**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Utilização de técnicas multivariadas na análise da divergência
genética via modelo AMMI com reamostragem “*bootstrap*”**

Priscila Neves Faria

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

**Piracicaba
2012**

Priscila Neves Faria
Licenciado em Matemática

**Utilização de técnicas multivariadas na análise da divergência genética via
modelo AMMI com reamostragem “bootstrap”**

Orientador:
Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Tese apresentada para obtenção do título de
Doutor em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2012**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Faria, Priscila Neves

Utilização de técnicas multivariadas na análise da divergência genética via modelo AMMI com reamostragem "*bootstrap*" / Priscila Neves Faria.- - Piracicaba, 2012.

128 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Análise multivariada 2. Divergência genética 3. Interação genótipo ambiente 4. Reamostragem I. Título

CDD 519.53
F224u

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

A

Deus

e à Nossa Senhora,

sempre guiando meus caminhos.

Aos meus pais,

Mozart Neves de Faria e

Elena Maria de Faria,

que me ensinaram a importância dos estudos

e em todos os momentos de dificuldade,

sempre me aconselharam.

Ao meu irmão Leandro,

meu exemplo de perseverança e

solidariedade.

Ao Filipe

pela companhia constante, amizade,

paciência e amor, DEDICO.

AGRADECIMENTOS

Há oportunidades em que além do trabalho em equipe, existem alguns e privilegiados casos, onde acontece a amizade:

Ao meu orientador, o Prof. Dr. Carlos Tadeu dos Santos Dias, por me ajudar nos meus primeiros passos no doutorado, pelo conhecimento compartilhado, confiança no meu trabalho e apoio;

Ao Prof. Dr. Lúcio Borges Araújo, que além de colega de trabalho se tornou um grande amigo, que com seu profissionalismo e conhecimento tornou possível o aprimoramento do meu aprendizado tanto na parte teórica quanto na computacional.

À Prof. Ms. Mirian Fernandes C. Araújo, colega de trabalho e amiga, obrigada pelas conversas, companhia, por me aturar nos momentos de estresse! Você e o Lúcio tornaram meus dias melhores, podem ter certeza disso.

Ao Prof. Dr. Marcelo Cirillo, serei eternamente agradecida pela paciência, pelos ensinamentos e pela valiosa colaboração. Muito obrigada por tudo!

Ao Prof. Dr. José Baldin Pinheiro, obrigada pela valiosa colaboração cedendo dados e informações na parte genética.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, obrigada pela amizade e contribuição na minha formação.

Às secretárias Luciane e Solange, sempre me ajudando com as pendências burocráticas e documentações, obrigada também pelas longas conversas e apoio.

Aos meus colegas do doutorado pelo constante apoio e amizade, a todos vocês que fizeram parte deste meu aprendizado e de uma forma muito especial agradeço a: Elizabeth Mie Hashimoto, Ricardo Olinda, Ana Patrícia Peixoto e Simone Sartório.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

E finalmente ao Cookie, meu cachorrinho de estimação, pela companhia de sempre e durante todo o tempo que passei diante do computador... estive ao meu lado!!!

EPÍGRAFE

.....Muitas pessoas têm buscado sonhos que não são seus. A sociedade quer definir o que é certo, instituir que todos têm de ter sucesso, como se ele não tivesse significados individuais. A loucura de que você tem de estar feliz todos os dias, que você tem que comprar tudo o que puder e fazer as coisas do jeito certo. Jeito certo não existe, não há um caminho único para se fazer as coisas. As metas são interessantes para o sucesso, mas não para a felicidade. Felicidade não é uma meta, mas um estado de espírito. Aprenda que a felicidade é feita das pequenas coisas!!!

Roberto Shinyashik

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	21
2 REVISÃO BIBLIOGRÁFICA.....	25
A soja no Brasil e no mundo	25
Interação genótipos $\times$ ambientes	26
Modelo de Efeitos Aditivos com Interação Multiplicativa – AMMI.....	28
Vantagens do uso do modelo AMMI	30
Análise da Divergência Genética utilizando métodos multivariados	32
Análise de agrupamento	35
Soma de quadrados progressiva (Ward 1963, Orlóci 1978)	39
Análise de componentes principais (ACP)	39
Reamostragem <i>bootstrap</i>	43
3 MATERIAL E MÉTODOS	47
Material experimental	47
Análise de Variância Conjunta.....	49
Análise da interação $G \times E$ e o enfoque estatístico	51
O modelo AMMI.....	55
A seleção do modelo AMMI.....	56
Análise da interação $G \times E$ via reamostragem <i>bootstrap</i> AMMI	59
4 RESULTADOS E DISCUSSÃO	61
Experimento 1, parte 1: População PCI - ano 2000	61

Experimento 1, parte 2: População PCI - ano 2001.....	72
Experimento 2, parte 1: População PRIS - ano 2000	80
Experimento 2, parte 2: População PRIS - ano 2001	88
Experimento 3, parte 1: População PRIM - ano 2000	94
Experimento 3, parte 2: População PRIM - ano 2001	101
5 CONCLUSÕES	107
REFERÊNCIAS	109
APÊNDICES	115
ANEXOS	125

RESUMO

Utilização de técnicas multivariadas na análise da divergência genética via modelo AMMI com reamostragem “*bootstrap*”

Em estudos de divergência genética por métodos multivariados, a distância euclidiana é a medida de distância mais amplamente utilizada e essa distância é a mais recomendada quando as unidades de cálculos são escores de componentes principais, como é o caso da análise AMMI (additive main effects and multiplicative interaction analysis). Tal análise permite a obtenção de estimativas mais precisas das respostas genotípicas e possibilita a análise da divergência genética por métodos aglomerativos. A análise dos modelos AMMI combina, num único modelo, componentes aditivos para os efeitos principais (genótipos e ambientes) e componentes multiplicativos para os efeitos da interação genótipos $\times$ ambientes. Os melhoristas de plantas compreendem que a interação genótipos $\times$ ambientes é de suma importância para a obtenção de variedades superiores e as estimativas de dissimilaridade atendem aos objetivos do melhorista, por quantificarem e informarem sobre o grau de semelhança ou de diferença entre pares de indivíduos. Entretanto, quando o número de indivíduos é grande, torna-se inviável o reconhecimento de grupos homogêneos pelo exame visual das estimativas de distância. Portanto, é importante proceder à análise de agrupamentos, obter dendrogramas por meio de métodos hierárquicos e posteriormente, analisar os grupos formados. A fim de determinar e classificar os grupos formados na clusterização hierárquica foram utilizados comandos específicos do programa computacional R que desenha no dendrograma os retângulos de cada grupo e os numera. Desta forma, o objetivo deste trabalho foi analisar a divergência genética via modelo AMMI, utilizando-se de técnicas multivariadas e reamostragem “*bootstrap*”.

Palavras-chave: Interação genótipo ambiente; Análise Multivariada; Reamostragem

ABSTRACT

Use of multivariate techniques in the analysis of genetic diversity through ammi model with bootstrap resampling

In studies of genetic diversity using multivariate approaches, the Euclidean distance is the most common measure used. This method is recommended when data are scores of principal components, such as in AMMI analysis (additive main effects and multiplicative interaction analysis). The AMMI method allows obtaining more precise estimates for genotypic results and also permits the use of genetic diversity analysis by using agglomerative approaches. Furthermore, this method combines additive components for the main effects (genotypes and environments) and multiplicative components for genotypes x environment interaction effects in a unique model. Plant breeders understand the importance of genotype and environment interaction to obtain superior varieties and the dissimilarity estimation meets breeders' objectives since it quantifies and determines the similarity or the divergence between pairs of individuals. However, when the number of individuals is large it is unfeasible to recognize the group homogeneity by using a visual analysis of the distances estimation. Therefore, is important to use cluster analysis to obtain dendograms based on hierarchical methods and then analyze the groups obtained. In order to determine and classify the obtained groups from hierarchical cluster analysis specific commands in the R software were used which shows in the dendrogram rectangles and numbers for each group. In this way, the objective of this work was to analyze the genetic divergence through AMMI model, by using multivariate approaches and "bootstrap" resampling.

Keywords: Genotype and Environment interaction; Multivariate analysis; Resampling

LISTA DE FIGURAS

Figura 1 - Gráfico do comportamento de 44 genótipos de soja da população PCI (ano 2000) em relação ao primeiro componente da interação e ao caráter PG	66
Figura 2 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PCI, ano 2000	69
Figura 3 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₂ , para dados de produtividade de grãos, em kg/ha.....	71
Figura 4 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PCI, ano 2001	75
Figura 5 - Gráfico do comportamento de 23 genótipos de soja da população PCI (ano 2001) em relação ao primeiro componente da interação e ao caráter PG	76
Figura 6 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₁ , para dados de produtividade de grãos, em kg/ha.....	78
Figura 7 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PRIS, ano 2000	83
Figura 8 - Gráfico do comportamento de 44 genótipos de soja da população PRIS (ano 2000) em relação ao primeiro componente da interação e ao caráter PG	84
Figura 9 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₁ , para dados de produtividade de grãos , em kg/ha.....	87
Figura 10 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PRIS, ano 2001	91
Figura 11 - Gráfico do comportamento de 23 genótipos de soja da população PRIS (ano 2001) em relação ao primeiro componente da interação e ao caráter PG	91

Figura 12 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₁ , para dados de produtividade de grãos , em kg/ha	93
Figura 13 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PRIM, ano 2000	97
Figura 14 - Gráfico do comportamento de 44 genótipos de soja da população PRIM (ano 2000) em relação ao primeiro componente da interação e ao caráter PG.....	98
Figura 15 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₁ , para dados de produtividade de grãos , em kg/ha	100
Figura 16 - Biplot AMMI ₂ para dados de produtividade de grãos em kg/ha, população PRIM, ano 2001	102
Figura 17 - Gráfico do comportamento de 23 genótipos de soja da população PRIM (ano 2001) em relação ao primeiro componente da interação e ao caráter PG.....	105
Figura 18 - Dendrograma das distâncias euclidianas entre os escores “ <i>bootstrap</i> ” de marcadores de genótipos AMMI ₁ , para dados de produtividade de grãos, em kg/ha	106

LISTA DE TABELAS

Tabela 1 - Esquema da ANOVA Conjunta para experimentos de um mesmo grupo de genótipos avaliados em e ambientes com r blocos.....	50
Tabela 2 - Disposição dos dados em linhas (genótipos) e colunas (ambientes)	51
Tabela 3 - Esquema da análise AMMI, pelo critério de Gollob (1968) para atribuição de graus de liberdade aos componentes principais da interação.....	57
Tabela 4 - Análise de variância conjunta do Experimento 1 no ano 2000, incluindo o desdobramento da interação $G \times E$ para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja avaliados em 4 ambientes com 2 blocos.....	62
Tabela 5 - Matriz de Resíduos da interação $G \times E$, população 1 (PCI - população com controle de insetos), no ano agrícola 2000	63
Tabela 6 - Porcentagem responsável por cada eixo da interação e porcentagem da soma de quadrados acumulada (PA) por eixo singular, Teste F para os componentes. População PCI, ano 2000.....	65
Tabela 7 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 1 (PCI), no ano agrícola 2000.....	67
Tabela 8 - Agrupamentos de quarenta e quatro genótipos de soja da população 1 referente ao experimento realizado no ano 2000, obtidos pelo método hierárquico de Ward	71
Tabela 9 - Análise conjunta da variância do Experimento 1 no ano 2001, incluindo o desdobramento da interação $G \times E$ para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja	72
Tabela 10 - Matriz de Resíduos da interação $G \times E$, população 1 (PCI - população com controle de insetos), no ano agrícola 2001	74
Tabela 11 - Porcentagem da soma de quadrados acumulada (PA) por eixo singular, Teste F para os IPCA referente aos dados de produtividade de grãos em kg/ha de 23 genótipos de soja. População PCI, ano 2001	75

Tabela 12 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 23 genótipos de soja da população 1 (PCI), no ano agrícola 2001	77
Tabela 13 - Agrupamentos de vinte e três genótipos de soja da população 1 referente ao experimento realizado no ano 2001, obtidos pelo método hierárquico de Ward	79
Tabela 14 - Análise conjunta da variância do Experimento 2 no ano 2000, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja	80
Tabela 15 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2000	81
Tabela 15 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2000	82
Tabela 16 - Porcentagem da soma de quadrados acumulada por eixo singular, Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 44 genótipos de soja. População PRIS, ano 2000.....	82
Tabela 17 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 2 (PRIS), no ano agrícola 2000.....	85
Tabela 18 - Análise conjunta da variância do Experimento 2 no ano agrícola 2001, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja.....	88
Tabela 19 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2001	89
Tabela 20 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 23 genótipos de soja. População PRIS, ano 2001	90
Tabela 21 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 23 genótipos de soja da população 2 (PRIS), no ano agrícola 2001.....	92

Tabela 22 - Análise conjunta da variância do Experimento 3 no ano 2000, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja	95
Tabela 23 - Matriz de Resíduos da interação G×E, população 3 (PRIM - população com controle de insetos), no ano agrícola 2000	95
Tabela 24 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente, referente aos dados de produtividade de grãos em kg/ha de 44 genótipos de soja. População PRIM, ano 2000	97
Tabela 25 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001	98
Tabela 25 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001	99
Tabela 26 - Análise conjunta da variância do Experimento 3 no ano 2001, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja	101
Tabela 27 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 23 genótipos de soja. População PRIM, ano 2001	102
Tabela 28 - Matriz de Resíduos da interação G×E, população 3 (PRIM - população com controle de insetos), no ano agrícola 2001	103
Tabela 29 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001	104
Tabela 30 - Agrupamentos de vinte e três genótipos de soja da população 3 referente ao experimento realizado no ano 2001, obtidos pelo método hierárquico de Ward.....	106

1 INTRODUÇÃO

Os programas de melhoramento de espécies demandam cálculos precisos. Esses cálculos, fundamentados em base de dados de experimentos genéticos proporcionam tomadas de decisões eficientes, quanto ao planejamento das ações futuras desses referidos programas de melhoramento.

A inclusão do melhoramento para resistência a insetos-pragas em programas de melhoramento é uma tendência nova, sendo que tal abordagem deve-se ao comprometimento da produção com a ocorrência de insetos que atacam a soja por exemplo, e que até então tinham como principal método de controle o uso de produtos químicos.

Dada a importância e magnitude de perdas quantitativas e qualitativas que os ataques de insetos provocam na cultura de soja, Pinheiro (1998) destacou que a utilização da resistência genética a insetos em cultivares de soja constitui-se numa estratégia no mínimo complementar ao uso de inseticidas, por apresentar vantagens adicionais, como: maior equilíbrio ecológico, ausência de gastos adicionais para os agricultores e desenvolvimento de tecnologia nacional.

Em relação ao cultivo extensivo da soja, quando contempla ambientes diversificados, este provoca uma resposta diferencial dos genótipos. Neste sentido, a interação entre genótipos e ambientes ($G \times E$) representa aspecto relevante no contexto do melhoramento. Os melhoristas de plantas compreendem que o estudo da interação $G \times E$ é de suma importância para a obtenção de variedades superiores. No entanto, as posições críticas dos estatísticos que atuam em programas de melhoramento genético, referem-se à falta de uma análise criteriosa da estrutura da interação $G \times E$ como sendo um dos principais problemas para a recomendação de cultivares.

Em experimentos agrícolas nos quais os ensaios são realizados em vários ambientes, a análise de interação $G \times E$ isola o componente de variabilidade devido à interação entre genótipos e ambientes por meio de uma análise estatística adequada. Neste sentido, os programas de melhoramento têm como principal objetivo buscar técnicas capazes de detectar de maneira aprofundada essas interações para a seleção de genótipos consistentes e de elevada produtividade.

Existem vários métodos para o estudo da interação $G \times E$ sendo os tradicionais aqueles que consistem em realizar o zoneamento agronômico, permitindo recomendações regionalizadas, ou métodos que utilizam regressão linear simples e múltipla com o objetivo de identificar genótipos estáveis que podem ser recomendados de forma generalizada, desde que eles mostrem uma “performance” desejável. A crítica feita à técnica de zoneamento agronômico é que o agrupamento é feito com base nas diferenças macro ambientais, tornando-o vulnerável às oscilações imprevisíveis que possam ocorrer no ambiente, além de não permitir a avaliação da interação genótipo e anos (ROSSE, 2000). Já em relação à análise de regressão linear, Crossa (1990) argumenta a possibilidade da linearidade falhar, além da técnica ser altamente dependente do tipo de genótipos e ambientes incluídos na análise e tender a simplificar padrões de resposta.

Outro procedimento que pode ser utilizado nas análises da interação $G \times E$ é avaliação do padrão de similaridade (ou dissimilaridade) de diferentes genótipos. Estes podem ser agrupados por métodos de conglomeração (métodos multivariados), que exigem, preliminarmente, a obtenção de uma medida de dissimilaridade entre os ambientes e, posteriormente, a utilização de uma técnica adequada de agrupamento.

No entanto, uma análise que ganha destaque quando o assunto é interação genótipos e ambientes é a AMMI. Segundo Pereira (2009), apesar da seleção de poucos componentes que expliquem pequena proporção da variação, a informação fornecida pela análise AMMI pode ser considerada de melhor qualidade do que a fornecida por métodos tradicionais.

Sabe-se que uma baixa eficiência na análise da interação $G \times E$ pode representar um problema aos melhoristas, por reduzir a precisão de seleção de genótipos de um ambiente para outro. Neste sentido, a análise AMMI (additive main effects and multiplicative interaction analysis) permite a obtenção de estimativas das respostas genotípicas possibilitando, por meio de seus marcadores genotípicos, a análise da divergência genética por métodos aglomerativos. Além disso, pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na realização do zoneamento agronômico, com fins de recomendação regionalizada e seleção de locais de teste, permitindo também a análise detalhada da interação ($G \times E$) e fácil interpretação gráfica dos resultados da análise estatística (ZOBEL et al., 1988).

A análise AMMI surge, então, com a finalidade de estudar detalhadamente as interações $G \times E$, uma vez que a soma de quadrados da interação ($SQ_{G \times E}$) é importante para

explicar o padrão adjacente à interação $G \times E$. Este estudo é feito por meio da decomposição da soma de quadrados dessas interações, fato que o torna vantajoso se comparado aos métodos tradicionais, além da sua boa capacidade preditiva.

No entanto, tem-se observado em trabalhos científicos que alguns resultados obtidos na utilização do procedimento AMMI são caracterizados pela falta de precisão nas análises gráficas dos biplots para as estabilidades. Assim, uma metodologia complementar é necessária para avaliar a precisão dos escores de genótipos e de ambientes, permitindo uma melhoria na qualidade das inferências para as estabilidades fenotípicas. Como os escores AMMI têm distribuição de probabilidade desconhecida, isso dificulta a obtenção de fórmulas via argumentos analíticos para medir a precisão destes escores. Os recentes avanços computacionais permitiram o desenvolvimento de sistemas interativos de processamento de dados, com algoritmos rápidos e precisos. Dentro desse enfoque, buscou-se ferramentas estatísticas que permitissem uma análise criteriosa da precisão dos escores AMMI e optou-se, neste estudo, pelo uso da reamostragem "*bootstrap*". Como os coeficientes dos escores de genótipos e de ambientes são obtidos a partir da matriz de resíduos, procurou-se enquadrar a ferramenta estatística "*bootstrap*" à necessidade da análise AMMI. Em concordância com o modelo AMMI, que permite a estimação dos genótipos associados aos ambientes livres de "ruídos", optou-se pela reamostragem na matriz de resíduos, obtida dos valores estimados. Assim, o procedimento AMMI *bootstrap* tem caráter de refinamento para análises de estabilidade fenotípica, mostrando grande potencial complementar à análise AMMI. No entanto, na maioria das vezes, apenas uma estimativa pontual a partir de uma amostra não é informação suficiente para se fazer inferências sobre o parâmetro de interesse em uma reamostragem *bootstrap*. Também é interessante que se tenha idéia da precisão ou possível erro dessa estimativa e os intervalos de confiança possuem essas características.

Assim, o objetivo deste trabalho é estudar o efeito da interação $G \times E$, avaliar a adaptabilidade e estabilidade de genótipos em diferentes ambientes através do modelo AMMI e, através da reamostragem *bootstrap* não-paramétrica, obter os intervalos de confiança interpercentílicos, desenvolvendo assim procedimentos estatísticos que possibilitem análises gráficas e numéricas dos genótipos e dos ambientes envolvidos, bem como a classificação dos genótipos em grupos por meio de técnicas multivariadas, avaliando a divergência genética destes e identificando aqueles que reúnam características de alta produtividade e tolerância a insetos mastigadores e sugadores.

2 REVISÃO BIBLIOGRÁFICA

A soja no Brasil e no mundo

A soja foi introduzida no sul do país, expandiu-se do estado do Rio Grande do Sul rumo ao norte a partir de 1960, abrangendo o oeste do estado de Santa Catarina e do Paraná, as antigas regiões cafeeicultoras do norte paranaense, sul e norte paulista, triângulo mineiro, toda a Região Centro-Oeste, e avança para regiões do Norte e do Nordeste (ALMEIDA, 2000).

Por suas características nutricionais, a soja é mais completa como alimento e prestou-se de modo mais adequado à difusão do modo capitalista de produção no campo, como agente catalisador de diversos segmentos do setor agropecuário e industrial (ALMEIDA, 2000).

Os 5 maiores países em produção de soja são responsáveis por 90% da produção mundial: EUA, Brasil, Argentina, China e Índia. O Brasil é o segundo maior produtor, respondendo por quase 25% do total mundial, o que mostra a grande importância econômica desta cultura no país e no mundo. Além disso, o complexo agroindustrial da soja movimenta aproximadamente US\$ 6 bilhões (10% das receitas cambiais) e ocupa uma área de 22 milhões de ha (cerca de 220.000 km²) - área equivalente à do estado de SP.

O aumento progressivo da demanda mundial e a disponibilidade de vastas áreas agrícolas no país têm servido de estímulo e reforço ao aumento da produção, muitas vezes sem considerar a capacidade ecológica de suporte em determinados ambientes e as consequências sociais e econômicas de um crescimento desequilibrado.

Por ser uma das espécies mais cultivadas no Brasil a partir dos anos 70, a soja ocupa posição de destaque na economia agrícola do país. No período entre 1986 e 2003, a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), em parceria com outras empresas, listaram mais de 200 cultivares recomendadas para plantio na maioria das regiões sojícolas do Brasil, representando uma maior produção ano após ano em função do desenvolvimento de novas cultivares com resistência a fatores bióticos e abióticos (EMBRAPA, 2003).

Os programas de melhoramento genético da soja estão voltados à criação de cultivares mais produtivas, mais estáveis e mais bem adaptadas às diversas regiões ecológicas e aos

vários sistemas de cultivo existentes no Brasil, incluindo a resistência às possíveis doenças e pragas presentes. Embora o número de cultivares disponíveis no Brasil seja grande, a base genética do germoplasma de soja brasileiro está extremamente reduzida, aumentando, dessa forma, a vulnerabilidade para perdas com doenças e pragas (BECELAERE et al. 2005). BONETTI (1983) estimou que cerca de 70% das cultivares desenvolvidas para o Rio Grande do Sul, naquela data descendiam das cultivares americanas Hill, Hood ou ambas.

A cultura da soja é comumente submetida a inúmeras variações ambientais e a interação $G \times E$ assume papel importante na manifestação fenotípica (OLIVEIRA et al., 2003). Dessa forma, estudos de divergência genética entre genótipos de soja são de importância econômica e têm merecido atenção cada vez maior, uma vez que provê informações sobre recursos genéticos disponíveis e auxilia na localização e no intercâmbio desses.

Interação genótipos $\times$ ambientes

As condições edafoclimáticas, associadas às práticas culturais, a ocorrência de patógenos e outras variáveis que afetam o desenvolvimento das plantas, são coletivamente denominadas ambiente (BOREM; MIRANDA, 2005). A importância do ambiente é incontestável no processo de melhoramento genético devido a influência que o ambiente exerce sobre cada genótipo. Além disso, a resposta diferenciada dos genótipos nos vários ambientes é um fenômeno natural conhecida como interação de genótipos com ambientes ($G \times E$) (Eberhart & Russel, 1966) e seus efeitos permitem a identificação de genótipos aptos a um ambiente específico (adaptabilidade), assim como de comportamento geral, aptos a vários ambientes (estabilidade).

Em um determinado ambiente, a manifestação fenotípica é o resultado da ação do genótipo sob a influência do meio. Entretanto, quando se consideram uma série de ambientes, detecta-se, além dos efeitos genéticos e ambientais, um efeito adicional, proporcionado pela interação destes.

A avaliação da interação genótipos $\times$ ambientes ($G \times E$) torna-se de grande importância no melhoramento, pois, no caso de sua existência, há possibilidade de o melhor genótipo em um determinado ambiente não o ser em outro (CRUZ et al., 2004).

A interação entre genótipo e ambiente existe toda vez que a expressão de determinado genótipo for dependente do ambiente onde ele é avaliado. A existência desta interação resulta na expressão $P = G + E + GE$ em que P representa o efeito do fenótipo, G do genótipo, E do ambiente e GE representa a interação genótipo $\times$ ambiente.

Para contornar os inconvenientes proporcionados pela interação genótipos $\times$ ambientes, recomenda-se a estratificação da região de adaptação da cultura em subregiões mais homogêneas. Entretanto, observa-se que, mesmo com esse procedimento, uma fração da interação ainda permanece, em razão da ocorrência de fatores incontrolláveis dos ambientes contra os quais a estratificação não oferece eficácia (CRUZ et al., 2004).

Apesar de ser básico, do ponto de vista genético, quer seja concernente aos parâmetros e/ou às mudanças genéticas, quer seja relativo ao progresso genético, torna-se importante a variação existente nas características. A detecção de significância para a interação não esclarece as conclusões que se possa ter sobre o melhoramento, sendo necessário um estudo detalhado deste componente de variação.

