO, Y., KORZUN, V., BOTHE, R., JIANG, Y. Genomic selection in a commercial winter wheat population. **Theoretical and Applied Genetics**, v. 129, n. 3, p. 641–651, 2016.

HEFFNER, E. L., A. J. LORENZ, J. JANNINK, AND M. E. SORRELLS. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. **Crop Sci**, v. 50, p. 1681–1690, 2010. doi:10.2135/cropsci2009.11.0662

HEFFNER, E. L., J. JANNINK, AND M. E. SORRELLS. Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. **Plant Genome**, v. 4, p. 65–75, 2011. doi:10.3835/plantgenome2010.12.0029

HICKEY, J. M., DREISIGACKER, S., CROSSA, J., HEARNE, S., BABU, R., PRASANNA, B. M., GORJANC, G. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. **Crop Science**, v. 54, n. 4, p. 1476–1488, 2014.

http://www.fao.org/docrep/016/ap106e/ap106e.pdf (accessed 29 Apr. 2017).

HUANG, M., CABRERA, A., HOFFSTETTER, A., GRIFFEY, C., VAN SANFORD, D., COSTA, J., SNELLER, C. Genomic selection for wheat traits and trait stability. **Theoretical and applied genetics**, v. 129, n. 9, p. 1697–1710, 2016.

JAMSHIDMOGHADDAM, M.; POURDAD, S.S. Genotype x environment interactions for seed yield in rainfed winter safflower (*Carthamus tinctorius* L.) multi-environment trials in Iran. **Euphytica**, v. 190, p. 357–369, 2013. DOI: 10.1007/s10681-012-0776-z.

KARMAKAR, P. G., AND P. S. BHATNAGER, 1996: Genetic improvement of soybean varieties released in India from 1969 to 1993. **Euphytica**, v. 90, p. 95–103, 1996.

KASTER, M.; FARIAS, J.R.B. Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja: terceira aproximação. 2012. Embrapa Soja (in Portuguese).

KOESTER, R.P.; SKONECZKA, J. A.; CARY, T.R.; DIERS, B.W.; AINSWORTH, E.A. Historical gains in soybean (Glycine max Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. **J Exp Bot**; v. 65, n. 12, p. 3311–3321, 2014. doi: 10.1093/jxb/eru187

KRISHNAMURTHY, S.L.; SHARMA, P.C.; SHARMA, D.K.; RAVIKIRAN, K.T.; SINGH, Y.P.; MISHRA, V.K.; BURMAN, D.; MAJI, B.; MANDAL, S.; SARANGI, S.K.;

GAUTAM, R.K. Identification of mega-environments and rice genotypes for general and specific adaptation to saline and alkaline stresses in India. **Scientific Reports**, v. 7, 2017. DOI:10.1038/s41598-017-08532-7

LIN, C.S.; BINNS, M.R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, v. 68, p. 193–198, 1988. DOI: 10.4141/cjps88-018.

LOPES, M.S.; REYNOLDS, M.P.; MANES, Y.; SINGH, R.P.; CROSSA, J.; BRAUN, H.J. Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a “historic” set representing 30 years of breeding. **Crop Science**, v. 52, p. 1123–1131, 2012. DOI: 10.2135/cropsci2011.09.0467.

MA, Y., REIF, J. C., JIANG, Y., WEN, Z., WANG, D., LIU, Z., QIU, L. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). **Molecular Breeding**, v. 36, p. 113, 2016. http://doi.org/10.1007/s11032-016-0504-9

MA, Y., REIF, J. C., JIANG, Y., WEN, Z., WANG, D., LIU, Z., Wang, H. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). **Molecular Breeding**, v. 36, n. 8, p. 113, 2016.

MEUWISSEN, T. Genomic selection: marker assisted selection on a genome wide scale. **Journal of animal Breeding and genetics**, v. 124, n. 6, p. 321–322, 2007.

MEUWISSEN, T.H.E., HAYES, B.J., GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819–1829, 2001.

MICHEL, S., AMETZ, C., GUNGOR, H., EPURE, D., Grausgruber, H., Löschenberger, F., Buerstmayr, H. Genomic selection across multiple breeding cycles in applied bread wheat breeding. **Theoretical and Applied Genetics**, v. 129, n. 6, p. 1179–1189, 2016.

