

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

**Comparação dos modelos AMMI e AMMI ponderado na análise  
da interação genótipo  $\times$  ambiente e interação QTL  $\times$  ambiente**

**Tatiana Oliveira Gonçalves de Assis**

Dissertação apresentada para obtenção do título de  
Mestra em Ciências. Área de concentração: Esta-  
tística e Experimentação Agronômica

**Piracicaba  
2015**

Tatiana Oliveira Gonçalves de Assis  
Bacharel em Matemática

**Comparação dos modelos AMMI e AMMI ponderado na análise da interação  
genótipo  $\times$  ambiente e interação QTL  $\times$  ambiente**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

Dissertação apresentada para obtenção do título de Mes-  
tra em Ciências. Área de concentração: Estatística e Ex-  
perimentação Agrônômica

**Piracicaba  
2015**

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

Assis, Tatiana Oliveira Gonçalves de

Comparação dos modelos AMMI e AMMI ponderado na análise da interação genótipo x ambiente e interação QTL x ambiente / Tatiana Oliveira Gonçalves de Assis. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

74 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Interação entre genótipo e ambiente 2. Interação entre QTL e ambiente 3. Modelo AMMI 4. Modelo W-AMMI 5. Quantitative trait locus I. Título

CDD 575.10212  
A848c

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

## Dedicatória

A Deus,

*"Tu és o meu Deus; graças te darei! Ó meu Deus, eu te exaltarei! Deem graças ao Senhor, porque ele é bom; o seu amor dura para sempre."(Salmos 118:28-29).*

*Ao meu amado esposo e amigo **Frédéric Assaag Gonzaga de Assis**, pelo apoio e incentivo em todas as etapas desse trabalho. Aos meus filhos **Lécio Oliveira Gonçalves de Assis** e **Lavinia Oliveira Gonçalves de Assis** pela compreensão e por tornar meus dias mais felizes. Aos meus pais **Lécio Gonçalves de Matos** e **Eliana Nazaret de Oliveira Gonçalves** pelo estímulo constante e todo amor que sempre me fez ir mais longe.*



## AGRADECIMENTOS

A Deus, autor e consumidor de minha fé, a esse Deus poderoso e digno de toda a honra, poder e majestade, pelos séculos dos séculos.

Ao professor Dr. Carlos Tadeu dos Santos Dias pela orientação, dedicação, ensinamentos e principalmente pela confiança a mim depositada. Ao professor Dr. Paulo Canas Rodrigues que, sempre prestativo, muito contribuiu para o desenvolvimento desse trabalho de pesquisa.

Aos professores e amigos mais chegados que irmãos, Lúcio Borges de Araújo e Mirian Fernandes Carvalho Araújo que foram mais que meus orientadores desde a graduação, foram agentes motivadores de todas as etapas da minha formação, os quais tem lugar especial em minha vida.

Ao professor e coordenador do curso de Matemática da Universidade Federal de Uberlândia Victor Gonzalo Lopes Neumann que sem sua competência e profissionalismo eu não teria chegado até aqui, minha admiração.

Aos professores do departamento de Ciências Exatas da ESALQ/USP que contribuíram com minha formação e desenvolvimento desse trabalho, com seu conhecimento e compromisso com o ensino de qualidade desse país.

A profa. Dra. Sonia Maria de Stefano Piedade e ao professor César Gonçalves de Lima pelo suporte e solicitude, além de suas atribuições como professores.

A professora Renata Alcarde Sermarini pelo seu exemplo de profissionalismo e carisma que alegrou muitos dos meus dias na ESALQ.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP Solange de Assis Paes Sabadin, Eduardo Bonilha, Jorge Alexandre Wiendl, Rosni Honofre Aparecido Pinto e Mayara Segatto pelo apoio permante. Em especial à secretária Luciane Brajão que além de todo apoio dentro de sua competência ainda me presenteou com sua amizade e companherismo em todos os momentos, sempre com um sorriso cativante renovando minha motivação.

Aos colegas do curso de Pós-Graduação em Estatística e Experimentação Agrônômica da ESALQ/USP e os colegas de outros departamentos da ESALQ/USP que tive o prazer de conviver durante esse período. Em especial agradeço o apoio da colega e amiga Natália Martins, que considero um exemplo de pessoa e profissional.

A CAPES pela concessão de bolsa de estudos, sem a qual seria inviável o

desenvolvimento de pesquisas como essa no Brasil.

A todos que de alguma maneira contribuíram para a elaboração desse trabalho e que também sempre me apoiaram.

*"É fundamental diminuir a distância entre o que se diz e o que se faz, de tal forma que, num dado momento, a tua fala seja a tua prática. "*

Paulo Freire





## SUMÁRIO

RESUMO . . . . .	11
ABSTRACT . . . . .	13
LISTA DE FIGURAS . . . . .	15
LISTA DE TABELAS . . . . .	17
1 INTRODUÇÃO . . . . .	19
2 REVISÃO BIBLIOGRÁFICA . . . . .	23
2.1 Interação entre Genótipo e Ambiente (IGA) . . . . .	23
2.2 Modelo de Efeitos Principais Aditivos e Interação Multiplicativa . . . . .	26
2.3 QTL Scans . . . . .	28
2.4 Análise AQ . . . . .	33
2.5 Análise AQ Ponderada . . . . .	33
2.6 Interação entre QTL e Ambiente . . . . .	33
3 MATERIAL E MÉTODOS . . . . .	37
3.1 Característica dos Dados . . . . .	37
3.2 Fundamentação Geométrica do Biplot . . . . .	37
3.3 Decomposição em Valores Singulares (DVS) . . . . .	38
3.4 Modelo AMMI e AMMI Ponderado . . . . .	39
4 RESULTADOS . . . . .	43
4.1 IGA: AMMI vs. AMMI Ponderado . . . . .	43
4.2 IQA: AMMI vs. AMMI Ponderado . . . . .	46
4.2.1 Análise da Interação entre Genótipo e Ambiente . . . . .	46
4.2.2 QTL scans e Análise da Interação QTL $\times$ Ambiente . . . . .	49
5 CONCLUSÕES . . . . .	57
REFERÊNCIAS . . . . .	59



## RESUMO

### **Comparação dos modelos AMMI e AMMI ponderado na análise da interação genótipo $\times$ ambiente e interação QTL $\times$ ambiente**

As características genéticas das culturas agrícolas podem ser influenciadas pelo ambiente, interferindo na produtividade. Sendo assim, esta pesquisa visa entender como ocorre a interação entre genótipo e ambiente (IGA) e a interação entre quantitative trait locus (QTL) e ambiente (IQA), a fim de fornecer instrumentos que possam melhorar a produtividade. Em destaque, vemos que o modelo de efeitos principais aditivos e interação multiplicativa (AMMI), que considera como aditivos os efeitos principais de genótipo e ambiente e como multiplicativos os efeitos da interação, é uma importante ferramenta que permite estudar a interação com dados de ensaios multi-ambientais e apresentar boas previsões na detecção de QTL para novos ambientes. Para se levar em consideração a heterogeneidade da variância do erro ao longo dos ambientes, uma generalização do modelo AMMI é utilizada, o modelo AMMI ponderado ou W-AMMI. Nesta pesquisa, observando os dados resultantes de um experimento com 9 genótipos de milho conduzidos em 20 ambientes e 4 blocos, foram comparadas as análises da IGA utilizando o modelo AMMI e W-AMMI. Com um segundo conjunto de dados, resultantes do cruzamento das variedades de cevada Harrington e TR306, com 141 genótipos conduzidos em 12 ambientes foram comparados os resultados das análises da IGA e IQA utilizando os modelos AMMI e W-AMMI, sendo que foram propostas ponderações por linha (todos os ambientes para determinado genótipo ficam com o mesmo peso) e coluna (todos os genótipos para determinado ambiente ficam com o mesmo peso).

Palavras chave: Interação entre genótipo e ambiente; Interação entre QTL e ambiente; Modelo AMMI; Modelo W-AMMI; Quantitative Trait Locus



## ABSTRACT

### **Comparison between the AMMI and weighted AMMI models to analyze genotype-by-environment interaction and QTL-by-environment interaction**

The genetic characteristics of crops can be influenced by the environment, interfering with productivity. This research intends to understand the genotype-by-environment interaction (IGA) and quantitative trait locus (QTL)-by-environment interaction (IQA), in order to provide tools that can improve productivity. That the Additive Main Effects and Multiplicative Interactions (AMMI) model, has been widely used to study and understand these interactions and has shown to provide good interpretations of both IGA and IQA, as well as QTL detection. In order to take into account of the heterogeneity of error variance over the environments, a generalization of the AMMI model is used, the weighted AMMI or W-AMMI. In this study, observing the data resulting from an experiment with 9 maize genotypes conducted in 20 environments and 4 blocks, was used to compare the results between the AMMI and W-AMMI models. A second set of data, resulting a cross between the barley varieties Harrington and TR306, with 141 genotypes conducted in 12 environments, was used to compare the AMMI and two versions of the W-AMMI (equal weights per row; and equal weights per column) models in terms of IGA, IQA and QTL detection.

Keywords: Genotype-by-environment interaction; QTL-by-environment interaction; AMMI model; W-AMMI model; Quantitative Trait Locus



## LISTA DE FIGURAS

Figura 1 - Biplot dos modelos AMMI2 e W-AMMI2 para dados de produção de milho	44
Figura 2 - Biplot dos modelos AMMI2 e W-AMMI2 (ponderação por linhas) para dados de cevada Harrington/TR306 . . . . .	48
Figura 3 - Biplot dos modelos AMMI2 e W-AMMI2 (ponderação por colunas) para dados de cevada Harrington/TR306 . . . . .	48
Figura 4 - Mapa genético dos dados de cevada Harrington/TR306 . . . . .	49
Figura 5 - QTL scan para os primeiros 16 ambientes dos dados de rendimento para população Harrington/TR306 . . . . .	50
Figura 6 - QTL scan para os últimos 9 ambientes dos dados de rendimento para população Harrington/TR306 . . . . .	51





## LISTA DE TABELAS

Tabela 1 - Análise de variância para o modelo completo AMMI dos dados referentes a experimentos com genótipos de milho . . . . .	43
Tabela 2 - Análise de variância para o modelo completo AMMI dos dados referentes ao cruzamento Harrington/TR306 . . . . .	46
Tabela 3 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 nos cinco primeiros ambientes	51
Tabela 4 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do sexto ao nono ambiente .	52
Tabela 5 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo ao décimo terceiro ambiente . . . . .	53
Tabela 6 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo quarto ao décimo oitavo ambiente . . . . .	54
Tabela 7 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo nono ao vigésimo segundo ambiente . . . . .	54
Tabela 8 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 nos três últimos ambientes .	55
Tabela 9 - Variância do erro para cada um dos 20 ambientes referentes ao experimento com 9 genótipos de milho . . . . .	67



## 1 INTRODUÇÃO

As interações entre genótipo e ambiente (IGA) e entre QTL e ambiente (IQA) são fenômenos comuns em ensaios multi-ambientais (MET's) e representam um grande desafio para melhoristas que pretendem desenvolver genótipos mais adaptados a diferentes condições dos ambientes. Visto que as características genéticas das culturas agrícolas podem ser influenciadas pelo ambiente, interferindo na produtividade, é importante a pesquisa buscar o aprimoramento das ferramentas capazes de analisar a IGA e IQA do cultivar.

Em destaque, vemos que o modelo de efeitos principais aditivos e interação multiplicativa (AMMI) é uma importante ferramenta utilizada por pesquisadores especializados em melhoramento. Ele aplica a decomposição em valores singulares (DVS) para o resíduo da análise de variância (ANOVA).

Quando IGA nos dados é alta, é de se esperar heterogeneidade de variância do erro ao longo dos ambientes e este modelo não está preparado para ter em conta esta heterogeneidade. Neste caso é indicado uma generalização do modelo AMMI capaz de levar em conta a heterogeneidade de variância do erro usando uma decomposição em valores singulares (DVS) de posto inferior ponderado, o AMMI ponderado (W-AMMI, RODRIGUES et al., 2014). Embora esta generalização ocorra em um modelo fixo, o W-AMMI oferece uma razoável aproximação ao modelo misto para a IGA, o que é considerado mais apropriado nestes casos (RODRIGUES et al., 2014).

No seguimento de Gauch et al. (2011), o uso da metodologia AMMI é aqui utilizada para estudar a IQA, visando identificar locos potencialmente envolvidos no controle da interação, ou seja, identificar genótipos adaptados especificamente a cada ambiente. Este trabalho tem importante aplicação no uso de dados de marcadores moleculares para seleção assistida. Muitos outros trabalhos encontrados na literatura relatam a localização de QTL's envolvidos no controle dos componentes da interação, estando estes envolvidos com fatores como disponibilidade de água, altitude, temperatura e outros.

Visando fornecer previsões bem sucedidas de testes de QTL para novos ambientes, a estratégia proposta é aplicar o modelo AMMI em dados fenotípicos para obter precisão e conseqüentemente aumentar a probabilidade logarítmica (LOD) das medidas dos QTL detectados. Sendo os ambientes ordenados por parâmetros do modelo AMMI, o qual assume informações da interação genótipo  $\times$  ambiente, são revelados padrões consistentes e tendências sistemáticas que frequentemente tem interpretações ecológicas ou

biológicas evidentes. A combinação de uma maior precisão para dados fenotípicos e tendências sistemáticas para os ambientes prevê resultados mais consistentes e compreensíveis de QTL.

Neste trabalho serão identificados entre 9 genótipos de milho, conduzidos em 20 ambientes com 4 blocos, os que apresentam maior adaptabilidade e estabilidade de rendimento em diferentes condições ambientais em regiões para qual são recomendados, com a utilização do modelo AMMI e, em busca de resultados mais precisos, o uso do modelo W-AMMI, verificando assim qual o impacto da heterogeneidade da variância do erro ao longo dos ambientes nos resultados do modelo AMMI. Como proposta de ponderação dos dados, a contribuição de um dado ambiente ao ajuste do modelo W-AMMI será proporcional ao inverso da variância do erro e neste caso os blocos serão considerados repetições na parte multiplicativa do modelo.

Na segunda etapa deste trabalho serão utilizados os dados observados resultantes do cruzamento das variedades de cevada Harrington e TR306, com 141 genótipos conduzidos em 25 ambientes e serão analisadas as informações de 140 fenótipos e 127 marcadores para obter os resultados com QTL's. Com esse conjunto de dados serão feitas além das análises da IGA, as análises da IQA. Será apresentada uma proposta de ponderação dos dados, considerando os pesos por linha e por coluna e posteriormente observado o impacto dessas ponderações na análise dos resultados e assim poder definir qual foi a melhor proposta e finalmente fazer a varredura de QTL's.