Essas variações, expressas como variâncias, podem ser de origem genética, fenotípica, ambiental, ou ainda, oriunda de combinações entre estas. Assim, ao se medir os componentes de variância, tal componente de interação cria uma variação adicional (VGE) resultante da expressão $VP = VG + VE + VGE$. A interação pode se expressar de diferentes formas e com diferentes intensidade, lembrando que a interação entre genótipos e ambientes pode interferir negativamente ou positivamente nos resultados finais.

Crossa (1990), revisando alguns métodos de análise estatística para ensaios de produtividade em multi-ambientes, coloca três principais objetivos agrícolas a serem atingidos nestes ensaios: (a) estimar com precisão e prever a produtividade, baseado em um número reduzido de dados experimentais, (b) determinar a estabilidade dos rendimentos e o padrão de resposta dos genótipos ou dos procedimentos agronômicos nos diferentes ambientes e (c) permitir uma orientação segura na seleção dos melhores genótipos ou procedimentos agronômicos. Na prática, para verificar a significância da interação de genótipos com ambientes, é necessário repetir o experimento várias vezes, pois se o experimento for realizado somente em um ambiente, poderá ocorrer uma superestimação dos ganhos genéticos.

Assim, a interação $G \times E$ deve ser considerada como um fenômeno biológico natural, que deve ser bem conhecido para melhor aproveitá-lo no processo de seleção (CHAVES,

2001). Logo os genótipos que interagem positivamente com os ambientes podem fazer a diferença entre um bom e um ótimo cultivar (DUARTE; VENCOVSKY, 1999).

Modelo de Efeitos Aditivos com Interação Multiplicativa – AMMI

Segundo Gauch (1988), o procedimento AMMI faz a decomposição da soma de quadrados original ($SQ_{G \times E}$) em eixos singulares ortogonais, que podem ser separados em duas partes, padrão e ruído. Os primeiros eixos associados ao modelo AMMI detêm praticamente todo o padrão devido à interação, e os eixos de mais alto grau captam, praticamente, somente ruídos.

Sendo assim, o padrão possibilita a identificação de fatores genotípicos e ambientais mais associados à interação, e o ruído, sem interesse agrônômico é descartado, propiciando uma melhoria na capacidade preditiva das respostas fenotípicas (Lavoranti, 2003).

Em Gauch e Zobel (1996) é feita uma revisão da metodologia AMMI no período de 1992 a 1996, na qual mostram ser útil para apresentar o melhor entendimento de interações complexas genótipo $\times$ ambiente e comentam também sobre a análise gráfica “biplot”, a capacidade do modelo AMMI em construir um modelo rico em padrão, deixando o resíduo rico em ruído, o que faz com que o modelo tenha uma maior precisão. Para estes autores a utilização do modelo AMMI significa um ganho de precisão equivalente ao aumento de 5 a 20 repetições.

O primeiro passo para entender os Modelos Aditivos com Interação Multiplicativa (AMMI) é o tipo de dados para os quais esses modelos são aplicados. Primeiramente, deve-se estabelecer um Fatorial de dupla entrada, com repetições ou não. Neste caso, os dados devem estar organizados em uma tabela de dupla entrada, tais como nos delineamentos fatoriais genótipos $\times$ ambientes (G $\times$ E). Os dois fatores podem ser chamados de linhas e colunas, e a combinação entre seus níveis de tratamentos. Os fatoriais de uma e tripla entrada não são permitidos, a razão desta restrição é que embora a parte da Análise da Variância (ANOVA) do AMMI seja flexível, a parte da Análise de Componentes Principais (ACP) precisa de uma estrutura de dupla entrada. No entanto, essa abordagem pode ser feita pelos modelos de Tucker ou Parafac (Bro, 1998). Uma tabela de tripla entrada, por exemplo, genótipos, locais e

anos, pode se aproximar através do modelo AMMI como um ou vários subproblemas de dupla entrada. Os dados podem ter repetições ou não quando o AMMI é usado para estimação, seleção, modelamento e outros propósitos. Em resumo, o AMMI pode ser usado para dados com estrutura de dupla entrada com pelo menos três linhas e três colunas, com ou sem repetições.

Em relação ao tipo dos dados para um modelo AMMI, estes devem ser de uma única variável resposta. Não é permitida uma matriz de várias linhas ou colunas que contenham variáveis. Padronizar cada linha de matriz para a média zero e variância unitária não tem sentido para o modelo AMMI pois isso elimina o significado das linhas ou colunas desse modelo.

A respeito do ajuste do modelo, com cada análise estatística, o modelo AMMI tem um modelo específico, e só se este modelo é usado para ajustar dados é que então os resultados serão realmente úteis. Em geral, AMMI é o modelo selecionado quando os dados apresentam efeitos principais e interação significativa. (GAUCH JÚNIOR, 1992).

Desta forma, com os dados organizados em uma tabela de dupla entrada, com os genótipos nas linhas e os ambientes nas colunas, a interação $G \times E$ é facilmente estudada. Segundo Barros (2007), o processo usual de investigação da interação entre os genótipos e os ambientes é feito através da análise da variância conjunta (ANOVA) envolvendo vários experimentos, e por essa análise é possível determinar a magnitude da interação, normalmente através do teste F que determina a magnitude da interação usando a razão entre o Quadrado Médio da Interação ($QM_{G \times E}$) e o Quadrado Médio do Resíduo (QM_{Res}).

Uma consideração importante a ser feita para realização do teste F e interpretação dos resultados da análise conjunta da variância, diz respeito à natureza fixa ou aleatória dos efeitos do modelo de análise (CHAVES, 2001). A natureza fixa ou aleatória da interação é determinada pelos efeitos principais. Se genótipos e ambientes são fixos, a interação será fixa. Se pelo menos um dos fatores for aleatório, a interação será aleatória. Na análise de ensaios multi-ambientais, consideram-se, em geral, os efeitos de genótipos como fixos e os efeitos de ambientes como aleatórios, de tal forma que o efeito da interação genótipos $\times$ ambientes é aleatória nesse caso. Isto se deve ao fato de normalmente se estar interessado em verificar qual o melhor dos genótipos avaliados em relação à produtividade e se existe uma variabilidade de sua produção em relação ao ambiente onde foi cultivado, assim tem-se um modelo misto com genótipos fixo e ambiente aleatório.

Os cálculos com álgebra de matrizes e modelos lineares, demandados por essas técnicas, são realizados facilmente devido a popularização do uso de computadores e de modernos pacotes estatísticos e desta forma, as técnicas multivariadas ganharam espaço no estudo da interação $G \times E$ nos últimos anos.

Vantagens do uso do modelo AMMI

O modelo de efeitos aditivos com interação multiplicativa (AMMI) é uma análise mais recente no estudo da interação genótipos e ambientes ($G \times E$) que tem como objetivo selecionar modelos que expliquem o padrão de resposta da interação, deixando fora o ruído presente nos dados (DUARTE; VENCOSKY, 1999).

A análise AMMI (“additive main effects and multiplicative interaction model”), como o próprio nome diz é um método uni-multivariado, que engloba uma análise de variância (ANOVA) para os efeitos principais, que são os efeitos dos genótipos (G) e os ambientes (E) e ainda leva em conta uma análise multivariada (neste caso, a análise de componentes principais - ACP) dos efeitos provocados pela interação genótipo $\times$ ambiente ($G \times E$). Na ANOVA, as estimativas dos efeitos dos genótipos e dos ambientes são somadas à média geral, enquanto que na ACP os efeitos são multiplicados entre si e depois são somados à média geral (GAUCH JÚNIOR, 1992).

A análise AMMI combina num único modelo componentes aditivos para os efeitos principais e componentes multiplicativos para os efeitos da interação $G \times E$, podendo a ACP ser diretamente tratada via decomposição por valores singulares (DVS) da matriz de dados. Essas técnicas multivariadas são baseadas no uso dos autovalores e autovetores provenientes da matriz de interação $G \times E$, por meio da DVS. Fazendo a DVS da matriz residual de interação pode-se obter o Biplot, que é uma representação gráfica em que as linhas e as colunas são apresentadas em um gráfico com duas ou três dimensões (GABRIEL, 1971).

Utilizando a técnica de biplots (KEMPTON, 1984), percebe-se que o método AMMI permite uma análise mais detalhada da interação $G \times E$ com as seguintes vantagens: garante a seleção de genótipos mais produtivos (capazes de capitalizar interações positivas com ambientes); propicia estimativas mais precisas das respostas genotípicas e possibilita uma

fácil interpretação dos resultados, por meio dos biplots em uma representação gráfica simultânea dos genótipos e ambientes.

Segundo Lavoranti (2004), essa representação fundamentada nos escores dos efeitos da interação para cada fator permite que a interpretação possa ser feita por meio da análise da magnitude e do sinal dos escores IPCA (interaction principal component analysis).

O ponto forte da análise é sumarizar grande parte da interação ($G \times E$) em apenas uns poucos eixos ($SQ_{G \times E \text{ Padrão}}$), consumindo assim, poucos graus de liberdade e resultando num modelo reduzido e parcimonioso, que descarta um resíduo adicional ($SQ_{G \times E \text{ Ruídos}}$) (ROCHA, 2002).

Os escores AMMI têm distribuição de probabilidade desconhecida, fato que dificulta a obtenção de fórmulas via argumentos analíticos. Para medir a precisão desses escores, Lavoranti (2003) apresentou uma metodologia complementar para avaliar a precisão dos escores de genótipos e de ambientes, por meio de técnicas de reamostragem '*bootstrap*', permitindo uma melhoria na qualidade das inferências sobre as adaptabilidades e estabilidades fenotípicas estimadas pelo modelo AMMI.

A interação $G \times E$ é representada pelos termos multiplicativos do modelo AMMI e a validação cruzada um procedimento estatístico livre de distribuição teórica que tem seu objetivo na determinação do número desses termos a serem retidos no modelo AMMI. O número ótimo de IPCA (eixo da interação da análise de componentes principais) a ser retido no modelo AMMI ou o melhor modelo da família AMMI ($AMMI_0$, $AMMI_1$, $AMMI_2$, ..., $AMMI_n$) pode ser determinado por duas abordagens: *predictiva* e *posdictive* (Fox et al., 1997).

Na abordagem *predictiva*, a técnica mais usada é a validação cruzada. Nela, os dados originais de repetições são divididos em dois grupos: dados de modelagem (dados de uma repetição) e os dados de validação (dados das demais repetições); os dois grupos são analisados, paralelamente, várias vezes, para cada modelo da família AMMI por reamostragem aleatória dos dados de repetições, que são então comparados tomando-se por base a diferença preditiva média (DPM). O modelo mais preditivo e parcimonioso é aquele com menor DPM (Gauch, 1992; Gauch e Zobel, 1996).

Na abordagem *posdictive* utiliza-se de teste de hipótese para avaliar as significâncias dos IPCA's ou seus resíduos (resíduo AMMI). O teste F de Gollob (1968) é o mais adotado,

por ser de fácil análise. Este teste tende a ser mais liberal em selecionar mais eixos do que o verdadeiro modelo contempla.

A seleção do número ótimo desses termos através de outros testes de hipóteses pode ser feita como os apresentados por Gollob (1968), Mandel (1971), Cornelius (1993), Corsten e van Eijnsberger (1972), Johnson e Graybill (1972) e Yochmowitz e Cornell (1978) entre outros, ou através de formas apresentadas por Stone (1974), Wold (1976) e Krzanowski (1987). Dias e Krzanowski (2003) descrevem duas metodologias preditivas para validação cruzada, as mesmas apresentadas por Krzanowski (1987) e Gabriel (2002), baseadas em dados “leave-one-out”, com vantagens sobre os testes de Gollob e Cornelius usados para esse mesmo fim.

Piepho (1995) investigou a robustez (para as suposições de homogeneidade e normalidade dos resíduos) de alguns destes testes. Ele confirma que o teste F aplicado de acordo com o critério de Gollob (1968) é liberal, pois ele seleciona muitos termos multiplicativos. Dos quatro métodos que ele estudou o teste proposto por Cornelius foi o mais robusto. O autor então recomenda que avaliações preliminares devam ser conduzidas para verificar a validade das suposições se algum dos outros testes for usado.

Vale ressaltar que, a escolha de um teste mais rigoroso para a seleção do modelo AMMI implica, consequentemente, na diminuição do número de eixos utilizados para explicar as somas de quadrados dos efeitos multiplicativos e isto pode levar a uma menor porcentagem da soma de quadrados, explicada por esses eixos. Entretanto, diferentemente da maneira convencional, em que se busca uma explicação máxima dessa soma de quadrados, no modelo AMMI, busca-se absorver apenas o “padrão” devido à interação.

Análise da Divergência Genética utilizando métodos multivariados

Estudos de divergência genética são importantes para o conhecimento da variabilidade genética das populações e possibilitam o monitoramento de acessos (CRUZ; CARNEIRO, 2006), pois geram informações úteis para preservação e uso destes (TOQUICA et al., 2003). O conhecimento da divergência genética tem proporcionado um contínuo aprimoramento das técnicas de análises, essencialmente aquelas envolvendo métodos multivariados com cálculos mais exatos e com fundamentos matemáticos mais sólidos. Além disso, a partir da análise da

variabilidade genética é possível estabelecer estratégias de melhoramento e a seleção de genótipos superiores.

A escolha apropriada de genitores possibilita a obtenção de genótipos superiores e é essencial para o sucesso de qualquer programa de melhoramento. Dentro de determinada faixa de diversidade, é aceito que cruzamentos de pais mais distantemente relacionados mostram maior heterose que os de pais mais estreitamente relacionados. Na produção de híbridos, o fenômeno da heterose é muito importante porque o híbrido resultante mostra superioridade agrônômica sobre os seus genitores. O termo heterose foi introduzido por Shull, em 1914, (Shull, 1948) para explicar o aumento de vigor, tamanho, produção de frutos, tempo de crescimento, resistência a doenças e pragas, ou a fatores climáticos, manifestada pelo cruzamento dois organismos diferentes. O relacionamento entre os pais divergentes e a heterose para características complexas, como é o caso da produção de grãos, pode ser visto como consequência da complementação de características componentes (GHADERI, A. et al., 1984).

Assim, o sucesso em programas de melhoramento de plantas depende da diversidade genética do material utilizado, pois quanto maior esta diversidade, maior a probabilidade de se detectar indivíduos superiores, ou seja, com as características comerciais de interesse. Dias (1998) alerta que muitas distâncias genéticas são essencialmente medidas de distâncias geométricas, sem qualquer conteúdo genético, e por isso são denominadas de coeficientes de dissimilaridade ou divergência. Outro ponto considerado importante por certos autores é a distância geográfica como indicadora de divergência genética. Ram e Panwar (1970), trabalhando com arroz, perceberam que os padrões de agrupamento foram relacionados com a distância geográfica.

Assim, para avaliar a diversidade genética de forma simultânea em relação a todas as características de interesse, recomenda-se a utilização de medidas de dissimilaridade (CRUZ; CARNEIRO, 2006). Uma forma prática e eficiente de se obter estas medidas é por meio da análise de agrupamentos (Análise de Cluster – AC), a qual tem por finalidade reunir as variáveis em grupos, de forma que exista máxima homogeneidade dentro do grupo e máxima heterogeneidade entre os grupos (JOHNSON; WICHERN, 1992; CRUZ et al., 2004).

Atualmente, diversos coeficientes de similaridade são propostos para a análise das divergências genéticas, principalmente devido à utilização das técnicas multivariadas (MEYER, 2002). Quando um experimento depende de muitas variáveis, não basta conhecer

informações estatísticas isoladas para cada variável, mas é necessário também conhecer a totalidade destas informações fornecida pelo conjunto das variáveis. As relações existentes entre as variáveis não são percebidas e assim efeitos antagônicos ou sinérgicos de efeito mútuo entre variáveis complicam a interpretação do fenômeno a partir das variáveis consideradas. A necessidade de entender as relações entre várias variáveis faz com que a análise multivariada seja de grande importância. Com esta análise pode-se reduzir os dados ou simplificar a sua estrutura sem muita perda da informação contida nos dados obtendo assim uma fácil interpretação dos resultados, principalmente quando as variáveis são correlacionadas.

Desta forma, para determinar quão distante geneticamente uma população ou genótipo é de outra são utilizados os métodos de agrupamentos, com os quais se quantifica ou se estima a heterose, que são analisados pela estatística multivariada permitindo unificar múltiplas informações de um conjunto de caracteres.

Vários métodos podem ser utilizados, dentre eles estão a análise por componentes principais, variáveis canônicas e métodos aglomerativos. Mais especificamente a técnica de componentes principais, apresenta entre seus principais objetivos a possibilidade de eliminar os ruídos presentes nos dados, possibilitando dessa forma uma interpretação segura do padrão de resposta presente nos dados. O ruído sugere que as respostas são imprevisíveis e não interpretáveis sendo parte integrante da variabilidade estranha contida nos dados, enquanto que o padrão responde de forma sistemática, significativa e interpretável pois permite quantificar e informar o grau de semelhança ou de diferença entre pares de indivíduos.

Além disso, a aplicação dos métodos multivariados é tida como uma técnica que melhor explora as informações contidas nos dados e, em estudos de divergência genética por métodos aglomerativos, a medida de distância mais amplamente utilizada é a distância euclidiana. A vantagem dessa estimativa de dissimilaridade é devida ao universo euclidiano e ao fato de minimizar os erros da classificação nos agrupamentos. Essa distância é a mais recomendada quando as unidades de cálculos são escores de componentes principais (CRUZ et al., 2004), como é o caso da análise AMMI (additive main effects and multiplicative interaction analysis).

Análise de agrupamento

A análise de agrupamento permite classificar n itens (populações, clones, variedades, indivíduos, etc.) avaliados por um conjunto de m variáveis, cujo objetivo é identificar e separar os itens em grupos, de forma que os mais semelhantes permaneçam no mesmo grupo. Em agrupamentos hierárquicos, tem-se que em todos os casos se desconhece “a priori” o número e a composição dos diferentes grupos ou “clusters” a serem formados.

Inicia-se o processo definindo-se os indivíduos e os objetivos desejados para a aplicação da análise, além dos critérios que irão definir as similaridades entre eles. Assim os dados são dispostos na forma de uma matriz, em que as linhas representam os indivíduos de interesse e as colunas representam as variáveis. A matriz de distância é gerada a partir de amostras de n itens, totalizando $n(n-1)/2$ pares de distâncias.

Existem inúmeros métodos de agrupamento, diferenciados nas formas de especificar a medida de proximidade entre indivíduos. Cada critério de análise de agrupamento impõe certo grau de estrutura nos dados. Portanto, recomenda-se diferentes critérios de agrupamento sejam aplicados e comparados entre si (DIAS, 1998; CRUZ ; CARNEIRO, 2006).

Em se tratando dos métodos de agrupamento, as delimitações podem ser estabelecidas por um exame visual do dendrograma, em que se avaliam pontos de alta mudança de nível, tomando-os em geral como delimitadores do número de indivíduos para determinado grupo (CRUZ; CARNEIRO, 2006), contextualizados neste trabalho pelos genótipos.

É importante ressaltar que o dendrograma ilustra as fusões ou partições efetuadas em cada nível sucessivo do processo de agrupamento, no qual um eixo representa os indivíduos e o outro eixo representa as distâncias obtidas após a utilização de uma metodologia de agrupamento. Os ramos da árvore fornecem a ordem das $(n-1)$ ligações, em que o primeiro nível representa a primeira ligação, o segundo a segunda ligação, e assim sucessivamente, até que todos se juntem.

Entretanto, a falta de critérios objetivos para se determinar o ponto de corte no dendrograma (número ótimo de grupos) ainda é um problema em estudos que utilizam a análise de agrupamentos. Um método considerado como “objetivo”, dentre os poucos existentes, é o Método de Mojena (1977). Este Método é um procedimento baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma.

A análise de agrupamento hierárquico consiste no tratamento matemático de cada amostra como um ponto no espaço multidimensional descrito pelas variáveis escolhidas (MOITA NETO, J. M.; MOITA, G. C., 1998). Também é possível, nesta técnica, tratar cada variável como um ponto no espaço multidimensional descrito pelas amostras, ou seja, podemos ter agrupamento de amostras ou de variáveis de acordo com o interesse em cada situação. Quando uma determinada amostra é tomada como um ponto no espaço das variáveis, é possível calcular a distância deste ponto a todos os outros pontos, constituindo-se assim uma matriz que descreve a proximidade entre todas as amostras estudadas.

Existem várias maneiras de calcular a distância entre dois pontos, a mais conhecida e utilizada é a distância euclidiana. Além disso, existem vários algoritmos a serem utilizados na formação dos agrupamentos hierárquicos, contudo todos eles utilizam as informações da matriz de proximidade para criar um dendrograma de similaridade.

A interpretação de um dendrograma de similaridade entre amostras fundamenta-se na intuição: duas amostras próximas devem ter também valores semelhantes para as variáveis medidas, portanto, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas. O gráfico denominado por dendrograma hierarquiza esta similaridade de modo que podemos ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras utilizado no estudo. Um caso particular é verificado quando o dendrograma construído é das variáveis, a similaridade entre duas variáveis aponta forte correlação entre estas variáveis do conjunto de dados estudado.

Convém ressaltar que a aplicação da análise de agrupamento hierárquico, quando temos variáveis de escalas diferentes, deve ser precedida por um tratamento prévio dos dados. Quando não é feito o pré-tratamento, as variáveis com valores numéricos mais altos serão mais importantes no cálculo que as variáveis com valores numéricos mais baixos. O pré-tratamento mais comumente empregado é a transformação Z, que transforma as medidas de cada variável de tal modo que o conjunto de dados tenha média zero e variância unitária. A finalidade deste procedimento é equalizar a importância estatística de todas as variáveis utilizadas. Tendo por base estas informações, o algoritmo tradicional da estrutura de formação de '*clusters*' do tipo hierárquica, de maneira geral, tem os seguintes passos:

1. Considere-se uma base inicial de N *clusters* iniciais. Em geral, esses agrupamentos correspondem simplesmente às unidades a serem agrupadas, cada um desses N

clusters contém apenas uma unidade inicialmente. A cada unidade i , está associado um vetor de m características $x_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$.

2. Calcula-se a distância entre todos os pares formados por elementos dentre esses N *clusters* iniciais. A distância nesse caso, pode ser qualquer medida de dissimilaridade entre o conjunto de atributos $x_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$. Para uma discussão sobre as diversas medidas de dissimilaridade, vide Khattree e Naik (2000).
3. Sejam I e J os dois *clusters* apresentando a menor distância, ou dissimilaridade, entre eles. Agrupa-se então o par I e J em um único novo cluster. O número de *clusters* agora passa a ser $N-1$.
4. Para os $N-1$ novos *clusters*, depois da junção descrita no passo 3, calculam-se as distâncias entre todos os pares. Para o par com a menor distância, agrupam-se os elementos em um único novo cluster, de forma que o número de *clusters* existentes passe a ser $N-2$.
5. Repetem-se os passos 2 a 4 até se obter um único cluster, que deverá conter todos os N *clusters* iniciais.

Ao final do processo, o analista terá em mãos uma árvore descrevendo a sequência de agrupamentos em cada passo do algoritmo. Para um número inicial de N unidades observacionais na base de dados, ao todo ocorrem $N-1$ junções.

Existem na literatura vários métodos para AAH. No agrupamento hierárquico denominado Método do Vizinho Mais Próximo ou Método de Ligação Simples (“Single Linkage Method”), forma-se um grupo inicial de indivíduos mais similares e em seguida são calculadas as distâncias daquele grupo em relação aos demais indivíduos usando-se a menor distância do conjunto. Neste método, as conexões entre indivíduos e grupos são feitas por ligações simples entre pares de indivíduos, ou seja, a distância entre os grupos é definida como sendo aquela entre os indivíduos mais parecidos entre esses grupos (CRUZ ; CARNEIRO, 2006). Este método leva a grupos longos e os dendrogramas resultantes deste procedimento são geralmente pouco informativos, devido à informação dos indivíduos intermediários que não são evidentes.

Quando a semelhança entre dois grupos é definida como a máxima distância entre pares de indivíduos tem-se o Método do Vizinho Mais Distante ou Método de Ligação Completa (“Complete Linkage Method”). A similaridade entre dois grupos é definida como

aquela apresentada pelos indivíduos de cada grupo que menos se parecem, formando-se todos os pares com um membro de cada grupo. Este método, geralmente leva a grupos compactos e discretos, tendo os seus valores de similaridade relativamente pequenos (MEYER, 2002).

No entanto, se a distância entre os pares de grupos for estimada com base na média aritmética dos indivíduos, tem-se o método hierárquico UPGMA (“Unweighted Pair-Group Method Using Arithmetic Averages”) ou Ligação média entre grupo, que evita caracterizar a dissimilaridade por valores extremos entre os genótipos, não considera a estrutura de subdivisão do grupo, dando pesos iguais a cada indivíduo do grupo calculando a similaridade média de um indivíduo que pretende se juntar ao grupo existente.

Já o método da Mediana ou WPGMA (“Weighted Pair-Group Method Using Arithmetic Averages”) difere do UPGMA pelo fato de atribuir aos indivíduos admitidos mais recentemente a um grupo, peso igual aos dos demais indivíduos já pertencentes ao grupo (ROMESBURG, 1990).

Dentre os métodos de agrupamento hierárquicos, o método de Ward é o que forma grupos de maneira a atingir sempre o menor erro interno entre os vetores que compõe cada grupo e o vetor médio do grupo. Isto equivale a buscar o mínimo desvio padrão entre os dados de cada grupo.

No método de Ward, os grupos de dados são formados em etapas. No princípio, têm-se m grupos; ou seja, um grupo para cada vetor componente da base de dados. Neste estágio inicial o erro interno é nulo para todos os grupos pois cada vetor que compõe cada grupo é o próprio vetor médio do grupo. Igualmente o desvio padrão para cada grupo é nulo. Na etapa subsequente, cada possibilidade de aglutinação entre os grupos 2 a 2 é verificada, e é escolhido o agrupamento que causa o menor aumento no erro interno do grupo. São $m \times m$ verificações. Nota-se que a cada iteração tem-se $m - i$ grupos (i = número de iterações), no entanto, como o número de elementos pertencentes a cada grupo aumenta, é maior o número de cálculos para o erro interno de cada grupo.