MORRISON, M. J., H. D. VOLDENG, AND E. R. COBER. Physiological Changes from 58 Years of Genetic Improvement of Short-Season Soybean Cultivars in Canada. **Agron. J**, v. 91, p. 685–689, 1999. doi:10.2134/agronj1999.914685x

OLIVEIRA, A.C.B. de; ROSA, A.P.S.A. da (Ed.). Indicações técnicas para a cultura da soja no Rio Grande do Sul e em Santa Catarina, safras 2014/2015 e 2015/2016. Pelotas: Embrapa Clima Temperado, 2014. (Embrapa Clima Temperado. Documentos, 382). XL Reunião de Pesquisa de Soja da Região Sul.

OLIVEIRA, E. J., DE RESENDE, M. D. V., DA SILVA SANTOS, V., FERREIRA, C. F., OLIVEIRA, G. A. F., DA SILVA, M. S., DE OLIVEIRA, L. A. AGUILAR-VILDOSO, C. I. Genome-wide selection in cassava. Euphytica, v. 187, n. 2, p. 263–276, 2012.

ONOGI, A., WATANABE, M., MOCHIZUKI, T., HAYASHI, T., NAKAGAWA, H., HASEGAWA, T., IWATA, H. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. **Theoretical and Applied Genetics**, v. 129, v.4, p. 805–817, 2016.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. "Updated world map of the Köppen–Geiger climate classification". **Hydrol. Earth Syst. Sci**, v. 11, p. 1633–1644, 2007.

PEIXOUTO, L.S.; NUNES, J.A.R.; FURTADO, D.F. Factor analysis applied to the G+ GE matrix via REML/BLUP for multi-environment data. **Crop Breeding and Applied Biotechnology**, v. 16, p. 1–6, 2016. DOI: 10.1590/1984-70332016v16n1a1.

PÉREZ, P., DE LOS CAMPOS, G. Genome-wide regression & prediction with the BGLR statistical package. **Genetics**, v. 114, 2014.

PINHEIRO, L.C. de M.; GOD, P.I.V.G.; FARIA, V.R.; OLIVEIRA, A.G.; HASUI, A.A.; PINTO, E.H.G.; ARRUDA, K.M.A.; PIOVESAN, N.D.; MOREIRA, M.A. Parentesco na seleção para produtividade e teores de óleo e proteína em soja via modelos mistos. **Pesquisa Agropecuária Brasileira**, v. 48, p. 1246–1253, 2013. DOI: 10.1590/S0100-204X2013000900008.

POLAND J, ENDELMAN J, DAWSON J, RUTKOSKI J, WU S, MANES Y, DREISIGACKER S, CROSSA J, SÁNCHEZ-VILLEDA H, SORRELLS M. Genomic selection in wheat breeding using genotyping-by-sequencing. **Plant Genome**. v. 5, n. 3, p.103–113, 2012, doi: 10.3835/plantgenome2012.06.0006

QIN, J.; XU, R.; LI, H.; YANG, C.; LIU, D.; LIU, Z.; ZHANG, L.; LU, W.; FRETT, T.; CHEN, P.; ZHANG, M.; QIU, L. Evaluation of productivity and stability of elite summer soybean cultivars in multi-environment trials. **Euphytica**, v. 206, p. 759–773, 2015. DOI: 10.1007/s10681-015-1513-1.

QIN, J.; XU, R.; LI, H.; YANG, C.; LIU, D.; ZHANG, L.; LU, W.; FRETT, T.; CHEN, P.; ZHANG, M.; QIU, L. Evaluation of productivity and stability of elite summer soybean cultivars in multi-environment trials. 2015. **Euphytica,** v. 206 p. 759–773. DOI:10.1007/s10681-015-1513-1

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**. Vienna. Austria, 2016. URL<https://www.R-project.org/>.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2015. Available from: <http://www.R-project.org/>. Accessed: Feb. 26, 2016.

RAMALHO, M.A.P. et al. Experimentation in genetics and plant breeding. Lavras: UFLA, 2000. 326p.

RAY, D.K., N.D. MUELLER, P.C. WEST, AND J.A. FOLEY. 2013. Yield trends are insufficient to double global food production by 2050. **PLOS One** v. 8, 2013. e66428. doi:10.1371/journal.pone.0066428

RESENDE, M.D.V. de; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropica**l, v. 37, p. 182 –194, 2007.

RESENDE, M.D.V. de. Métodos estatísticos ótimos na análise de experimentos de campo. Colombo: Embrapa Florestas, 2004. 65p. (Embrapa Florestas. Documentos, 100).