Entre os objetivos desta dissertação estão:

- 1.) fazer o levantamento dos avanços acerca das pesquisas sobre o modelo AMMI na verificação e entendimento da IGA e IQA de dados agrônômicos;
- 2.) identificar entre os genótipos de milho os que apresentam maior adaptabilidade e estabilidade de rendimento em diferentes condições de ambiente em regiões para as quais são recomendados, com a utilização do modelo AMMI e do modelo W-AMMI visando verificar qual o impacto da heterogeneidade da variância do erro ao longo dos ambientes nos resultados do modelo AMMI e posteriormente apresentar os resultados da análise da IGA.
- 3.) em um conjunto de dados resultantes das variedades de cevada Harrington e TR306 pontuar os genótipos que apresentam maior adaptabilidade e estabilidade de rendimento em diferentes condições de ambiente em regiões para as quais são recomendados e apresentar os resultados da IGA e IQA.

4.) comparar os resultados obtidos com o AMMI e W-AMMI, verificando qual a melhor proposta de ponderação dos dados (por linha ou por coluna) e qual o mais indicado para a varredura de QTL's e posterior análise da IQA, observando que será proposta uma ponderação para o caso de dados sem repetição.

O modelo AMMI será usado na detecção de QTL e no estudo da interação entre QTL e ambiente, com o uso do software R/qtl. Nomeadamente, aplicaremos a análise AQ (obtenção dos QTL scans tendo por base os valores preditos pelo ajustamento do modelo AMMI aos dados fenotípicos) que permite obter picos mais elevados (em termos de LOD scores) para os QTL nos dados observados.



## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Interação entre Genótipo e Ambiente (IGA)

O efeito da IGA visa verificar, estatisticamente, o efeito do genótipo na presença dos ambientes ou o efeito do ambiente na presença dos genótipos. Quando se verifica de forma significativa evidências da interação do genótipo com o ambiente, pode-se concluir que há um efeito diferenciado do genótipo na presença de pelo menos um ambiente ou vice-versa. Neste caso, o procedimento adequado para interpretar o efeito da IGA é por meio de sua decomposição, que consiste em estudar o efeito de cada fator na presença de cada nível do outro fator.

Para Chaves (2001) a IGA não deve ser vista somente como um simples efeito estatístico e sim como um fenômeno biológico em suas aplicações no melhoramento de plantas e é resultante da resposta diferencial dos genótipos na variação ambiental.

A IGA é uma fonte de variação fenotípica, que na maioria dos casos é inseparável da variância ambiental, segundo afirma Falconer (1987). Para Crossa (1990), na prática, para verificar a significância da IGA, é necessário ter repetição do experimento, pois poderia ocorrer uma superestimação dos ganhos genéticos caso o experimento fosse realizado em um único ambiente.

Segundo Allard e Bradshaw (1964) as variações ambientais que contribuem para a interação com genótipos podem ser classificadas como previsíveis e não previsíveis. Sendo o primeiro grupo os que incluem todas as características do clima cuja variação é previsível como local, solo, comprimento do dia, intensidade dos raios solares e as determinadas pela ação do homem, tais como época e densidade de semeadura, níveis de adubação e outros tratamentos às culturas.

No segundo grupo, Allard e Bradshaw (1964) corresponde as variações ambientais imprevisíveis como sendo aquelas referentes as atuações climáticas tais como quantidade e distribuição de chuvas, variações da temperatura e diversos outros fatores que não podem ser controlados pelo homem.

No entanto, pode-se dizer que é com instrumentos estatísticos juntamente com conhecimento do fenômeno ao nível biológico que se pode tirar informações mais precisas para entender a complexidade da interação. Por exemplo, uma suposta população que ocupa uma área geográfica com grandes variações ambientais estará sujeita a pressões de seleção diferentes ao longo de sua extensão territorial. Ocorrendo variabilidade de



adaptação dos indivíduos, serão beneficiados aqueles mais adaptados na condição ambiental particular. Esta adaptação específica a determinados ambientes é considerada a causa evolutiva básica da IGA.

Pelas vastas publicações de pesquisas relacionadas ao assunto, pode-se sugerir que existe uma concordância geral entre os melhoristas de que IGA têm um importante significado para a obtenção de variedades superiores. Entretanto, é muito difícil encontrar concordâncias sobre o que se deve conhecer em relação a IGA e como utilizá-la (ALLARD, 1971).

Os melhoristas encontram dificuldade na identificação de genótipos superiores, tanto no processo de seleção quanto no processo de recomendação de cultivares quando há existência evidente de IGA, pois essa interação indica que o comportamento dos genótipos nos experimentos depende principalmente das condições ambientais a que são submetidos. Desta forma, a resposta obtida de um genótipo em comparação a outro é variável, sendo que essas variações se apresentam devido a mudança de ambientes (OLIVEIRA; DUARTE; PINHEIRO, 2003; KANG, 1998).

Vale ressaltar que a IGA deve ser vista como um fenômeno biológico natural, que deve ser bem conhecido para melhor aproveitá-lo no processo de seleção e não como um problema ou um fator indesejável (CHAVES, 2001). Logo os genótipos que interagem positivamente com os ambientes podem fazer a diferença na classificação do cultivar como bom ou ótimo (DUARTE; VENKOVSKY, 1999).

A interação pode ser de considerada de natureza simples ou complexa, sendo que a natureza simples ou complexa da IGA deve ser considerada e não somente a verificação de sua existência (VENKOVSKY; BARRIGA, 1992). A interação de natureza simples indica a presença de genótipos adaptados em um grande número de ambientes, sendo possível fazer uma recomendação generalizada de cultivares. A interação de natureza complexa mostra que existem genótipos adaptados a apenas alguns ambientes, o que traz uma complicação ao pesquisador, quando se trata da recomendação de cultivar.

Para González (1988), a interação entre genótipos e ambientes pode interferir negativamente ou positivamente nos resultados finais, sendo assim, a influência que o ambiente exerce sobre cada genótipo merece atenção especial por parte dos melhoristas no processo de seleção genotípica.

A referência mais antiga, com enfoque estatístico, da IGA é a apresentada por Fisher e Mackenzie em 1923, segundo publicado por Freeman em 1973. A partir desta

data muitos outros trabalhos têm sido feitos para análises estatísticas da IGA.

Existindo interação, os métodos mais simples que utilizam componentes de variância para a interpretação dos resultados incluem estudos de regressão, métodos baseados em análises modificadas e métodos envolvendo variáveis ambientais externas.

A regressão foi utilizada pela primeira vez por Yates e Cochran (1938) em análise de grupos de experimentos. Neste caso, o grau de associação entre diferenças varietais pode ser verificado pelo cálculo da regressão dos rendimentos das variedades isoladas sobre os rendimentos médios de todas as variedades, mostrando assim que a regressão explica grande parte da interação em uma série de experimentos. O método de regressão foi também usado por Perkins e Jinks (1968) para estimação de parâmetros em um modelo genético aplicado à biometria.

A ANOVA também é um método utilizada no processo de investigação da interação sendo que neste caso os dados são organizados em uma tabela de dupla entrada. Esta análise envolve vários experimentos e por isso é possível determinar a magnitude da interação por meio da razão do quadrado médio da interação ( $QM_{G \times A}$ ) pelo quadrado médio do resíduo ( $QM_{Res}$ ). Porém estudos de detalhamento deste componente de variação são geralmente necessários pois a detecção de significância para a interação não esclarece as implicações que estas possam ter sobre o melhoramento.

Em suma, a importância do estudo da IGA é devido a sua capacidade de possibilitar a identificação de cultivares mais adaptados a determinadas regiões, onde as mesmas poderão expressar o seu potencial genético. Assim, estudos sobre a magnitude de tais interações podem ser úteis na regionalização de cultivares, ou seja, direcionando os cultivares a áreas onde as possam expressar o máximo que as condições ambientais particulares permitam, com respeito a respostas de genótipos, e possibilitar a exploração de efeitos específicos de adaptação para determinadas regiões.

Os fenômenos de adaptabilidade e estabilidade de cada genótipo estão relacionados, porém não são iguais uma vez que a adaptabilidade refere-se a capacidade de os genótipos aproveitarem vantajosamente o estímulo do ambiente e a estabilidade diz respeito a capacidade de os genótipos mostrarem comportamento altamente previsível em razão do estímulo do ambiente.

Em 1986 Lin, Binns e Lefkovicth classificaram conceito de estabilidade em três categoria. Na primeira o genótipo é considerado estável se sua variância entre os ambientes for pequena. Ela pode ser medida pela variância de cada genótipo nos diferentes

ambientes e é útil para características que devem ser mantidas, tal como resistência a patógenos e pragas (NUNES, 2000).

Na segunda categoria o genótipo é considerado estável se sua resposta ao ambiente for paralela à resposta média de todos os materiais avaliados no experimento ocorrendo somente quando ocorre quando o genótipo possui interações mínimas com o ambiente. Já na terceira categoria o genótipo é considerado estável se o quadrado médio dos desvios de regressão for pequeno, sendo então classificado como de alta confiabilidade de resposta.

## 2.2 Modelo de Efeitos Principais Aditivos e Interação Multiplicativa

O modelo de efeitos principais aditivos e interação multiplicativa (AMMI) é uma ferramenta largamente utilizada na análise de ensaios multi-ambientais que permite particionar a interação em  $N = \min(I-1, J-1)$  termos, em que  $I$  é o número de genótipos e  $J$  o número de ambientes. O modelo AMMI combina a análise de variância (ANOVA) e a decomposição em valores singulares (DVS), sendo que a ANOVA é aplicada primeiramente e permite extrair os efeitos principais aditivos, e a DVS é aplicada aos resíduos da ANOVA, ou seja, na matriz da IGA, de forma a decompor a interação em  $N$  partes (GAUCH, 1988, 1992).

Para Zobel et al. (1988), o modelo AMMI traz quatro pontos importantes: permite uma análise mais detalhada da IGA, garante a seleção de genótipos mais produtivos, propicia estimativas mais precisas das respostas genotípicas e possibilita fácil interpretação gráfica dos resultados da análise estatística por meio do Biplot.

Foi introduzido por Gollob (1968) e Mandel (1969,1971) um modelo (FANOVA) com fator analítico para estudar a interação em tabelas de dupla entradas completas. Zobel; Wright e Gauch Júnior (1988), Gauch (1988); Gauch e Zobel (1988) renomearam o modelo FANOVA como “Additive Main Effects and Multiplicative Interactions (AMMI) model”.

Tanto a Análise de Variância (ANOVA) quanto a Análise de Componentes Principais (ACP), que estão relacionados diretamente com a análise AMMI, buscam resumir os dados contidos em uma matriz de interação genótipo  $\times$  ambiente em um modelo. Do ponto de vista conceitual, diferença apresentada entre elas é que 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 somente depois são somados a média geral

(GAUCH, 1992).

A ACP foi criada por Pearson em 1901, enquanto que a ANOVA foi proposta por Fisher em 1918. Vale ressaltar que a ANOVA se tornou mais popular devido a simplicidade dos cálculos e a fácil interpretação do modelo aditivo, apesar da ACP ser mais antiga e ser considerada mais eficiente. Sendo assim, a ACP foi pouco utilizada pois a exigência de grandes recursos computacionais não foram supridos na época em que a teoria foi proposta e também pela dificuldade de interpretação do modelo multiplicativo.

Artigos utilizando metodologia semelhante ao modelo AMMI, de acordo com Gauch (1992), foram publicados por Williams (1952) e Pike e Silverberg (1952), sob a forma de soma de funções mais o produto de funções, onde essas funções podem ser substituídas por números. Estes autores utilizaram a ACP apenas na matriz de resíduos de IGA resultante da ANOVA.

O primeiro artigo foi publicado em um jornal estatístico e analisava uma experiência de química, já o segundo foi publicado em um jornal de engenharia e interessava-se pelo uso de computadores analógicos. Trabalhos envolvendo o modelo AMMI permaneceram parados até 1968 sendo retomado por Gollob. Somente após a disponibilidade de computadores digitais foi consumado o interesse ativo em modelos AMMI.

Em 1971 Gabriel apresentou o conceito de Biplot como uma representação gráfica que dispõe ambas as entradas de conjunto de dados em uma tabela de dupla entrada. Ele permite a visualização dos dados como a inter-relação entre as entradas (por exemplo os genótipos), a inter-relação entre os testadores (por exemplo os ambientes) e a inter-relação entre as entradas e os testadores.

Em 1984 a metodologia foi utilizada por Kempton em dados de produtividade enfatizando a possibilidade de utilizar a técnica na interpretação dos dados. Kempton apresenta a utilização e interpretação de Biplots para os modelos de ANOVA, RL (regressão linear), ACP, e AMMI e considera o Biplot uma interessante ferramenta para a apresentação dos resultados das diversas técnicas, com especial interesse para a análise da IGA.

Gauch e Zobel (1996) disponibilizaram uma revisão da metodologia AMMI no período de 1992 a 1996, na qual afirmam que esta metodologia é 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.

Zobel, Wright e Gauch (1988) compararam as metodologias ANOVA, ACP, Regressão Linear (RL) com o modelo AMMI, identificando assim algumas falhas nos métodos. A ANOVA falhou em detectar a significância da IGA enquanto que a ACP falhou em identificar e separar os componentes principais de genótipos e de ambientes e a RL em reter apenas uma pequena parcela da soma de quadrados da IGA, misturando o efeito da IGA com os efeitos principais. Por outro lado, a análise AMMI mostrou uma alta e significativa IGA com sentido agrônômico, permitindo assim separar os ambientes quanto à latitude e os genótipos quanto à precocidade.

Em 1997 Annicchiarico analisa quatro grupos de dados de diferentes culturas, comparando dois procedimentos baseados em validação cruzada para testar os eixos dos componentes principais da IGA. O procedimento de validação cruzada consiste em dividir o conjunto de dados em duas partes de forma aleatória, sendo uma parte destinada a ajustar o modelo e a outra a validá-lo.

Desta forma o modelo AMMI é obtido com dados destinados a ajustar o modelo e a soma de quadrados residual (RMSPD - "root mean squared predictive difference") é determinada nos dados destinados à validação. Para experimentos realizados em vários ambientes, os ajustes devem ser feitos para as repetições dentro dos ambientes, sendo que o melhor modelo será aquele que apresentar menor RMSPD (DIAS; KRZANOWSKI, 2003).

A abordagem AMMI não busca recuperar toda a soma de quadrados da interação genótipo  $\times$  ambiente ( $SQ_{G \times A}$ ), mas apenas uma parcela que melhor a representa, determinada por genótipos e ambientes. Essa parcela é formada por eixos que captam, sucessivamente, porções cada vez menores da variação presente na matriz de IGA. Por isso, o método AMMI é considerado um modelo capaz de separar a resposta padrão e ruído da  $SQ_{G \times A}$  (WEBER; WRICKE; WESTERMANN, 1996).

### 2.3 QTL Scans

Quantitative trait loci designados por QTL's são regiões do genoma responsáveis pela expressão de caracteres fenotípicos, que possuem distribuição contínua, tais como, altura e peso de plantas e de animais, produção de grãos, teor de óleo etc. Assim, QTL's podem ser definidos como localizações genéticas que contribuem para a variação de uma característica quantitativa.

Com o advento dos marcadores moleculares (LANDER; BOTSTEIN, 1989) tornou-se possível mapear regiões cromossômicas (QTL's) que afetam esses caracteres quantitativos. Desta forma, o conhecimento sobre o número, localização, efeito e identidade de localizações genéticas (QTL) pode ser usado para auxiliar na seleção e melhoramento de produção de cultivares (BROMAN; SEN, 2009).