Ressalta-se que o agrupamento conduz à perda de informações ao nível de indivíduos, restando apenas informações sobre grupos similares. Por esse motivo, o estudo da divergência pode ser conduzido também simultaneamente, por componentes ou coordenadas principais, isto é particularmente importante quando o número de indivíduos é grande (CRUZ; CARNEIRO, 2006).

Soma de quadrados progressiva (Ward 1963, Orlóci 1978)

O critério de agrupamento de Ward minimiza o aumento na soma de quadrados dentro do grupo formado a cada passo de agrupamento, isto é, $Q_{PQ} = Q_{P+Q} - Q_P - Q_Q$ em que Q_{P+Q} é a soma de quadrados total no grupo $P+Q$; Q_P e Q_Q são as somas de quadrados dentro dos grupos P e Q . Tem-se que $Q_{P+Q} = \frac{1}{n_p + n_q} \sum_h \sum_i d_{hi}^2$ para $h=1, \dots, n-1$ e $i=h+1, \dots, n$ objetos, desde

que h e i pertençam ao grupo P ou Q ;

$$Q_P = \frac{1}{n_p} \sum_h \sum_i d_{hi}^2 \text{ para } h=1, \dots, n-1 \text{ e } i=h+1, \dots, n \text{ objetos, desde que } h \text{ e } i \text{ pertençam ao}$$

grupo P e $Q_Q = \frac{1}{n_q} \sum_h \sum_i d_{hi}^2$ para $h=1, \dots, n-1$ e $i=h+1, \dots, n$ objetos, desde que h e i pertençam ao grupo Q .

Análise de componentes principais (ACP)

Podemos definir a Análise de Componentes Principais como uma técnica multivariada que consiste em realizar uma transformação linear nas variáveis em estudo, transformação esta que produz novas variáveis não correlacionadas. Em geral, é utilizada com o objetivo de redução da dimensionalidade, considerando-se que algumas das novas variáveis são suficientes para representar adequadamente os dados originais, ou seja, a estrutura de variância e covariância de um conjunto de variáveis (originais) é explicada através de algumas combinações lineares destas variáveis (componentes).

Algebricamente a técnica de componentes pode ser descrita da maneira a seguir. Considere as variáveis aleatórias $X_1, X_2, \dots, X_p$ e o vetor aleatório $\mathbf{X}' = [X_1 \ X_2 \ \dots \ X_p]$ com matriz de variâncias e covariâncias Σ cujos autovalores associados são $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Considere as seguintes combinações lineares

$$C_1 = a_1' \mathbf{X}$$

$$C_2 = \mathbf{a}_2' \mathbf{X}$$

.

.

.

$$C_p = \mathbf{a}_p' \mathbf{X}$$

em que $\mathbf{a}_j' = [\mathbf{a}_{1j} \ \mathbf{a}_{2j} \ \dots \ \mathbf{a}_{pj}]$. $C_1, C_2, \dots, C_p$ são chamadas de componentes principais. Podemos observar que $V(C_j) = \mathbf{a}_j' \Sigma \mathbf{a}_j$ e $\text{Cov}(C_j, C_k) = \mathbf{a}_j' \Sigma \mathbf{a}_k$, para $j, k=1, \dots, p$.

As componentes principais são determinadas de forma a atender às seguintes condições:

- (i) A primeira componente principal (C_1) é a combinação linear $\mathbf{a}_1' \mathbf{X}$ que maximiza $V(C_1)$, sujeita à restrição $\mathbf{a}_1' \mathbf{a}_1 = 1$;
- (ii) a segunda componente principal (C_2) é a combinação linear $\mathbf{a}_2' \mathbf{X}$ que maximiza $V(C_2)$, sujeita às restrições $\mathbf{a}_2' \mathbf{a}_2 = 1$ e $\text{Cov}(\mathbf{a}_1' \mathbf{X}, \mathbf{a}_2' \mathbf{X}) = 0$;
- (iii) a j -ésima componente principal (C_j) é a combinação linear $\mathbf{a}_j' \mathbf{X}$ que maximiza $V(C_j)$, sujeita às restrições $\mathbf{a}_j' \mathbf{a}_j = 1$ e $\text{Cov}(\mathbf{a}_j' \mathbf{X}, \mathbf{a}_k' \mathbf{X}) = 0 \ \forall \ k < j$.

Considerando-se as condições definidas nos itens (i)-(iii) acima a j -ésima componente principal é dada por $C_j = \mathbf{e}_j' \mathbf{X} = \mathbf{e}_{1j} X_1 + \mathbf{e}_{2j} X_2 + \dots + \mathbf{e}_{pj} X_p$, $j=1, \dots, p$, em que $(\lambda_j, \mathbf{e}_j)$ é o j -ésimo par de autovalor e autovetor ortonormalizado associados à matriz de variâncias e covariâncias Σ . Dessa forma, temos que $V(C_j) = \mathbf{e}_j' \Sigma \mathbf{e}_j = \lambda_j$ e $\text{Cov}(C_j, C_k) = \mathbf{e}_j' \Sigma \mathbf{e}_k = 0$, para $j, k=1, \dots, p$. Então, as componentes principais são combinações lineares não correlacionadas de $X_1, X_2, \dots, X_p$ e suas variâncias são iguais aos autovalores de Σ .

Além disso, temos que, como $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p V(X_i) = \text{tr}(\Sigma) = \lambda_1 + \lambda_2 + \dots + \lambda_p$,

então $\sum_{j=1}^p V(C_j) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p V(X_i)$, sendo $\sigma_{ii} = V(X_i)$. Como consequência desse

resultado temos que a proporção da variabilidade total devida à j -ésima componente é dada

por $\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$, $j=1, \dots, p$. É neste fato que se baseia um dos critérios adotados para

decidir quantas componentes devem ser consideradas para representar as variáveis originais sem muita perda de informação. As primeiras componentes que acumulem uma grande parte da variabilidade total (recomenda-se acima de 80%) serão retidas e substituirão as variáveis inicialmente consideradas (ANDERBERG, 1973).

Componentes Principais também podem ser obtidas após a padronização das variáveis, ou seja, consideram-se as variáveis $Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$, $i=1, \dots, p$. O correspondente vetor aleatório é

dado por $\mathbf{Z} = (\Sigma^{1/2})^{-1} (\mathbf{X} - \boldsymbol{\mu})$, em que $\Sigma^{1/2}$ e $\boldsymbol{\mu}$ são, respectivamente, a matriz diagonal dos desvios padrão e vetor de médias das variáveis que compõem $\mathbf{X}$. $E(\mathbf{Z}) = \mathbf{0}$ e $V(\mathbf{Z}) = \boldsymbol{\rho}$, sendo $\boldsymbol{\rho}$ matriz de correlações associada ao vetor $\mathbf{X}$. Dessa forma, as componentes principais de $\mathbf{Z}$ devem ser obtidas dos autovetores da matriz de correlações $\boldsymbol{\rho}$ de $\mathbf{X}$.

Os resultados anteriores também se aplicam neste caso, tornando-se um pouco mais simples já que as variâncias de todos os Z_i 's são iguais a 1, $i=1, \dots, p$, e conseqüentemente a variância total é simplesmente p (soma dos elementos da diagonal de $\boldsymbol{\rho}$) e assim, a proporção da variância total explicada pela j -ésima componente principal é dada por $\frac{\lambda_j}{p}$, $j=1, \dots, p$, onde

os λ_j 's são autovalores de $\boldsymbol{\rho}$.

A abordagem acima considerou as características populacionais das variáveis em estudo, cujos parâmetros são, em geral, desconhecidos. Para uma abordagem baseada na amostra, a Análise de Componentes Principais é realizada de forma equivalente. Suponha que $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ represente uma amostra aleatória de n indivíduos de uma população p -dimensional com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias Σ . Esses dados produzem um vetor de médias amostrais $\bar{\mathbf{x}}$, uma matriz de covariâncias amostrais $\mathbf{S}$ e uma matriz de

correlações amostrais $\mathbf{R}$. Essas matrizes $\mathbf{S}$ e $\mathbf{R}$ representam a estrutura de interdependência das variáveis.

Buscamos construir combinações lineares não correlacionadas das características medidas que expliquem muito da variação amostral. Essas combinações lineares com o máximo da variância amostral serão chamadas de *componentes principais amostrais*. Se $\mathbf{S} = \{s_{ik}\}$ ($p \times p$) é uma matriz de covariâncias amostrais com pares de autovalores e autovetores $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$, $i = 1, 2, \dots, p$, o i -ésimo *componente principal amostral* é dado por

$$\hat{C}_i = \hat{\mathbf{e}}_i^t \mathbf{x} = \sum_{j=1}^p \hat{e}_{ij} x_j, \text{ para } i = 1, 2, \dots, p$$

em que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ e $\mathbf{x}$ é uma matriz de observações das variáveis $X_1, X_2, \dots, X_p$.

Também,

- Variância amostral $(\hat{C}_i) = \hat{\lambda}_i$, $i = 1, 2, \dots, p$
- Covariância amostral $(\hat{C}_i, \hat{C}_k) = 0$, $i \neq k$
- Variância amostral total $= \sum_{i=1}^k s_{ii} = \sum_{i=1}^p \hat{\lambda}_i$

Correlação amostral entre a k -ésima variável original e o i -ésimo componente

principal amostral: $r_{\hat{y}_i, x_k} = \frac{\hat{e}_{ik} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}} \quad i, k = 1, 2, \dots, p$

Os valores das novas variáveis para cada unidade da amostra são chamados de coordenadas principais e podem ser escritos na forma $\hat{C}_{ij} = e_{1j}x_{11} + e_{2j}x_{12} + \dots + e_{pj}x_{1p}$.

Na forma matricial temos:

$$\mathbf{C} = \begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} & \cdots & \hat{c}_{1p} \\ \hat{c}_{21} & \hat{c}_{22} & \cdots & \hat{c}_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{c}_{n1} & \hat{c}_{n2} & \cdots & \hat{c}_{np} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}$$

$$\mathbf{C}_{n \times p} = \mathbf{X}_{n \times p} \mathbf{V}_{p \times p},$$

em que $\mathbf{V}$ é a matriz cujas colunas são os autovetores ortonormalizados de $\mathbf{S}$.

Como o objetivo da Análise de Componentes Principais é, em geral, a redução da dimensionalidade, há sempre a discussão de quantas componentes devem ser retidas. Pode-se levar em consideração a variabilidade acumulada total explicada pelas componentes a serem retidas, como também o tamanho dos autovalores, entre outros critérios, o que leva sempre a uma decisão subjetiva e com uma parcela de arbitrariedade. Levados por estes fatos, alguns autores propuseram procedimentos mais objetivos para a decisão de quantas componentes reter em uma Análise de Componentes Principais. Entre eles estão Stone (1974), Velicer (1976) e Wold (1976 e 1978).

Reamostragem *bootstrap*

Existem diversas técnicas de reamostragem que visam estimar parâmetros de uma distribuição de interesse. O método de reamostragem *bootstrap* foi originalmente proposto por Efron (1979) em um influente artigo publicado no *Annals of Statistics*. É um método de reamostragem que se baseia na construção de distribuições amostrais empíricas de uma estatística de interesse. Uma vantagem em utilizar a técnica de reamostragem *bootstrap* é a generalidade com que pode ser aplicada, pois requer que menos suposições sejam feitas.

Hesterberg et al. (2003) afirmam que a amostra original representa a população da qual foi retirada. Desta forma, tratando a amostra como se ela fosse a “população” realizando sucessivas reamostragens com reposição. A partir destas reamostras, é possível estimar características da população, tais como média, variância, percentis, etc.

A distribuição empírica de uma estatística, gerada pelo método *bootstrap* tem aproximadamente a mesma forma e amplitude que a distribuição amostral da estatística.

Barros et al. (2008) descreve o método *bootstrap* não paramétrico para obtenção de intervalos de confiança dos coeficientes de um modelo de regressão múltipla. Com esse propósito, o autor considera a seguinte notação:

Seja $U = (y, x_1, x_2, \dots, x_k)$, em que, $y = (y_1, y_2, \dots, y_n)$ é o vetor de observações da variável dependente e $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})$, $j = 1, \dots, k$, são os vetores de observações das variáveis independentes, x_j é uma variável contínua para todo j .

Intervalos p-*bootstrap*

- I. Amostrar, com reposição de U , uma amostra *bootstrap* $(y_1^*, x_{11}^*, x_{12}^*, \dots, x_{1k}^*), \dots, (y_n^*, x_{n1}^*, x_{n2}^*, \dots, x_{nk}^*)$.
- II. Da amostra *bootstrap* $(y_1^*, x_{11}^*, x_{12}^*, \dots, x_{1k}^*), \dots, (y_n^*, x_{n1}^*, x_{n2}^*, \dots, x_{nk}^*)$, obter o estimador de mínimos quadrados de β , representado por $\hat{\beta}^*$.
- III. Repetir os passos I e II um número B grande de vezes.
- IV. Ordenam-se os valores obtidos, do menor ao maior de forma que $\hat{\beta}^* = (\hat{\beta}_{(1)}^* \leq \hat{\beta}_{(2)}^* \leq \dots \leq \hat{\beta}_{(B)}^*)$.
- V. Determinam-se limites de confiança para uma especificada probabilidade α , igual ao nível de significância, sendo $0 < \alpha < 1$, de forma que o intervalo de confiança p-*bootstrap* $100 \times (1 - \alpha)\%$ é dado por $(\hat{\beta}_{(q_1)}^*; \hat{\beta}_{(q_2)}^*)$ em que, $q_1 = \lceil B \times (\alpha / 2) \rceil$ e $q_2 = B - q_1$, em que $\lceil x \rceil$ indica o menor número inteiro maior ou igual ao argumento x .

Como exemplo, em V, para $(\alpha = 0.05 \text{ e } B=1000) \Rightarrow (q_1 = 25 \text{ e } q_2 = 975)$. Logo, o intervalo de confiança p-*bootstrap* de 95% é dado por $(\hat{\beta}_{(25)}^*; \hat{\beta}_{(975)}^*)$. Os intervalos de confiança para quaisquer outros parâmetros de interesse são obtidos de maneira similar.

Pode-se então afirmar, com uma probabilidade a de se estar errado, que o intervalo de confiança construído tem alta probabilidade de conter o verdadeiro valor do parâmetro sobre o qual a estimação foi baseada.

Quando uma variável independente não é contínua, deve-se fazer o processo de reamostragem (I, II e III) dentro de cada nível da variável (Wu, 1986; Tibshirani, 1988). Outras alternativas ao intervalo de confiança p-*bootstrap* são discutidas, por exemplo, em Efron & Tibshirani (1993), Davison & Hinkley (1997) e Carpenter & Bithell (2000).

Para que a aplicação da técnica resulte em valores confiáveis devem ser feitas, a partir da amostra mestre, centenas ou até milhares de reamostras do mesmo tamanho n . A maioria dos autores recomenda a utilização de 1000 reamostras.

3 MATERIAL E MÉTODOS

Material experimental

O material genético utilizado envolve três tipos de populações obtidas a partir de um dialelo parcial 4×4 desenvolvido com a finalidade de reunir os genes de tolerância/resistência a insetos presentes em quatro genitores (IAC-100, Crockett, Lamar e D72-9601-1), com genes para desempenho agrônômico favorável – principalmente, produtividade de grãos e precocidade – presentes em quatro cultivares adaptadas (BR-6, IAS-5, Davis, Ocepar-4).

Os experimentos que originaram os dados do presente estudo envolveram linhagens dos três tipos de população (PCI, PRIS e PRIM) e foram conduzidos em dois locais do município de Piracicaba, SP (Estação Experimental Anhembi e Fazenda Areão), durante dois anos agrícolas (1999/00, 2000/01) e em dois sistemas de manejo (com controle intensivo de insetos e com controle ecológico de insetos).

Em relação às três populações tem-se que foram empregados três procedimentos distintos de seleção de plantas individuais, conforme relatado a seguir: PCI (populações com controle de insetos) – originadas da seleção para produtividade de grãos, aplicou-se inseticidas durante o ciclo total para o controle de insetos mastigadores e sugadores; PRIS (populações selecionadas para resistência a insetos sugadores) – plantas que receberam aplicações de inseticidas apenas durante o ciclo vegetativo, de maneira que os insetos sugadores pudessem desenvolver-se, multiplicar-se e atacar as vagens e grãos produzidos pelas plantas F₂; PRIM (populações selecionadas para resistência a insetos mastigadores) – iniciou-se a aplicação de inseticidas após o florescimento, com a finalidade de controlar insetos sugadores.

Os experimentos desenvolvidos em 1999/2000 incluíram 120 linhagens experimentais F₁₀, (40 de cada população); os experimentos de 2000/01 envolveram 60 linhagens F₁₁, sendo 20 linhagens selecionadas em cada população. As repetições foram estratificadas em conjuntos experimentais, cada um deles com três testemunhas comuns: IAC-100, OCEPAR-4 e IAS-5.

Nos experimentos com controle intensivo de insetos (CII), com base em monitoramento frequente, promoveu-se a aplicação de inseticidas sempre que se detectaram insetos mastigadores e ou dois percevejos/m² de pano (método do pano de batida). Nos experimentos com controle ecológico de insetos (CEI), promoveu-se aplicações de inseticidas apenas na ocorrência de grande quantidade de danos nas folhas causadas por insetos mastigadores e ou quando a infestação natural atingiu quatro percevejos/m² de pano. O caráter produtividade de grãos (PG) em gramas/parcela foi avaliado pelo peso dos grãos maduros colhidos na área útil de cada parcela, após um mínimo de três semanas de secagem em ambiente de laboratório para a padronização da umidade. Maiores informações sobre a geração F₂, método de condução das populações segregantes e avanço de gerações, podem ser obtidas no trabalho de Pinheiro (1998).

É importante ressaltar que os dados ideais para aplicação da técnica proposta neste estudo são oriundos de gerações iniciais (F₁, F₂) de programas de melhoramento, uma vez que apresentam ampla segregação possuindo maior volume informacional em estudos de diversidade. No entanto, dados desta natureza podem ser influenciados por perdas de indivíduos ou repetições (parcelas perdidas). Porém, nestes casos, a técnica de imputação de dados auxilia na resolução deste problema e com isso possibilita a utilização da técnica.

Dados de gerações avançadas de melhoramento (F₅, F₆), já sofreram processos direcionais de seleção, minimizando os dados “indesejáveis” bem como favorecendo a homozigose do material. Sendo assim de pouco valor informacional para estudos de diversidade e formação de agrupamentos, pois os indivíduos em estudo foram favorecidos em relação aos caracteres estudados. Portanto, no presente estudo, consideram-se os dados utilizados somente para critério de aplicação da técnica proposta.

Os experimentos foram conduzidos no delineamento aleatorizado em blocos, com as repetições estratificadas em conjuntos experimentais com testemunhas comuns: IAC-100, Ocepar, IAS-5 e Primavera. A cultivar Primavera foi eliminada nos experimentos do ano agrícola 2000. Foram considerados fixos os efeitos de populações (linhagens e testemunhas comuns), visto que se originaram de diferentes cruzamentos, não representando uma mesma população. O efeito de ambientes foi considerado aleatório.

Antes da obtenção da ANOVA foram analisadas as pressuposições do modelo em cada um dos quatro ambientes. Desta forma, verificou-se a normalidade dos resíduos, a homogeneidade entre as variâncias residuais, erros independentes e identicamente distribuídos

e efeitos aditivos no modelo. Tendo o modelo atendido a todas as pressuposições, em todos os ambientes, a análise de variância individual (em cada ambiente) foi obtida para avaliação estatística da variabilidade genética entre os tratamentos (linhagens de soja) e da precisão experimental. Uma vez detectada diferenciação entre as cultivares, realizou-se análise conjunta de variância, e assim, foram feitas as interpretações relativas às significâncias do teste F.

As coordenadas das linhagens e de ambientes nos eixos principais de interação (IPCA) foram representadas através do gráfico biplot.

Todas as análises estatísticas foram feitas com o software estatístico R (The R Development Core Team, 2008) usando os pacotes: fields (FURRER; NYCHKA, 2009) e agricolae (De MENDIBURU, 2009).

Análise de Variância Conjunta

Com o objetivo de verificar se a interação entre genótipos e ambientes foi significativa, realizou-se uma análise de variância conjunta (Tabela 1) que envolveu o estudo de todos os genótipos em todos os ambientes, sendo que em cada ambiente foi utilizado um delineamento aleatorizado em blocos. Assim, temos o seguinte modelo matemático:

$$Y_{ijk} = \mu + b_{k(j)} + g_i + e_j + (ge)_{ij} + \varepsilon_{ijk} \quad (1)$$

em que:

Y_{ijk} : é o valor observado do i -ésimo genótipo no j -ésimo ambiente e no k -ésimo bloco, com $i = 1, 2, \dots, g$, $j = 1, 2, \dots, e$ e $k = 1, 2, \dots, r$;

μ : é uma constante comum às respostas (normalmente a média geral);

$b_{k(j)}$: é o efeito do k -ésimo bloco dentro do j -ésimo ambiente;

$(ge)_{ij}$: é o efeito da interação específica do i -ésimo genótipo e o j -ésimo ambiente;

ε_{ij} : é o erro experimental associado ao i -ésimo genótipo no j -ésimo ambiente, e no k -ésimo bloco assumido ser independente e $\varepsilon_{ij} \sim N(0, \sigma^2)$;

Tabela 1 - Esquema da ANOVA Conjunta para experimentos de um mesmo grupo de genótipos avaliados em e ambientes com r blocos

Fonte de Variação	GL	QM
Blocos dentro do ambiente (B d. E)	$e(r-1)$	$QM_{B \text{ d. E}}$
Genótipos (G)	$(g-1)$	QM_G
Ambientes (E)	$(e-1)$	QM_E
Interação ($G \times E$)	$(g-1)(e-1)$	$QM_{G \times E}$
Resíduo	$e(g-1)(r-1)$	QM_{Res}
Total	$ger - 1$	

GL: Graus de liberdade; QM: Quadrado médio.

A estrutura da análise da variância (Tabela 1) permite a estimação da interação $G \times E$ que estaria confundida com a variância devida a genótipos na análise individual (por ambiente). Se os ambientes abrangidos pelos ensaios, representam, de fato, uma zona ecológica adequada à espécie, a variância genética devida a genótipos, sem a devida interação com ambientes, seria mais indicada para expressar de fato a variação disponível ao melhoramento.

Análise da interação G×E e o enfoque estatístico

Para a análise da interação considerou-se como ambiente a combinação local e manejo, para cada uma das três populações, sendo que cada população foi estudada em dois anos agrícolas (1999/2000 e 2000/2001). Assim, pela combinação dos dois locais (Anhembi e Areão) com os dois sistemas de manejo (controle intensivo de insetos = CII e controle ecológico de insetos = CEI), obteve-se quatro ambientes distintos (E1, E2, E3 e E4) para cada tipo de população (População 1 - PCI, População 2 - PRIS e População 3 - PRIM), quais sejam: [Anhembi-CII (E1), Anhembi-CEI (E2), Areão-CII (E3), Areão-CEI (E4)]. Esta situação se repetiu para cada ano agrícola. A análise foi realizada com as médias das duas repetições em cada ambiente e população.

Para a aplicação da metodologia AMMI (Zobel et al. 1988) empregou-se o teste F_{Gollob} proposto por Gollob (1968), feita a associação entre “*bootstrap*” e AMMI (Lavoranti, 2003) com base na reamostragem da matriz de resíduos (livres de ruídos), obtida dos valores estimados pelo método AMMI. Nos dois casos, foram esquematizadas as análises de variância pelo quadro completo da análise conjunta, a partir das médias (Tabela 2).

Tabela 2 - Disposição dos dados em linhas (genótipos) e colunas (ambientes)

Genótipos (G)	Ambientes					Média
	1	2	3	...	e	
G ₁	Y ₁₁	Y ₁₂	Y ₁₃	...	Y _{1e}	$\bar{Y}_{1.}$
G ₂	Y ₂₁	Y ₂₂	Y ₂₃	...	Y _{2e}	$\bar{Y}_{2.}$
G ₃	Y ₃₁	Y ₃₂	Y ₃₃	...	Y _{3e}	$\bar{Y}_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
G _g	Y _{g1}	Y _{g2}	Y _{g3}	...	Y _{ge}	$Y_{g.}$
Médias	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{.3}$	...	$\bar{Y}_{.e}$	$\bar{Y}_{..}$

Suponha que um conjunto de g genótipos tenha sido testado experimentalmente em e ambientes, que podem ser locais, anos etc. As médias de cada combinação de genótipo e

ambiente $(\bar{Y}_{ij})$, obtidas de r repetições de um experimento que consiste de dados balanceados, podem ser representadas como na da Tabela 2.

Nessa tabela de dupla entrada de dimensão $g \times e$, associada aos genótipos G_i colocados nas linhas ($i=1, 2, \dots, g$) e os ambientes E_j nas colunas ($j=1, 2, \dots, e$), assume-se que as n observações fenotípicas de uma variável Y_{ij} seguem o modelo matemático:

$$Y_{ij} = \mu + G_i + E_j + (GE)_{ij} + \bar{\varepsilon}_{ij} \quad (2)$$

em que:

Y_{ij} : é a observação fenotípica do i -ésimo genótipo no j -ésimo ambiente, podendo representar a média de r repetições;

μ : é a média geral;

G_i : é o efeito do i -ésimo genótipo ($i=1, \dots, g$);

E_j : é o efeito do j -ésimo ambiente ($j=1, \dots, e$);

$(GE)_{ij}$: é o efeito da interação i -ésimo genótipo com o j -ésimo ambiente,

$\bar{\varepsilon}_{ij}$: é o erro experimental médio associado ao i -ésimo genótipo no j -ésimo ambiente, assumido ser independente e pressupondo $\bar{\varepsilon}_{ij} \sim N(0, \sigma^2/r)$, sendo σ^2 a variância do erro dentro de ambiente, assumida como constante;

A análise de variância feita sob o modelo acima com base nos dados da Tabela 2 é simples, sendo calculados os efeitos principais pelas marginais desta tabela e a interação como desvios do modelo, em que $ge_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$ são os correspondentes elementos da tabela para o estudo da interação. O erro experimental $\bar{\varepsilon}_{ij}$ é calculado para cada experimento, utilizando-se o resíduo na análise conjunta.