RESENDE, M.D.V. de. Software Selegen – REML/BLUP. Colombo: Embrapa Florestas, 2002. 67p. (Embrapa Florestas. Documentos, 77).

RESENDE, M.D.V.; DUARTE, J.B. Precision and quality control in variety trials. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182–194, 2007. Available from: <https://revistas.ufg.br/pat/article/viewFile/1867/1773>. Accessed: June 24, 2016.

ROCHA, F. da; VIEIRA, C.C.; FERREIRA, M.C.; OLIVEIRA, K.C. de; MOREIRA, F.F.; PINHEIRO, J.B. Selection of soybean lines exhibiting resistance to stink bug complex in distinct environments. **Food and Energy Security**, v. 4, p. 133–143, 2015. DOI: 10.1002/fes3.57.

ROSADO, A.M.; ROSADO, T.B.; ALVES, A.A.; LAVIOLA, B.G.; BHERING, L.L. Seleção simultânea de clones de eucalipto de acordo com produtividade, estabilidade e adaptabilidade. **Pesquisa Agropecuária Brasileira**, v. 47, p. 964–971, 2012. DOI: 10.1590/S0100-204X2012000700013.

RUTKOSKI, J. E., J. A. POLAND, R. P. SINGH, J. HUERTA-ESPINO, S. BHAVANI, H. BARBIER, M. N. ROUSE, J. JANNINK, AND M. E. Sorrells. Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. **Plant Genome,** v. 7, 2014. doi:10.3835/plantgenome2014.02.0006

SHU, Y. J., YU, D. S., WANG, D., BAI, X., ZHU, Y. M.; GUO, C. H. Genomic selection of seed weight based on low-density SCAR markers in soybean. **Genetics and Molecular Research**, v. 12 n. 3, p. 2178–2188, 2013.

SOUZA, E.A. et al. Experimental alternatives for evaluation of progenies and clones in eucalyptus breeding programs. **Revista Árvore**, v. 27, n. 4, p. 427–434, 2003. Available from: <http://www.scielo.br/pdf/rarv/v27n4/a02v27n4.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-67622003000400002.

SPIEGELHALTER D. J., BEST N. G., CARLIN B. P., VAN DER LINDE A., Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. **Methodol,** v. 64, n. 4, p. 583–639, 2002.

SPINDEL J, BEGUM H, AKDEMIR D, VIRK P, COLLARD B, REDOÑA E, et al. Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. **PLoS Genet**, v. 11, n. 2, 2015. e1004982. doi:10.1371/journal.pgen.1004982

SPINELLI, V.M.; DIAS, L.A.S.; ROCHA, R.B.; RESENDE, M.D.V. Estimates of genetic parameters with selection within and between half-sib families of *Jatropha curcas* L. Industrial **Crops and Products**, v. 69, p. 355–361, 2015. DOI: 10.1016/j.indcrop.2015.02.024.

STEEL, R.G.D. et al. Principles and procedures of statistics: a biometrical approach. New York: McGraw Hill Book, 1997. 672p.

STORCK, L. et al. DelPapa – Software tools for analyzing data for experiments in randomized block design using the Papadakis method. **Ciência Rural**, v. 45, n. 5, p. 774–781, 2015. Available from: <http://www.scielo.br/pdf/cr/v45n5/0103-8478-cr-45-05-00774.pdf>. Accessed: June 24, 2016. doi: 10.1590/0103-8478cr20131470.

STORCK, L. et al. The Papadakis method and the replications number in soybean experiments. **Ciência Rural**, v. 39, n. 4, p. 977–982, 2009. Available from: <http://www.scielo.br/pdf/cr/v39n4/a133cr557.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0103-84782009005000027.

STORCK, L. et al. Use of the Papadakis method to improve soybean trial quality. **Pesquisa Agropecuária Brasileira**, v. 43, n. 5, p. 581–587, 2008. Available from: <http://www.scielo.br/pdf/pab/v43n5/a05v43n5.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0100-204X2008000500005.

STORCK, L.; SILVA, M.S. Experimental precision in wheat cultivars trials using the Papadakis method. **Ciência Rural**, v. 44, n. 3, p. 407–413, 2014. Available from: <http://www.scielo.br/pdf/cr/v44n3/a6314cr2012-1127.pdf>. Accessed: June 24, 2016. doi: 10.1590/S0103-84782014000300004.