O mapeamento de QTL é uma técnica que busca identificar QTL numa população experimental resultante do cruzamento de dois progenitores. Mapear um QTL significa identificar sua posição no genoma e estimar seus efeitos genéticos, tais como: o efeito aditivo, efeito de dominância e outros efeitos presentes no modelo adotado.

Para realizar o mapeamento de QTL algumas informações são indispensáveis como a discriminação do caractere quantitativo de interesse e os dados de marcadores moleculares os quais são obtidos em laboratórios especializados. Uma vez que esses dados estejam disponíveis é necessário distribuir os marcadores utilizados em grupos, chamados grupos de ligação, ordená-los no grupo e dentro de cada um estimar a distância entre os marcadores adjacentes ou marcadores flanqueadores, tem-se então o mapa genético.

Tendo o mapa em mão pode-se dar continuidade ao mapeamento de QTL's. Vale destacar que o princípio básico do mapeamento é que os marcadores moleculares devem estar ligados aos locos que controlam a característica de interesse. Isto porque é de suma importância que na segregação esses marcadores sejam passados de uma geração para outra junto com o gene que afeta a característica selecionada (ARÚS; MORENO-GONZÁLES, 1993).

A identificação, mapeamento e quantificação dos efeitos de seus locos é permitido devido aos estudos desse locos. Porém, a detecção eficiente de QTL's depende de vários fatores, como do seu número, magnitude de seus efeitos, características de herdabilidade, interações entre genes, frequência de recombinação entre os QTL's e os tipos de marcadores e grau de saturação do mapa (TANKSLEY, 1993; YOUNG, 1996; LIU, 1998).

Existem, na literatura, vários métodos estatísticos para o mapeamento de QTL's. A maioria desses métodos envolvem análises de marcadores individuais, utilizando-se teste t simples, regressão linear, análise da variância, razão de verossimilhança e estimação da máxima verossimilhança, mapeamento por intervalo usando novamente a abordagem de verossimilhança, regressão e uma combinação das abordagens de verossimilhança e regressão (LIU, 1998) e por fim o mapeamento por intervalo composto (CIM).

O CIM é uma combinação de mapeamento de intervalo simples e regressão

linear múltipla (ZENG, 1993, 1994). Neste trabalho, após o uso do modelo AMMI na análise da IGA, a abordagem utilizada na detecção de QTL será o mapeamento por intervalo composto.

A ideia básica do mapeamento de QTL consiste em associar genótipo e fenótipo em uma população apresentando variação genética. Uma opção é fazer o mapeamento em experimentos cruzados nos quais se tem um maior controle sobre a composição genética de uma população, permitindo assim ampliar o efeito genético de um suposto QTL por escolher criteriosamente as linhagens para cruzamento. Cruzando duas ou mais linhagens, também efetivamente aleatorizamos a variação genética em uma população progenitora, podendo então concluir que as associações genéticas detectadas em um cruzamento são causais.

A estrutura genética simples de experimentos cruzados fornece uma estrutura relativamente fácil para estudar o conceito e princípios estatísticos de mapeamento genético. Conceitualmente, o estudo mais simples seria em uma população natural dos organismos de interesse. Por exemplo, para compreender a hipertensão, podemos estudar associações genéticas com hipertensão em um grande grupo.

Embora extremamente útil, a abordagem de estudar essas associações genéticas apresenta três problemas que se destacam: o alto custo relacionados a estudos com pessoas; possíveis ruídos apresentados na caracterização fenotípica (devido à impossibilidade de controlar o meio ambiente e as história de vida das pessoas); por último as associações não necessariamente implicam a causa devido à possibilidade de distorções dada à estrutura da população (BROMAN; SEN, 2009).

Mapeamento de QTL's em experimentos cruzados fornece uma excelente alternativa ao escolher adequadamente um modelo de organismo o qual seja possível isolar em um aspecto particular o fenótipo de interesse. Por exemplo, estudando a hipertensão em animais que são usualmente analisados em laboratório e prosseguindo para a ligação evolutiva entre os seres humanos e este determinado animal.

Desta maneira é possível um controle maior sobre a precisão da fenotipagem, meio ambiente e história de vida do que em populações naturais. Para isso, podemos citar como exemplo a alimentação de todos os animais envolvidos no estudo com a mesma dieta e a permanência destes em locais com clima controlado e desta forma manter idênticas as condições experimentais. Seguindo esses procedimentos podemos realizar seguramente o processo de fenotipagem o qual pode ser invasivo, impraticável ou até mesmo antiético

nos seres humanos.

Além disso, temos também um maior controle sobre a composição genética de uma população em experimentos cruzados. Isso nos permite ampliar o efeito genético de um suposto QTL por escolher criteriosamente as linhagens para cruzamento. Cruzando duas ou mais linhagens, também efetivamente aleatorizamos a variação genética em uma população progenitora.

Isto nos permite concluir que as associações genéticas detectadas em um cruzamento são causais, isto porque em experimentos cruzados segregados a variação genética é mais simples em relação a uma população natural, podendo assim realizar mais modelagem estatística complexa para identificar epistasia (QTL  $\times$  interações QTL) e QTL  $\times$  interações ambientais.

A região que se estendeu por um QTL pode conter dezenas e até milhares de genes, por este motivo não podemos afirmar que a identificação de um QTL nos ajuda certamente a identificar um gene. Um QTL pode não ser replicado com sucesso numa linhagem congênita (idêntico a uma deformação em toda a região, mas o QTL derivado de uma linhagem diferente) por causa da interação epistática de fundo genético (BROMAN; SEN, 2009). Desenvolver soluções estatísticas e experimentais para estas deficiências é uma área de pesquisa importante neste contexto e que está em constante progresso.

Vale ressaltar que o mapeamento de QTL's além de determinar a posição destes no genoma, também quantifica e caracteriza seus efeitos. Sendo conhecido os genótipos desses locos controladores de caracteres quantitativos, a alternativa é estimar os efeitos de QTL's fixados em posições intermediárias de marcadores moleculares genotipados que não têm efeito sobre o caráter em estudo, utilizando a informação do mapa desses marcadores.

Portanto, o procedimento de mapeamentos de QTL's consiste em avaliar os genótipos dos indivíduos da população de interesse. Se houver diferença entre as médias fenotípicas do caráter para cada uma das classes genotípicas de um dado marcador, pode-se inferir que existe um QTL ligado aquela marca (LYNCH; WALSH, 1998).

Os pré-requisitos para o mapeamento de QTL's são o delineamento genético adequado (populações do tipo  $F_2$ , de retrocruzamentos, linhagens recombinantes, dentre outras), genotipagem, mapa genético e dados dos caracteres de interesse (TOLEDO, 2006).

O tipo de delineamento genético considerado é um fator muito importante para a precisão e qualidade do mapeamento de QTL's. As populações derivadas de cru-



zamentos controlados são interessantes devido à flexibilidade na escolha dos pais na determinação do esquema de cruzamento, pois além da variabilidade encontrada nos genes controladores de um caráter e do ambiente, as variações genéticas do caráter entre os pais também devem ser consideradas.

Geralmente são usados os delineamentos de Populações  $F_2$ , de retrocruzamentos e linhagens recombinantes, sendo um dos mais utilizados na construção de mapas genéticos vegetais é representado pela população de indivíduos da geração  $F_2$ , resultante do cruzamento entre duas linhagens homozigotas contrastantes para os caracteres de interesse.

A principal vantagem nesse tipo de delineamento está na maximização da variabilidade genotípica, pois são obtidos três genótipos para cada loco marcador. Isso permite estimar o grau médio de dominância associado ao QTL detectado, cuja estimativa está diretamente associada a uma escolha eficiente nos métodos de melhoramento a serem adotados na cultura. Além disso, garante o equilíbrio de ligação, fator indispensável para o mapeamento de QTL's.

O mapeamento genético de organismos superiores é possível porque o genoma é organizado e transmitido como unidades lineares denominadas cromossomos. Os mapas genéticos apresentam a ordem dos locos ao longo do cromossomo e a distância relativa entre eles. Como citado anteriormente, estes mapas são essenciais para a localização de QTL's. Ressaltando, para a construção do mapa genético devem ser estabelecidas a posição dos marcadores e a distância entre eles (LYNCH; WALSH, 1998). Os mapas são construídos com base em análises de segregação dos marcadores em populações que apresentam desequilíbrio de ligação.

Não é padronizada universalmente a relação entre distância de mapa e a distância física entre locos. Essa relação dentro de um mesmo cromossomo pode sofrer uma diferença marcante. A taxa de recombinação está sob controle genético e alguns genes influenciam essa taxa ao longo do todo o genoma e outros em regiões específicas do cromossomo.

Também não é definido um número necessário de marcadores para construir um mapa genético, pois este depende do tamanho do genoma, do número de cromossomos e da frequência de recombinação. Quando o número de grupos de ligação obtidos pela análise dos marcadores for igual ao número de cromossomos gaméticos do organismo e todos os marcadores genéticos mapeados estiverem ligados, o mapa pode ser finalmente

considerado como satisfatório.

## 2.4 Análise AQ

O modelo AMMI é usado na detecção de QTL e no estudo da IQA. Primeiramente é aplicado um modelo AMMI parcimonioso aos dados fenotípicos e posteriormente os valores preditos com esse modelo AMMI são usados para obter os QTL scans, permitindo então incluir informação de outros ambientes no scan de um determinado ambiente (JIANG e ZENG, 1995).

Esta técnica é denominada de análise AQ (GAUCH et al., 2011) e permite obter picos mais elevados (em termos de LOD scores) para os QTL quando efetuado um QTL scan. Ademais, a ordenação dos ambientes por parâmetros AMMI que sumarizem a IGA revela padrões consistentes e tendências sistemáticas, que muitas vezes têm interpretação ecológica ou biológica relevantes (GAUCH et al., 2011).

## 2.5 Análise AQ Ponderada

Gauch et al. (2011) sugerem uma nova abordagem para a detecção e compreensão da IGA, a análise AQ, onde os QTL scans são feitos com base em previsões AMMI em vez de QTL scans significativos na IGA.

A análise AQ ponderada, ou WAQ, consiste na realização de exames de QTL no modelo W-AMMI para valores previstos para cada ambiente separadamente. Esta abordagem pode, potencialmente, melhorar a detecção de QTL's. Os ambientes podem ser ordenados por parâmetros AMMI e W-AMMI que resumem a IGA e a IQA com informações para revelar padrões consistentes e tendências sistemáticas que muitas vezes podem ser explicados em termos de condições ambientais (GAUCH, 1992; GAUCH et al., 2011).

## 2.6 Interação entre QTL e Ambiente

Com o desenvolvimento de marcadores moleculares e técnicas de mapeamento supracitadas, os pesquisadores podem analisar todo o genoma para detectar localizações específicas para os genes que influenciam a característica quantitativa. Sendo estas localizações os QTL's, a IQA ocorre quando estes QTL's apresentam uma expressão diferente ao longo dos ambientes, representando a base da IGA.

A compreensão destas interações é de suma importância, pois somente após

este entendimento os pesquisadores têm capacidade de selecionar melhores genótipos para diferentes gradientes ambientais e desta forma melhorar a produção de determinado cultivar, baseando-se no clima e características dos solos em determinados locais.

Muitos fenótipos (características) avaliados em experiências agrícolas são quantitativos sendo que a variação nessas características quantitativas é usualmente devida ao efeito de várias localizações genéticas e a fatores ambientais. O conhecimento sobre o número, localização, efeito e identidade de tais localizações genéticas, ou seja, QTL, poderão conduzir a novas descobertas biológicas. Esta informação sobre QTL pode ser usada para ajudar na seleção e melhoramento de colheitas agrícolas (BROMAN; SEN, 2009).

Em resumo, o mapeamento de QTL é uma técnica para tentar identificar QTL numa população experimental resultante do cruzamento de dois progenitores. Um grande desafio em estatística genética é encontrar melhores genótipos ao longo de uma ampla variação de condições agroecológicas e também ao longo dos anos. Este é também um desafio para agricultores, melhoradores de plantas e geneticistas.

No entanto, vale citar que os agricultores e os melhoradores de plantas têm interesses diferentes: os agricultores pretendem identificar genótipos melhores para determinado clima e tipo de solo e os melhoradores de plantas pretendem desenvolver genótipos que tenham uma boa performance em localizações diferentes e heterogêneas. Para alcançar estes objetivos, são conduzidos ensaios multi-ambientais em que uma série de genótipos são avaliados em diferentes condições ambientais e em diferentes anos.

Os dados provenientes destes MET's são usualmente sumarizados numa tabela de dupla entrada com genótipos nas linhas e ambientes (combinações de localização e ano) nas colunas. Na maior parte destas tabelas de dupla entrada é possível encontrar diferenças na estabilidade dos fenótipos ao longo dos ambientes, isto é, os efeitos genotípicos e ambientais não são simplesmente aditivos e a interação entre genótipo e ambiente (IGA) está presente nos dados.

IGA é definida como a alteração do ranking genético dos genótipos para diferentes ambientes, isto é, por exemplo, um genótipo com boa performance em condições úmidas pode apresentar uma má performance em condições de seca. A IGA pode ser expressa como crossovers, quando dois genótipos diferentes apresentam alteração no ranking de performance quando avaliados em ambientes diferentes ou respostas inconsistentes de alguns genótipos ao longo dos ambientes sem mudança em termos de ranking.

O estudo e entendimento destas interações representam um grande desafio com o objetivo de melhorar características complexas (rendimento) para diferentes gradientes ambientais. Com o desenvolvimento de marcadores moleculares e técnicas de mapeamento, os pesquisadores podem ir mais além e analisar todo o genoma para detectar localizações específicas para os genes que influenciam a característica quantitativa. Estas localizações, como citado anteriormente, são os chamados QTL's e quando estes QTL's apresentam uma expressão diferente ao longo dos ambientes, estamos perante interação entre QTL e ambiente (IQA), que representa a base da IGA.

Um bom entendimento destas interações permite aos pesquisadores selecionar melhores genótipos para diferentes gradientes ambientais e conseqüentemente melhorar colheitas e as suas produções em países desenvolvidos e em desenvolvimento, com base no seu clima e características dos solos.

O uso da metodologia AMMI pode ser estendido para o estudo da IQA, a fim de identificar locos potencialmente envolvidos no controle da IGA, ou seja, locos responsáveis pela adaptação geral ou específica dos genótipos. Em um trabalho pioneiro, Romagosa et al. (1996) mapearam QTL's para os componentes da IGA, simplesmente utilizando como característica fenotípica os escores dos genótipos para os componentes de maior importância.

Desta maneira, foram mapeados locos com diferentes graus de influência sobre a IGA, o que permitiu a identificação de genótipos adaptados especificamente a cada ambiente utilizado no estudo. Este tipo de trabalho tem importante aplicação no uso de dados de marcadores moleculares para seleção assistida. Muitos outros trabalhos encontrados na literatura relatam a localização de QTL's envolvidos no controle dos componentes da interação, que por sua vez estão relacionados a fatores como a disponibilidade de água, altitude, temperatura, teor de nitrogênio no solo etc.