Na solução do modelo exposto acima e visando encontrar os estimadores dos parâmetros pelo processo de mínimos quadrados, o sistema de equações normais na forma matricial é descrito a seguir:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1e} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2e} \\ \vdots \\ y_{g1} \\ y_{g2} \\ \vdots \\ y_{ge} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ g_1 \\ g_2 \\ \vdots \\ g_g \\ e_1 \\ \vdots \\ e_e \\ (ge)_{11} \\ (ge)_{12} \\ \vdots \\ (ge)_{ge} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{11} \\ \vdots \\ \varepsilon_{1e} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2e} \\ \vdots \\ \varepsilon_{g1} \\ \varepsilon_{g2} \\ \vdots \\ \varepsilon_{ge} \end{bmatrix}$$

Têm-se que a matriz X não é de posto coluna completo e, portanto, $(X'X)$ não possui inversa regular. Uma possível solução paramétrica do sistema de equações normais $X'X\hat{\beta} = X'Y$ é dada por $\hat{\beta} = (X'X)^- X'Y$ em que $(X'X)^-$ é uma inversa generalizada. Uma vez que o sistema é indeterminado, admite-se condições marginais:

$$\sum_{i=1}^g g_i = \sum_{j=1}^e e_j = \sum_{i=1}^g (ge)_{ij} = \sum_{j=1}^e (ge)_{ij} = \sum_{i=1}^g \sum_{j=1}^e (ge)_{ij} = 0 \quad (3)$$

Assim, a solução do sistema de equações normais é:

$$\hat{\mu} = \frac{\sum_{ij} Y_{ij}}{ge} = \bar{Y}_{..};$$

$$\hat{g}_i = \frac{\sum_j Y_{ij}}{e} - \hat{\mu} = \bar{Y}_{i.} - \bar{Y}_{..};$$

$$\hat{e}_j = \frac{\sum_i Y_{ij}}{g} - \hat{\mu} = \bar{Y}_{.j} - \bar{Y}_{..};$$

$$(ge)_{ij} = Y_{ij} - \frac{\sum_j Y_{ij}}{e} - \frac{\sum_i Y_{ij}}{g} + \frac{\sum_{ij} Y_{ij}}{ge} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..};$$

Esses são estimadores não viesados dos parâmetros μ , g_i , e_j e $(ge)_{ij}$, sujeitos às condições marginais dadas acima. Se essas ou quaisquer outras condições não forem impostas sobre os parâmetros, então os estimadores dos parâmetros individuais são viesados.

A aproximação de mínimos quadrados $\hat{Y}_{ij}$ (ou valor predito) e seu respectivo resíduo, correspondente ao termo geral de interação $(ge)_{ij}$, são dados por:

$$\hat{Y}_{ij} = \frac{\sum_j Y_{ij}}{e} + \frac{\sum_i Y_{ij}}{g} - \frac{\sum_{ij} Y_{ij}}{ge} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..};$$

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij}$$

Com esta especificação, a matriz de interações $GE_{(g \times e)} = (ge_{ij})$, é definida por:

$$GE_{(g \times e)} = \begin{pmatrix} ge_{11} & ge_{12} & \cdots & ge_{1e} \\ ge_{21} & ge_{22} & \cdots & ge_{2e} \\ \vdots & \vdots & \cdots & \vdots \\ ge_{g1} & ge_{g2} & \cdots & ge_{ge} \end{pmatrix}$$

Um dos procedimentos usuais na determinação dos graus de liberdade associados à parcela da $SQ_{G \times E}$ é obtido pelo teste F avaliando-se a significância de cada componente em relação ao QM(erro médio). Isso resulta num quadro de análise de variância semelhante ao tradicional, com desdobramento para a fonte de variação $G \times E$. Os quadrados médios são obtidos dividindo-se cada soma de quadrados (dos efeitos principais e da interação) pelos respectivos graus de liberdade. Já quadrado médio do resíduo ou do erro médio (QM_{Erro Médio}) é calculado pela média ponderada dos QM's do resíduo (obtido nas ANOVA's individuais de experimentos), assumidos homogêneos, usando seus respectivos graus de liberdade como pesos, ou seja:

$$QM_{Erro\ Médio} = \frac{\sum_j SQ_{Res\ j}}{\sum_j GL_{Res\ j}}, \quad (4)$$

em que $j=1,2,\dots, e$ ambientes ou experimentos.

O modelo AMMI

A análise AMMI foi realizada em duas etapas sequenciais: os efeitos principais, na parte aditiva (média geral, efeitos de genótipos e ambientes), foram ajustados por análise de variância (ANOVA), resultando em um resíduo de não aditividade (interação G×E); a interação (parte multiplicativa do modelo) foi analisada pela Análise de Componentes Principais (ACP). O modelo geral usado foi de acordo com aquele apresentado por Duarte e Vencovsky (1999), segundo o modelo matemático já mencionado.

O efeito da interação do i-ésimo genótipo com o j-ésimo ambiente, $(ge)_{ij}$, no modelo matemático foi modelado de acordo com a equação:

$$(ge)_{ij} = \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} \quad (5)$$

em que:

λ_k : é o valor singular do k-ésimo componente principal da interação (CPI) retido no modelo AMMI;

γ_{ik} é o vetor singular do i-ésimo genótipo no k-ésimo CPI;

α_{jk} é o vetor singular do j-ésimo ambiente no k-ésimo CPI;

ρ_{ij} o resíduo da interação G×E ou resíduo AMMI (ruído presente nos dados);

k: raízes características não nulas, ou seja, o número de CPI's retidos no modelo, sendo que $k = (1, 2, \dots, p)$, em que $\text{posto} = \min(g-1, e-1)$ é o posto da matriz GE.

Dessa forma, a matriz de interação GE é modelada pela equação de $(ge)_{ij}$ descrita acima, sob as restrições (condições marginais). Portanto, o termo GE (interação no modelo tradicional) na metodologia AMMI é representado pela soma de p parcelas, cada uma resultante da multiplicação de λ_k , expresso na mesma unidade de Y_{ij} , por um efeito genotípico (γ_{ik}) e um efeito ambiental (α_{jk}), ambos adimensionais, ou seja, $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$ (n: termos da interação). O termo λ_k traz uma informação relativa a k-ésima parcela da interação G×E ; os efeitos γ_{ik} e α_{jk} representam os pesos do genótipo i e do ambiente j, naquela parcela da interação.

λ_k : k-ésimo valor singular (escalar) da matriz de interação original (denotada por

$$GE_{G \times E} = \left| (ge) \right|_{ij}, \text{ com } (ge)_{ij} = (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..});$$

γ_{ik} : elemento correspondente ao i-ésimo genótipo no k-ésimo vetor singular coluna da matriz GE e;

α_{jk} : o elemento correspondente ao j-ésimo ambiente no k-ésimo vetor singular linha da matriz GE. Os termos: λ_k , γ_{ik} e α_{jk} foram obtidos pela decomposição por valores singulares (DVS) da matriz de interação GE.

Assim, os componentes aditivos e multiplicativos do modelo AMMI podem ser escritos por:

$$Y_{ij} = \underbrace{\mu + g_i + e_j}_{\text{parte aditiva}} + \underbrace{\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}}_{(ge)_{ij}; \text{ parte multiplicativa}} + \rho_{ij} + \bar{\varepsilon}_{ij} \quad (6)$$

em que a parte multiplicativa pode ser desdobrada da seguinte forma:

$$Y_{ij} = \mu + g_i + e_j + \underbrace{\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}}_{\text{padrão}} + \underbrace{\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}}_{\text{ruído}} + \rho_{ij} + \bar{\varepsilon}_{ij} \quad (7)$$

com $n < p$ e $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk} = \rho_{ij}$.

A seleção do modelo AMMI

Para definição do número de eixos a serem retidos a fim de explicar o padrão relacionado à interação, foram adotados os critérios propostos por Gauch e Zobel (1996), levando em consideração a proporção da soma de quadrados da interação ($SQ_{G \times E}$)

acumulada até o n-ésimo eixo $\sum_{k=1}^n \lambda_k^2 / SQ_{G \times E}$.

Na metodologia AMMI (GAUCH & ZOBEL, 1996), a $SQ_{G \times E}$ é decomposta pela Análise de Componentes Principais (ACP) e compreende uma família de modelos: AMMI₀ (nenhum eixo da ACP da interação G×E é considerado no modelo, ou seja, há ausência de interação G×E); AMMI₁ (apenas o primeiro eixo da ACP da interação G×E é considerado no modelo); AMMI₂ (apenas o primeiro e segundo eixos da ACP da interação G×E são considerados no modelo),..., AMMI_k (apenas os k primeiros eixos da ACP da interação G×E são considerados no modelo). Assim, k, corresponde ao número de eixos ou componentes principais da interação G×E incorporados pelo modelo selecionado. A seleção do número de eixos a ser incorporado pelo modelo pode ser realizada por um teste paramétrico, existindo diferentes métodos para isso, todos concordando que o número de componentes retidos no modelo deve ser o menor possível, para assim explicar a estrutura da interação.

Duarte e Vencovsky (1999) apresentam o método de Gollob como um dos mais usados, o qual distribui graus de liberdade às somas de quadrados $SQ_k = m\lambda_k^2$, contando o número de parâmetros no k-ésimo termo multiplicativo. Para o k-ésimo componente de interação IPCA_k, $GL(IPCA_k) = g+e-1-2k$, logo o teste F é calculado como na análise da variância para modelos lineares.

Tabela 3 - Esquema da análise AMMI, pelo critério de Gollob (1968) para atribuição de graus de liberdade aos componentes principais da interação

Graus de Liberdade (GL)	Soma de Quadrados (SQ)	Quadrado Médio (QM)	F
g-1	SQ(G)	QM _G	-
e-1	SQ(E)	QM _E	-
(g-1) (e-1)	SQ(G×E)	QM _{G×E}	-
g+e+1-(2x1)	λ_1^2	QM _{IPCA1}	QM _{IPCA1} / QM(erro médio)
g+e+1-(2x2)	λ_2^2	QM _{IPCA2}	QM _{IPCA2} / QM(erro médio)
g+e+1-(2x3)	λ_3^2	QM _{IPCA3}	QM _{IPCA3} / QM(erro médio)
...	...	...	...
g+e+1-(2xp)	λ_p^2	QM _{IPCAp}	QM _{IPCAp} / QM(erro médio)
ge (r - 1)	SQ(erro médio)	QM(erro médio)	
ger - 1	SQ (total)	QM (total)	

g: número de genótipos; e: número de ambientes; λ : valor singular (escalar) da matriz de interação GE; posto: mínimo (g-1, e-1); IPCA: eixo da interação da análise de componentes principais.

Já o teste F_R (Cornelius et al., 1992; Piepho, 1995) é descrito pela equação

$$F_{R,n} = (SQ_{G \times E} - \sum_{k=1}^n \lambda_k^2) / (f_2 \cdot QM_{\text{ErroMédio}}) \text{ em que o número de graus de liberdade associados}$$

a cada parcela da interação $G \times E$ é dado por $f_2 = (g-1-n)(a-1-n)$, com n sendo o número de termos multiplicativos incluídos no modelo. Sendo assim, o ruído AMMI, reunindo os demais termos da $SQ_{G \times E}$ (não retidos no modelo selecionado), é testado de maneira a assegurar o seu caráter desprezível, através da significância do teste aplicado.

Assim sendo, após feita a seleção por meio da significância do IPCA a 1% de probabilidade, de acordo com os níveis de significância do teste F , foram ser feitas as demais análises a fim de obter a matriz GE de interação.

Os escores de genótipos (G) e de ambientes (H') foram obtidos pela decomposição aproximada por n componentes:

$GE = USV' = U s^{1/2} s^{1/2} V'$; $G = U s^{1/2}$ e $H' = s^{1/2} V'$, em que U é a matriz com vetores singulares-coluna γ_k ; S é a matriz com os valores singulares λ_k e V' é a matriz com vetores singulares- linha α'_k .

A opção por DVS é por ser possível estimar simultaneamente os parâmetros multiplicativos do modelo a partir da matriz GE , enquanto por ACP são necessárias duas análises, aumentando o trabalho e as chances de equívocos na interpretação dos resultados. Isso ocorre porque a extração separada dos componentes principais para genótipos e ambientes pode resultar em escores com sinais trocados, o que compromete seriamente a interpretação das relações adaptativas dos genótipos aos ambientes.

Os resultados da análise foram apresentados graficamente em biplot (GABRIEL, 1971). Como lembra Duarte e Vencovsky (1999), a interpretação de um biplot quanto à interação $G \times E$ é feita observando-se a magnitude e o sinal dos escores de genótipos e ambientes para o(s) eixo(s) que representa(m) a interação. Assim, os escores baixos (próximos de zero) são próprios de genótipos e ambientes que contribuem pouco para a interação, caracterizando-os como estáveis.

Considerando o biplot $AMMI_1$, a estabilidade é avaliada inspecionando-se as ordenadas (PC_1). Logo, os pontos situados na faixa horizontal em torno de zero em relação ao PC_1 correspondem aos genótipos e ambientes mais estáveis.

No caso do biplot AMMI₂, genótipos e ambientes estáveis são aqueles cujos pontos situam-se próximo à origem, ou seja, com escores praticamente nulos para os dois eixos da interação (PC₁ e PC₂), que são próprios de genótipos e ambientes que contribuíram pouco ou quase nada para a interação (DUARTE; VENCovsky, 1999). A análise AMMI foi realizada utilizando o programa computacional R.

Análise da interação G×E via reamostragem *bootstrap* AMMI

O tipo de reamostragem realizada foi *bootstrap* não paramétrico. Seguindo o procedimento descrito por Lavoranti (2003), visando aprimorar as predições das estabilidades fenotípicas. Neste contexto, a execução do método AMMI seguindo a proposta de Gollob (1968), garante pelo menor rigor dos graus de liberdade, um número maior de fatores multiplicativos e consequentemente, capta todo o padrão devido à interação G×E. A partir desse modelo, foram determinadas as estimativas dos genótipos i nos ambientes j ($\hat{Y}_{ij}$) livres de interferência de ruídos, e, a partir desse novo conjunto de dados, foram obtidas as matrizes de resíduos $\left| GE_{G \times E} = (\hat{ge}_{ij}) \right|$ representadas por:

$$GE_{G \times E} = \begin{vmatrix} \hat{ge}_{11} & \hat{ge}_{12} & \cdots & \hat{ge}_{1e} \\ \hat{ge}_{21} & \hat{ge}_{22} & \cdots & \hat{ge}_{2e} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{ge}_{g1} & \hat{ge}_{g2} & \cdots & \hat{ge}_{ge} \end{vmatrix}$$

com $\hat{ge}_{ij} = \hat{y}_{ij} - \hat{\bar{y}}_{i.} - \hat{\bar{y}}_{.j} + \hat{\bar{y}}_{..}$, sendo que $\hat{\bar{y}}_{i.}$ é a média dos valores estimados do genótipo i ; $\hat{\bar{y}}_{.j}$ é a média dos valores estimados do ambiente j ; $\hat{\bar{y}}_{..}$ é a média geral dos valores estimados.

Das matrizes $GE_{G \times E}$, foram reamostradas 2000 matrizes “*bootstrap*” considerando a unidade amostral multivariada sorteando no sentido ‘linha’ da matriz GE . Assim, obteve-se:

$$GE_{kl}^* = \begin{vmatrix} ge_{11}^* & ge_{12}^* & \cdots & ge_{1e}^* \\ ge_{21}^* & ge_{22}^* & \cdots & ge_{2e}^* \\ \vdots & \vdots & \cdots & \vdots \\ ge_{g1}^* & ge_{g2}^* & \cdots & ge_{ge}^* \end{vmatrix}$$

com : $k = g$ ou $k = e$, para genótipos ou ambientes, respectivamente e $l = (1, 2, \dots, 2000)$.

Dessa forma, foram obtidos para cada genótipo e ambiente, 2000 escores associados a cada matriz “*bootstrap*” por eixo da interação $G \times E(IPCA_1^*, IPCA_2^*)$. A adoção dessa representação gráfica foi ajustada pelo critério de avaliação “pos-dictiva” (Gauch e Zobel, 1996) que visa a obtenção de um alto coeficiente de determinação.

Foi estimado o traço da matriz de covariância da matriz GE uma vez que o traço dessa matriz contém toda a informação a respeito da variação contida na interação. Isto foi feito para cada população em cada ano e então obteve-se um intervalo inter-percentílico representado respectivamente pelos 10º e 90º percentis.

Computou-se também a medida de distância euclidiana aplicada entre escores de genótipos de cada matriz “*bootstrap*” reamostrada, considerando X_{ij} a média no i -ésimo indivíduo (genótipo, cultivar, linhagem etc.) para a j -ésima característica, com base na distância Euclidiana entre o par de indivíduos i e i' por meio da expressão:

$$d_{ii'} = \sqrt{\sum_j (X_{ij} - X_{i'j})^2} \quad (8)$$

Desta forma, foram obtidas 2000 matrizes de distância euclidiana que posteriormente foram utilizadas para o cálculo da matriz média aritmética das matrizes de distância euclidiana ($\bar{d}$) que representa a média dessas matrizes de distância, definida da seguinte maneira:

$$\bar{d} = \frac{\sum_{k=1}^{n_{Rboot}} d_{(ii')k}}{n_{Rboot}} \quad (9)$$

em que $d_{(ii')k}$ é a k -ésima matriz de distância euclidiana obtida, $k = 1, 2, \dots, 2000$; n_{Rboot} é o número de reamostragens *bootstrap* realizadas.

Após a obtenção de $\bar{d}$, foi analisada a divergência genética mediante o emprego do método de agrupamento hierárquico de Ward, utilizando o software estatístico R. Esta análise resultou em um dendrograma e a partir deste, foram feitas as discussões a respeito dos grupos formados.

4 RESULTADOS E DISCUSSÃO

Para o desenvolvimento dessa pesquisa, cujo objetivo principal foi implementar uma metodologia para avaliar a diversidade genética de linhagens de soja por meio da análise AMMI com reamostragem *bootstrap*, utilizou-se os dados referentes à produtividade de grãos, em dois anos agrícolas (1999/2000 e 2000/2001), sob dois sistemas de manejo (com controle intensivo de insetos e com controle ecológico de insetos) aplicados em dois locais distintos (Estação Experimental Anhembi e Fazenda Areão) para três populações: PCI, PRIS e PRIM, respectivamente.

Foram verificadas as pressuposições do modelo para todos os ambientes, nos dois anos, antes que a análise de variância individual fosse realizada. Os p-valores obtidos foram todos superiores a 5%. Assim sendo, o teste de Shapiro-Wilk não significativo evidenciou a normalidade dos erros, o que implica que o conjunto de dados atendeu a pressuposição de normalidade. O Teste de Bartlett não significativo confirmou a homogeneidade das variâncias residuais.

A análise de variância individual dos genótipos dentro de cada ambiente foi realizada com o objetivo de obter o QM_{EM} e, posteriormente, a análise de variância conjunta. Os dados de produtividade de grãos dos ensaios foram analisados separadamente para cada ano (2000 e 2001), sendo que neste estudo a combinação de manejo e local identificava um ambiente. Para estudar o efeito da interação entre genótipos $\times$ ambientes, utilizou-se a análise AMMI.

A maioria dos coeficientes de variação experimental no presente estudo esteve abaixo de 20%, o que indica boa precisão no controle das causas de variação de ordem sistemática dos ambientes experimentais, para a produtividade de grãos, um caráter quantitativo muito influenciado pelo ambiente. Além disso, a existência de variabilidade genética entre linhagens foi constatada tanto nas análises individuais quanto na análise conjunta.

Experimento 1, parte 1: População PCI - ano 2000

Na população 1 (PCI - população com controle de insetos), no ano agrícola 2000, os genótipos utilizados nas análises estão identificados no Anexo A.

Foi realizada a análise de variância conjunta para o experimento envolvendo a população 1 (Tabela 4 e 6) sendo a presença significativa da interação genótipo por ambiente (G×E), diagnosticada pelo teste F (COCHRAN, 1954).

Tabela 4 - Análise de variância conjunta do Experimento 1 no ano 2000, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja avaliados em 4 ambientes com 2 blocos

Fonte de Variação	GL	SQ	QM	F	Pr (>F)
Amb	3	38278785	12759595	242.8261	5.58e-05**
Blocos(Amb)	4	210185	52546	2.1966	0.070626
Gen	43	2392397	55637	2.3259	4.54e-05**
G×E	129	4542867	35216	1.4722	0.006804**
Resíduo	204	4879923	23921		

F : teste F de Gollob (1968).

** : significativo ao nível de probabilidade, $p < 0,01$.

* : significativo ao nível de probabilidade, $p < 0,05$. ns: não significativo.

A princípio, a interação G×E é a condição necessária para se proceder com as análises. Verificou-se pela análise da variância (Tabela 4) diferenças significativas ao nível de 1% para genótipos (G), ambientes (E) e interação entre eles (G×E). No entanto, o efeito de blocos dentro de ambientes foi não significativo.

Em termos do teste F, a magnitude da fonte de variação de ambientes foi muito superior às demais, sendo ela responsável pela maior parte da variação ocorrida. Pode-se, assim, inferir que os efeitos de locais contribuíram de modo mais acentuado para a variação na produtividade. A significância para genótipos indicou que eles são formados por grupos geneticamente distintos, mostrando uma suficiente disponibilidade de variabilidade para a seleção. Além disso, juntamente com a análise de variância foi obtida a média para a variável produtividade de grãos (PG em kg/ha) igual a 991,5312 com um coeficiente de variação baixo (15,6), o que indica boa precisão experimental.

Uma das premissas da análise AMMI é a de que nos primeiros eixos concentra-se maior porcentagem de “padrão”. Assim, à medida que se eleva o número de eixos selecionados, aumenta-se a porcentagem de “ruído”, reduzindo o poder de predição da análise

AMMI. Gauch (1988) e Gauch & Zobel (1996) mostram que a inclusão excessiva de termos multiplicativos pode reduzir seriamente a acurácia da análise.

Um indicativo do comportamento diferenciado dos genótipos frente aos ambientes e condição primária para se estabelecer a análise de estabilidade é observar a significância da interação do tipo genótipos com ambientes. Assim, o resultado da interação G×E indicou a necessidade de identificar e selecionar genótipos que consigam capitalizar os ganhos genéticos em locais específicos ou que sejam menos influenciados pelas variações ambientais, podendo ser explorados em vários outros ambientes.

Em outras palavras, essa significância sugeriu que as variâncias dos genótipos são diferentes de um ambiente para outro, o que permitiu inferir que há genótipos ou grupos de genótipos com adaptação específica para determinados ambientes e outros com adaptação geral a todos os ambientes. Cruz & Regazzi (2004) comentaram que, quando for observada a significância da interação G×E, é necessário que a seleção de genótipos seja efetuada para locais específicos, ou grupos afins. Com base nesses resultados, procedeu-se o estudo mais detalhado da interação significativa.

A matriz de interações G×E é apresentada na Tabela 5 e a próxima etapa da análise corresponde ao ajuste da interação pela decomposição por valores singulares aplicada a esta matriz. O posto da matriz de interações foi posto = $\min(g-1, e-1) = \min(44-1, 4-1) = 3$, conseqüentemente a $SQ_{G \times E}$ pode ser decomposta em até três componentes. Neste caso, na Tabela 6 tem-se que o componente $IPCA_4$ apresentou porcentagem nula de explicação da variação observada na matriz de interação G×E, sendo assim, desnecessário.

Tabela 5 - Matriz de Resíduos da interação G×E, população 1 (PCI - população com controle de insetos), no ano agrícola 2000

(continua)				
Genótipos	Amb1	Amb2	Amb3	Amb4
97-8001	-86,804	46,5142	95,41193	-55,1222
97-8002	185,446	-130,236	-54,3381	-0,87216
97-8004	60,69602	-21,9858	1,911932	-40,6222
97-8007	265,446	119,7642	-351,338	-33,8722
97-8010	-141,054	-73,7358	174,1619	40,62784
97-8012	-129,554	103,7642	122,6619	-96,8722
97-8014	-265,304	140,0142	4,911932	120,3778

Tabela 5 - Matriz de Resíduos da interação G×E, população 1 (PCI - população com controle de insetos), no ano agrícola 2000

Genótipos	(conclusão)			
	Amb1	Amb2	Amb3	Amb4
97-8015	72,19602	-27,4858	76,41193	-121,122
97-8016	-189,804	42,5142	55,41193	91,87784
97-8017	-65,304	45,0142	-195,088	215,3778
97-8018	-92,429	78,8892	-112,213	125,7528
97-8019	-250,304	71,0142	226,9119	-47,6222
97-8020	48,82102	-130,361	40,03693	41,50284
97-8021	17,94602	69,2642	-132,838	45,62784
97-8022	-29,554	-9,2358	89,66193	-50,8722
97-8023	-134,554	10,7642	82,66193	41,12784
97-8024	-117,304	-19,9858	83,91193	53,37784
97-8025	44,19602	-219,486	147,4119	27,87784
97-8026	-156,804	254,5142	-112,588	14,87784
97-8027	129,446	70,7642	-156,338	-43,8722
97-8029	-8,30398	27,0142	-31,0881	12,37784
97-8031	-180,054	77,2642	99,16193	3,627841
97-8033	-79,804	138,5142	-21,5881	-37,1222
97-8034	97,69602	-40,9858	-114,088	57,37784
97-8035	198,446	-107,236	-46,3381	-44,8722
97-8037	151,696	-46,9858	-286,088	181,3778
97-8039	90,69602	130,0142	-158,088	-62,6222
97-8043	-204,054	-22,7358	187,1619	39,62784
97-8044	136,446	-119,236	105,6619	-122,872
97-8050	43,94602	67,2642	-25,8381	-85,3722
97-8054	104,696	2,014205	-66,0881	-40,6222
97-8055	121,196	-28,4858	-84,5881	-8,12216
97-8056	141,196	-177,486	-97,5881	133,8778
97-8058	125,446	-195,236	50,66193	19,12784
97-8060	104,446	-189,236	10,66193	74,12784
IAC-100	-73,679	93,1392	-12,4631	-6,99716
IAS-5	-74,929	87,8892	-14,2131	1,252841
OC-4	-126,679	117,6392	69,53693	-60,4972
PRIMAVERA	-35,804	49,0142	-40,5881	27,37784

Os componentes principais da interação e as porcentagens da variância acumulada encontram-se apresentados na Tabela 6. O efeito multiplicativo da interação $G \times E$ foi diagnosticado por meio da decomposição em valores singulares (DVS), pela decomposição da soma de quadrados da interação ($SQ_{G \times E}$) em quatro eixos ou componentes principais da interação (IPCA).