THAVAMANIKUMAR, S., DOLFERUS, R., THUMMA, B. R. Comparison of Genomic Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations. G3: **Genes|Genomes|Genetics**, v. 5, n. 10, p. 1991–1998, 2015. http://doi.org/10.1534/g3.115.019745

VILLUMSEN, T. M., JANSS, L., LUND, M. S. The importance of haplotype length and heritability using genomic selection in dairy cattle. **Journal of Animal Breeding and**

Genetics, v. 126, n. 1, p. 3–13, 2009.

VISSCHER, P. M., HILL, W. G., WRAY, N. R. Heritability in the genomics era—concepts and misconceptions. **Nature Reviews Genetics**, p. 9, n. 4, p. 255–266, 2008.

VOLLMANN, J. et al. Spatial field variations in soybean (*Glycine max* [L.] Merr.) performance trials affect agronomic characters and seed composition. **European Journal of Agronomy**, v. 12, p. 13–22, 2000. Available from: <http://www.sciencedirect.com/science/article/pii/S1161030199000428>. Accessed: June 24, 2016. doi: 10.1016/S1161-0301(99)00042-8.

WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

XAVIER, A., MUIR, W. M., RAINEY, K. M. Assessing predictive properties of genome-wide selection in soybeans. G3: **Genes| Genomes| Genetics**, v. 6, n. 8, p. 2611–2616, 2016.

XAVIER, A., MUIR, W. M., RAINEY, K. M. Assessing Predictive Properties of Genome-Wide Selection in Soybeans. G3: **Genes|Genomes|Genetics**, v. 6, n. 8, p. 2611–2616, 2016. http://doi.org/10.1534/g3.116.032268

YAN, W. Analysis and handling of G× E in a practical breeding program. 2016 **Crop Science** v. 56 p. 2106–2118. DOI:10.2135/cropsci2015.06.0336

YAN, W. Crop Variety Trials: Data Management and Analysis. 2014. Chichester, West Sussex, UK: Wiley Blackwell.

YAN, W. GGEbiplot—a Windows application for graphical analysis of multi-environment trial data and other types of two-way data. **Agronomy Journal**, v. 93, p. 1111–1118, 2001. DOI: 10.2134/agronj2001.9351111x.

YAN, W. GGEbiplot—A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. 2001. **Agronomy Journal** v. 93 p. 1111–1118. DOI:10.2134/agronj2001.9351111x

YAN, W. Mega-environment analysis and test location evaluation based on unbalanced multiyear data. 2015. **Crop Science** v. 55 p. 113–122. DOI:10.2135/cropsci2014.03.0203

YAN, W. Mega-environment analysis and test location evaluation based on unbalanced multiyear data. **Crop Science**, v. 55, p. 113–122, 2015. DOI: 10.2135/cropsci2014.03.0203.

YAN, W.; FRÉGEAU-REID, J.; PAGEAU, D.; MARTIN, R.; MITCHELL-FETCH, J.; ETIENNE, M.; ROWSELL, J.; SCOTT, P.; PRICE, M.; DE HAAN, B.; CUMMISKEY, A. Identifying essential test locations for oat breeding in eastern Canada. 2010. **Crop Science** v. 50 p. 504–515. DOI:10.2135/cropsci2009.03.0133

YAN, W.; HOLLAND, J.B. A heritability-adjusted GGE biplot for test environment evaluation.2010. **Euphytica** v. 171 p. 355–369. DOI:10.1007/s10681-009-0030-5

YANG, R.-C.; CROSSA, J.; CORNELIUS, P.L.; BURGUEÑO, J. Biplot analysis of genotype × environment interaction: proceed with caution. **Crop Science**, v. 49, p. 1564–1576, 2009. DOI: 10.2135/cropsci2008.11.0665.

ZHANG, X., PÉREZ-RODRÍGUEZ, P., BURGUEÑO, J., OLSEN, M., BUCKLER, E., ATLIN, G., CROSSA, J. Rapid cycling genomic selection in a multi-parental tropical maize population. **G3: Genes, Genomes, Genetics**, v. 7; p. 3–117, 2017.