O modelo AMMI será aqui usado na detecção de QTL's e no estudo da interação entre QTL e ambiente. Primeiramente será aplicado um modelo AMMI parcimonioso aos dados fenotípicos e seguidamente os valores preditos com esse modelo AMMI serão usados para obter os QTL scans, o que permite incluir informação de outros ambientes no scan de um determinado ambiente (JIANG; ZENG, 1995).

Esta técnica é denominada de AQ analysis (GAUCH et al., 2011) e permite obter picos mais elevados em termos de LOD scores para os QTL's quando efetuado um QTL scan. Além disso a ordenação dos ambientes por parâmetros AMMI que sumari-

zem a IGA revela padrões consistentes e tendências sistemáticas, que muitas vezes têm interpretação ecológica ou biológica (GAUCH et al., 2011).

Um grande ponto de interesse é se a IQA pode ser detectada para característica fenotípica de interesse e verificar se podemos analisar esta IQA em termos de QTL's subjacentes para parâmetros fisiológicos ou marcadores moleculares. O modelo de QTL que estamos interessados é explicitamente com informações derivadas de marcadores para descrever a IGA em termos de QTL's em suas dependências ambientais, isto é, IQA. A inclusão de informações de marcadores, precisões genéticas, permite testar se as características fenotípicas (por exemplo, produção) são afetadas pelo DNA numa posição de um genoma em particular e se os efeitos dependem dos ambientes.

### 3 MATERIAL E MÉTODOS

#### 3.1 Característica dos Dados

O primeiro conjunto de dados utilizado é relativo a experimentos com 9 genótipos de milho que foram conduzidos em 20 ambientes, com 4 blocos. Os dados foram obtidos pelo CIMMYT (Centro Internacional de Mejoramiento de Maiz y Trigo) em experimentos realizados em vários países, caracterizando assim, experimentos multi-ambientais. Têm-se ainda que em cada experimento foi utilizado um delineamento aleatorizado em blocos. A variável utilizada foi a produtividade, em t/ha. Para ajustar o modelo W-AMMI os blocos foram considerados repetições na parte multiplicativa do modelo para o cálculo da variância do erro por ambiente.

O segundo conjunto de dados é referente à produção de cevada (*Hordeum Vulgare* L.) pelo cruzamento Harrington/TR306. Harrington (HARVEY; ROSSNAGEL, 1984) e TR306 são dados em duas linhagens de cevada selecionadas com hábito de crescimento na primavera e cultivadas na América do Norte. TR306 é uma linhagem experimental desenvolvida na Universidade de Saskatchewan, tem algumas semelhanças com a Harrington, porém não é de tão boa qualidade para maltagem como a Harrington.

Neste trabalho foram analisados os resultados referentes a 141 genótipos conduzidos em 25 ambientes, os quais não houve replicações. Para as análises com QTL foram consideradas informações de 140 fenótipos e 127 marcadores.

#### 3.2 Fundamentação Geométrica do Biplot

O termo Biplot refere-se a um tipo de gráfico contendo duas categorias de pontos ou marcadores: um referindo-se a genótipos e outro a ambientes, segundo Duarte e Vencovsky (1999).

A interpretação quanto à IGA é feita observando, para o(s) eixo(s) da interação, a magnitude e o sinal dos escores de genótipos e ambientes. Desta maneira, escores baixos, ou seja, próximos a zero, representam genótipos e ambientes que contribuirão pouco para a interação sendo caracterizados como estáveis.

A análise Biplot é utilizada tanto com variáveis contínuas quanto discretas e cuja finalidade é aproximar os dados originais e realizar uma análise simultânea das relações entre estas variáveis e é baseada na aproximação de uma matriz  $Z$  de ordem  $g \times a$ , sendo que  $g$  é o número de genótipos e  $a$  o número de ambientes, e posto  $r$  por

outra matriz de posto  $q$  ( $q < r$ ), por meio de sua DVS.

Seja:

$$Z_{(r)g \times a} \approx Z_{(q)g \times a} = U_{(q)g \times q} D_{(q)q \times q} V_{(q)q \times a}^T = G_{(q)g \times q} A_{(q)q \times a}^T$$

sendo que:

$$G_{(q)g \times q} = U_{(q)g \times q} D_{(q)q \times q}^s$$

$$A_{(q)q \times a}^T = D_{(q)q \times q}^{1-s} V_{(q)q \times a}^T$$

$s$  é uma constante que pertence ao intervalo  $[0,1]$ .

O gráfico Biplot é originado quando as  $g$  linhas de  $G$  e as  $a$  colunas de  $A$  são plotadas na forma de coordenadas. Cada  $g \times a$  valores pode ser recuperado pela observação dos  $g + a$  pontos do Biplot. A informação da produção do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente é dado por (YAN; KANG, 2002):

$$Z_{ij} = \overline{OG}_i \cos(\alpha_{ij}) \overline{OA}_j = \overline{OP}_{ji} \overline{OA}_j$$

sendo que:

$\overline{OG}_i$  é a distância absoluta da origem do Biplot até o marcador do genótipo  $i$ ;

$\overline{OA}_j$  é a distância absoluta da origem do Biplot até o marcador do ambiente  $j$ ;

$\alpha_{ij}$  é o ângulo entre os vetores  $\overline{OG}_i$  e  $\overline{OA}_j$  ;

$\overline{OP}_{ji}$  é a projeção do marcador do genótipo  $i$  no vetor ambiente  $j$ .

### 3.3 Decomposição em Valores Singulares (DVS)

Conceituando algebricamente a fatoração de matrizes, tem-se a DVS. Seja  $Z \in \mathfrak{R}^{g \times a}$  uma matriz de posto  $r$ , em que  $r \leq \min(g; a)$ . A DVS de  $Z$  consiste em escrever  $Z$  como produto de três matrizes (JOHNSON; WICHERN, 1998), como segue:

$$Z_{(r)g \times a} = U_{(r)g \times a} D_{(r)r \times r} V_{(r)r \times a}^T$$

sendo que:

$U$  é uma matriz de autovetores da matriz  $ZZ^T$  que caracteriza os  $g$  genótipos e  $V$  é uma matriz de autovetores da matriz  $Z^T Z$  que caracteriza os  $a$  ambientes;

As colunas das matrizes  $U$  e  $V$  são ortogonais, isto é:

$$U_{(r)r \times g}^T U_{(r)g \times r} = V_{(r)r \times a}^T V_{(r)a \times r} = I_{(r)}$$

$D$  é uma matriz diagonal formada pelos  $r$  maiores valores singulares de  $Z$ ,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ , os quais correspondem às raízes quadradas dos autovalores da matriz  $ZZ^T$  ou  $Z^T Z$ .

Desta maneira, vemos então que a DVS decompõem  $Z$  em  $r$  componentes principais ( $CP_l$ ), em que  $l = 1, \dots, r$  e cada  $CP$  contém um vetor do genótipo ( $\gamma_i$ ), um vetor do ambiente ( $\delta_j^T$ ) e um valor singular ( $\lambda$ ).

Assim, cada elemento de  $Z$  pode ser expresso como:

$$Z_{ij} = \sum_{k=1}^r \lambda_k \gamma_{ik} \delta_{jk}$$

sendo que:

$Z_{ij}$  é a produção do genótipo  $i$  no ambiente  $j$ ;

$\lambda_{ik}$  e  $\gamma_{jk}$  são os elementos dos vetores ( $\lambda_i$ ) e ( $\gamma_j^T$ ) respectivamente;

$i = 1, \dots, n, j = 1, \dots, p$  e  $k = 1, \dots, r$  sendo  $r$  o posto da matriz de dupla entrada.

### 3.4 Modelo AMMI e AMMI Ponderado

O modelo AMMI (*Additive Main Effect and Multiplicative Interaction*) é um método uni-multivariado que engloba uma análise de variância para os efeitos principais, que são os efeitos dos genótipos ( $G$ ) e ambientes ( $A$ ) e para efeitos multiplicativos IGA utiliza-se a DVS, que é uma técnica da álgebra matricial, onde uma aplicação imediata resulta na construção de uma aproximação para uma matriz real qualquer, de posto  $r$ , por  $r$  parcelas ortogonais entre si e de posto unitário. Para se verificar a existência da IGA utilizando o modelo AMMI, realiza-se uma análise de variância. Sendo significativa, é feita a decomposição da soma de quadrados da IGA ( $SQ_{(G \times A)}$ ) para descartar um resíduo adicional presente nessa soma de quadrados. O modelo é escrito como segue:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \theta_{jk} + \sum_{n=1}^N \lambda_n \gamma_{ni} \delta_{nj} + \varepsilon_{ijk} \quad (1)$$

sendo que:

$Y_{ijk}$  : é o valor observado do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente e no  $k$ -ésimo bloco;

$\mu$ : uma constante;

$\alpha_i$  : é o efeito do  $i$ -ésimo genótipo;



$\beta_j$  : é o efeito do  $j$ -ésimo ambiente;

$\theta_{jk}$  : é o efeito  $k$ -ésimo bloco no  $j$ -ésimo ambiente;

$\lambda_n$  : é o valor singular para interação do componente principal (IPC) no eixo  $n$ ;

$\gamma_{ni}$  e  $\delta_{nj}$  : são os escores do IPC do  $i$ -ésimo genótipo e  $j$ -ésimo ambiente para o eixo  $n$ ;

$N \leq \min(I - 1, J - 1)$ , com  $I$  o número de genótipos e  $J$  o número de ambientes;

$\varepsilon_{ijk}$ : é erro experimental.

Dependendo do número  $n$  de termos (eixos ou componentes principais) retidos para descrever o padrão da interação, o modelo é denotado por AMMI0, AMMI1, ..., AMMIF. Em AMMI0 nenhum eixo é ajustado, em AMMI1 é ajustado apenas o primeiro eixo da interação, e segue desta maneira até AMMIF o qual é o modelo onde todos os componentes principais são ajustados (DUARTE; VENCOSKY, 1999).

A formulação matricial da equação (1) é dada por:

$$\mathbf{Y} = \mathbf{1}_I \mathbf{1}_J^T \mu + \alpha \mathbf{1}_I^T + \mathbf{1}_J^T \beta + \mathbf{U} \mathbf{D} \mathbf{V}^T + \varepsilon \quad (2)$$

Quando a tabela de dupla entrada  $\mathbf{Y}$  tem dados faltantes e/ou a variância do erro não é constante, as células da tabela devem ter diferentes pesos para os quadrados dos resíduos na estimação dos parâmetros do modelo. Para dar conta dessa heterogeneidade de variância do erro, a proposta é substituir a DVS na equação (2) por uma DVS ponderada (RODRIGUES et al., 2014). A abordagem é baseada no algoritmo EM (E: esperança, M: maximização), enquanto a soma de quadrados da diferença entre duas iterações sucessivas,  $X^{(t+1)}$  e  $X^{(t)}$ , é maior do que um pequeno valor, por exemplo,  $10^{-9}$ , executamos iterativamente:

$$\mathbf{X}^{(t+1)} = DVS(\mathbf{W} \odot \mathbf{Y} + (\mathbf{1} - \mathbf{W}) \odot \mathbf{X}^{(t)}) \quad (3)$$

no qual  $\mathbf{W}$  é uma matriz de ponderações ( $I \times J$ ),  $W_{i,j}$ ,  $0 \leq W_{i,j} \leq 1$ ,  $\mathbf{1}$  é uma matriz de uns,  $\odot$  é o produto de Hadamard e  $t$  é o número da interação (Srebro and Jaakkola, 2003). Os resultados desse processo são as matrizes  $\mathbf{U}_N$ ,  $\mathbf{D}_N$  e  $\mathbf{V}_N$ , de tal modo que a parte multiplicativa do modelo (2) irá ser aproximada por  $\tilde{\mathbf{Y}} \approx \mathbf{U}_N \mathbf{D}_N \mathbf{V}_N'$ , sendo  $N$  o grau de aproximação.

Aplicando a DVS ponderada (3) na matriz  $\tilde{\mathbf{Y}}$  e substituindo em (2), resulta então no modelo W-AMMI. Esta generalização do modelo AMMI leva em conta as diferenças de variâncias do erro em ambientes e/ou células que faltam, e pode ser aplicado a todos

os conjuntos de dados em que o modelo AMMI é usualmente utilizado. É necessário ser capaz de estimar a variância do erro para um ambiente, por isso precisamos de repetições por ambiente.



## 4 RESULTADOS

### 4.1 IGA: AMMI vs. AMMI Ponderado

As análises AMMI e W-AMMI abaixo são referentes ao primeiro conjunto de dados, ou seja, relativo a experimentos com 9 genótipos de milho conduzidos em 20 ambientes, com 4 blocos. No ajuste do modelo W-AMMI os blocos foram considerados repetições na parte multiplicativa do modelo para o cálculo da variância do erro por ambiente e a variável utilizada foi a produtividade (t/ha).

Pela análise de variância percebe-se que os efeitos de genótipos, ambientes e IGA são significativos. A próxima etapa da análise, uma vez que a interação é significativa, corresponde ao ajuste da interação pela decomposição em valores singulares, aplicada a matriz de interações. O desdobramento da somas de quadrados da interação é apresentado na Tabela 1.

Tabela 1 - Análise de variância para o modelo completo AMMI dos dados referentes a experimentos com genótipos de milho

FV	GL	SQ	QM	PE	PA
Ambientes	19	989593772	52083883		
Blocos(Ambientes)	60	118813054	1980218		
Genótipos	8	79828575	9978572		
IGA	152	249704162	1642791		
IPC1	26	35429503	1362673.2	33.9	33.9
IPC2	24	25079060	1044960.8	24.0	57.9
IPC3	22	17621493	800977.0	16.8	74.7
IPC4 - IPC8	80	26469809	1546932.6	25.3	100.0

GL: graus de liberdade; SQ: somas de quadrados; QM: quadrados médio; PE: proporção de SQ explicada; PA: proporção de SQ acumulada;  $IPCA_k$ : componentes principais de interação, modelo com k componentes, k = 1, 2, ..., 8.

Observa-se que o primeiro eixo singular da interação captura 33,9% da soma de quadrados da IGA, o segundo 24,0%, o terceiro 16,8% e os demais eixos capturam 25,3% da interação. Consequentemente, o modelo AMMI com dois componentes explica 57,9% da soma de quadrados da IGA como resposta padrão e apesar da porção pequena da  $SQ_{G \times A}$  explicada pelos dois primeiros eixos, espera-se capturar a maior parte da explicação padrão devido à IGA. 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. Vários autores utilizaram o mesmo tipo de representação apesar de terem verificado uma menor proporção da  $SQ_{G \times A}$  explicada (todos estes inferiores a 60%): Eyhéabide et al., 1997;

Flores et al., 1996; Crossa et al., 1990; Pereira e Costa, 1998; Arias, 1996; Crossa et al., 1991; Garcia e Dias (2009).

Para o conjunto de dados em questão, o Biplot AMMI2 (Figura 1 a esquerda) de duas dimensões representa 57,9% da soma de quadrados da interação, pois este gráfico utiliza os dois primeiros componentes para obter as coordenadas dos genótipos e ambientes (DIAS, 2005; GARCIA; DIAS, 2009).