Tabela 6 - Porcentagem responsável por cada eixo da interação e porcentagem da soma de quadrados acumulada (PA) por eixo singular, Teste F para os componentes. População PCI, ano 2000

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA₁	49.1	49.1	45	2162507	48055.71	2.01	0.0006
IPCA₂	34.8	83.9	43	1534622	35688.88	1.49	0.036
IPCA₃	16.1	100	41	710244.2	17323.03	0.72	0.8945
IPCA₄	0	100	39	0	0	0	1

A análise da interação $G \times E$, por componentes principais, mostrou os dois primeiros eixos (IPCA₁ e IPCA₂) significativos, com $p < 0,05$. Assim, os escores de genótipos foram obtidos seguindo o modelo AMMI₂ (Figura 1).

De acordo com Gauch (1988), o primeiro eixo singular da análise AMMI captura a maior porcentagem de “padrão” e, com acumulação subsequente das dimensões dos eixos, há uma diminuição na porcentagem de “padrão” e um incremento de “ruídos”. Obteve-se que a porção de $SQ_{G \times E}$ explicada pelos dois primeiros eixos foi de 83,9%.

Na Figura 1, o gráfico que visa demonstrar o comportamento dos genótipos mediante a produtividade média de grãos, importante na recomendação de cultivares. Dentre todos os genótipos em estudo, verificou-se que aqueles à direita da linha vertical tracejada são os genótipos que ultrapassaram a média geral de produtividade de grãos obtida anteriormente (991,5312 kg/ha). O genótipo 97-8056 obteve a maior média de produtividade de grãos (1166,500) e, no entanto, a testemunha IAC-100 registrou o pior resultado para esse caráter.

Segundo Duarte e Vencovsky (1999), por meio das análises das magnitudes e dos sinais dos escores biplots correspondentes aos modelos AMMI selecionados, são considerados estáveis aqueles genótipos e/ou ambientes cujos escores foram considerados baixos (próximos de zero), que são próprios de genótipos e ambientes que contribuíram pouco ou quase nada para a interação.

No primeiro eixo da interação, o mais elevado valor de escore para genótipos (11,470) foi observado no genótipo 97-8007 (5) caracterizando-o assim, como o mais instável, pela sua maior contribuição às variações ambientais.

Tabela 7 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 1 (PCI), no ano agrícola 2000

(continua)

Genótipos	PG	IPCA ₁	IPCA ₂	IPCA ₃
97-8001	1109,500	-4,198	-0,635	-2,304
97-8002	934,250	6,553	-3,305	0,384
97-8004	1090,000	1,596	-1,236	-1,755
97-8006	1010,500	-3,220	-5,623	-2,138
97-8007	956,250	11,470	8,449	-4,107
97-8009	1071,250	3,256	-1,084	-3,423
97-8010	1080,750	-5,774	-4,335	3,221
97-8011	917,000	-1,209	2,059	-8,092
97-8012	1028,250	-6,228	0,232	-4,411
97-8014	1002,000	-7,625	5,857	4,845
97-8015	1036,500	0,570	-3,521	-5,063
97-8016	1073,500	-5,856	1,698	4,327
97-8017	1029,000	1,574	7,004	8,708
97-8018	927,125	-0,877	5,715	4,805
97-8019	1020,000	-10,644	-1,688	-1,388
97-8020	1008,875	1,638	-4,145	2,856
97-8021	1053,750	2,178	4,692	1,028
97-8022	1046,250	-2,229	-2,246	-1,832
97-8025	1026,500	0,419	-8,681	3,286
97-8026	1011,500	-4,020	9,818	-1,308
97-8027	915,250	5,207	3,941	-3,260
97-8029	919,000	0,115	1,459	0,260
97-8031	979,750	-6,742	1,025	0,274
97-8033	987,500	-2,776	4,188	-2,599
97-8034	832,000	4,740	1,109	2,279
97-8035	938,250	6,499	-3,262	-1,742
97-8039	1025,000	3,776	5,586	-4,465
97-8043	980,750	-7,965	-2,937	2,964
97-8044	1002,250	2,406	-6,849	-4,485
97-8050	926,750	0,872	1,399	-4,470
97-8055	968,500	4,642	0,236	-0,673
97-8056	1166,500	6,736	-2,439	6,667
97-8057	1077,250	2,784	-6,146	-3,747
97-8058	885,250	3,851	-6,621	2,230

Tabela 7 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 1 (PCI), no ano agrícola 2000

Genótipos	PG	IPCA ₁	IPCA ₂	(conclusão)
				IPCA ₃
97-8060	937,250	4,025	-5,148	4,537
IAC-100	774,875	-2,374	3,015	-0,903
IAS-5	1095,625	-2,322	2,981	-0,500
PRIMAVERA	998,500	-0,564	2,479	0,785
Amb1	1408,602	25,441	-6,471	-6,893
Amb2	1161,284	-8,678	21,016	-10,165
Amb3	824,386	-17,783	-19,244	-3,710
Amb4	578,921	1,020	4,700	20,768

IPCA: eixo da interação da análise de componentes principais.

Os primeiros eixos associados ao modelo AMMI, detém praticamente todo o padrão devido à interação, e os eixos de mais alto grau captam, praticamente, somente os ruídos (Gauch, 1988). Normalmente os ambientes são classificados como favoráveis quando apresentam médias superiores à média geral da experimentação (alta produtividade), ou desfavoráveis, quando suas médias são inferiores à média geral (baixa produtividade). Neste caso, a média geral é 991,5312 kg/ha e os ambientes favoráveis são então o ambiente 1 e ambiente 2.

Assim, por meio da Tabela 7 observou-se que o genótipo 97-8056 (37) foi o que apresentou o maior escore de produtividade de grãos (1166,5) e está alta e positivamente relacionado ao ambiente 1 (141,196) conforme se observa no biplot (Figura 2).

Sabe-se que a diversidade das condições ambientais a que a cultura é submetida contribui para que ocorra a interação genótipo $\times$ ambiente, ou seja, a alteração no desempenho relativo dos genótipos em virtude das diferenças entre ambientes. Portanto, os materiais genéticos precoces, além de alta produtividade de grãos, devem ter um comportamento previsível e responder aos estímulos propostos pelo ambiente (FARIA et al., 2009).

Na Figura 2, tem-se a representação gráfica do comportamento dos genótipos no sistema de eixos (biplot), representando o primeiro e o segundo componentes principais obtidos nas análises para a população PCI no ano 2000. Tal apresentação baseia-se na dispersão espacial dos genótipos no plano bidimensional e na sua capacidade de produtividade em relação a cada um dos quatro ambientes, representados pelos vetores.

Obeve-se que os genótipos localizados à esquerda de PC₁ (coeficientes negativos) apresentam os menores valores de produtividade de grãos no ambiente 1 (Local: Anhembi, Manejo: controle intensivo de insetos - CII). No entanto, nesta mesma região do gráfico biplot, englobam-se os genótipos com a melhor característica de produtividade de grãos nos ambientes 2 (Local: Anhembi, Manejo: controle ecológico de insetos - CEI) e 3 (Local: Areão, Manejo: controle intensivo de insetos - CII).

Nota-se que, genótipos com escores baixos, próximos de zero, contribuem pouco ou quase nada para a interação, e dessa forma, são considerados estáveis (Duarte e Vencovsky, 1999). Pode-se perceber por meio da Figura 2, que os genótipos que atendem a essa situação são 97-8011 (8), 97-8029 (24), 97-8050 (33) e a testemunha IAS-5 (44). Esses, pelos critérios adotados por Lavoranti (2003), podem ser considerados com alta estabilidade e desta forma, se adaptam a qualquer um dos quatro ambientes.

Figura 2 - Biplot AMMI₂ para dados de produtividade de grãos em kg/ha, população PCI, ano 2000

Escore baixo indicam genótipos e/ou ambientes que contribuem pouco ou quase nada para a interação G×E, sendo, portanto, estáveis. Tais genótipos podem ser recomendados amplamente desde que tenham médias elevadas, fato este ocorreu uma vez que verificou-se o genótipo 97-8011 (8) com média 917, o genótipo 97-8029 (24) com média 919, o genótipo 97-8050 (33) com média 926,75 e a testemunha IAS-5 (44) com média 1095,625.

Todas as médias destes genótipos ficaram acima da média geral de PG obtida inicialmente. Já o ambiente 1 apresentou a maior média para o caráter produtividade de grãos (1408,602 kg/ha).

Além disso, os genótipos 97-8012 (9), 97-8014 (10), 97-8026 (22) apresentaram interação positiva no ambiente 2, bem como a testemunha IAS-5 (43). Já os genótipos com interação positiva no ambiente 3 são: 97-8010 (7), 97-8012 (9), 97-8031 (25), 97-8019 (15) e 97-8043 (31).

Observa-se que no biplot os aspectos mais relevantes da matriz de interação dupla, ficam bem apresentados. Por exemplo, a interação positiva dos genótipos 97-8017 (13) e 97-8021 (17) com o ambiente 4 e a interação negativa dos genótipo 97-8006 (4) e 97-8015 (11) com este mesmo ambiente.

A divergência entre os genótipos foi avaliada pelo método hierárquico aglomerativo de Ward, com o emprego da distância Euclidiana. A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se um dendrograma (Figura 3) representativo dos agrupamentos formados entre os genótipos em estudo.

Observando o dendrograma percebe-se a formação de seis grupos distintos, dos quais três apresentam testemunhas dentre seus materiais, demonstrando que cada testemunha possui características diversificadas entre si a ponto de serem alocadas em grupos distintos quanto à similaridade. No entanto, os grupos que continham as testemunhas se destacaram dos demais quanto à produtividade de grãos e estabilidade observados nos genótipos neles alocados.

Foi obtido que o grupo da testemunha IAS-5 (42) apresentou média de produção superior a média geral, sendo o genótipo 97-8017 (13) o de maior estabilidade, podendo o mesmo ser recomendado para qualquer um dos ambientes.

Já o grupo formado pelas testemunhas Primavera (44) e OC-4 (43) apresentou média produtiva similar à média geral, no entanto foi o grupo que apresentou o maior número de genótipos estáveis, 97-8004 (3) e 97-8015 (11). Por outro lado, quando se analisa a redução relativa da produção quando submetido ao ataque de insetos percebe-se que estes materiais apresentaram um dos menores valores médios de redução, justificando assim sua proximidade com a testemunha tolerante (IAC-100).

Figura 3 - Dendrograma das distâncias euclidianas entre os escores “bootstrap” de marcadores de genótipos AMMI₂, para dados de produtividade de grãos, em kg/ha

A partir do dendrograma formado, obteve-se a formação de seis grupos (Tabela 8). O primeiro grupo alocou a testemunha IAC-100 (41) e pode ser classificado como o grupo com produção média nos quatro ambientes, abaixo da produção média geral. No entanto, tal grupo apresenta genótipos contrastantes quanto a adaptabilidade devido a presença do genótipo mais instável, 97-8007 (5), e um dos mais estáveis, 97-8025 (21).

Tabela 8 - Agrupamentos de quarenta e quatro genótipos de soja da população 1 referente ao experimento realizado no ano 2000, obtidos pelo método hierárquico de Ward

Grupos	Genótipos
1	25 19 41 5 21 29 33
2	9 32 7 23
3	18 24 27 43 44 3 36 39 4 38 11 35
4	6 34 12 30 40
5	16 22 10 15
6	42 14 1 28 13 20 2 26 17 31 8 37

Pela observação de alguns histogramas feitos, percebeu-se a assimetria da distribuição por amostragem das estimativas em foco, não semelhante com a curva normal, o que indica assumir o intervalo de confiança percentílico para a estimação do traço matriz de covariância da matriz de resíduos da interação $G \times E$.

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico

baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz G×E, mostrou que tal estimativa (igual a 2.203.686) está contida no intervalo definido pelo percentil 10 (igual a 1.881.002), o limite inferior, e o percentil 90 (igual a 2.738.849), o limite superior.

Experimento 1, parte 2: População PCI - ano 2001

Na população 1 (PCI - população com controle de insetos), no ano agrícola 2001, os genótipos utilizados nas análises foram identificados no Anexo B.

O teste F na análise de variância conjunta para o experimento envolvendo a população 1 indicou que a interação dos genótipos com os ambientes mostrou-se altamente significativa (Tabela 9).

Tabela 9 - Análise conjunta da variância do Experimento 1 no ano 2001, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja

Fonte de Variação	GL	SQ	QM	F	Pr (>F)	
Amb	3	9956056	3318685	71.2493	0.000629	**
Blocos(Amb)	4	186314	46578	1.3691	0.249213	
Gen	22	1735346	78879	2.3185	0.002221	**
G×E	66	4342211	65791	1.9338	0.001052	**
Resíduo	112	3810481	34022			

F : teste F de Gollob (1968).

**: significativo ao nível de probabilidade, $p < 0,01$.

*: significativo ao nível de probabilidade, $p < 0,05$. ns: não significativo.

Pela Tabela 9 obteve-se que os genótipos sofreram influências diferenciadas dos ambientes, o que dificulta a recomendação de cultivares para a região abrangida pelo estudo. As interações significativas indicam resultados das alterações no posicionamento do genótipo ou mudanças na magnitude das diferenças entre genótipos de um ambiente para outro. Assim, é necessário estudo mais detalhado dessa interação de modo que ela possa ser controlada, e não interfira negativamente na recomendação.

Verificou-se, em termos do teste F que a magnitude da fonte de variação de ambientes foi muito superior às demais, sendo ela responsável pela maior parte da variação ocorrida uma vez que o maior valor de quadrado médio foi para ambiente, em relação aos demais quadrados médios, indicando haver grande diferença no potencial produtivo entre os locais. Pode-se, assim, inferir que os efeitos de locais contribuíram de modo mais acentuado para a variação na produtividade. Em relação aos genótipos (G), ambientes (E) e interação dos genótipos com os ambientes (G×E) foram detectadas diferenças significativas, ao nível de 1% de probabilidade.

Juntamente com a análise de variância foi obtida a média para a variável produtividade de grãos (PG) igual a 1001,219 kg/ha com coeficiente de variação baixo (18,42%), indicando boa precisão experimental. O efeito de ambientes foi responsável pela maior parte da variação dos tratamentos, seguida pelo efeito da interação G×E e por último o efeito de genótipos, enquanto que o efeito de blocos dentro de ambientes foi não significativo.

A significância para genótipos indicou que eles são formados por grupos geneticamente distintos, mostrando uma suficiente disponibilidade de variabilidade para a seleção. Com relação à significância de ambientes, em princípio era de se esperar a sua existência, devido a grande diversidade de condições ambientais onde esses experimentos foram instalados.

As interações G×E significativas indicaram que os genótipos sofreram influências diferenciadas dos ambientes, ou seja, responderam de forma diferente aos estímulos ambientais, dificultando assim a indicação generalizada dos genótipos superiores. A presença de variabilidade entre genótipos indica que esses poderão ser selecionados pelos seus desempenhos médios nos vários locais de avaliação.

A matriz de interações GE é apresentada na Tabela 10 e a próxima etapa da análise corresponde ao ajuste da interação pela decomposição em valores singulares aplicada à matriz GE. A matriz GE terá posto = $\min(g-1, e-1) = \min(23-1, 4-1) = 3$, consequentemente a $SQ_{G \times E}$ pode ser decomposta em até três componentes.

Tabela 10 - Matriz de Resíduos da interação G×E, população 1 (PCI - população com controle de insetos), no ano agrícola 2001

Genótipos	Amb1	Amb2	Amb3	Amb4
97-8001	88,93804	-248,805	66,58587	93,28152
97-8004	-125,062	12,19457	128,5859	-15,7185
97-8007	73,43804	68,69457	-149,914	7,781522
97-8009	-18,562	-50,3054	-70,9141	139,7815
97-8010	56,18804	-69,5554	49,83587	-36,4685
97-8012	-33,562	-25,3054	-33,9141	92,78152
97-8014	232,688	-96,0554	-122,664	-13,9685
97-8016	10,43804	23,69457	-43,9141	9,781522
97-8019	-204,812	208,4446	119,8359	-123,468
97-8022	-137,562	-168,305	93,08587	212,7815
97-8024	-24,312	13,94457	12,33587	-1,96848
97-8026	-7,56196	151,6946	6,08587	-150,218
97-8033	184,688	43,94457	-199,664	-28,9685
97-8039	-86,062	30,19457	209,5859	-153,718
97-8043	26,93804	11,19457	59,58587	-97,7185
97-8051	85,18804	172,4446	-41,1641	-216,468
97-8054	-143,062	104,1946	-68,4141	107,2815
97-8055	-54,812	188,4446	-85,1641	-48,4685
97-8056	-131,062	270,1946	-89,4141	-49,7185
97-8057	85,43804	-43,3054	-55,9141	13,78152
IAC-100	40,48804	171,3446	-110,264	-101,568
IAS-5	-166,562	-271,805	190,5859	247,7815
OC-4	248,563	-497,18	135,2109	113,4065

Em concordância com o modelo AMMI, a interação G×E original pôde ser decomposta em três componentes (posto da matriz GE), dos quais apenas o primeiro termo (IPCA₁) foi julgado como altamente significativo. Os resíduos do AMMI₂ mostraram-se não-significativos, a 1% de probabilidade, confirmando a seleção de AMMI₁. Este modelo explicou aproximadamente 58,4 % da $SQ_{G \times E}$ (Tabela 11). Os demais eixos representam variações da $SQ_{G \times E}$ mais predominantemente ricas em ruído.

Tabela 11 - Porcentagem da soma de quadrados acumulada (PA) por eixo singular, Teste F para os IPCA referente aos dados de produtividade de grãos em kg/ha de 23 genótipos de soja. População PCI, ano 2001

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA ₁	58.4	58.4	24	1866129	77755.35	2.29	0.002
IPCA ₂	27.8	86.2	22	887880.6	40358.21	1.19	0.2712
IPCA ₃	13.8	100	20	439734.5	21986.72	0.65	0.8658
IPCA ₄	0	100	18	0	0	0	1

A representação gráfica do comportamento dos genótipos no sistema de eixos (biplot), contendo o primeiro e o segundo componentes principais encontra-se na Figura 4. Tal apresentação baseia-se na dispersão espacial dos genótipos no plano bidimensional e na sua capacidade de produtividade em relação a cada um dos quatro ambientes (setas).

Figura 4 - Biplot AMMI₂ para dados de produtividade de grãos em kg/ha, população PCI, ano 2001

Conforme os critérios adotados, e de acordo com a Figura 4, pode-se considerar como de alta estabilidade fenotípica para a população PCI no ano 2001, os genótipos 97-8016 (8), 97-8024 (11) e 97-8012 (6). Os mais instáveis foram as testemunhas IAS-5 (22) e OC-4 (23), bem como os genótipos 97-8019 (9) e 97-8056 (19).

De acordo com Duarte e Vencovsky (1999), os genótipos instáveis, apesar de sua adaptabilidade restrita, podem ser aproveitados vantajosamente na seleção e recomendação para um ambiente específico. Sendo assim, através da Figura 4 observa-se que as testemunhas

S-5 e OC-4 podem ser recomendadas para os ambientes 3 e 4 (Areão-CII e Areão CEI, respectivamente) pois apresentaram interação positiva com os mesmos. As mesmas temunhas apresentaram interação negativa em relação ao ambiente 2 (Anhembi-CEI), não sendo portanto adaptadas a este ambiente. No entanto, os genótipos instáveis 97-8019 (9) e 97-8056 (19) apresentaram interação altamente positiva com o ambiente 2, sendo portanto, bem adaptados a este ambiente.

Sabendo-se que genótipos com escores baixos, próximos de zero, contribuem pouco ou quase nada para a interação, ou seja, são estáveis, percebe-se que nesta situação temos no plot da Figura 4 os genótipos 97-8010 (5), 97-8016 (8) e 97-8024 (11). Esses, pelos critérios adotados por Lavoranti (2003), podem ser considerados com alta estabilidade. Escores baixos indicam genótipos e/ou ambientes que contribuem pouco ou quase nada para a interação $G \times E$, sendo, portanto, estáveis.

Figura 5 - Gráfico do comportamento de 23 genótipos de soja da população PCI (ano 2001) em relação ao primeiro componente da interação e ao caráter PG

Na Figura 5 apresenta-se um importante gráfico, pois, para fins de recomendação de cultivares, genótipos estáveis devem também apresentar desempenho desejável, o que é aliado por suas médias (abscissa), sendo o genótipo 97-8010 (5) aquele que apresenta essa propriedade. Assim, este genótipo apresentou efeito da interação pequeno e produção alta, apresentando a melhor combinação de estabilidade e produtividade.

Na mesma figura observou-se à direita da linha vertical tracejada os genótipos que apresentaram média de produtividade de grãos superior à média geral uma vez que esta foi igual a 1001,219 kg/ha. Desta forma, dentre todos os genótipos desta população, verificou-se

que aqueles mais promissores para o caráter PG analisado foram 97-8007 (3), 97-8057 (20), 97-8051 (16), e 97-8026 (12).

Da Tabela 12 percebe-se que o genótipo 97-8007 (3) foi o que apresentou o maior escore de produtividade de grãos (1171,0). Este genótipo está alta e positivamente relacionado aos ambientes 1 e 2 conforme se observa no biplot (Figura 4).

Tabela 12 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 23 genótipos de soja da população 1 (PCI), no ano agrícola 2001

Genótipos	PG	IPCA₁	IPCA₂	IPCA₃
97-8001	1067,500	-9,138	-1,524	-1,252
97-8004	976,500	0,249	6,520	-2,990
97-8007	1171,000	2,544	-5,119	4,366
97-8009	920,000	-2,625	0,082	6,657
97-8010	997,250	-2,125	-1,277	-3,683
97-8012	1058,000	-1,503	1,024	4,294
97-8014	995,750	-2,806	-10,321	0,226
97-8016	965,000	0,804	-1,077	1,664
97-8019	894,250	7,633	8,652	-4,343
97-8022	1062,000	-7,477	7,417	4,740
97-8024	917,750	0,451	1,054	-0,100
97-8026	1155,000	6,212	-0,335	-4,212
97-8033	925,750	2,124	-10,041	3,333
97-8039	906,500	1,766	5,905	-10,288
97-8043	1016,500	1,018	-0,373	-5,236
97-8051	1155,250	7,530	-4,706	-5,844
97-8054	887,500	2,751	4,327	7,689
97-8055	1014,250	6,821	0,220	2,590
97-8056	1096,500	9,551	2,808	3,940
97-8057	1160,000	-1,424	-3,889	0,989
IAC-100	1033,350	6,731	-3,795	0,489
IAS-5	925,500	-11,401	10,278	2,696
OC-4	885,875	-17,685	-5,830	-5,724
Amb1	1198,691	-5,560	-21,136	-4,721
Amb2	1251,435	26,473	2,171	2,856
Amb3	696,044	-7,670	13,667	-13,844
Amb4	886,348	-13,244	5,298	15,709

Em relação ao caráter produtividade de grãos, os genótipos 97-8026 e 97-8051 possuíram aproximadamente a mesma média, o que pode ser confirmado pelo gráfico da Figura 5. Os genótipos que apresentaram menor escore de produtividade de grãos foram 97-8054 (17) e 97-8019 (9), com aproximadamente 900 kg/ha.

Observa-se que o ambiente 2 foi aquele que mais contribuiu para a interação $G \times E$ uma vez que seu escore (26,473) foi o maior associado $IPCA_1$, primeiro eixo da interação e que mais retém porcentagem do desempenho real “padrão” associada à interação. Desta forma, caracteriza-se assim, como o ambiente mais instável.

Após a obtenção de resultados a respeito da matriz de interação $G \times E$, a divergência entre os acessos foi avaliada pelo método hierárquico aglomerativo de Ward, com o emprego da distância Euclidiana. A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se um dendrograma representativo dos agrupamentos formados entre os genótipos em estudo (Figura 6).

Figura 6 - Dendrograma das distâncias euclidianas entre os escores “bootstrap” de marcadores de genótipos $AMMI_1$, para dados de produtividade de grãos, em kg/ha

O dendrograma uniu os materiais semelhantes como pode ser identificado ao observar simultaneamente os gráficos anteriormente obtidos (Figuras 4 e 5). No dendrograma obteve-

se que o genótipo 97-8051 (16) pertenceu ao grupo do genótipo 97-8026 (12), genótipos estes que apresentaram o mesmo valor médio de produtividade de grãos na Figura 5 e, além disso, apresentaram alto escore para o primeiro eixo da interação bem como interação positiva com o ambiente 2.