ZHAO, Y., J. ZENG, R. FERNANDO, AND J. C. REIF. Genomic Prediction of Hybrid Wheat Performance. **Crop Sci.** v. 53, p. 802–810, 2013. doi:10.2135/cropsci2012.08.0463

MARK A. MIKEL,* BRIAN W. DIERS, RANDALL L. NELSON, AND HEBRON H. SMITH. Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. **Crop Sci.** v. 50, p. 1219–1229 , 2010. doi: 10.2135/cropsci2009.08.0456

GLOSSÁRIO

INDEX OF APPENDICES AND ATTACHMENT

APPENDICE A – Identification of Core Locations for Soybean Breeding in Southern Brazil – Abstract 89

APPENDICE B – Methods of analysis and number of replicates for trials with large numbers of soybean genotypes – Abstract 90

APPENDICE C – Agronomic performance of modern soybean cultivars in multi-environment trials – Abstract..... 91

APPENDICE D – Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection– Abstract..... 92

APPENDICE E– Scatter plots obtained for the five characters using the GWS Bayes B methodology in the analysis of the genotypic and phenotypic data of the 324 soybean rows. Analyzes performed with correction for intrapopulation structuring.. 93

APPENDICE F– Imagens of the phenotyping experiment..... 94

ATTACHMENT A – Map of brazil's soybeans macroregions, Kaster, 2011. 96

ATTACHMENT B – Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. MARK A. et al., 2010. 97

ATTACHMENT B Cont. – Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. MARK A. et al., 2010..... 98

ATTACHMENT C – Methodology of the Wide Genomic Selection (GWS) – Reference Population and Parameters of the model, Castro et. Al, 2012. 99

APPENDICES

APPENDICE A – Identification of Core Locations for Soybean Breeding in Southern Brazil – Abstract

In plant breeding, the identification of ideal test locations is essential for selection in generations with low seed availability. This study aimed to identify core locations for evaluation and selection of segregating soybean populations in macroregions 1 (M1) and 2 (M2), in Brazil. Value of Cultivation and Use (VCU) trials from the 2012–13 to 2015–16 growing seasons were used. The datasets consisted of 22 soybean genotypes cultivated in 23 locations within soybean M1 and M2. Trials were conducted in a randomized complete block design with three replications. All analyses were performed used GGEbiplot software. The analyses GGL + GGE ((genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)) and G (Genotypic effect) vs. GE (genotype x environment effects) were used to identify core locations (i.e., locations with high representativeness and consistence of results). The locations Chapada-RS, and Maracaju-MS turned out as core locations in M1 and M2, respectively. These locations were characterized by long vectors and a high genetic correlation with the mean environment (lowest angle); thus, they were the most representative and consistent over the years of study within each macroregion. Furthermore, these locations offer the possibility for selecting superior genotypes in some environments (year + location); this selection occurred through genotypic effect, which is desirable. Identification of a core location is crucial and most effective for evaluating segregating populations because the number of trial locations may reduce to only one per macroregion.

APPENDICE B – Methods of analysis and number of replicates for trials with large numbers of soybean genotypes – Abstract

The aim of this study was to evaluate the experimental precision of different methods of statistical analysis for trials with large numbers of soybean genotypes, and their relationship with the number of replicates. Soybean yield data (nine trials; 324 genotypes; 46 cultivars; 278 lines; agricultural harvest of 2014/15) were used. Two of these trials were performed at the same location, side by side, forming a trial with six replicates. Each trial was analyzed by the randomized complete block, triple lattice design, and use of the Papadakis method. The selective accuracy, least significant difference, and Fasoulas differentiation index were estimated, and model assumptions were tested. The resampling method was used to study the influence of the number of replicates, by varying the number of blocks and estimating the precision measurements. The experimental precision indicators of the Papadakis method are more favorable as compared to the randomized complete block design and triple lattice. To obtain selective accuracy above the high experimental precision range in trials with 324 soybean genotypes, two repetitions can be used, and data can be analyzed using the randomized complete block design or Papadakis method.