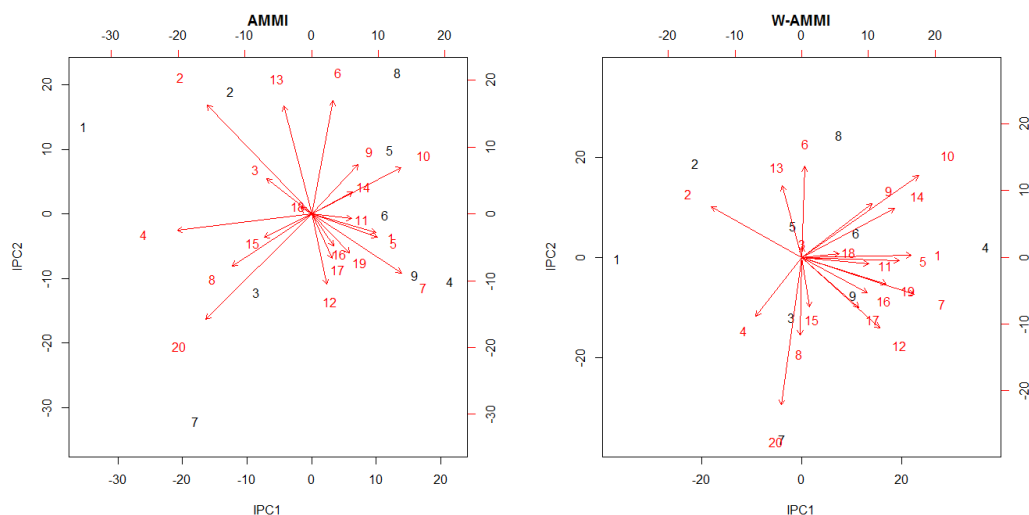


Figura 1 - Biplot dos modelos AMMI2 e W-AMMI2 para dados de produção de milho. Os pontos enumerados representam os genótipos e as setas os ambientes.

A análise AMMI ponderada é utilizada para evitar que ambientes com variância do erro elevada tenham demasiada importância na modelagem (RODRIGUES et al., 2014) ou que sejam considerados como outliers e removidos da análise (GAUCH et al., 2011). Neste novo modelo, a contribuição de um dado ambiente ao ajuste do modelo (W-AMMI) é proporcional ao inverso da variância do erro.

Entre a porcentagem de variância explicada pela parte multiplicativa do modelo W-AMMI2, 65,2% é explicada pelo primeiro eixo singular da interação e 34,8% é explicada pelo segundo eixo principal da interação.

Comparando os dois Biplots da Figura 1 (Biplot AMMI à esquerda e Biplot W-AMMI à direita), observamos que os ângulos formados entre os ambientes e a disposição dos genótipos sofre uma certa alteração, isto porque a contribuição de cada ambiente para o modelo varia devido às ponderações no modelo W-AMMI, ou seja, em alguns casos, essa angulação diminui, apresentando uma maior correlação entre genótipos e ambientes. Por exemplo, no Biplot W-AMMI, os genótipos 3 e 7 apresentaram uma correlação maior com

o ambiente 20 do que no AMMI. O mesmo aconteceu com o genótipo 5 em relação ao ambiente 13, o genótipo 6 em relação ao ambiente 14, e o genótipo 9 com os ambientes 17 e 12. Já o genótipo 9 apresentava uma correlação maior no Biplot AMMI com o ambiente 7 do que no W-AMMI. O mesmo ocorre com o genótipo 5 em relação ao ambiente 9.

Ainda na Figura 1 é possível observar que existe correlação inversa do genótipo 5 com o ambiente 20, do genótipo 2 com o ambiente 7 e do genótipo 9 com o ambiente 2. Ou seja, estes genótipos não são recomendados para estes ambientes.

Os ambientes mais importantes para os modelos, ou seja, que mais influenciam para a interação, são: 2, 4, 6, 7, 10, 13 e 20. Porém, o ambiente 13, tem um diminuição considerável na magnitude apresentada no Biplot W-AMMI.

O ambiente 3 apresenta uma grande variância do erro. Vemos que, se compararmos o Biplot AMMI2 com o Biplot W-AMMI2, a magnitude da flecha diminui consideravelmente, passando o ambiente 3 a ter a menor magnitude, podendo então ser considerado o ambiente mais estável.

O ambiente 9 é o que tem uma maior variância do erro dentre todos os estudados, o que é indicativo que a sua influência no ajuste do modelo AMMI é superior à recomendada. Ao comparar o comportamento e a influência deste ambiente nos Biplots da Figura 1, observamos que a influência do ambiente 9 no Biplot do modelo AMMI2 era indicativa que o genótipo 8 era indicado para esse ambiente (e para o ambiente 6) e que o genótipo 5 também era indicado para esse ambiente (e para o ambiente 10). No entanto, ao fazer a correção com o modelo W-AMMI2, o Biplot indica que o genótipo 8 é essencialmente adaptado às características do ambiente 6 (e não às do ambiente 9) e que o genótipo 5 é adaptado às características do ambiente 13 (e não às do ambiente 9).

O ambiente 10 é o que apresenta uma menor variância do erro (e consequentemente aquele que tem mais precisão/homogeneidade nas suas repetições/blocos). Ao comparar a influência/comportamento deste ambiente nos dois modelos/Biplots, verificamos o descrito no parágrafo anterior. Ou seja, estes dois ambientes (9 e 10), ao receberem uma ponderação mais coerente na modelação W-AMMI2, permitem que os genótipos 8 e 5 passem a ser associados aos ambientes nos quais têm mais possibilidade de ser mais produtivos.

O ambiente 20 é um dos que apresenta uma menor variância do erro. Comparando os dois Biplots da Figura 1 verifica-se que por meio da ponderação com o modelo W-AMMI, a interpretação de que este ambiente poderia ser adequado para os genótipos 3

e 7, é claramente confirmada porque estes genótipos passam a ser modelados exatamente sobre a seta do ambiente 20.

## 4.2 IQA: AMMI vs. AMMI Ponderado

### 4.2.1 Análise da Interação entre Genótipo e Ambiente

Os resultados a seguir são referentes ao segundo conjunto de dados relativo ao cruzamento Harrington/TR306 com 141 genótipos conduzidos em 25 ambientes, incluindo as informações de 140 fenótipos e 127 marcadores para as análises com QTL. Não houve replicação dos dados.

Por meio da estatística  $W$  três componentes são selecionados para serem retidos no modelo. A próxima etapa corresponde ao ajuste da interação pela decomposição em valores singulares, aplicada à matriz de interações. O desdobramento das somas de quadrados da interação é apresentado na Tabela 2.

Observa-se que o primeiro eixo singular da interação captura 18,9% da soma de quadrados da IGA, o segundo 15,0%, o terceiro 8,4% e os demais eixos capturam 57,7% da interação. Consequentemente, o modelo AMMI com dois componentes explica 33,9% da soma de quadrados da interação entre genótipos e ambientes como resposta padrão. Assim, o Biplot AMMI2 (Figura 2 e Figura 3 à esquerda) de duas dimensões representa 33,9% da soma de quadrados da interação.

Tabela 2 - Análise de variância para o modelo completo AMMI dos dados referentes ao cruzamento Harrington/TR306

FV	GL	SQ	QM	PE	PA
Ambientes	24	2088855278	87035637		
Genótipos	140	71283814	509170		
IGA	3360	468475524	139427		
IPC1	163	88493105.7	542902.49	18.9	18.9
IPC2	161	70221295.4	436157.11	15.0	33.9
IPC3	159	39386395.7	247713.18	8.4	42.3
IPC4 - IPC24	2877	270374720.6	1860131,3	57.7	100.0

GL: graus de liberdade; SQ: somas de quadrados; QM: quadrados médio; PE: proporção de SQ explicada; PA: proporção de SQ acumulada;  $IPCA_k$ : componentes principais de interação, modelo com  $k$  componentes,  $k = 1, 2, \dots, 24$ .

Visando amenizar possíveis distorções nos resultados caso haja elevada variância do erro por ambiente, será incorporada uma análise ponderada da IGA com o intuito de produzir resultados potencialmente mais confiáveis em termos de visualização e

detecção de QTL's. Para isto, foram dados pesos menores para os ambientes com maior variância do erro.

Considerando que na tabela de dupla entrada dos dados encontram-se os genótipos nas linhas e os ambientes nas colunas, foram consideradas duas ponderações para análise no modelo W-AMMI. Na primeira análise considerou-se os pesos iguais por linha, ou seja, foi calculada a variância para cada genótipo, sendo assim todos os ambientes para determinado genótipo ficam com o mesmo peso, resultando no Biplot W-AMMI “rows” (Figura 2 a direita).

Na segunda análise considerou-se os pesos iguais por coluna, desta forma todos os genótipos para determinado ambiente ficam com o mesmo peso, resultando no Biplot W-AMMI “columns” (Figura 3 a direita). No novo modelo para a análise W-AMMI a contribuição de um dado ambiente ao ajuste do modelo (W-AMMI) é a razão da menor variância pela variância de determinado ambiente.

A Figura 2 que refere-se aos Biplots resultantes do modelo AMMI e W-AMMI “rows” e apesar de haver 3 componentes significativos, com o intuito de uma melhor visualização, foram utilizados apenas dois componentes nos Biplots. Comparando estes Biplots vemos que os ângulos formados entre os ambientes e a disposição de alguns genótipos sofre alteração insignificante, quase imperceptível. Podemos citar, por exemplo, uma pequena diminuição do ângulo entre o genótipo 59 e o ambiente AB92c, quando comparamos os dois Biplots, isto devido a variação causada pelas ponderações no modelo W-AMMI.

Porém, visualizando a Figura 3 e comparando os Biplots referentes ao modelo AMMI e W-AMMI “columns” notamos que essa alteração angular entre genótipos e ambientes é significativamente mais evidente. Por exemplo, o genótipo 4 é recomendado ao ambiente SK92c pelo modelo AMMI, já depois da ponderação dos dados, Biplot W-AMMI “columns” o genótipo 4 apresenta uma forte interação positiva com o ambiente AB92c, já para o ambiente SK92c, o genótipo mais recomendado passa a ser o 80. Também notamos que os genótipos 25 e 92 apresentam uma alta correlação com o ambiente MB93 no modelo W-AMMI “columns” e estes genótipos não são recomendados a esse ambiente se analisamos o modelo AMMI.

Considerando que os resultados apresentados na Figura 3 é perceptível mais alterações que os apresentados na Figura 2, como comentado anteriormente, as análises WAQ serão feitas baseadas nas ponderações por colunas.



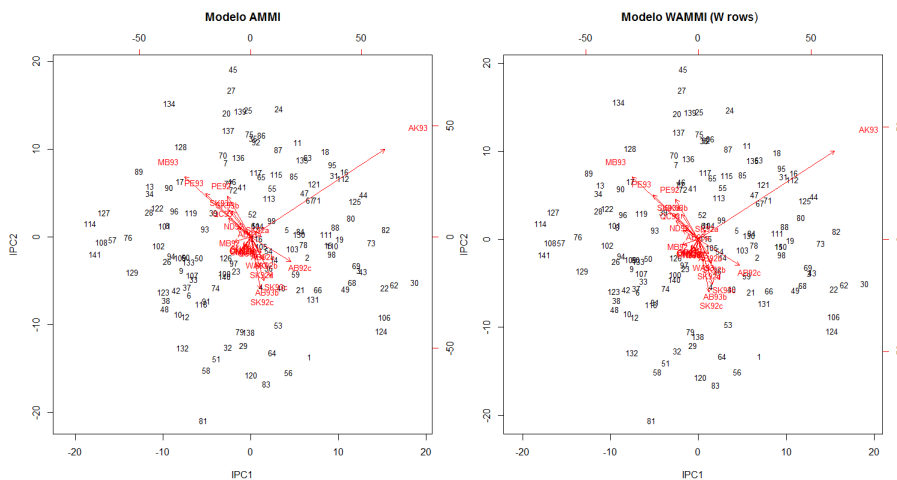


Figura 2 - Biplot dos modelos AMMI2 e W-AMMI2 (ponderação por linhas) para dados de cevada Harrington/TR306

Os pontos enumerados representam os genótipos e as setas os ambientes.

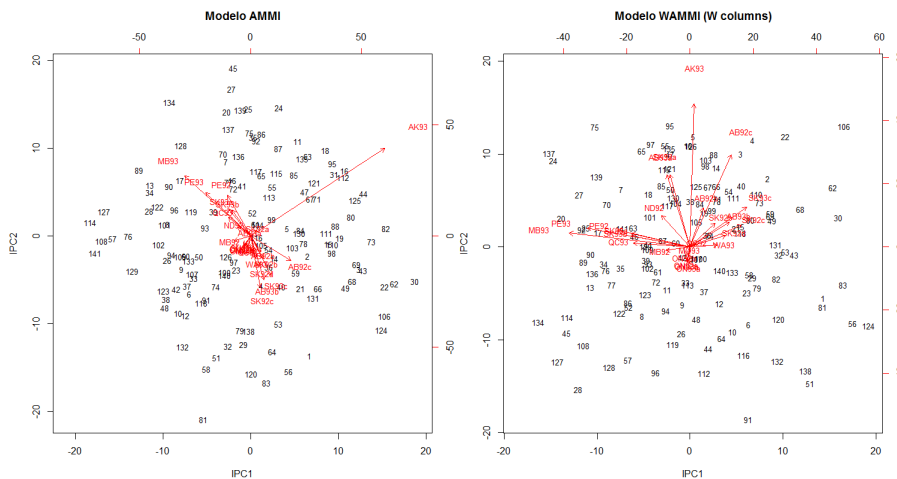


Figura 3 - Biplot dos modelos AMMI2 e W-AMMI2 (ponderação por colunas) para dados de cevada Harrington/TR306

Os pontos enumerados representam os genótipos e as setas os ambientes.

#### 4.2.2 QTL scans e Análise da Interação QTL × Ambiente

Os dados fenotípicos resultantes dos ensaios multi-ambientais permite a análise da complexidade da IGA e consequentemente da IQA.

A Figura 4 representa o mapa dos marcadores utilizados para análise de QTL nos 7 cromossomos, os nomes dos 127 marcadores e suas posições. A Figura 5 e a Figura 6 mostram os QTL scans para cada um dos 25 ambientes individualmente para os dados de rendimento para população Harrington/TR306.