Observa-se que os genótipos 97-8010 (5) e 97-8014 (7) pertenceram ao mesmo grupo, o que faz sentido quando são verificados os gráficos das Figuras 4 e 5. Nestes, percebe-se que ambos os genótipos possuem interação negativa com o ambiente 2, escores negativos com relação aos dois primeiros eixos da interação (não contribuem para a interação) e possuem aproximadamente a mesma média de produtividade de grãos.

O genótipo 97-8009 (4) permaneceu isolado dos demais e isso pode ser justificado pelo fato do mesmo ter se destacado como genótipo estável no gráfico da Figura 4, apresentar boa média para o caráter produtividade de grãos (920 kg/ha, conforme Tabela 12), e não contribuir para a interação $G \times E$ uma vez que seu escore para o primeiro eixo da interação foi negativo e para o segundo eixo foi praticamente nulo, evidenciando não estar relacionado com os primeiros eixos que captam a maior parte da variação devido à interação.

A partir do dendrograma obtido obteve-se a formação de 6 grupos os quais podem ser descritos segundo a Tabela 13.

Tabela 13 - Agrupamentos de vinte e três genótipos de soja da população 1 referente ao experimento realizado no ano 2001, obtidos pelo método hierárquico de Ward

Grupos	Genótipos
1	1 17
2	20 5 7
3	19 3 23 21 2 14
4	11 12 18 22 6 8 9 10 16
5	4
6	13 15

Pela observação de alguns histogramas feitos, percebeu-se a assimetria da distribuição por amostragem das estimativas em foco, não semelhante com a curva normal, o que indica assumir o intervalo de confiança percentílico para a estimação do traço matriz de covariância da matriz de resíduos da interação $G \times E$.

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz $G \times E$, mostrou que tal estimativa (igual a 1.596.872) está contida no intervalo definido pelo percentil 10 (igual a 1.065.747), o limite inferior, e o percentil 90 (igual a 2.044.560), o limite superior.

Experimento 2, parte 1: População PRIS - ano 2000

Na população 2 (PRIS - populações selecionadas para resistência a insetos sugadores), no ano agrícola 2000, os genótipos utilizados nas análises foram identificados no Anexo C.

Na Tabela 14 é apresentado o resumo da análise de variância conjunta onde foram apresentados o teste de F e a probabilidade associada. Todos os efeitos foram significativos, o que é indicativo da presença de variabilidade para seleção entre os genótipos, variabilidade entre os ambientes utilizados e a ocorrência de resposta diferencial dos genótipos aos ambientes, o que determina a necessidade de análises de estabilidade.

Observou-se diferença significativa de 5% de probabilidade entre genótipos, entre ambientes e para a interação genótipos $\times$ ambientes. Juntamente com a análise de variância foi obtida a média para a variável produtividade de grãos (PG) igual a 978,5807 e coeficiente de variação igual a 14,18%, indicando boa precisão experimental.

Tabela 14 - Análise conjunta da variância do Experimento 2 no ano 2000, incluindo o desdobramento da interação $G \times E$ para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja

Fonte de Variação	GL	SQ	QM	F	Pr (>F)	
Amb	3	25398550	8466183	61.6330	0.0008348	**
REP(Amb)	4	549458	137364	7.1373	2.132e-05	**
Gen	43	5977577	139013	7.2229	< 2.2e-16	**
G×E	129	3412075	26450	1.3743	0.0213067	*
Resíduo	204	3926209	19246			
Média	978.5807					
C.V.	14.17669					

F : teste F de Gollob (1968).

** : significativo ao nível de probabilidade, $p < 0,01$. * : significativo ao nível de probabilidade, $p < 0,05$.

ns: não significativo.

A matriz de interações G×E é apresentada na Tabela 15 e a próxima etapa da análise correspondeu ao ajuste da interação pela decomposição em valores singulares aplicada a esta matriz.

Tabela 15 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2000

(continua)				
Genótipos	Amb 1	Amb 2	Amb 3	Amb 4
97-8063	50.22443	-75.65057	-54.7642	80.19034
97-8064	-128.0256	147.0994	41.98580	-61.05966
97-8065	-21.27557	79.84943	-32.26420	-26.30966
97-8067	142.7244	-50.15057	-89.2642	-3.309659
97-8068	37.97443	66.09943	83.9858	-188.0597
97-8070	88.47443	218.5994	-209.5142	-97.55966
97-8071	15.72443	-9.150568	63.7358	-70.30966
97-8072	-12.27557	-18.15057	66.7358	-36.30966
97-8073	-78.77557	-15.65057	270.2358	-175.8097
97-8074	-80.02557	-47.90057	-27.01420	154.9403
97-8076	-64.02557	46.09943	50.9858	-33.05966
97-8078	259.4744	-177.4006	-188.5142	106.4403
97-8080	-29.77557	-76.65057	23.23580	83.19034
97-8081	18.97443	-98.90057	-108.0142	187.9403
97-8083	-15.52557	-29.40057	97.4858	-52.55966
97-8084	7.224432	-65.65057	37.23580	21.19034
97-8085	-20.52557	-26.40057	207.4858	-160.5597
97-8086	-8.275568	-125.1506	22.73580	110.6903
97-8087	-138.7756	182.3494	63.2358	-106.8097
97-8088	-89.27557	4.849432	-13.26420	97.69034
97-8090	23.22443	13.34943	-124.7642	88.19034
97-8091	74.22443	77.34943	-180.7642	29.19034
97-8093	-49.27557	254.8494	-28.26420	-177.3097
97-8095	142.2244	-173.6506	-43.76420	75.19034
97-8096	-91.52557	-20.40057	52.4858	59.44034
97-8097	-133.2756	104.8494	47.7358	-19.30966
97-8100	65.22443	-13.65057	-108.7642	57.19034
97-8101	30.22443	-8.650568	34.23580	-55.80966
97-8105	-64.02557	169.0994	-202.0142	96.94034
97-8108	149.4744	-309.4006	53.4858	106.4403

Tabela 1516 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2000

(conclusão)				
Genótipos	Amb 1	Amb 2	Amb 3	Amb 4
97-8109	25.59943	-42.27557	-5.389205	22.06534
97-8110	-109.5256	-53.40057	168.4858	-5.559659
97-8111	29.22443	-14.65057	-20.76420	6.190341
97-8112	11.97443	-69.90057	-8.014205	65.94034
97-8114	-64.77557	54.34943	-17.76420	28.19034
97-8116	-20.52557	-109.4006	50.4858	79.44034
97-8117	-45.52557	40.59943	112.4858	-107.5597
97-8118	-157.0256	52.09943	166.9858	-62.05966
97-8119	43.22443	-21.65057	28.23580	-49.80966
97-8120	102.7244	-53.15057	-95.2642	45.69034
IAC-100	-33.02557	-50.40057	69.4858	13.94034
IAS-5	81.59943	63.22443	-66.8892	-77.93466
OC-4	42.22443	95.84943	-72.7642	-65.30966
Primavera	13.09943	86.22443	-115.3892	16.06534

O posto da matriz foi posto = $\min(43,3) = 3$ e consequentemente, a $SQ_{G \times E}$ pode ser decomposta em até três componentes. Pela Tabela 16 observa-se que o modelo AMMI com dois componentes explicou 84,5% da soma de quadrados da interação entre genótipos e ambientes como resposta-padrão, porcentagem significativa em se tratando do caráter produtividade de grãos em estudo, e 15,5% dessa soma de quadrados é de ruídos presentes nos dados.

Tabela 17 - Porcentagem da soma de quadrados acumulada por eixo singular, Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 44 genótipos de soja. População PRIS, ano 2000

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA₁	50.3	50.3	45	1647127.3	36602.83	1.9	0.0014
IPCA₂	34.2	84.5	43	1119230.7	26028.62	1.35	0.0879
IPCA₃	15.5	100	41	506585.3	12355.74	0.64	0.9550
IPCA₄	0	100	39	0	0	0	1

É importante notar que o $IPCA_4$ (Tabela 16) teve porcentagem nula na explicação da variabilidade observada na matriz de interação, o que é justificado pelo posto, que indicava a necessidade de somente 3 componentes. Assim, pelo critério de Gollob, o modelo $AMMI_1$ foi selecionado como aquele que melhor descreve o padrão inerente à interação $G \times E$ uma vez que o teste só foi significativo para $IPCA_1$ (0,0014).

A última etapa da análise AMMI consiste na representação gráfica dos genótipos e ambientes em um gráfico denominado de biplot, onde escores de genótipos e de ambientes foram representados seguindo o modelo $AMMI_2$ (Figura 7).

Segundo Silva et al. (2002), escores altos não importando se positivos ou negativos indicam grandes interações, ou seja, baixa estabilidade. Por outro lado, escores próximos a zero, refletem baixas interações, acarretando consequentemente, alta estabilidade.

De acordo com a Figura 7 pode-se considerar como de alta estabilidade fenotípica para a população PRIS, os genótipos 97-8088 (20), 97-8096 (25), 97-8101 (28), 97-8111 (33), 97-8119 (39) e a testemunha IAC-100 (41). Esses genótipos também foram mais estáveis que as testemunhas comuns; entre essas, a IAS-5 (42) foi a mais estável.

Os genótipos mais afastados da origem são os que mais contribuíram para a interação, sendo estes 97-8078 (12) e o 97-8108 (30), ambos apresentando interação positiva com os ambientes 3 e 4, que correspondem a Areão-CII e Areão-CEI, respectivamente.

Por outro lado, genótipos posicionados no gráfico biplot do lado oposto de um determinado local, indicam interação negativa, chamando a atenção a baixa adaptação do genótipo 97-8108 (30) ao ambiente 2 (Anhembi-CEI) e do genótipo 97-8087 (19) ao ambiente 1 (Anhembi - CII).

Figura 7 - Biplot $AMMI_2$ para dados de produtividade de grãos em kg/ha, população PRIS, ano 2000

Deve ser ressaltado que a análise gráfica baseia-se nos desvios da interação e não no desempenho “per se” de genótipos ou de ambientes. Ainda de acordo com a Figura 7, como os genótipos mais instáveis, constata-se os genótipos 97-8078 (12), 97-8108 (30), 97-8073 (9), 97-8087 (19), 97-8093 (23) e 97-8070 (6).

Na Figura 8 apresenta-se um importante gráfico para fins de recomendação de cultivares, o que é avaliado por suas médias (abscissa). Pelo gráfico, verifica-se que o genótipo com maior média para produtividade de grãos (PG) foi o 97-8080 (13).

Dentre todos os genótipos desta população, observou-se que os mais promissores para o caráter PG analisado foram 97-8085 (17), 97-8088 (20), 97-8096 (25), 97-8064 (2), 97-8112 (34) e as testemunhas IAS-5 (42) e Primavera (44).

Os genótipos estáveis 97-8111 (33) e 97-8119 (39) tiveram suas produtividades semelhantes entre si e próximos da média geral (978,5807 kg/ha). Por outro lado, o material genético 97-8064 (2) apresentou alta produtividade, porém baixa estabilidade, não se enquadrando dentro das características desejáveis de um bom material genético pelo método AMMI.

Os genótipos estáveis, mas com média de produtividade de grãos acima da média geral foram 97-8088 (20), 97-8096 (25) e a testemunha IAS-5 (42), apresentando assim boa média para o caráter produtividade de grãos, não contribuindo para a interação, podendo os mesmos serem recomendados para qualquer um dos ambientes.

Figura 8 - Gráfico do comportamento de 44 genótipos de soja da população PRIS (ano 2000) em relação ao primeiro componente da interação e ao caráter PG

Na Tabela 17, com as análises das magnitudes e dos sinais dos escores biplots correspondentes ao modelo AMMI selecionado, foram confirmados como estáveis (escores

considerados baixos, próximos de zero) os genótipos 97-8088 (20), 97-8096 (25), 97-8101 (28), 97-8119 (39) e a testemunha IAC-100 (41), detectados desta forma na Figura 7.

Tabela 18 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 2 (PRIS), no ano agrícola 2000

(continua)

Genótipos	PG	IPCA ₁	IPCA ₂	IPCA ₃
97-8063	1025,5000	-4,3712247	-0,449720977	0,603964640
97-8064	1127,7500	6,2675616	2,249970646	2,84862945
97-8065	874,0000	1,6648315	2,829136731	0,341992928
97-8067	894,0000	-4,4262851	1,617284459	-4,702199621
97-8068	1086,7500	5,216669	0,001686208	-6,838316355
97-8070	768,2500	0,8253447	11,52159959	-4,25326679
97-8071	969,0000	1,8529744	-1,692181906	-2,846776939
97-8072	1008,0000	1,5751517	-2,173562631	-0,962656898
97-8073	881,5000	8,3374573	-7,394623691	-3,565576123
97-8074	902,7500	-2,7640072	-1,123436156	7,126139181
97-8076	965,7500	3,1788085	-0,345598106	1,142907944
97-8078	1036,2500	-11,9451771	1,221340057	-5,470235948
97-8080	1186,5000	-1,9712537	-2,823189473	3,066820288
97-8081	897,7500	-7,0418452	0,015903239	4,909129472
97-8083	976,2500	2,2053359	-3,239695195	-1,482605837
97-8084	984,5000	-1,0328661	-2,643989717	0,008663301
97-8085	1140,2500	5,9884136	-5,835911869	-4,867929037
97-8086	1012,0000	-3,6370928	-4,007482197	2,984062961
97-8087	959,5000	8,1898742	2,636307111	1,916154406
97-8088	1122,0000	-0,4772576	-0,070678498	5,897187306
97-8090	1023,5000	-3,6646901	3,504031205	2,311242849
97-8091	1027,5000	-3,1871222	6,895477096	-0,679870987
97-8093	981,0000	7,8614909	7,339513715	-2,53326034
97-8095	1043,5000	-7,2244866	-2,847866468	-2,996347181
97-8096	1119,2500	0,8390137	-2,355383654	4,522142974
97-8097	869,0000	4,9695626	0,929066160	4,073075098
97-8100	785,5000	-3,9780966	2,628442929	-0,149438012
97-8101	677,5000	0,9138817	-0,883380133	-2,795642913
97-8105	906,7500	-1,0011426	9,099496325	6,311821673

Tabela 17 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 2 (PRIS), no ano agrícola 2000

Genótipos	PG	IPCA ₁	IPCA ₂	(conclusão)
				IPCA ₃
97-8108	845,2500	-8,6982667	-8,847728761	-3,164214688
97-8109	870,1250	-1,6026536	-0,872903135	-0,338161410
97-8110	1055,2500	3,5337121	-6,15348793	2,677333458
97-8111	963,5000	-1,1431047	0,271582981	-0,768427295
97-8112	1115,7500	-2,6885611	-1,642458884	1,310246066
97-8114	1068,5000	1,1600383	1,53037559	3,241073035
97-8116	898,2500	-2,1827673	-4,318007384	2,432358013
97-8117	1078,2500	5,0908676	-1,868191413	-1,897683093
97-8118	941,7500	7,0688955	-3,519319207	2,988325132
97-8119	956,5000	0,2870100	-1,01723179	-3,0774812
97-8120	866,0000	-4,8231054	1,445978245	-1,925209541
IAC-100	706,7500	0,4914901	-3,240305024	1,073358608
IAS-5	1147,1250	0,1544785	3,811519854	-4,527045046
OC-4	1035,0000	1,0103231	4,611996965	-2,708378446
PRIMAVERA	1141,6250	-0,8221804	5,205625097	0,764094920
Amb1	1312,2159	-14,0101664	2,687415454	-16,24029873
Amb2	1127,0909	15,7023999	18,70361575	2,317285112
Amb3	778,2045	14,3726567	-19,70214898	-1,322200462

Pela mesma tabela verificou-se que os ambientes que mais contribuíram para a interação G×E foram os ambientes 2 e 4 possuindo os maiores escores em relação ao primeiro eixo da interação. Já o ambiente 1 obteve a maior média para o caráter produtividade de grãos (1312.2159 kg/ha). Além disso, confirmou-se o que foi verificado pelo gráfico da Figura 8: que o genótipo 97-8080 (13) possui maior média 1186,5 kg/ha para produtividade de grãos (PG).

A divergência entre os acessos foi avaliada pelo método hierárquico aglomerativo de Ward, com o emprego da distância Euclidiana. A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se um dendrograma representativo dos agrupamentos formados entre os genótipos em estudo (Figura 8).

A partir do dendrograma obtido obteve-se a formação de 8 grupos, os quais foram delimitados no próprio gráfico. Observa-se que os genótipos 97-8076 (11) e 97-8119 (39) pertenceram ao mesmo grupo, o que faz sentido quando é verificado o gráfico da Figura 7. Nesta, percebe-se que ambos os genótipos são estáveis (não contribuem para a interação) por estarem próximos da origem no Biplot, possuindo também valores aproximados para média de produtividade de grãos.

Figura 9 - Dendrograma das distâncias euclidianas entre os escores “bootstrap” de marcadores de genótipos AMMI₁, para dados de produtividade de grãos, em kg/ha

Além disso, os genótipos 97-8081 (14), 97-8108 (30), 97-8086 (18) e 97-8090 (21) também pertenceram ao mesmo grupo segundo o dendrograma, o que também pode ser confirmado pela proximidade destes no gráfico da Figura 7, todos apresentando interação positiva com os ambientes 1 e 4.

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz $G \times E$, mostrou que tal estimativa (igual a 1.636.472) está contida no intervalo definido pelo percentil 10 (igual a 1.367.044), o limite inferior, e o percentil 90 (igual a 2.159.879), o limite superior.

Experimento 2, parte 2: População PRIS - ano 2001

Na população 2 (PRIS - populações selecionadas para resistência a insetos sugadores), no ano agrícola 2001, os genótipos utilizados nas análises foram apresentados no ANEXO D.

Na Tabela 18 é apresentado o resumo da análise de variância conjunta onde foram apresentados o teste de F e a probabilidade associada. Observou-se efeito significativo a 1% de probabilidade de genótipos, de ambientes e para a interação genótipos $\times$ ambientes.

Tabela 19 - Análise conjunta da variância do Experimento 2 no ano agrícola 2001, incluindo o desdobramento da interação G $\times$ E para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja

Fonte de Variação	GL	SQ	QM	F	Pr (>F)	
Amb	3	15689519	5229840	374.8519	0.00002353	**
REP(Amb)	4	55807	13952	0.5981	0.6648	
Gen	22	2488907	113132	4.8500	1.043e-08	**
G$\times$E	66	5704592	86433	3.7054	5.249e-10	**
Resíduo	112	2612560	23326			
Média	999.9519					
C.V.	5.27373					

F : teste F de Gollob (1968).

**: significativo ao nível de probabilidade, $p < 0,01$. *: significativo ao nível de probabilidade, $p < 0,05$. ns: não significativo.

Pelo fato da interação ter se mostrado altamente significativa, pode-se dizer que os genótipos sofreram influências diferenciadas dos ambientes, dificultando a recomendação de cultivares para a região abrangida pelo estudo. Assim, é necessário estudo mais detalhado dessa interação de modo que ela possa ser controlada e não interfira negativamente na recomendação.

Em termos do teste F, a magnitude da fonte de variação de ambientes foi superior às demais, sendo ela responsável pela maior parte da variação ocorrida. Pode-se, assim, inferir que os efeitos de locais contribuíram de modo mais acentuado para a variação na produtividade. A significância para genótipos indicou que eles são formados por grupos geneticamente distintos, mostrando disponibilidade de variabilidade para a seleção.

Juntamente com a análise de variância foi obtida a média para a variável produtividade de grãos (PG) igual a 999,9519 kg/ha com coeficiente de variação baixo (15,27%), indicando boa precisão experimental.

A matriz GE (Tabela 19) terá $\text{posto} = \min(g-1, e-1) = \min(23-1, 4-1) = 3$, conseqüentemente a $SQ_{G \times E}$ pode ser decomposta em até três componentes. A próxima etapa da análise corresponde ao ajuste da interação pela decomposição em valores singulares aplicada à matriz GE.

Tabela 20 - Matriz de Resíduos da interação G×E, população 2 (PRIS - população com controle de insetos), no ano agrícola 2001

Genótipos	Amb 1	Amb 2	Amb 3	Amb 4
97-8063	97,059780	38,951090	-37,44022	-98,5707
97-8064	-57,940220	-140,048900	22,55978	175,4293
97-8067	63,434780	-40,673910	14,93478	-37,6957
97-8068	215,059800	-50,048910	-94,44022	-70,5707
97-8072	20,434780	-9,673913	155,93480	-166,696
97-8073	-1,440217	188,451100	-137,94020	-49,0707
97-8080	108,559800	49,451090	-101,9402	-56,0707
97-8085	-171,19020	24,701090	239,3098	-92,8207
97-8088	-56,94022	80,951090	-142,4402	118,4293
97-8091	-16,44022	182,451100	-199,9402	33,92935
97-8093	-64,94022	127,951100	-79,44022	16,42935
97-8095	-153,44020	115,451100	33,05978	4,929348
97-8096	-32,69022	203,201100	-18,19022	-152,321
97-8100	77,05978	-4,048913	-101,4402	28,42935
97-8105	65,05978	87,951090	-64,44022	-88,5707
97-8108	-125,81520	149,0761	14,68478	-37,9457
97-8109	161,18480	176,0761	-147,3152	-189,946
97-8114	117,05980	-233,0489	60,55978	55,42935
97-8118	90,05978	-29,04891	-38,44022	-22,5707
97-8119	-235,4402	108,4511	15,05978	111,9293
IAC-100	36,05978	39,45109	-44,94022	-30,5707
IAS-5	-132,9402	-507,0489	384,5598	255,4293
OC-4	-1,815217	-558,9239	267,6848	293,0543

Pela Tabela 20 observa-se que o modelo AMMI com dois componentes explicou 88,7% da soma de quadrados da interação entre genótipos e ambientes como resposta-padrão, porcentagem significativa em se tratando do caráter produtividade de grãos em estudo, e 11,3% dessa soma de quadrados é de ruídos presentes nos dados.

Tabela 21 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 23 genótipos de soja. População PRIS, ano 2001

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA ₁	69.9	69.9	24	2733142.4	113880.93	4.88	0.0000
IPCA ₂	18.8	88.7	22	734813.2	33400.60	1.43	0.1158
IPCA ₃	11.3	100	20	440150.6	22007.53	0.94	0.5390
IPCA ₄	0	100	18	0	0	0	1

Assim, pelo critério de Gollob, o modelo AMMI₁ foi selecionado como aquele que melhor descreve o padrão inerente à interação G×E uma vez que o teste só foi significativo para o IPCA₁.

É importante notar na Tabela 20 que o IPCA₄ teve porcentagem nula na explicação da variabilidade observada na matriz de interação, o que é justificado pelo posto, que indicava a necessidade de somente 3 componentes.

A representação gráfica desses resultados é difícil, bem como a associação dos escores dos eixos IPCA₃ e superiores, com atributos identificáveis dos genótipos e dos ambientes. Segundo Gauch (1988), os primeiros eixos AMMI captam maior porcentagem do desempenho real (“padrão”) e com subsequente acumulação de dimensões (eixos), ocorre diminuição na porcentagem de “padrão” e um acréscimo de informações imprecisas (“ruídos”). Com isso, apesar da seleção de um só eixo, com pequena porção da SQ_{G×E} original, espera-se captar maior porcentagem do “padrão”.

Foram utilizados, para ilustrar o efeito de cada genótipo e ambiente, os gráficos biplot AMMI 2 (PC1 e PC2) (Figura 10) e AMMI 1 (médias×PC1) que pode ser visto na Figura 11.

Observando a Figura 10, os genótipos 97-8067 (3), 97-8072 (5) e 97-8118 (19) contribuíram menos para a interação, e foram considerados, por esse método, os mais estáveis agronomicamente. Dentre os estáveis, aquele que possuiu a maior média para produtividade

de grãos foi 97-8067 (5) com 962,875 kg/ha, não ultrapassando a média geral de 999,9519 kg/ha. Os genótipos mais instáveis foram as testemunhas IAS-5 e OC-4.

Figura 10 - Biplot AMMI₂ para dados de produtividade de grãos em kg/ha, população PRIS, ano 2001

Na Figura 11, a abscissa indica os efeitos principais (médias), e a ordenada os efeitos da interação (IPCA 1). Quanto mais distante da origem do eixo AMMI de interação (AMMI 1), maior é a contribuição para a interação.

Figura 11 - Gráfico do comportamento de 23 genótipos de soja da população PRIS (ano 2001) em relação ao primeiro componente da interação e ao caráter PG

A adaptação específica indica maior média de produtividade de grãos em um determinado ambiente. A identificação de genótipos, com adaptação específica a uma região, pode levar a recomendação mais precisa das cultivares, com melhor aproveitamento do seu potencial genético, pois capitaliza a interação genótipos com ambientes a seu favor (RAMALHO et al., 1993).

Nesse caso, pela Figura 11 observa-se que os genótipos 97-8080, 97-8088 e 97-8095 se adaptaram muito bem ao ambiente 1, com média de produtividade em torno de 1181 kg/ha.

Na Tabela 21, por meio das análises das magnitudes e dos sinais dos escores biplots correspondentes aos modelos AMMI selecionados, foram confirmados como estáveis (escores considerados baixos, próximos de zero), aqueles já detectados desta forma na Figura 10. Eles são os genótipos 97-8067 (3), 97-8072 (5) e 97-8118 (19).

Pela Tabela 21 é verificado que o ambiente 3 apresentou a menor média dos genótipos ($602,3 \text{ kg ha}^{-1}$), e o ambiente 2 a maior média ($1327,9 \text{ kg ha}^{-1}$), e podem ser considerados exemplos de ambiente desfavorável e favorável, respectivamente.

Pela mesma tabela verificou-se que os ambientes que mais contribuíram para a interação G×E foram os ambientes 2 e 3 possuindo os maiores escores em relação ao primeiro eixo da interação.