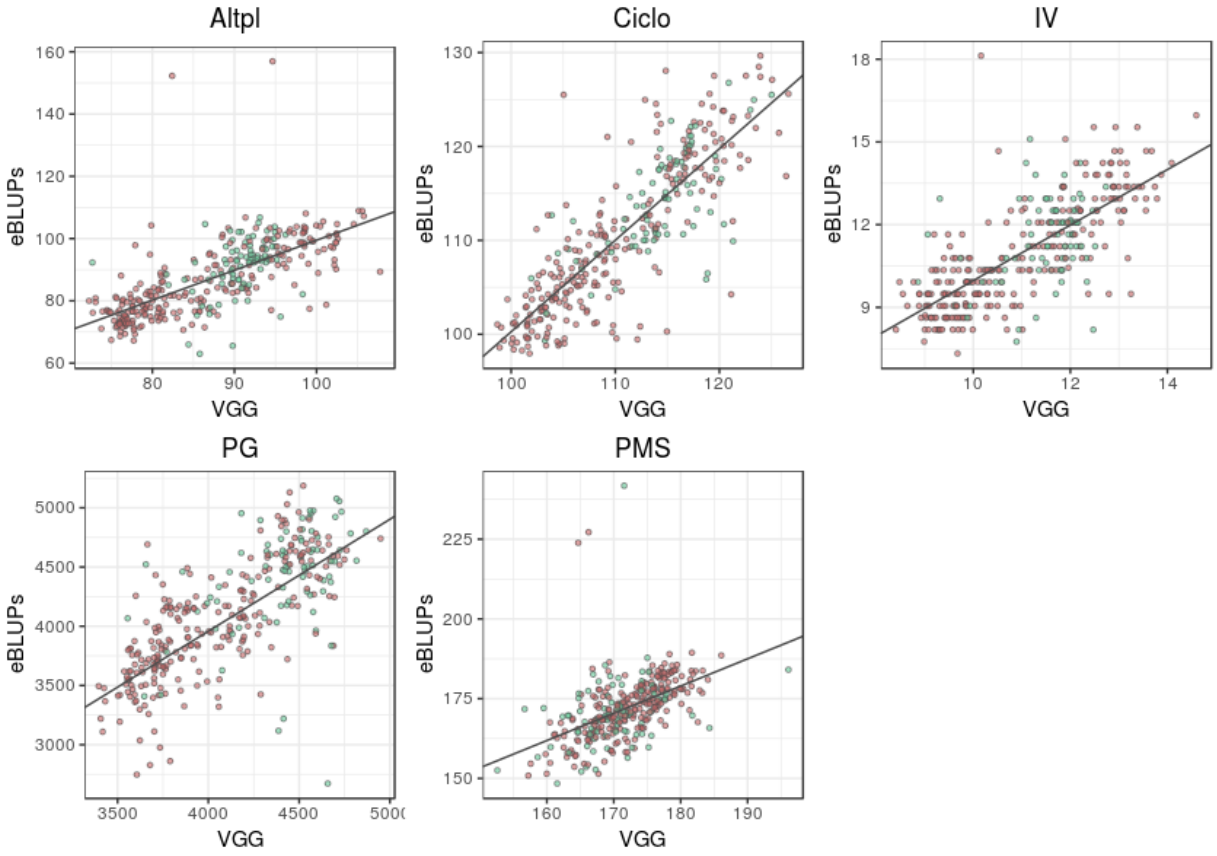
APPENDICE C – Agronomic performance of modern soybean cultivars in multi-environment trials – Abstract

This study aimed to evaluate the crop performance, adaptability, and stability of modern soybean cultivars in multi-environment trials, and to identify the ideal genotypes for eight growing environments in Brazil. A complete randomized block design, with three replicates, was used to evaluate 46 cultivars in eight environments, in the microregions of adaptation 102, 201, and 202, in the 2014/2015 crop season. Complex genotype × environment interaction occurred with cultivar ranking differing between sites. The cultivars NA 5909 RG, M6410IPRO, NS 5959 IPRO, NS6823RR, M5917IPRO, NS 6767, and 6563RSF IPRO had the highest average production rate. The cultivars NA 5909 RG, NS6823RR, M6410IPRO and NS 5959 IPRO showed high adaptability, stability, and grain yield in the environments tested and had values close to those of the ideal genotypes for the study environments. Thus, modern soybean cultivars that are adapted, stable, and have high yield are available for cultivation in the regions of adaptation 101, 201, and 202 in Brazil.

APPENDICE D – Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection– Abstract

Genomic selection (GS) can potentially accelerate genetic improvement of soybean [Glycine max L. (Merrill)] by reducing the time to complete breeding cycles. The objectives of this study were to (1) explore the accuracy of GS in soybean; (2) evaluate the contribution of intrapopulation structure to the accuracy of GS; and (3) compare the efficiencies of phenotypic selection and GS in soybean. The BayesB model with cross validation was applied to phenotypic and genotypic data from 324 soybean genotypes (243 recombinant inbred lines and 81 cultivars). The GS accuracy was evaluated according to grain yield, plant height, insertion of first pod, days to maturity, and 1000-grain weight at multiple locations; genotyping of 5403 single nucleotide polymorphism markers was also performed. We found that genotypic accuracy was similar irrespective of population structure, which did not affect the accuracy of the model for the evaluated traits. Thus, GS can reduce the time required to complete a selection cycle in soybean, which can lead to increased production of this commercially important crop.