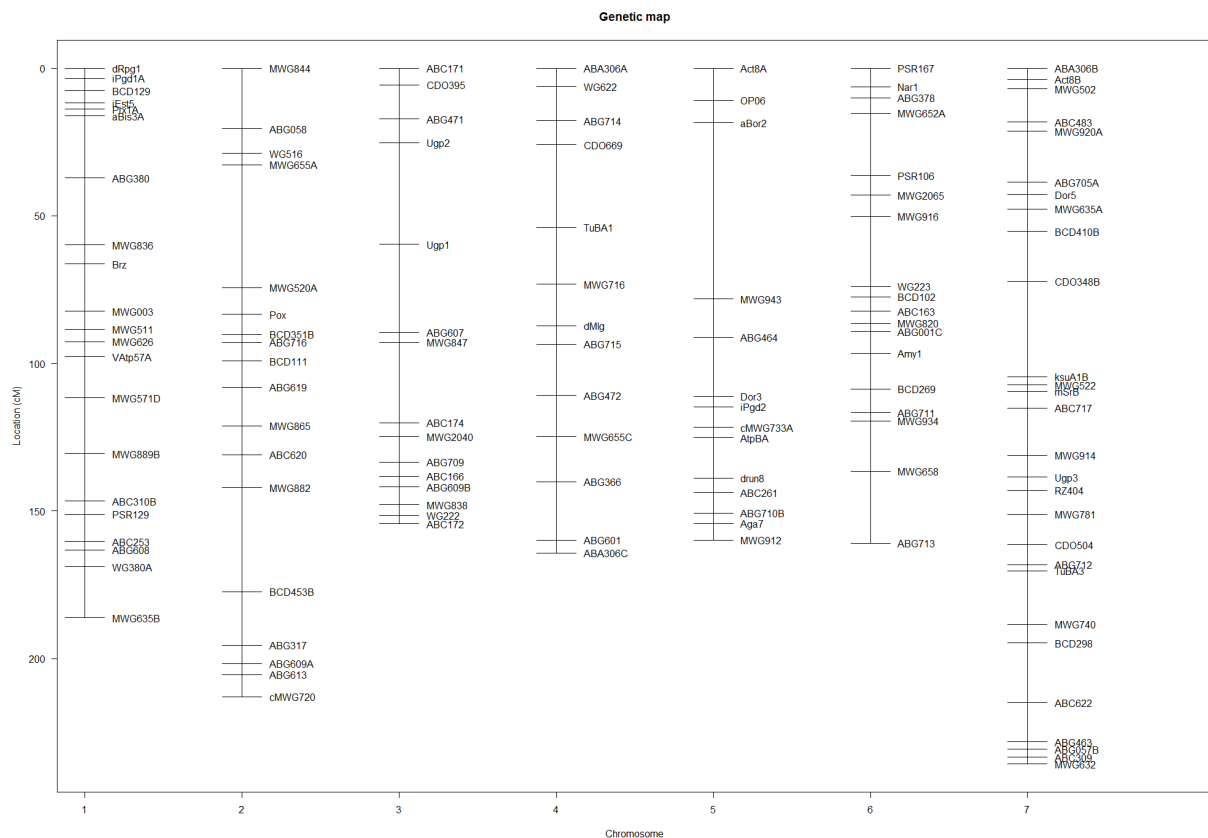


Figura 4 - Mapa genético dos dados de cevada Harrington/TR306

A análise de QTL ou análise AQ é obtida ajustando o modelo AMMI para a IGA e posteriormente fazendo varreduras de QTL pelos valores previstos para cada ambiente (GAUCH et al., 2011). Ou seja, é a combinação da análise AMMI seguido por QTL scans. A WAQ é uma generalizaçã da análise AQ, sendo que a análise AMMI é substituída pela análise W-AMMI.

A Figura 5 e a Figura 6 mostram os QTL scans, sendo que os LOD scores foram obtidos por meio da AQ (linha cinza) e WAQ (linha preta) para modelos com dois IPC's. A linha cinza tracejada refere-se aos dados originais. Observando a altura desses QTL scans (LOD scores) vemos que há uma visível melhoria nas análises WAQ, sejam

em números de QTL's detectados como na pontuação dos LOD scores (Tabela 3 à Tabela 8) quando comparados aos dados originais e aos referentes à análise AQ. Porém, os LOD scores nos dados originais são menores que para os valores preditos com o modelo AMMI. Por exemplo, o ambiente AK93 que é o mais significativamente associado à interação, os cromossomos 2 e 3 apresentam QTL's somente quando o ajuste é pelo modelo W-AMMI. Já no ambiente AB92b, o cromossomo 2 apresenta um alto LOD score quando a análise é obtida ajustando o modelo AMMI ou W-AMMI.

Os maiores valores de LOD scores são encontrados com o ajuste do modelo W-AMMI, sendo que o maior deles é aproximadamente 9.46, no cromossomo 3 do ambiente AK93.

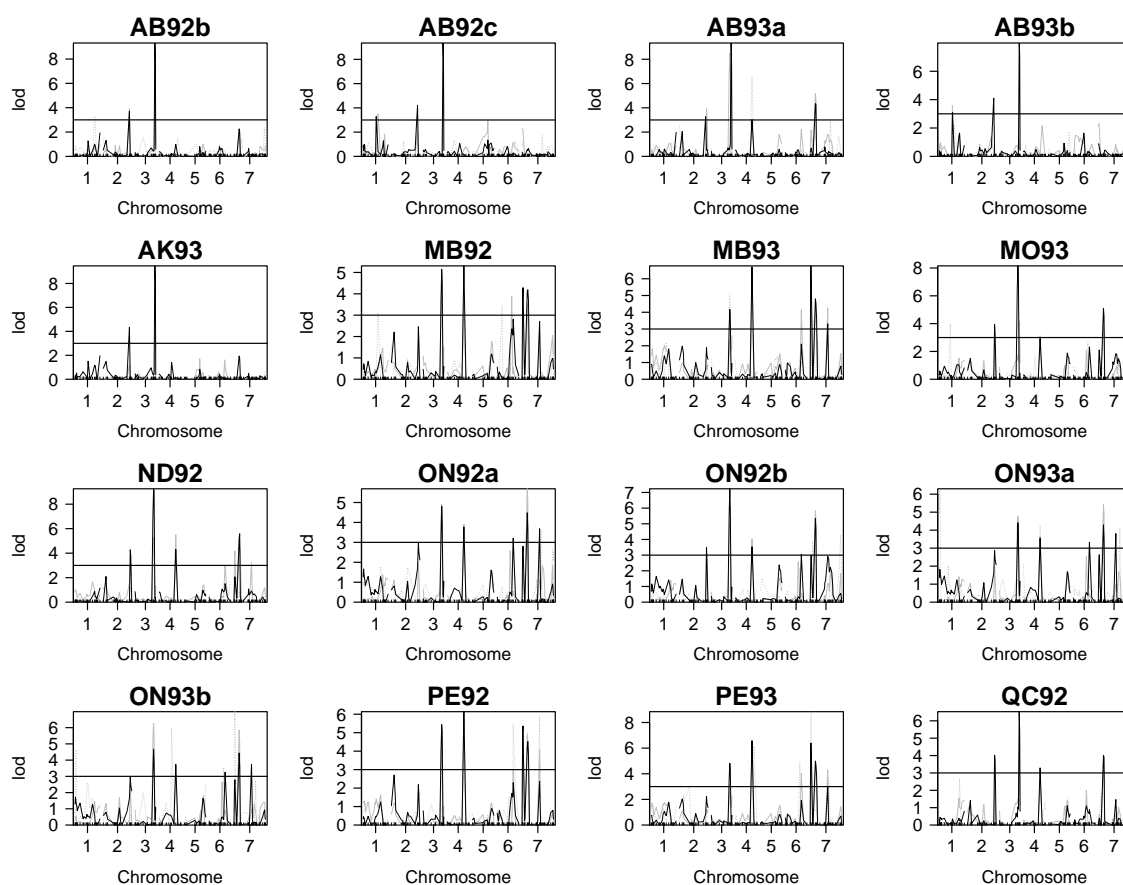


Figura 5 - QTL scan para os primeiros 16 ambientes dos dados de rendimento para população Harrington/TR306

Os segmentos pontilhados cinzas pontilhados representam as análises dos dados reais, as linhas cinzas representam as varreduras dos valores previstos pelo modelo AMMI e as linhas pretas representam as análises obtidas pelo modelo W-AMMI "columns". Todas as análises são com base em mapeamento por intervalo-composto. Os códigos acima de cada scan individual representam os ambientes.

O algoritmo W-AMMI proposto é viável para trabalhos com outros sistemas

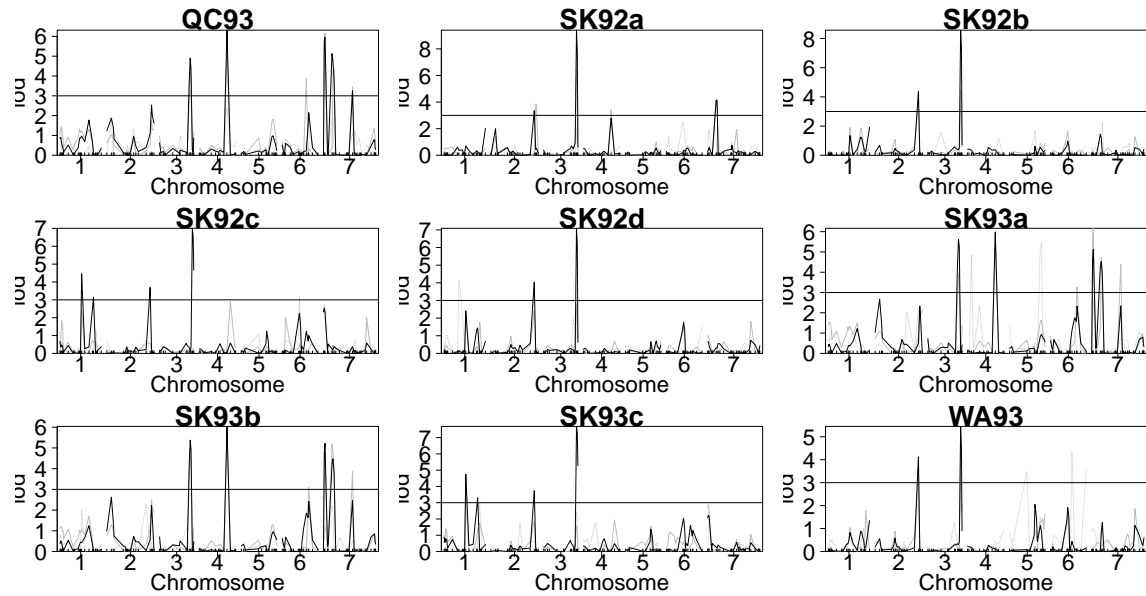


Figura 6 - QTL scan para os últimos 9 ambientes dos dados de rendimento para população Harrington/TR306

Os segmentos pontilhados cinzas pontilhados representam as análises dos dados reais, as linhas cinzas representam as varreduras dos valores previstos pelo modelo AMMI e as linhas pretas representam as análises obtidas pelo modelo W-AMMI “columns”. Todas as análises são com base em mapeamento por intervalo-composto. Os códigos acima de cada scan individual representam os ambientes.

Tabela 3 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 nos cinco primeiros ambientes

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
AB92b	1	146.6	3.32	186.2	0.99	186.2	1.92
	2	0	1.46	195.4	3.94	195.4	3.72
	3	59.7	1.57	147.9	4.93	147.9	9.32
AB92c	1	151.1	1.11	111.7	3.46	97.7	3.28
	2	195.4	2.94	195.4	2.01	195.4	4.22
	3	141.9	2.64	151.5	4.51	147.9	9.35
AB93a	2	205.4	3.12	201.7	3.98	195.4	3.27
	3	147.9	4.87	138.2	8.48	147.9	9.34
	4	124.8	6.57	124.8	3.02	124.8	3.02
	7	151.1	3.05	38.6	5.18	38.6	4.34
AB93b	1	111.7	2.23	97.7	3.58	97.7	3.08
	2	177.3	0.35	177.3	1.39	195.4	4.11
	3	5.7	0.40	147.9	1.15	147.9	7.99
AK93	2	195.4	0.78	195.4	1.26	195.4	4.34
	3	151.5	2.92	151.5	3.74	147.9	9.46

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.

de ponderação. Os pesos para o algoritmo do modelo W-AMMI pode ser escolhido de acordo com prescrições estabelecidas em Smith et al. (2001), Möhring Piepho (2009) e

Tabela 4 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do sexto ao nono ambiente

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
MB92	1	111.7	3.11	146.6	1.46	130.6	1.14
	3	154.3	0.28	138.2	4.35	138.2	5.15
	4	124.8	2.21	124.8	3.72	124.8	5.31
	6	36.4	3.45	108.7	3.88	119.5	2.81
	7	228.2	1.75	38.6	3.54	7	4.29
MB93	3	138.2	5.06	154.3	2.58	138.2	4.17
	4	124.8	3.40	124.8	5.97	124.8	6.71
	6	108.7	1.25	119.5	4.15	119.5	2.10
	7	21.4	3.50	7	5.27	3.8	6.77
MO93	1	82.3	3.99	146.6	1.52	130.6	1.05
	2	74.3	0.40	201.7	2.69	201.7	3.94
	3	124.8	1.01	147.9	4.19	138.2	8.15
	4	87.2	1.66	124.8	3.12	124.8	2.94
	7	0	2.64	38.6	3.81	38.6	5.10
ND92	2	205.4	1.23	201.7	2.96	201.7	4.27
	3	124.8	0.43	138.2	5.29	138.2	9.25
	4	124.8	2.97	124.8	5.50	124.8	4.30
	6	160.9	1.81	119.5	3.02	119.5	1.51
	7	47.7	1.66	38.6	5.36	42.8	5.58

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.

Welham et al. (2010). Vale ressaltar que os pesos no algoritmo W-AMMI aqui utilizado exige um (re)dimensionamento que os leva a ter valores entre 0 e 1, como foi feito com os dois conjuntos de dados apresentados.

Assim, este algoritmo W-AMMI é generalização da análise AMMI padrão (GAUCH, 1992) que é capaz de explicar a heterogeneidade de variância do erro, podendo assim estimar escores da IGA em MET's. Em casos específicos com pouca heterogeneidade de variâncias de erro (GAUCH et al., 2011), o modelo AMMI padrão é totalmente apropriado.

Ao direcionar a abordagem que será utilizada em determinado experimento, AMMI ou W-AMMI, deverá ser calculado o erro ou a variância residual para cada ambiente e verificar se é homogênea em todos os ambientes. Se a variância de erro por meio dos ambientes for homogênea, os resultados do modelo AMMI serão semelhantes aos da abordagem W-AMMI. Sendo assim, a estratégia do modelo AMMI padrão já se apresenta como suficiente. Quando as variações do erro apresentam alta heterogeneidade entre os ambientes o uso do modelo AMMI não é aconselhável e o algoritmo de W-AMMI deve ser utilizado.

Tabela 5 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo ao décimo terceiro ambiente

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
ON92a	2	121.1	1.75	201.7	2.01	201.7	3.01
	3	17.2	0.92	138.2	4.90	138.2	4.81
	4	124.8	3.40	124.8	3.90	124.8	3.77
	6	119.5	3.08	96.6	2.58	119.5	3.20
	7	143.3	3.58	38.6	5.69	38.6	4.48
ON92b	2	74.3	0.58	201.7	2.05	201.7	3.50
	3	141.9	1.21	138.2	5.02	138.2	7.23
	4	124.8	1.36	124.8	4.02	124.8	3.54
	6	108.7	2.20	96.6	2.58	119.5	3.04
	7	228.2	4.29	38.6	5.84	38.6	5.34
ON93a	1	3.4	6.30	146.6	1.41	3.4	1.82
	3	133.4	1.29	138.2	4.77	138.2	4.12
	4	124.8	4.27	124.8	3.65	124.8	3.56
	6	160.9	0.97	96.6	2.56	119.5	3.32
	7	161.4	4.12	38.6	5.42	38.6	4.30
ON93b	1	7.6	4.64	146.6	1.46	3.4	1.72
	3	59.7	2.02	138.2	6.25	138.2	4.68
	4	93.5	5.97	124.8	3.69	124.8	3.74
	6	108.7	3.56	96.6	2.64	119.5	3.26
	7	7	6.98	38.6	5.84	38.6	4.44

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.