Tabela 22 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 23 genótipos de soja da população 2 (PRIS), no ano agrícola 2001

(continua)

Genótipos	PG	IPCA ₁	IPCA ₂	IPCA ₃
97-8063	1132,25	-2,83262	3,868132	2,772729
97-8064	1141,25	5,661359	-1,23973	-5,81369
97-8067	962,875	0,513049	2,909669	2,003443
97-8068	1043,25	-1,6758	9,886056	0,528872
97-8072	898,875	0,49603	0,04466	10,56121
97-8073	973,75	-6,79277	-1,00077	-2,26387
97-8080	1181,75	-3,54604	4,614795	-0,58089
97-8085	971,5	2,334303	-8,66604	9,674723
97-8088	1180,25	-2,35295	-2,28378	-8,57273
97-8091	1045,75	-6,55631	-1,21245	-7,08446
97-8093	1135,25	-3,64874	-3,40427	-3,12506
97-8095	1182,75	-1,61629	-7,5888	0,246996

Tabela 21 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 23 genótipos de soja da população 2 (PRIS), no ano agrícola 2001

(conclusão)				
Genótipos	PG	IPCA ₁	IPCA ₂	IPCA ₃
97-8096	1078	-6,4896	-3,11007	4,7973
97-8100	943,25	-1,27247	3,800331	-3,69528
97-8105	877,25	-4,11711	2,250204	1,518008
97-8108	995,125	-3,21142	-6,65046	1,341064
97-8109	996,125	-8,74211	5,887896	3,066267
97-8114	1040,25	6,414369	6,62946	0,191993
97-8118	921,25	-0,42352	4,207786	0,004301
97-8119	952,75	-0,25064	-10,8114	-4,38634
IAC-100	957,75	-1,99326	1,384965	-0,08844
IAS-5	928,25	20,28691	-3,03878	1,593403
OC-4	786,625	19,81563	3,522572	-2,68954
Amb1	1260,87	-3,55911	21,04285	2,013069
Amb2	1327,978	-26,4557	-9,56677	0,367019
Amb3	602,3696	16,73743	-7,46519	14,01032
Amb4	865,5	13,27741	-4,01089	-16,3904

A divergência entre os acessos foi avaliada pelo método hierárquico aglomerativo de Ward, com o emprego da distância Euclidiana. A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se um dendrograma representativo dos agrupamentos formados entre os genótipos em estudo (Figura 12).

Figura 12 - Dendrograma das distâncias euclidianas entre os escores “bootstrap” de marcadores de genótipos AMMI₁, para dados de produtividade de grãos, em kg/ha

A partir do dendrograma obtido obteve-se a formação de 4 grupos, os quais foram delimitados no próprio gráfico. De acordo com a identificação recebida por cada genótipo (ANEXO D) temos que os grupos formados no dendrograma podem ser descritos: no grupo 1 os genótipos 1, 2, 11 e 12; pertencentes ao grupo 2 os genótipos 3, 4, 6, 7, 8, 10, 13, 15, 17, 19, 20, 21, 22 e 23; pertencentes ao grupo 3 tem-se os genótipos 5, 14, 16; pertencentes ao grupo 4 os genótipos 9 e 18.

Além disso, observa-se na Figura 12 que as testemunhas 21, 22 e 23 ocorreram no mesmo grupo. Em relação aos genótipos 1 e 11, eles permaneceram no mesmo grupo o que pode ser justificado pelo fato de apresentarem praticamente a mesma média para produtividade de grãos e ambos podem ser considerados estáveis por se localizarem próximo à origem no gráfico Biplot da Figura 10.

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz $G \times E$, mostrou que tal estimativa (igual a 1.954.053) está contida no intervalo definido pelo percentil 10 (igual a 1.095.376), o limite inferior, e o percentil 90 (igual a 2.926.320), o limite superior.

Experimento 3, parte 1: População PRIM - ano 2000

Na população 3 (PRIM (populações selecionadas para resistência a insetos mastigadores), no ano agrícola 2000, os genótipos utilizados nas análises foram identificados no ANEXO E.

Foi realizada a análise de variância conjunta para o experimento envolvendo a população 3 (Tabela 22 e 24) sendo a presença significativa da interação genótipo por ambiente ($G \times E$), diagnosticada pelo teste F (Cochran, 1954).

A existência de variabilidade genética entre linhagens foi constatada tanto nas análises individuais quanto na análise conjunta. Na análise conjunta, todos os efeitos foram significativos, indicando presença de variabilidade para seleção entre os genótipos,

variabilidade entre os ambientes em estudo e ocorrência de resposta diferencial dos genótipos aos ambientes, o que determina a necessidade de análises de estabilidade (Tabela 22).

Tabela 23 - Análise conjunta da variância do Experimento 3 no ano 2000, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 44 genótipos de soja

Fonte de Variação	GL	SQ	QM	F	Pr (>F)	
Amb	3	26158609	8719536	98.6721	0.0003318	**
REP(Amb)	4	353475	88369	4.4726	0.00002132	**
Gen	43	3700526	86059	4.3557	5.047e-13	**
G×E	129	3763202	29172	1.4765	0.0064550	**
Resíduo	204	4030571	19758			
Média	1004.964					
C.V.	13.98678					

F : teste F de Gollob (1968).

** : significativo ao nível de probabilidade, $p < 0,01$. * : significativo ao nível de probabilidade, $p < 0,05$.

ns: não significativo.

A matriz de interações GE é apresentada na Tabela 23 e a próxima etapa da análise corresponde ao ajuste da interação pela decomposição em valor singular aplicada à matriz GE.

Tabela 24 - Matriz de Resíduos da interação G×E, população 3 (PRIM - população com controle de insetos), no ano agrícola 2000

(continua)				
Genótipos	Amb 1	Amb 2	Amb 3	Amb 4
97-8121	51,28409	146,3182	-108,409	-89,1932
97-8122	279,7841	-147,182	-195,909	63,30682
97-8123	129,7841	-119,182	-164,909	154,3068
97-8124	-95,9659	210,0682	-105,659	-8,44318
97-8125	100,2841	68,31818	-24,4091	-144,193
97-8129	-164,716	38,31818	-27,4091	153,8068
97-8130	-42,7159	20,31818	61,59091	-39,1932
97-8132	-74,9659	-70,9318	94,34091	51,55682
97-8133	-150,466	132,5682	48,84091	-30,9432
97-8134	67,28409	-186,682	-88,4091	207,8068
97-8135	156,0341	-138,932	-45,6591	28,55682
97-8136	-12,9659	-84,9318	10,34091	87,55682

Tabela 23 - Matriz de Resíduos da interação G×E, população 3 (PRIM - população com controle de insetos), no ano agrícola 2000

Genótipos	Amb 1	Amb 2	Amb 3	(conclusão)
				Amb 4
97-8137	3,034091	170,0682	-118,659	-54,4432
97-8139	-72,9659	-115,932	28,34091	160,5568
97-8142	54,78409	32,81818	68,09091	-155,693
97-8143	-24,9659	-20,9318	194,3409	-148,443
97-8144	-96,2159	115,8182	-176,909	157,3068
97-8145	31,28409	36,31818	74,59091	-142,193
97-8146	-64,4659	73,56818	-44,1591	35,05682
97-8147	-106,716	-55,6818	113,5909	48,80682
97-8148	-141,966	75,06818	-91,6591	158,5568
97-8149	-10,9659	152,0682	-67,6591	-73,4432
97-8150	-71,7159	-69,6818	67,59091	73,80682
97-8151	-112,466	-67,4318	236,8409	-56,9432
97-8153	14,28409	17,31818	-92,4091	60,80682
97-8154	32,28409	81,31818	-96,4091	-17,1932
97-8156	73,53409	-102,432	70,84091	-41,9432
97-8158	14,53409	-94,4318	-21,1591	101,0568
97-8159	-121,216	-41,1818	71,09091	91,30682
97-8160	89,78409	-57,1818	-75,9091	43,30682
97-8163	73,28409	-79,6818	-83,4091	89,80682
97-8164	-30,7159	82,31818	-101,409	49,80682
97-8165	-88,7159	-9,68182	80,59091	17,80682
97-8166	-60,4659	-110,432	243,8409	-72,9432
97-8167	-75,9659	27,06818	34,34091	14,55682
97-8172	51,78409	-98,1818	117,0909	-70,6932
97-8173	98,53409	33,56818	-67,1591	-64,9432
97-8175	150,5341	-190,432	1,840909	38,05682
97-8176	99,28409	-143,682	146,5909	-102,193
97-8178	32,53409	82,56818	-16,1591	-98,9432
IAC-100	-11,2159	56,31818	-1,90909	-43,1932
IAS-5	19,28409	112,8182	11,59091	-143,693
OC-4	58,28409	153,8182	-46,4091	-165,693
PRIMAVERA	-48,9659	86,06818	85,84091	-122,943

A matriz GE terá posto = $\min(g-1, e-1) = \min(44-1, 4-1) = 3$, consequentemente a $SQ_{G \times E}$ pode ser decomposta em até três componentes, dos quais apenas o primeiro termo

(IPCA1) foi julgado como altamente significativo ao nível de 5% de probabilidade. Os resíduos do AMMI₂ mostraram-se não-significativos, confirmando a seleção de AMMI₁. Este modelo explicou apenas 38,5 % da $SQ_{G \times E}$ (Tabela 24) e os demais eixos representam variações da $SQ_{G \times E}$ mais predominantemente ricas em ruído.

Para a identificação das linhagens mais estáveis por AMMI, foram utilizadas as informações dos dois primeiros componentes, para a obtenção da média dos escores absolutos de cada genótipo, ponderada pela porcentagem de explicação de cada componente (Tabela 24). Quanto menor o valor da média dos escores, menor a contribuição do genótipo para a interação e, consequentemente, mais estável é o genótipo. Assim, os genótipos mais estáveis foram 97-8159 (29) e 97-8173 (37) (Figura 13).

Tabela 25 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente, referente aos dados de produtividade de grãos em kg/ha de 44 genótipos de soja. População PRIM, ano 2000

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA ₁	38.5	38.5	45	1355025.6	30111.68	1.52	0.0273
IPCA ₂	34.5	73.0	43	1215390.2	28264.89	1.43	0.0534
IPCA ₃	27.0	100	41	953286.1	23250.88	1.18	0.2272
IPCA ₄	0	100	39	0	0	0	1

Pela Figura 13 observa-se que o genótipo 97-8135 (11) é altamente adaptado ao ambiente 1 e além disso possui alta média de produtividade (Figura 14), o que ocorre também com os genótipos 97-8163 (31) e 97-8175 (38). Já os genótipos 97-8122 (2), 97-8123 (3) e 97-8134 (10) são bem adaptados ao ambiente 4.

Figura 13 - Biplot AMMI₂ para dados de produtividade de grãos em kg/ha, população PRIM, ano 2000

Figura 14 - Gráfico do comportamento de 44 genótipos de soja da população PRIM (ano 2000) em relação ao primeiro componente da interação e ao caráter PG

Desta forma, no primeiro eixo da interação (Tabela 25) observou-se que o ambiente que menos contribuiu para a interação foi o ambiente 1 com escore baixo e igual a 9,91403 sendo este ambiente o que também possui maior valor em produtividade (1361,8182). Já o ambiente que mais contribuiu para a interação G×E foi o ambiente 2 (Amb2) possuindo o maior escore (15,62852).

Tabela 26 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001

(continua)

Genótipos	PG	IPCA₁	IPCA₂	IPCA₃
97-8121	922,50	2,490717	6,025108	-3,87031
97-8122	937,00	-10,5022	1,547836	-8,48259
97-8123	968,00	-9,69791	1,855889	-1,68303
97-8124	947,75	3,7411	7,899613	2,526314
97-8125	1009,50	2,891722	1,707938	-6,24842
97-8129	1020,50	-1,08153	2,315962	8,326323
97-8130	1018,50	2,688016	-1,21961	0,816401
97-8132	1131,75	-0,15839	-3,94315	3,818789
97-8133	929,25	5,74404	1,949696	4,411532
97-8134	915,50	-10,2523	-1,54118	1,705662
97-8135	1166,75	-5,83163	-2,1999	-4,62751
97-8136	945,75	-3,22552	-2,01163	2,231047
97-8137	976,75	2,604828	7,029046	-1,64842
97-8139	865,75	-4,42536	-2,92935	5,713446
97-8142	995,00	4,409545	-1,57288	-4,66459

Tabela 27 25 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001

(conclusão)

Genótipos	PG	IPCA ₁	IPCA ₂	IPCA ₃
97-8143	927,75	6,088784	-6,08459	-1,46867
97-8144	1010,00	-2,76026	8,057446	5,623409
97-8145	966,50	4,560054	-1,59169	-3,60846
97-8146	953,25	0,738645	3,111443	2,623329
97-8147	1001,50	0,863519	-4,06057	4,859974
97-8148	987,75	-1,7284	4,883655	7,446249
97-8149	1153,75	3,616284	5,205037	-1,38929
97-8150	914,50	-1,06515	-3,13866	4,063779
97-8151	1013,25	4,887656	-7,93118	3,368097
97-8153	1038,50	-2,56951	3,035547	0,422431
97-8154	973,50	0,08323	4,403598	-1,7535
97-8156	969,25	-0,84328	-4,53953	-2,95473
97-8158	1097,25	-4,50758	-1,39373	1,498535
97-8159	945,00	-0,25985	-2,43824	6,028409
97-8160	933,00	-4,26061	0,676769	-2,30553
97-8163	1006,50	-5,62399	0,503499	-0,85951
97-8164	1019,50	-0,68689	4,843779	1,617374
97-8165	1091,50	1,705266	-2,20652	3,52679
97-8166	994,25	3,901856	-9,25413	1,392811
97-8167	891,75	1,625385	-0,12907	2,875158
97-8172	1156,00	0,826864	-5,72808	-2,66101
97-8173	1056,25	-0,13189	2,261554	-4,74776
97-8175	1139,25	-6,23771	-4,65144	-4,07844
97-8176	944,50	0,526868	-7,75013	-4,72025
97-8178	1056,25	3,107735	2,050735	-3,11495
IAC-100	738,50	2,125131	1,272632	-0,52039
IAS-5	1142,50	5,242653	1,912837	-3,48937
OC-4	1130,50	5,162739	4,313496	-5,41201
PRIMAVERA	1112,25	6,217289	-0,54785	-0,58712
Amb 1	1361,8182	-9,91403	-0,78454	-20,8521
Amb 2	1129,7841	15,62852	18,73656	1,420526
Amb 3	786,5114	12,37832	-20,5245	4,024882
Amb 4	732,2955	-18,0928	2,572497	15,40669

A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se o gráfico dendrograma (Figura 15) representativo dos agrupamentos formados entre os genótipos em estudo. A partir do dendrograma obtido obteve-se a formação de 4 grupos, os quais foram delimitados no próprio gráfico.

Figura 15 - Dendrograma das distâncias euclidianas entre os escores “*bootstrap*” de marcadores de genótipos AMMI₁, para dados de produtividade de grãos, em kg/ha

Observa-se que os genótipos 2 e 3 permaneceram no mesmo grupo, o que pode ser justificado pelo fato de apresentarem relativamente a mesma média para produtividade de grãos (abaixo da média geral) e serem ambos bem adaptados ao ambiente 4. Observa-se também o agrupamento dos genótipos 97-8121 (1), IAC-100 (41) e IAS-5 (42) que se mostraram bem adaptados ao ambiente 2 (Figura 13).

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz $G \times E$, mostrou que tal estimativa (igual a 1.761.851) está contida no intervalo definido pelo percentil 10 (igual a 1.515.193), o limite inferior, e o percentil 90 (igual a 2.078.966), o limite superior.

Experimento 3, parte 2: População PRIM - ano 2001

Na população 3 (PRIM) no ano agrícola 2001, os genótipos utilizados nas análises foram identificados no ANEXO F.

Os resultados da análise de variância conjunta mostraram haver diferenças altamente significativas ($p < 0,01$) para as fontes de variação de ambientes e da interação entre elas (Tabela 26).

Tabela 28 - Análise conjunta da variância do Experimento 3 no ano 2001, incluindo o desdobramento da interação G×E para dados de produtividade de grãos, em kg/ha, de 23 genótipos de soja

Fonte de Variação	GL	SQ	QM	F	Pr (>F)	
Amb	3	12059010	4019670	76.6195	0.0005454	**
REP(Amb)	4	209851	52463	1.3851	0.2436519	
Gen	22	902522	41024	1.0831	0.3755913	
G×E	66	4229993	64091	1.6921	0.0071357	**
Resíduo	112	4242096	9246			
Média	1019.226					
C.V.	19.09461					

F : teste F de Gollob (1968).

** : significativo ao nível de probabilidade, $p < 0,01$. * : significativo ao nível de probabilidade, $p < 0,05$.

ns: não significativo.

A significância da interação é um indicativo de que os materiais genéticos apresentaram comportamento diferenciado nos vários ambientes, sugerindo assim, a realização de uma análise mais criteriosa para identificar quais deles exibem uma menor oscilação nesses ambientes estudados.

O coeficiente de variação (CV%) do experimento mostrou-se de baixa magnitude, permitindo inferir que ele foi bem conduzido. Constatou-se que o primeiro componente principal (IPCA₁) foi altamente significativo ($p < 0,01$) (Tabela 27) e os demais componentes não foram significativos. De um modo geral, o primeiro componente explicou mais de 50% da variação total da interação, evidenciando que o mesmo conseguiu mostrar a maior parte da variação contida nesse conjunto de materiais genéticos.

Tabela 29 - Porcentagem da soma de quadrados acumulada por eixo singular (PA), Teste F para cada componente referente aos dados de produtividade de grãos em kg/ha, de 23 genótipos de soja. População PRIM, ano 2001

IPCA	Porcentagem	PA	GL	SQ	QM	F	P-valor
IPCA ₁	59.7	59.7	24	2236606.9	93191.95	2.46	0.0008
IPCA ₂	24.6	84.3	22	919533.7	41796.99	1.10	0.3577
IPCA ₃	15.7	100	20	587795.3	29389.77	0.78	0.7321
IPCA ₄	0	100	18	0	0	0	1

Na Figura 16 observa-se que o genótipo mais estável é 97-8132 (3), que apresentou escore praticamente nulo em relação ao IPCA₁. Mais associado ao ambiente 1 tem-se o genótipo 97-8175 (18), ambiente este que se destacou dentre os demais na média para produtividade de grãos.

Figura 16 - Biplot AMMI₂ para dados de produtividade de grãos em kg/ha, população PRIM, ano 2001

Os ambientes 2 e 3 apresentaram-se com maior escore em relação ao IPCA₁ (primeiro componente principal e maior responsável pela interação G×E). Mais associado ao ambiente 3 tem-se o genótipo 97-8176 (19) que apresentou média para produtividade de grãos (1052 kg/ha) acima da média geral observado no estudo (1019,226 kg/ha).

Devido ao procedimento de decomposição da interação materiais genéticos × ambientes do experimento por meio do modelo AMMI, foi obtida assim a matriz de interação G×E (Tabela 28).

Tabela 30 - Matriz de Resíduos da interação G×E, população 3 (PRIM - população com controle de insetos), no ano agrícola 2001

Genótipos	Amb 1	Amb 2	Amb 3	Amb 4
97-8121	-189 ,8777	68,25272	188,144	-66,51902
97-8122	181 ,3723	121,5027	-103,606	-199,269
97-8132	-5 ,752717	54,87772	134,769	-183,894
97-8135	31 ,37228	-127,497	148,394	-52,26902
97-8136	-8 ,127717	96,00272	-105,106	17,23098
97-8137	-137 ,8777	189,2527	-84,856	33,48098
97-8142	26 ,37228	122,5027	-59,606	-89,26902
97-8146	4 ,622283	154,7527	-337,356	177,981
97-8147	204 ,3723	76,50272	-214,606	-66,26902
97-8149	-124 ,6277	-216,497	153,394	187,731
97-8151	28 ,12228	68,25272	-80,856	-15,51902
97-8154	7 ,997283	37,12772	-82,981	37,85598
97-8158	39 ,37228	-266,497	94,39402	132,731
97-8159	-45 ,62772	154,5027	-98,606	-10,26902
97-8165	18 ,62228	83,75272	-59,356	-43,01902
97-8172	-219 ,3777	-100,247	53,64402	265,981
97-8173	-15 ,87772	-106,247	-26,356	148,481
97-8175	337 ,3723	57,50272	-128,606	-266,269
97-8176	106 ,1223	-408,747	261,144	41,48098
97-8178	110 ,1223	115,2527	-115,856	-109,519
IAC-100	7 ,497283	80,12772	-74,981	-12,64402
IAS-5	-188 ,5652	-142,935	278,4565	53,04348
OC-4	-167 ,6277	-111,497	260,394	18,73098

A significância dos componentes principais está relacionada com as magnitudes dos escores para materiais genéticos e para ambientes, explicando assim, parte da variação da interação. Pelos valores encontrados, pode-se inferir que há padrões de interação diferenciado de materiais genéticos e ambientes, ou seja, de estabilidade genotípica e ambiental. A estabilidade dos materiais genéticos dos vários grupos de ensaios pode ser identificada pela variabilidade na produtividade de grãos e pelos valores dos escores do primeiro componente principal e também pelo segundo componente (Tabela 29).

Escore alto não importando se positivos ou negativos, indicam grandes interações, ou seja, baixa estabilidade. Por outro lado, escores próximos a zero, refletem baixas interações, acarretando consequentemente, alta estabilidade. Nesse sentido, constatou-se que a amplitude de variação dos escores foi alta, atribuindo aos ambientes os seus maiores valores.

Tabela 31 - Médias e escores de genótipos e de ambientes para dados de produtividade de grãos (PG), em kg/ha, de 44 genótipos de soja da população 3 (PRIM), no ano agrícola 2001

Genótipos	PG	IPCA₁	IPCA₂	IPCA₃
97-8121	1065	-4,0138	1,636117	10,66678
97-8122	1114,75	8,274263	-5,9835	1,621309
97-8132	941,375	0,057685	-5,04409	8,346779
97-8135	1097,75	-4,37221	-5,62581	0,764411
97-8136	1041,25	3,55481	3,266176	-0,23523
97-8137	1101	3,195276	7,830525	4,458154
97-8142	1075,75	4,570265	-0,23321	3,066873
97-8146	1121,5	7,787878	10,42794	-7,6328
97-8147	1067,75	8,620078	-2,69467	-5,18351
97-8149	1035,75	-10,2148	2,385378	-3,58723
97-8151	995	3,308524	0,937614	-0,32138
97-8154	977,125	2,035171	2,24904	-1,8973
97-8158	881,75	-7,53487	-2,80822	-7,88585
97-8159	943,75	4,323521	4,274827	2,585124
97-8165	946,5	3,323441	0,477436	1,288582
97-8172	1075,5	-7,98879	9,373139	-2,95619
97-8173	1061,5	-3,03382	2,660782	-6,0355
97-8175	1008,75	10,0538	-11,8862	-1,86047
97-8176	1052	-11,7507	-10,6291	-6,87569
97-8178	1084	6,709803	-2,11849	0,566281
IAC-100	986,625	3,142636	1,595094	0,406184
IAS-5	1001,688	-10,7748	0,270245	4,993463
OC-4	917,75	-9,27337	-0,36101	5,707225
Amb1	1275,696	12,12175	-16,4532	-10,7159
Amb2	1241,565	19,13316	9,444051	12,14967
Amb3	656,6739	-20,9213	-8,61156	11,08672
Amb4	929,337	-10,3336	15,62067	-12,5205

Além disso, pode ser confirmado nesta tabela que o genótipo 97-8132 (3) foi o mais estável por apresentar o mais baixo escore em relação ao $IPCA_1$. Os altos valores de escores para ambientes indicaram que eles apresentaram níveis de produtividades bem distintos.

Na Figura 17, tem-se o gráfico que visa demonstrar o comportamento dos genótipos mediante a produtividade de grãos média, importante na recomendação de cultivares. Dentre todos os genótipos pertencentes à população PRIM no ano 2001, verificou-se que aqueles à direita da linha vertical tracejada são os genótipos que ultrapassaram a média geral de produtividade de grãos (1019,226 kg/ha).

Figura 17 - Gráfico do comportamento de 23 genótipos de soja da população PRIM (ano 2001) em relação ao primeiro componente da interação e ao caráter PG

Como destaque tem-se os genótipos 97-8122 (2) e 97-8146 (8) que apresentaram os maiores valores para o caráter produtividade de grãos avaliado (1114,75 e 1121,5 kg/ha, respectivamente). Já o genótipo 97-8158 (13) foi revelado no gráfico aquele com mais baixo valor para o caráter avaliado neste estudo. O ambiente 1 se revelou aquele com maior valor para produtividade de grãos (1275,696 kg/ha).

Após a obtenção de resultados a respeito da matriz de interação $G \times E$, a divergência entre os acessos foi avaliada pelo método hierárquico aglomerativo de Ward, com o emprego da distância Euclidiana. A partir da análise de agrupamentos utilizando a matriz de distâncias $\bar{d}$ obteve-se um dendrograma representativo dos agrupamentos formados entre os genótipos em estudo (Figura 18).

O dendrograma uniu os materiais semelhantes como pode ser identificado ao observar que os genótipos 1, 6, 16, 17 pertenceram ao mesmo grupo o que pode ser justificado pelo fato destes apresentarem valores próximos de produtividade (Tabela 31). Nota-se também que os indivíduos 10, 22 e 23 permaneceram no mesmo grupo o que pode ser justificado pelo fato

destes estarem altamente relacionados ao ambiente 3 (Figura 16) e aproximadamente possuírem mesma média de produtividade de grãos.

Figura 18 - Dendrograma das distâncias euclidianas entre os escores “*bootstrap*” de marcadores de genótipos AMMI₁, para dados de produtividade de grãos, em kg/ha

Outra observação que pode ser feita é em relação aos genótipos 5, 12 e 21 que ficaram no mesmo grupo. Estes estão altamente relacionados ao ambiente 2 e são aproximadamente estáveis. A partir do dendrograma formado, obteve-se a formação de seis grupos os quais podem ser descritos segundo a Tabela 30.