APPENDICE E– Scatter plots obtained for the five characters using the GWS Bayes B methodology in the analysis of the genotypic and phenotypic data of the 324 soybean rows. Analyzes performed with correction for intrapopulation structuring..

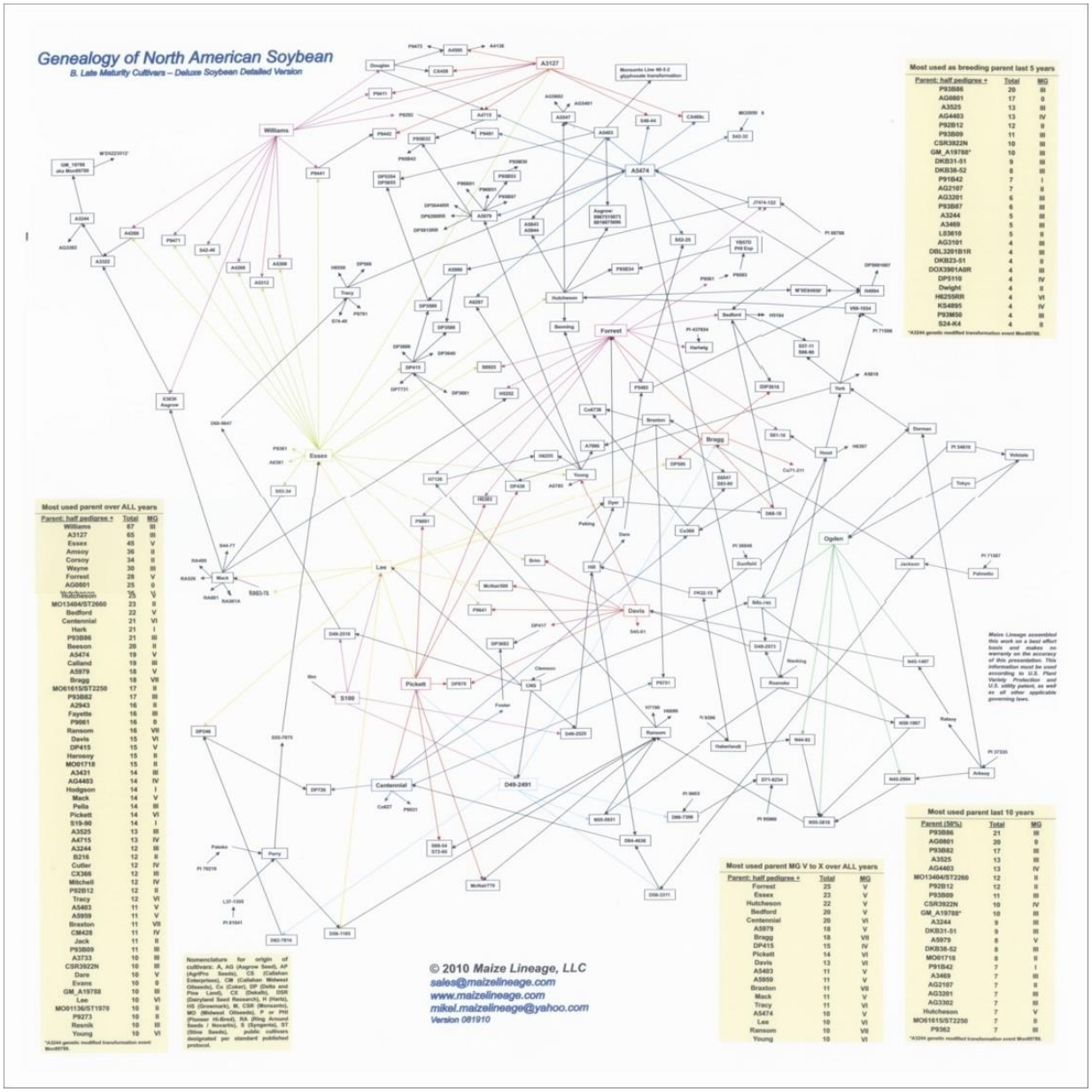


APPENDICE F– Imagens of the phenotyping experiment.