No segundo conjunto de dados foi proposta uma alternativa de pesos para quando há dados sem replicas. Em Rodrigues et al. (2014) foram sugeridos pesos somente para os casos onde há dados replicados.

Tabela 6 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo quarto ao décimo oitavo ambiente

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
PE92	3	59.7	1.09	138.2	4.59	138.2	5.44
	4	124.8	3.13	124.8	5.65	124.8	6.14
	6	119.5	5.46	119.5	3.01	119.5	2.29
	7	131.1	5.91	7	4.90	7	5.36
PE93	3	0	0.69	138.2	3.13	138.2	4.82
	4	140.1	6.14	124.8	6.17	124.8	6.58
	6	108.7	4.92	119.5	4.05	119.5	1.93
	7	7	8.84	7	5.33	7	6.39
QC92	2	0	0.74	201.7	2.67	201.7	4.02
	3	147.9	0.20	147.9	4.18	147.9	6.53
	4	159.8	1.25	124.8	2.59	124.8	3.29
	7	21.4	1.86	38.6	3.34	38.6	4.01
AB92b	1	146.6	3.32	186.2	0.99	186.2	1.92
	2	0	1.46	195.4	3.94	195.4	3.72
	3	59.7	1.57	147.9	4.93	147.9	9.32
AB92c	1	151.1	1.11	111.7	3.46	97.7	3.28
	2	195.4	2.94	195.4	2.01	195.4	4.22
	3	141.9	2.64	151.5	4.51	147.9	9.35

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.

Tabela 7 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 do décimo nono ao vigésimo segundo ambiente

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
AB93a	2	205.4	3.12	201.7	3.98	195.4	3.27
	3	147.9	4.87	138.2	8.48	147.9	9.34
	4	124.8	6.57	124.8	3.02	124.8	3.02
	7	151.1	3.05	38.6	5.18	38.6	4.34
AB93b	1	111.7	2.23	97.7	3.58	97.7	3.08
	2	177.3	0.35	177.3	1.39	195.4	4.11
	3	5.7	0.40	147.9	1.15	147.9	7.99
AK93	2	195.4	0.78	195.4	1.26	195.4	4.34
	3	151.5	2.92	151.5	3.74	147.9	9.46
MB92	1	111.7	3.11	146.6	1.46	130.6	1.15
	3	154.3	0.28	138.2	4.35	138.2	5.15
	4	124.8	2.21	124.8	3.72	124.8	5.31
	6	36.4	3.45	108.7	3.88	119.5	2.81
	7	228.2	1.75	38.6	3.55	7	4.29

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.

Tabela 8 - Valores de LOD score para os QTL's e suas respectivas posições referentes aos dados da população Harrington/TR306 nos três últimos ambientes

Ambientes	Cromossomos	Dados Originais		AMMI		W-AMMI	
		Posição	Load	Posição	Load	Posição	Load
MB93	3	138.2	5.06	154.3	2.58	138.2	4.17
	4	124.8	3.40	124.8	5.97	124.8	6.71
	6	108.7	1.25	119.5	4.15	119.5	2.10
	7	21.4	3.50	7	5.27	3.8	6.77
MO93	1	82.3	4	146.6	1.52	130.6	1.05
	2	74.3	0.40	201.7	2.69	201.7	3.94
	3	124.8	1.01	147.9	4.19	138.2	8.15
	4	87.2	1.66	124.8	3.12	124.8	2.94
	7	0	2.64	38.6	3.81	38.6	5.10
ND92	2	205.4	1.23	201.7	2.96	201.7	4.27
	3	124.8	0.43	138.2	5.29	138.2	9.25
	4	124.8	2.97	124.8	5.50	124.8	4.30
	6	160.9	1.81	119.5	3.02	119.5	1.51
	7	47.4	1.66	38.6	5.36	42.8	5.58

Os QTL's foram detectados segundo cada uma das três abordagens: QTL scans dos dados originais; QTL scans dos valores preditos pelo modelo AMMI; QTL scans dos valores preditos pelo modelo W-AMMI.





## 5 CONCLUSÕES

Na primeira etapa desta pesquisa, a ponderação dos dados de rendimento de milho para a utilização do modelo W-AMMI no caso de heterogeneidade de variância do erro se apresentou como uma boa alternativa, sendo que neste caso o modelo AMMI não é indicado e apresentou evidências substanciais para a precisão dos resultados da IGA. Já na segunda etapa, para os rendimentos da população Harrington/TR306, foi proposta uma alternativa de pesos para o caso de dados não replicados, sendo que a ponderação por colunas se mostrou eficiente na precisão dos resultados da IGA e IQA.

A utilização dos modelos AMMI assim como dos modelos W-AMMI nos casos indicados, permitiu melhorar a detecção de QTL, assim como alcançar um melhor entendimento da interação entre QTL e ambiente. Este objetivo foi alcançado por meio da análise WAQ, isto é, quando os QTL scans foram realizados nos valores preditos do modelo W-AMMI mais parcimônio em vez de os obter com base nos dados originais.

O entendimento de como os QTL's interagem com o ambiente permite aos melhoristas obterem informação para aumentar a produtividade em diferentes ambientes. Esta análise também permite aos melhoristas selecionar localizações que otimizem a detecção de um determinado QTL. Essa localização maximiza consistentemente o LOD score para o QTL ao longo dos anos.

As técnicas aqui apresentadas para detectar e entender IQA são baseadas em princípios estatísticos de aplicabilidade em populações microbiais e vegetais quando estudadas em vários ambientes e podem ser adaptadas a estudos genéticos em animais e humanos (GAUCH et al., 2011).



## REFERÊNCIAS

- ALLARD, R. W.; BRADSHAW, A. D. Implications of genotype-environment interaction in applied plant breeding. **Crop Science**, v. 4, n. 5, p. 503-508, 1964.
- ALLARD, R.W. **Princípios do melhoramento genético das plantas**. Rio de Janeiro:USAID/Edgard Blucher, 1971. 381p.
- ANNICCHIARICO, P. Additive main effects and multiplicative interaction (AMMI) analysis of genotype-location interactions in variety trials repeated over years. **Theoretical and Applied Genetics**, New York, v.94, n.8, p.1072-1077, 1997a.
- ARIAS, E. R. A. **Adaptabilidade e estabilidade das cultivares de milho avaliadas no Estado do Mato Grosso do Sul e avanço genético obtido no período de 1986/87 a 1993/94. 1996**. 118f. Tese (Doutorado em Estatística e Experimentação Agronômica) Universidade Federal de Lavras, Lavras, 1996.
- ARÚS, P.; MORENO-GONZÁLES, J. **Plant breeding: principles and prospects**. London: HAYWARD, 1993. 250p.
- BROMAN, K. W.; SEN, S. **A Guide to QTL Mapping with R/qtl**. New York: Dordrecht Springer, 2009. 396p.
- CROSSA, J. Statistical Analyses of multilocation trials. **Advances in Agronomy**, New York, v.44, p.55-85, 1990.
- CROSSA, J.; FOX, P. N. PFEIFER, W. H.; RAJARAM, S.; GAUCH, H. G. AMMI adjustment for statistical analysis of an international wheat yield trial. **Theoretical and Applied Genetics**, New York, v.81, n.1, p.27-37, 1991.
- CHAVES, L.J. 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.
- DIAS, C.T.S.; KRZANOWSKI, W.J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, Madison, v.43, p.865-873, 2003.
- DIAS, C.T.S. **Métodos para Escolha de Componentes em Modelo de Efeito Principal Aditivo e Interação Multiplicativa (AMMI)**. 2006. 73P. Tese (Livre Docência)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.
- DUARTE, J.B.; VENCOSKY, R. **Interação genótipos × ambientes: uma introdução à análise "AMMI "**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999, 60p. (Série Monografias, 9)

EYHÉRABIDE, G. H.; ALVAREZ, M. P.; PRESELLO, D.; COLAZO, J. C.; DAMILANO, A.; FERNÁNDEZ, A. Estabilidad del rendimiento de cultivares de maíz en la área de la EEA Pergamino en el trienio 1994/95, 1996/97. **Revista Tecnología Agropecuaria**, INTA Pergamino, may/ago, p.51-54, 1997.

FALCONER, D.S. **Introdução à genética quantitativa**. Tradução : Silva M.A., Viçosa: UFV, 1987. 279p.

FISHER, R.A. The correlation between relatives on the supposition of Mendelian inheritance. **R.A.;Trans. R. Soc. Edinb**, v.52, p.399-433, 1918.

FISHER, R.A.; MACKENZIE, W.A. Studies in crop variation. II. The manurial response of different potato varieties. **Journal of Agricultural Science**, Cambridge, v.23, p.311-320, 1923.

FLORES, F.; MORENO, M. T.; MARTINEZ, A. and CUBERO, J. I. Genotype environment interaction in faba bean: comparison of AMMI and principal coordinate models. **Field Crops Research**, Amsterdam, v.47, p.117-127, 1996.

FREEMAN, G.H. Statistical methods for the analysis of genotype-environment interactions. **Heredity**, London, v.31, p.339-354, 1973.

GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, Cambridge, v.58, n.3, p.453-467, 1971.

GARCÍA, P. M.; DIAS, C.T.S.; Análise dos modelos aditivos com interação multiplicativa (AMMI) bivariados. **Revista Brasileira de Biometria**, São Paulo, v.27, n.4, p.586-602, 2009.

GAUCH, H. G., RODRIGUES, P. C., MUNKVOLD, J. D., HEFFNER, E. L. SORRELLS, M. Two New Strategies for Detecting and Understanding QTL x Environment Interactions. **Crop Science**, Madison, v.51, p.96-113, 2011.

GAUCH Jr, H.G. Model selection and validation for yield trials with interactions. **Biometrics**, Washington, v.44, n.3, p.705-715, 1988.

\_\_\_\_\_. **Statistical analysis of regional yield trials: AMMI analysis of factorial designs**. Amsterdam: Elsevier, 1992. 277p.

GAUCH Jr, H.G.; ZOBEL, R.W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical and Applied Genetics**, New York, v.76, p.1-10, 1988

\_\_\_\_\_. AMMI analysis of yield trials. IN: KANG, M.D.; GAUCH Jr, H.G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. chap. 4, p.85-112.

GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

- GONZÁLES, L.M.A.R. **Estudo da interação genótipos × ambientes para a cultura do algodão (*Gossypium SP.*) BA região oriental do Paraguai.** 1988. 91p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1988.
- HARVEY, B.L.; ROSSNAGEL, B.G. Harrington Barley. **Canadian Journal of Plant Science**, Canada, v.64. p.193-194, 1984.
- JIANG, C. J.; ZENG, Z. B. Multiple-Trait Analysis of Genetic-Mapping for Quantitative Trait Loci. **Genetics**, Washington, v.140, n.3, p.1111-1127, 1995.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** Madison: Prentice Hall International, 1998. 816p.
- KANG, M.S. Using genotype-by-environment interaction for crop cultivar development. **Advances in Agronomy**, New York, v.62, p.199-252, 1998.
- KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **Journal of Agricultural Science**, Cambridge, v.103, p.123-135, 1984.
- LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Washington, v.121, 185-199, 1989.
- LIN, C.S.; BINNS, M.R.; LEFKOVICHTH, L.P. Stability analysis: where do we stand? **Crop Science**, Madison, v.26, n.5, p.894-900, 1986.
- LIU, B.H. **Statistical genomics: linkage, mapping and QTL analysis.** Cleveland: CRC Press, 1998. 648p.
- LYNCH, M.; WALSH, B. **Genetics And Analysis Of Quantitative Traits.** Sunderland: Sinauer Associates, 1998. 980p.
- MANDEL, J. The partitioning of interactions in analysis of variance. **Journal of Research of the National Bureau of Standards**, Washington, n.4, 04 out. 1969. Disponível em: <<http://www.nist.gov/nvl/journal-of-research-volume-73b.cfm>>. Acesso em: 7 ago. 2014.
- \_\_\_\_\_. A new analysis of variance model for non-additive data. **Technometrics**, Alexandria, v.13, p.1-18, 1971.
- MOHRING, J.; PIEPHO, H.P. Comparison of weighting in two-stage analysis of plant breeding trials. **Crop Science**, Madison, v.49, p.1977-1988, 2009.
- NUNES, G.H.S. **Interação genótipos × ambientes em eucalipto: implicações sobre a seleção e formas de atenuar seu efeito.** 2000. 160p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, 2000.

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

PEREIRA, A. S.; COSTA, D. M. Análise de estabilidade de produção de genótipos de batata no Rio Grande do Sul. **Pesquisa Agropecuária Brasileira**, Brasília, v.33, n.4, p.405-409, 1998.

PERKINS, J.M.; JINKS, J.L. Environmental and genotype-environmental components of variability III Multiple lines and crosses. **Heredity**, London, v.23, n.3, p.339-456, 1968.

PEARSON, Karl; On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, Series 6, vol. 2, no. 11, pp. 559-572, 1901.

PIKE, E.W.; SILVERBERG, T.R. Designing mechanical computers. **Machine Desing**, Cleveland, v.24, p.131-137,159-163, 1952.

RODRIGUES, P.C.; MALOSETTI, M.; GAUCH, H. G.; VAN EEUWIJK, F.A. A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. **Crop Science**, Madison, v.54, p.555-1570, 2014.

ROMAGOSA, I., ULLRICH, S. E., HAN, F. HAYES, P. M. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. **Theoretical and Applied Genetics**, New York, v.93, p.30-37, 1996.

SREBRO, N., JAAKKOLA, T.; Weighted low-rank approximations. **Twentieth International Conference on Machine Learning**, AAAI Press, pp. 720-727, 2003.

SMITH, A.; CULLIS, B.; GILMOUR, A. Applications: the analysis of crop variety evaluation data in Australia. **Australian & New Zealand Journal of Statistics**, New Jersey, v.43, p.129-145, 2001.

TANKSLEY, S.D. Mapping polygenes. **Annual Review of Genetics**, New York, v.27, p.205-233, 1993.

TOLEDO, E. R. **Mapeamento de QTL's Utilizando as Abordagens Clássicas e Bayesianas**. 2006. 99p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2006.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.

WEBER, W.E.; WRICKE, G.; WESTERMANN, T. Selection of genotypes and prediction of performance by analysis genotype-by-environment interactions. In: KANG, M.S; GAUCH, H.G. **Genotype by environment interactions**. Boca Raton: CRC Press, 1996. p.353-371.