Tabela 32 - Agrupamentos de vinte e três genótipos de soja da população 3 referente ao experimento realizado no ano 2001, obtidos pelo método hierárquico de Ward

Grupos	Genótipos
1	1 6 16 17
2	2 4 7 10 11 14 15 20 22 23
3	3 5 9 12 13 18 19 21
4	8

Por meio do uso do *bootstrap* podemos obter intervalos aproximados de $100(1-\alpha)\%$ de confiança para o parâmetro de interesse. A análise do intervalo de confiança interpercentílico baseado nos percentis *bootstrap*, referente à estimativa da função traço da matriz de covariâncias da matriz $G \times E$, mostrou que tal estimativa (igual a 1.871.968) está contida no intervalo definido pelo percentil 10 (igual a 1.479.452), o limite inferior, e o percentil 90 (igual a 2.548.456), o limite superior.

5 CONCLUSÕES

Os resultados obtidos nessa pesquisa permitem extrair as seguintes conclusões:

- a) As análises realizadas foram conclusivas no que se refere a selecionar os genótipos mais estáveis e com alta média para o caráter produtividade de grãos (PG) estudado. Na população PCI ano agrícola 2000 conclui-se que uma boa recomendação para o ambiente 1 é o genótipo 97-8056 que apresentou o maior escore de produtividade de grãos. Podem ser recomendados amplamente os genótipos 97-8011, 97-8029, 97-8050 (33) e a testemunha IAS-5 (44) uma vez que além de serem estáveis, possuíram ótima média para o caráter PG. Já o ambiente 1 apresentou a maior média para este caráter (1408,602 kg/ha).
- b) Na população PCI ano agrícola 2001 conclui-se que o genótipo 97-8010 (5) foi o que apresentou efeito da interação pequeno e produção alta, representando a melhor combinação de estabilidade e produtividade.
- c) Na população PRIS ano agrícola 2000 conclui-se que pode-se considerar como de alta estabilidade fenotípica os genótipos 97-8088, 97-8096, 97-8101, 97-8111, 97-8119 e a testemunha IAC-100. Esses genótipos também foram mais estáveis que as testemunhas comuns; entre essas, a IAS-5 (42) foi a mais estável. No entanto, para fins de recomendação, tem-se os genótipos 97-8088, 97-8096, 97-8111 e 97-8119 que, além de estáveis tiveram alta média de produtividade de grãos detectada para esta população e ano.
- d) Na população PRIS ano agrícola 2001 conclui-se que os genótipos 97-8067, 97-8072 e 97-8118 contribuíram menos para a interação, e foram considerados os mais estáveis agronomicamente. Dentre eles, aquele que possuiu a maior média para produtividade de grãos foi o genótipo 97-8067. Quanto à adaptabilidade, observou-se que os genótipos 97-8080, 97-8088 e 97-8095 se adaptaram muito bem ao ambiente 1, com ótima média de produtividade de grãos.
- e) Na população PRIM ano agrícola 2000 foi observado a adaptação do genótipo 97-8135 com o ambiente 1 e além disso possuiu alta média de produtividade, o que

ocorreu também com os genótipos 97-8163 e 97-8175. Já os genótipos 97-8122, 97-8123 e 97-8134 foram bem adaptados ao ambiente 4. Entre os genótipos de maior produtividade, mereceram destaque o 97-8172 e o 97-8173 que são também estáveis, o que é desejado do ponto de vista do melhoramento genético.

- f) Na população PRIM ano agrícola 2001 foi detectado que o genótipo 97-8132 foi o mais estável. Além disso, se destacaram os genótipos 97-8122 e 97-8146 que apresentaram os maiores valores para o caráter produtividade de grãos.
- g) Verificou-se, pelos casos estudados, a generalidade de aplicação da técnica de estimação por meio da reamostragem e que o método *bootstrap* permite que o cálculo do intervalo de confiança seja realizado de modo mais simples e abrangente para diversas estatísticas, mesmo quando as distribuições de probabilidades das mesmas são desconhecidas;

REFERÊNCIAS

ANDERBERG, M.R. Cluster Analysis for Applications. **Probability and Mathematical Statistics**, London, v. 19, p 142-148, 1973.

ALMEIDA, I.R. de **Variabilidade pluviométrica interanual e produção de soja no Estado do Paraná**. 2000. 130p. Dissertação (Mestrado em Geografia) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista “Julio de Mesquita Filho” Presidente Prudente, Presidente Prudente, 2000.

BARROS, H.B. **Adaptabilidade e estabilidade de soja por métodos paramétricos e não-paramétricos**. 2007. 82p. Tese (Doutorado em Fitotecnia) – Universidade Federal de Viçosa, Viçosa, 2007.

BECELAERE, G.V.; LUBBERS, E.L.; PATERSON, A.H.; CHEE, P.W. Pedigree-vs. DNA Marker-Based genetic similarity estimates in cotton, **Crop Science**, Madison, v. 45, p. 2281-2287, 2005.

BONETTI, L.P. Cultivares e seu melhoramento genético. In: VERNETTI, F. de J. (Coord.). **Soja, genética e melhoramento**. Campinas: Fundação Cargill, 1983. p.741- 800.

BOREM, A.; MIRANDA, G.V. **Melhoramento de plantas**. 4 ed. Viçosa: UFV, 2005. 525p.

BRO, R. **Multi-way analysis in the food industry. Models, algorithms, and applications**. 1998. 288p. Tese (Doutorado) - University of Amsterdam (NL) and Royal Veterinary and Agricultural University (DK), 1998.

CARPENTER, J.; BITHELL, J. Bootstrap Confidence Intervals: When, which, what? A Practical Guide for Medical Statistician, **Statistics in Medicine**, London, v. 19, p.1141–1164, 2000.

CAZAR, R.A. An Exercise on Chemometrics for a Quantitative Analysis Course. **Journal of Chemical Education**, Madison, WI, v. 80, n. 9, p. 1026-1029, 2003.

CHAVES, J.L. Interação de cultivares com ambientes. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES, M.C. Recursos genéticos e melhoramento - plantas. Rondonópolis: Fundação MT, 2001. p.673-713.

COCHRAN, W.G. The combination of estimates from different experiments. **Biometrics**, Washington, v.10, p.101-129, 1954.

COELHO-BARROS, EMÍLIO AUGUSTO et al. Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos. **Revista Colombiana de Estadística**, Bogotá, v.31, n.1, p.111-129, Jun 2008.

CORNELIUS, P.L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. **Crop Science**, Madison, v.33, n.6, p.1186-1193, 1993.

CORNELIUS, P.L.; SEYEDSADR, M.; CROSSA, J. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. **Theoretical and Applied Genetics**, Heidelberg, v.84, n.1/2, p.161-172, 1992.

CORSTEN, L.C.A.; C.A. van EIJSBERGEN. Multiplicative effects in two-way analysis of variance. **Statistica Neerlandica**, New York, v. 26, p.61-68, 1972.

CROSSA, J. Statistical analysis of multilocation trials. **Advance in Agronomy**, New York, v.44, p.55-85, 1990.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2006. 585p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 3rd ed. Viçosa: UFV, 2004. v.1 480p.

DAVISON, A.C.; HINKLEY, D.V. **Bootstrap Methods and their Application** Cambridge:. Cambridge University Press,. 1997. With 1 IBM-PC floppy disk (3.5 inch; HD). v.1: Cambridge Series in Statistical and Probabilistic Mathematics,

De MENDIBURU, F. **agricolae**: Statistical Procedures for Agricultural Research. R package version 1.0-7. Vienna, Austria: The R Foundation for Statistical Computing, 2009.

DIACONIS, P.; EFRON, B. **Computer-intensive methods in statistics**. **Scientific American**, 1983. p.116–130.

DIAS, C.T.S.; W.J. KRZANOWSKI. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. **Crop Science**, Madison, v. 43, p.865-873, 2003.

DIAS, L.A.S. Análises multidimensionais. In: ALFENAS, A.C. (Ed.). **Eletroforese de isoenzimas e proteínas afins**: fundamentos e aplicações em plantas e microorganismos. Viçosa: UFV, 1998. p. 405-475.

DUARTE, J.B.; R.VENCOVSKY. **Interação genótipo × ambiente**: uma introdução à análise "AMMI". Ribeirão Preto, S.P, 1999. 60p.

EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v.6, n.1, p.36-40, 1966.

EFRON, B. Bootstrap methods: another look at jakknife. **Annals of Statistics**, Madison, v.7, n.1, p.1-26, 1979.

- EFRON, B.; TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. **Statistical Science**, Beachwood, v. 1, n. 1, p. 55–77, Feb. 1986.
- EFRON, B.; TIBSHIRANI, R. Statistical data analysis in the computer age. **Science, London**, p. 390–395, 1991.
- EFRON, B.; TIBSHIRANI, R.J. **An Introduction to the Bootstrap**. New York: Chapman and Hall, 1993. v.57: Monographs on Statistics and Applied Probability.
- EFRON, B.; TIBSHIRANI, R.J. **An introduction to the bootstrap**. London: Champman & Hall, 1993. 579p,
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Tecnologias de Produção de Soja - Região Central do Brasil**, Londrina, 2003. p.11-24.
- EVERITT, B.S. **Cluster analysis**. 3rd ed. London: Heinemann Educational Books, 1993. 122p.
- FARIA, A.P.; Moda-Cirino, V.; BURATTO, J. S.; SILVA, C. F. B.; DESTRO, D. Interação genótipo $\times$ ambiente na produtividade de grãos de linhagens e cultivares de feijão. **Acta Scientiarum: Agronomy**, Maringá, v.31, n.4, p. 579-585, dez. 2009.
- FOX, P.N.; CROSSA, J.; ROMAGOSA, I. Multi-environmental testing and genotype $\times$ environment interaction. In: KEMPTON, R.A.; FOX, P.N. (Ed.). **Statistical methods for plant variety evaluation**. London: Chapman & Hall, 1997. p.117-138.
- FURRER, R; NYCHKA, D.; SAIN, S. **fields**: Tools for Spatial Data. R package version 6.01, URL <http://CRAN.R-project.org/package=fields>. 2009.
- GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, Oxford, v.58, n.3, p.453-467, 1971.
- GABRIEL, K.R. Le biplot - outil d'exploration de données multidimensionnelles. **Journal de la Societe Francaise de Statistique**, v.143, n.3/4, p.5-55, 2002.
- GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v. 44, p. 705-715, 1988.
- GAUCH JÚNIOR, H.G. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam: Elsevier, 1992. 278p.
- GAUCH JÚNIOR, H.G.; ZOBEL, R.W. AMMI analysis of yield trials. In: KANG, M.S.; GAUCH JÚNIOR, H.G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. cap.4. p.85-122.

- GHADERI, A.; ADAMS, M.W.; NASSIB, A.M. Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. **Crop Science**, Madison, v.24, p. 37-42, 1984.
- GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, Greensboro, v.33, n.1, p.73-115, 1968.
- GONZÁLEZ MANTEIGA, W.; PRADA SÁNCHEZ, J.M.; ROMO URROZ, J.J. The Bootstrap: a review. **Computational Statistics**, Berlin, v. 9, n. 1, p. 165-205, 1994.
- HESTERBERG, T.; MOORE, D.S.; MONAGHAN, S.; CLIPSON, A.; EPSTEIN, R. Bootstrap methods and permutation tests. 5th ed. New York: W.H. Freeman, 2003. cap. 18: The practice of business statistics: using data for decisions.. p. 14-1;14-70.
- JOHNSON, D.E.; GRAYBILL, F.A. An analysis of a two-way model with interaction and no replication. **Journal of the American Statistical Association**, Alexandria, v.67, p.862-868, 1972.
- JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. New Jersey-USA: Englewood Cliffs, 1992. 642p.
- KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **Journal of Agricultural Science**, Cambridge, v.103, p.123-135, 1984.
- KHATTREE, R.; NAIK, D.N. **Multivariate Data Reduction and Discrimination with SAS Software**. Wiley Inter-Science, 2000. 574p.
- KRZANOWSKI, W.J. Cross-validation in principal component analysis. **Biometrics**, Arlington, v. 43, p. 575-584, 1987.
- LANDIM, P.M.B. **Geologia Quantitativa: Introdução à análise estatística de dados geológicos multivariados**. Rio Claro - SP, 2001. 1 CD-ROM.
- LAVORANTI, O.J. **Estabilidade e adaptabilidade fenotípica através da reamostragem “Bootstrap” no modelo AMMI**. 2003. 166p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.
- LAVORANTI, O.J. **Modelagem AMMI para estudos de interação em modelos estatísticos de efeitos fixos**. Colombo: Embrapa Florestas, 2004. 7p. (Comunicado Técnico, 124).
- MANDEL, J. A new analysis of variance model for nonadditive data. **Technometrics**, Washington, v. 13, n. 1, p. 1-18, 1971.
- MEYER, A.S. **Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes**. 2002. 106p (Dissertação de Mestrado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.

MOITA NETO, J. M.; MOITA, G.C. Uma Introdução à Análise Exploratória de Dados Multivariados. **Química Nova**, São Paulo, SP: v. 21, n. 4, p. 467-469, 1998.

MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **The Computer Journal**, Rhode Island, , v.20, n.4, p. 359- 363,1977.

NAVIDI, W.C. **Statistics for engineers and scientists**. Boston: McGraw-Hill, 2006. 869p.

OLIVEIRA, A.B.; DUARTE, J.B.; PINHEIRO, J.B. Emprego da análise AMMI na avaliação da estabilidade produtiva em soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.38, n.3, p.357- 364, 2003.

ORLÓCI, L. **Multivariate analysis in vegetational research**. 2nded. The Hague: Dr. W.Junk B.V. Publishers, 1978. 451p.

PEREIRA, H.S. Adaptabilidade e estabilidade de genótipos de feijoeiro-comum com grãos tipo carioca na Região Central do Brasil. **Pesquisa Agropecuária Brasileira**, [online]. v.44, n.1 [cited 2012-02-15], p. 29-37,2009.

PIEPHO, H.P. Robustness of statistical test for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trial. **Theoretical Applied of Genetics**, Berlin, v.90, p.438-443, 1995.

PINHEIRO, J.B. **Seleção para caracteres agronômicos, em diferentes épocas de cultivo, de populações de soja com resistência a insetos**. 1998. 143p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.1998.

PRADO, P.I.; LEWINSOHN, T.M.;Carmo, R.L.; HOGAN, D.J. ordenação multivariada na ecologia e seu uso em ciências ambientais. **Ambiente e Sociedade**, Campinas, SP: v.10, p. 69-83, 2002.

R version 2.7.1. **The R Foundation for Statistical Computing**. ISBN 3-900051-07-0. 2008.

RAM, J. ; PANWAR, D.V.S. Intraspecific divergence in rice. **The Indian Journal of Genetics & Plant Breeding**, New Delhi, v. 30, n. 1, p. 1-10, 1970.

RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J. **Genética quantitativa em plantas autógamas**: aplicações ao melhoramento do feijoeiro. Goiânia: Ed. da UFG, 1993. 271p.

RAO, C.R.**An advanced statistical method in biometric research**. New York: Ed. John Wiley, 1952. p.390.

ROCHA, M.M. **Seleção de linhagens experimentais de soja para adaptabilidade e estabilidade fenotípica**. 2002. 173p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2002.

ROMESBURG, H.C. **Cluster analysis for researchers**. Malabar, Florida; Krieger Publishing, 1990. 334p.

ROSSE, L.N.; VENCOVSKY, R. Modelo de regressão não-linear aplicado ao estudo da estabilidade fenotípica de genótipos de feijão no Estado do Paraná. **Bragantia**, Campinas, v.59, n.1, p.99-107, 2000.

SHULL, G.H. What is "heterosis"? **Genetics**, Bethesda, v.33, p. 439-446, 1948.

SILVA, R.M.; ROSSE, L.N.; MÔRO, J.R. Estabilidade e adaptabilidade de híbridos duplos experimentais de milho. **Scientia Agrária**, Curitiba, v.3, n. 1/2, p.61-68, 2002.

STONE, M. Cross-validatory choice and assessment of statistical predictions (with Discussion). **Journal of the Royal Statistical Society**, Oxford, Series B 36, p.111-148, 1974.

TIBSHIRANI, R. Correction to Discussion of: Jackknife, Bootstrap and other Resampling Methods in Regression Analysis, **Annals of Statistics**, Madison, v.16, n.1, p.479, 1988.

TOQUICA, S.P.; RODRÍGUEZ, F.; MARTINEZ, E.; DUQUE, M.C.; TOHME, J. Molecular characterization by AFLPs of Capsicum germplasm from the Amazon department in Colombia. **Genetic Resources and Crop Evolution**, Bogotá, Colombia, v.50, n.6, p.639-647, 2003.

VELICER, W.F. Determining the number of components from the matrix of partial correlations. **Psychometrika**, Greensboro, v.41, p.321-327, 1976.

WOLD, S. Pattern recognition by means of disjoint principal component models. **Pattern Recognition**, Umeå, Sweden, v. 8, p. 127-139, 1976.

WOLD, S. Cross-validatory estimation of the number of components in factor and principal component models. **Technometrics**, Alexandria, v.20, p.397-405, 1978.

YOCHMOWITZ, M.G., CORNELL, R.G. Stepwise tests for multiplicative Components of interaction. **Technometrics**, Alexandria, v.20, p.79-84, 1978.

ZOBEL, R.W.; WRIGHT, M.J.; GAUCH, H.G. Statistical analysis of a yield trial. **Agronomy Journal**, Madison, v.80, p.388-393, 1988.

WARD, J.H.; Hierarchical grouping to optimize an objective function. **Journal of. American Statistical Association**, Alexandria, v. 58, p. 236-244, 1963.

WU, C.F.J. Jackknife, Bootstrap and other Resampling Methods in Regression Analysis, **Annals of Statistics**, Madison, v.14, n.4, 1261-1350, 1986.

APENDICES


```

library(agricolae)

model<-AMMI(envir, variety, block, yield, MSE, xlim=c(-25,25),ylim=c(-25,25),
number=TRUE, graph="biplot");

matrizGE<-model$genXenv

matrizGE

pc<- princomp(model$genXenv, cor = FALSE)

pc$loadings

summary(pc)

model$biplot

##### FAZENDO GRÁFICO DA PC1 VERSUS MÉDIA DA PROD.GRÃOS
#####

detach(x)

bplot<-model$biplot[,1:4]

attach(bplot)

#startgraph

par(cex=0.8)

plot(yield,PC1,cex=0.0,text(yield,PC1,labels=row.names(bplot),col="blue"), main="AMMI
BIPLOT",frame=TRUE)

MEANS<-mean(yield)

abline(h=0,v= MEANS,lty=2,col="red")

amb<-subset(bplot,type=="ENV")

detach(bplot)

attach(amb)

s <- seq(length(yield))

arrows(MEANS, 0, 0.9*(yield[s]-MEANS)+MEANS, 0.9*PC1[s], col = "brown", code=2,
lwd=1.8, length=0.1)

#####

```



```

cova<-cov(model$genXenv)

values<-eigen(cova) # fornece os autovalores e autovetores de matrizGE= model$genXenv #
total<-sum(values$values) #fornece o somatório dos autovalores, ou seja, a variação total #
round(values$values*100/total,2)

##### Começando a reamostragem Bootstrap #####

Primeiramente, deve-se inserir a matrizGE obtida no comando AMMI realizado. Essa matriz
GE foi resultado de model$means na coluna Residual.

#####

require(fields)

matrizGE<-matrix(c( ... ), nrow= ... , ncol= ..., byrow = FALSE, dimnames = NULL)

matrizGE

as.matrix(matrizGE)

class(matrizGE)

Cov<-(matrizGE)%*%(t(matrizGE)) ##matriz covariancia (GE)*t(GE) ##

Cov

traco<-sum(diag(Cov)) ## traco da matriz Cov = soma de quadrados total dos residuos ##

traco

ng=nrow(matrizGE)

nb=2000 ##### Numero de reamostragens #####

rescol=matrix(0,1,ncol(matrizGE))

dist_euclid=matrix(0,1,ng)

Cova=matrix(0,1,ng)

tracol=matrix(0,1,1)

for (i in 1:nb)

{

  boot_GE=matrix(0,1,ncol(matrizGE))

```



```

for (j in 1:ng)
{
  auxlin=round(runif(1,1,ng))
  for (k in 1:ncol(matrizGE))
  {
    auxcol=round(runif(1,1,ncol(matrizGE)))
    rescol[1,k]=matrizGE[auxlin,auxcol]
  }
  boot_GE=rbind(boot_GE,rescol)
  Boot=boot_GE[2:nrow(boot_GE),]
}
dist_euclid2=rdist(Boot )
dist_euclid=rbind(dist_euclid,dist_euclid2)
Cov2=(Boot)%*%(t(Boot))
Cova=rbind(Cova,Cov2)
traco2=sum(diag(Cov2))
traco1=rbind(traco1,traco2)
}
dist_euclid=dist_euclid[2:(ng*nb+1),]
dist_euclid
traco1
soma=matrix(0,ng,ng)
for(i in 1:nb)
{
  soma=soma+dist_euclid[(1+((i-1)*ng)):(ng + (i-1)*ng),]
}

```



```

soma

dist_euclid_m=(1/nb)*soma

dist_euclid_m

#### Obtendo os intervalos interpercentilcos:####

percentis=c(0.10,0.90)

qrp=quantile(traco1, probs = percentis)

qrp

##### Análise de Agrupamentos #####

class(dist_euclid_m)

dist_Pri <- as.dist(dist_euclid_m)

dist_Pri

class(dist_Pri)

hc1 <- hclust(dist_Pri, "ward")

plot(hc1, hang = -1)

##### Obtendo a separação dos grupos no dendrograma segundo o valor k fornecido #####

informacion<-rect.hclust(hc1, k=..., border="red")

```


ANEXOS

ANEXO A - Identificação dos genótipos analisados no Experimento 1, população PCI, ano 2000

1	97-8001	12	97-8016	23	97-8027	34	97-8051
2	97-8002	13	97-8017	24	97-8029	35	97-8054
3	97-8004	14	97-8018	25	97-8031	36	97-8055
4	97-8006	15	97-8019	26	97-8033	37	97-8056
5	97-8007	16	97-8020	27	97-8034	38	97-8057
6	97-8009	17	97-8021	28	97-8035	39	97-8058
7	97-8010	18	97-8022	29	97-8037	40	97-8060
8	97-8011	19	97-8023	30	97-8039	41	IAC-100
9	97-8012	20	97-8024	31	97-8043	42	IAS-5
10	97-8014	21	97-8025	32	97-8044	43	OC-4
11	97-8015	22	97-8026	33	97-8050	44	PRIMAVERA

ANEXO B - Identificação dos genótipos analisados no Experimento 1, população PCI, ano 2001

1	97-8001	7	97-8014	13	97-8033	19	97-8056
2	97-8004	8	97-8016	14	97-8039	20	97-8057
3	97-8007	9	97-8019	15	97-8043	21	IAC-100
4	97-8009	10	97-8022	16	97-8051	22	IAS-5
5	97-8010	11	97-8024	17	97-8054	23	OC-4
6	97-8012	12	97-8026	18	97-8055		

ANEXO C - Identificação dos genótipos analisados no Experimento 2, população PRIS, ano 2000

1	97-8063	9	97-8073	17	97-8085	25	97-8096	33	97-8111	41	IAC-100
2	97-8064	10	97-8074	18	97-8086	26	97-8097	34	97-8112	42	IAS-5
3	97-8065	11	97-8076	19	97-8087	27	97-8100	35	97-8114	43	OC-4
4	97-8067	12	97-8078	20	97-8088	28	97-8101	36	97-8116	44	PRIMAVERA
5	97-8068	13	97-8080	21	97-8090	29	97-8105	37	97-8117		
6	97-8070	14	97-8081	22	97-8091	30	97-8108	38	97-8118		
7	97-8071	15	97-8083	23	97-8093	31	97-8109	39	97-8119		
8	97-8072	16	97-8084	24	97-8095	32	97-8110	40	97-8120		

ANEXO D - Identificação dos genótipos analisados no Experimento 2, população PRIS, ano 2001

1	97-8063	7	97-8080	13	97-8096	19	97-8118
2	97-8064	8	97-8085	14	97-8100	20	97-8119
3	97-8067	9	97-8088	15	97-8105	21	IAC-100
4	97-8068	10	97-8091	16	97-8108	22	IAS-5
5	97-8072	11	97-8093	17	97-8109	23	OC-4
6	97-8073	12	97-8095	18	97-8114		

ANEXO E - Identificação dos genótipos analisados no Experimento 3, população PRIM, ano 2000.

1	97-8121	9	97-8133	17	97-8144	25	97-8153	33	97-8165	41	IAC-100
2	97-8122	10	97-8134	18	97-8145	26	97-8154	34	97-8166	42	IAS-5
3	97-8123	11	97-8135	19	97-8146	27	97-8156	35	97-8167	43	OC-4
4	97-8124	12	97-8136	20	97-8147	28	97-8158	36	97-8172	44	PRIMAVERA
5	97-8125	13	97-8137	21	97-8148	29	97-8159	37	97-8173		
6	97-8129	14	97-8139	22	97-8149	30	97-8160	38	97-8175		
7	97-8130	15	97-8142	23	97-8150	31	97-8163	39	97-8176		
8	97-8132	16	97-8143	24	97-8151	32	97-8164	40	97-8178		

ANEXO F - Identificação dos genótipos analisados no Experimento 3, população PRIM, ano 2001

1	97-8121	5	97-8136	9	97-8147	13	97-8158	17	97-8173	21	IAC-100
2	97-8122	6	97-8137	10	97-8149	14	97-8159	18	97-8175	22	IAS-5
3	97-8132	7	97-8142	11	97-8151	15	97-8165	19	97-8176	23	OC-4
4	97-8135	8	97-8146	12	97-8154	16	97-8172	20	97-8178		