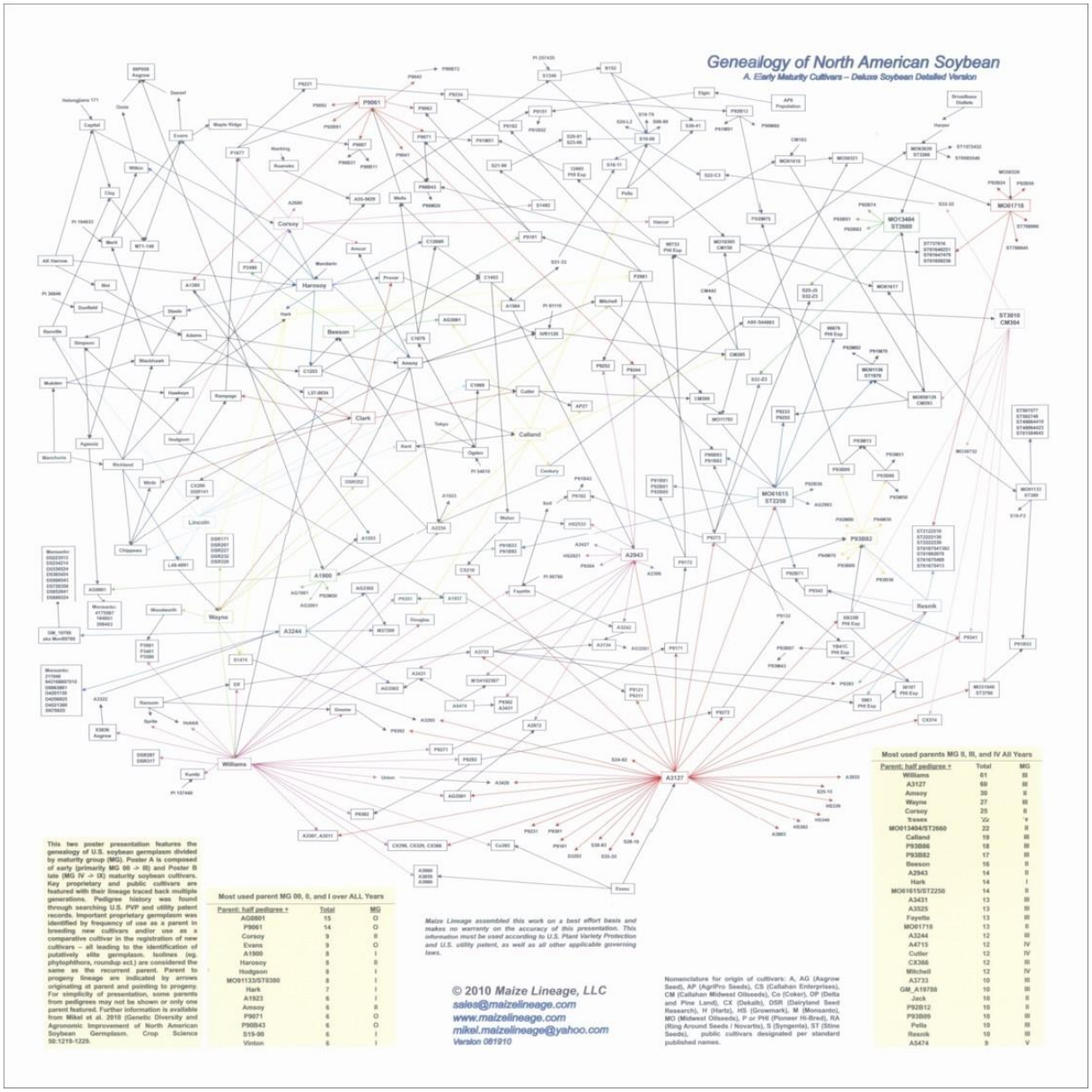
ATTACHMENT

ATTACHMENT A – Map of brazil's soybeans macroregions, Kaster, 2011.

ATTACHMENT B – Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. MARK A. et al., 2010.



ATTACHMENT B Cont. – Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. MARK A. et al., 2010.



ATTACHMENT C – Methodology of the Wide Genomic Selection (GWS) – Reference Population and Parameters of the model, Castro et. Al, 2012.