- WELHAM, S.J.; GOGEL, B.J.; SMITH, A.B.; THOMPSON, R.; CULLIS, B.R. A comparison of 869 analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, New Jersey, v.52, p.125-149, 2010.
- WILLIAMS, E.J. The interpretation of interactions in factorial experiments. **Biometrika**, Cambridge, v.39, p.65-81, 1952.
- YAN, W.; KANG, M. S. **GGE biplot analysis**: A graphical tool for breeders, geneticists, and agronomists. Boca Raton. Florida: CRC Press, 2002. 271p.
- YATES, F.; COCHRAN, W.G. The analysis of group of experiments. **Journal of Agricultural Science**, London, v.28, p.556-580, 1938.
- YOUNG, N.D. QTL mapping and quantitative disease resistance in plants. **Annual Review Phytopathology**, New York, v.34, p.479-501, 1996.
- ZENG, Z.B. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. **Proceedings of the National Academy of Sciences**, USA, v.90, p.10972-10976, 1993.
- \_\_\_\_\_. Precision mapping of quantitative trait loci. **Genetics**, Washington, vol.136, p.1457-1468, 1994.
- ZOBEL, R.W.; WRIGHT, M.J.; GAUCH, H.G. Statistical analysis of a yield trial. **Agronomy Journal**, Madison, v.80, p.388-393, 1988.





## APÊNDICES



**Apêndice A - Tabela com variância do erro por ambiente para o primeiro conjunto de dados**

Tabela 9 - Variância do erro para cada um dos 20 ambientes referentes ao experimento com 9 genótipos de milho

Ambientes	Variância do erro
1	889719.50
2	253830.20
3	1108294.44
4	245652.38
5	436980.61
6	431260.57
7	423186.36
8	389509.92
9	1319059.22
10	56631.17
11	870814.23
12	444932.14
13	675513.86
14	659519.25
15	564061.68
16	185786.97
17	588679.64
18	991864.04
19	1073884.70
20	344822.77

**Apêndice B - Script do *Software R* para o primeiro conjunto de dados**

```

#biblioteca
library(wammi)
source("wammi1.R")

#dados
data<- read.csv("milho1.csv")
data[,1]<- as.factor(data[,1])
data[,2]<- as.factor(data[,2])
data[,3]<- as.factor(data[,3])
x<- data

# AMMI e W-AMMI
ErrVar<- errVar(data)[,2] ammi<- wammi111(data, Ncomp=2, ErrVar=errVar,
outpath=NULL, save.csv=TRUE, type="unweighted")
wammi<- wammi111(data, Ncomp=2, ErrVar=ErrVar, outpath=NULL,
save.csv=TRUE, type="weighted")

# Componentes Principais
biplot.data<- read.csv("AMMI IPCsbiplot.csv")[, 2 : 5]
wbiplot.data<- read.csv("WAMMI IPCsbiplot.csv")[, 2 : 5]
row.names(wbiplot.data)
wbiplot.data.u<- cbind(wbiplot.data[1:9, 3], -wbiplot.data[1:9, 4])
wbiplot.data.v<- cbind(wbiplot.data[10:29,3], -wbiplot.data[10:29,4])

# Biplots
par(mfrow=c(1, 2)) biplot(biplot.data[1:9,3:4], biplot.data[10:29,3:4], var.axes=T,
col=c("black", "red"), xlab="IPC1", ylab="IPC2", main="AMMI", xlabs = paste(1:9), ylabs
= paste(1:20))

biplot(wbiplot.data.u, wbiplot.data.v, var.axes=T, col=c("black", "red"),
xlab="IPC1", ylab="IPC2", main="W-AMMI", xlabs = paste(1:9), ylabs = paste(1:20))

```

## Apendice C - Script do *Software R* para o segundo conjunto de dados

```

#bibliotecas
library(wammi)
library(qtl)

#dados
data<- read.csv2("fen.csv", sep=",")
sim<- sim0<- sim1<- sim2<- read.cross(format="csv", file="HxT.csv")

data[,1]<- as.factor(data[,1])
data[,2]<- as.factor(data[,2])
data[,3]<- as.factor(data[,3])
data[,4]<- as.numeric(data[,4])
x<- data
head(data)

Ngen<- 141
Nenv<- 25
NChrom <- length(sim0$geno)
Ncomp<- 2
par(mfrow=c(1,2))
Max<- min(Ngen-1, Nenv-1)

#Modelo AMMI
lm.GEI<- lm(Prod Gen+Amb, data=data)
residuals.GEI<- matrix(residuals(lm.GEI), nrow = Ngen, ncol = Nenv)
svd.GEI<- svd((residuals.GEI))

biplot.GEI.u<- cbind(svd.GEI$u[,1], svd.GEI$u[,2])*matrix(rep(svd.GEI$d[1:2]
^0.5, Ngen), ncol=2, byrow=T)

biplot.GEI.v<- cbind(svd.GEI$v[,1], svd.GEI$v[,2])*matrix(rep(svd.GEI$d[1:2]^0.5,
Nenv), ncol=2, byrow=T)

biplot(biplot.GEI.u, biplot.GEI.v, var.axes=T, col=c("black", "red"),
ylabs=levels(data$Amb), xlab="IPC1", ylab="IPC2", cex=0.8) title("Modelo AMMI",
line=3)

label<-c("autovalor", "Proporção", "Proporção acumulada")
Decomposicao<-cbind(svd.GEI$d,100*(svd.GEI$d^2)/sum(svd.GEI$d^2),
100*cumsum(svd.GEI$d^2)/sum(svd.GEI$d^2))
colnames(Decomposicao)<-label

```

```

# AMMI - Valores preditos parte multiplicativa
biplot<- NULL
for(i in 1:24) biplot[[i]]<- (cbind(svd.GEI$u[,i])*matrix(rep(svd.GEI$d[i]^0.5,
Ngen), ncol=1, byrow=T)) t(cbind(svd.GEI$v[,i])*matrix(rep(svd.GEI$d[i]^0.5, Nenv), ncol=1,
byrow=T))

gm<- matrix(rep(mean(data[,4]), Ngen*Nenv), ncol=Nenv)
gdev<- matrix(rep(tapply(data[,4], data[,c(1)], mean), Nenv), byrow=F,
ncol=Nenv)
edev<- matrix(rep(tapply(data[,4], data[,c(2)], mean), Ngen), byrow=T,
ncol=Nenv)

# Valores esperados incluindo a parte aditiva
expected<- NULL
expected[[1]]<- - gm + gdev + edev + biplot[[1]]
expected[[2]]<- - gm + gdev + edev + biplot[[1]] + biplot[[2]]

# calculando pesos (pesos iguais por linha)
dd<- tapply(data[,4], data[,c(1, 2)], mean)
Rvar<- apply(dd, 1, var)
w1<- min(Rvar)/Rvar
W <- matrix(rep(w1,Nenv), ncol=Nenv)
WGEI<- c(W)

# Modelo W-AMMI
w.lm.GEI<- lm(Prod Gen+Amb, weights=WGEI , data=data)
w.residuals.GEI<- matrix(residuals(w.lm.GEI), nrow = Ngen, ncol = Nenv)

D <- w.residuals.GEI
X <- matrix(0, Ngen, Nenv)
aux <- matrix(1, Ngen, Nenv)
Xold = Inf * aux
Err = Inf
eps <- 1e-10

while (Err > eps) Xold = X wsvd <- svd(W * D + (1 - W) * X) U <- wsvd$u
S <- diag(wsvd$d) V <- wsvd$v S[(Ncomp + 1):length(wsvd$d), (Ncomp + 1):length(wsvd$d)]
<- 0 X <- U Err = sum(sum((X - Xold)^2))

U <- U
L <- diag(S)
V <- V
L <- L[1:Max]

```

```

w.biplot.GEI.u<- cbind(-U[,1], U[,2])*matrix(rep(L[1:2]^ 0.5, Ngen), ncol=2, by-
row=T)

w.biplot.GEI.v<- cbind(-V[,1], V[,2])*matrix(rep(L[1:2]^ 0.5, Nenv), ncol=2, by-
row=T)

biplot(w.biplot.GEI.u, w.biplot.GEI.v, var.axes=T, col=c("black", "red"),
ylabs=levels(data[,2]), xlab="IPC1", ylab="IPC2", cex=0.8) title("Modelo WAMMI (W
rows)", line=3)

calculando os pesos (pesos iguais por coluna)
dd<- tapply(data[,4], data[,c(1, 2)], mean)
Cvar<- apply(dd, 2, var)
w1<- min(Cvar)/Cvar
W <- matrix(rep(w1, Ngen), byrow=T, ncol=Nenv)
WGEI<- c(W)

# Modelo W-AMMI
w.lm.GEI<- lm(Prod Gen+Amb, weights=WGEI , data=data)
w.residuals.GEI<- matrix(residuals(w.lm.GEI), nrow = Ngen, ncol = Nenv)

D <- w.residuals.GEI
X <- matrix(0, Ngen, Nenv)
aux <- matrix(1, Ngen, Nenv)
Xold = Inf * aux
Err = Inf
eps <- 1e-10

while (Err > eps) Xold = X wsvd <- svd(W * D + (1 - W) * X) U <- wsvd$u
S <- diag(wsvd$d) V <- wsvd$v S[(Ncomp + 1):length(wsvd$d), (Ncomp + 1):length(wsvd$d)]
<- 0 X <- U Err = sum(sum((X - Xold) ^ 2))

U <- U
L <- diag(S)
V <- V
L <- L[1:Max]

w.biplot.GEI.u<- cbind(U[,1], U[,2])*matrix(rep(L[1:2]^ 0.5, Ngen), ncol=2, by-
row=T)

w.biplot.GEI.v<- cbind(V[,1], V[,2])*matrix(rep(L[1:2]^0.5, Nenv), ncol=2, by-
row=T)

biplot(w.biplot.GEI.u, w.biplot.GEI.v, var.axes=T, col=c("black", "red"),
ylabs=levels(data[,2]), xlab="IPC1", ylab="IPC2", cex=0.8) title("Modelo WAMMI (W co-
lumns)", line=3)

```



```

# W-AMMI valores preditos parte multiplicativa
w.biplot<- NULL
for(i in 1:Ncomp) w.biplot[[i]]<- (cbind(U[,i])*matrix(rep(L[i]^0.5, Ngen), ncol=1,
byrow=T))
t(cbind(V[,i])*matrix(rep(L[i]^0.5, Nenv), ncol=1, byrow=T))
gm<- matrix(rep(mean(data[,4]), Ngen*Nenv), ncol=Nenv)
gdev<- matrix(rep(tapply(data[,4], data[,c(1)], mean), Nenv), byrow=F,
ncol=Nenv)
edev<- matrix(rep(tapply(data[,4], data[,c(2)], mean), Ngen), byrow=T,
ncol=Nenv)

# Valores esperados incluyendo parte aditiva
w.expected<- NULL
w.expected[[1]]<- - gm + gdev + edev + w.biplot[[1]]
w.expected[[2]]<- - gm + gdev + edev + w.biplot[[1]] + w.biplot[[2]]

#QTL Scans:
twoWay <- tapply(data[,4], data[,c(1,2)], mean)
dataAMMI <- expected[[Ncomp]]
dataWAMMI <- w.expected[[Ncomp]]

# Primeiro grupo de ambientes
par(mfrow = c(ceiling(4), ceiling(4)), oma = c(0, 0, 2, 0), mar = c(5.1, 4.1, 2.1,
2.1), xaxs = "i", yaxs = "i", cex.axis = 1, cex.lab = 1, cex.main = 1.5)

zz <- 1
ENVpeaks <- data.frame(NULL)
for (i in 1:16)
sim0$pheno <- as.data.frame(twoWay[, i])
sim1$pheno <- as.data.frame(dataAMMI[, i])
sim2$pheno <- as.data.frame(dataWAMMI[, i])
sim0 <- calc.genoprob(sim0, step = 0, error.prob = 0.001)
sim1 <- calc.genoprob(sim1, step = 0, error.prob = 0.001)
sim2 <- calc.genoprob(sim2, step = 0, error.prob = 0.001)
set.seed(seed)
out.cim0 <- cim(sim0, n.marcovar = 6, method = "em")
set.seed(seed)
out.cim1 <- cim(sim1, n.marcovar = 6, method = "em")
set.seed(seed)
out.cim2 <- cim(sim2, n.marcovar = 6, method = "em")

```

```

plot(out.cim0, out.cim1, out.cim2, chr = 1:NChrom, col = c("grey", "grey",
"black"), lty = c(3, 1, 1), lwd = 0.5, main = colnames(twoWay)[i])

if (is.null(threshold)) permo <- scanone(sim0, method = "hk", n.perm = 1000)
add.threshold(out.cim2, perms = permo, alpha = 0.05)

else abline(threshold, 0) out <- c(out.cim0, out.cim1, out.cim2) if
(is.null(threshold)) ENVpeaks <- rbind(ENVpeaks, as.data.frame(summary(out, thr =
3, format = "allpeaks"))) if (dim(summary(out, thr = 3, format = "allpeaks"))[1] > 0)
row.names(ENVpeaks)[zz] <- colnames(twoWay)[i] zz <- zz + dim(summary(out, thr = 3,
format = "allpeaks"))[1] else ENVpeaks <- rbind(ENVpeaks, as.data.frame(summary(out,
thr = threshold, format = "allpeaks"))) if (dim(summary(out, thr = threshold, format =
"allpeaks"))[1] > 0) row.names(ENVpeaks)[zz] <- colnames(twoWay)[i]

zz <- zz + dim(summary(out, thr = threshold, format = "allpeaks"))[1]
write.csv(ENVpeaks, file = "ENVpeaks1.csv")

# segundo grupo de ambientes
par(mfrow = c(ceiling(3), ceiling(3)), oma = c(0, 0, 2, 0), mar = c(5.1, 4.1, 2.1,
2.1), xaxs = "i", yaxs = "i", cex.axis = 2, cex.lab = 2, cex.main = 2.5)

zz <- 1
ENVpeaks <- data.frame(NULL)
for (i in 17:25)
sim0$pheno <- as.data.frame(twoWay[, i])
sim1$pheno <- as.data.frame(dataAMMI[, i])
sim2$pheno <- as.data.frame(dataWAMMI[, i])
sim0 <- calc.genoprob(sim0, step = 0, error.prob = 0.001)
sim1 <- calc.genoprob(sim1, step = 0, error.prob = 0.001)
sim2 <- calc.genoprob(sim2, step = 0, error.prob = 0.001)
set.seed(seed)
out.cim0 <- cim(sim0, n.marcovar = 6, method = "em")
set.seed(seed)
out.cim1 <- cim(sim1, n.marcovar = 6, method = "em")
set.seed(seed)
out.cim2 <- cim(sim2, n.marcovar = 6, method = "em")

plot(out.cim0, out.cim1, out.cim2, chr = 1:NChrom, col = c("grey", "grey",
"black"), lty = c(3, 1, 1), lwd = 0.5, main = colnames(twoWay)[i]) if (is.null(threshold)) permo
<- scanone(sim0, method = "hk", n.perm = 1000) add.threshold(out.cim2, perms = permo, alpha
= 0.05)

```

```
else abline(threshold, 0) out <- c(out.cim0, out.cim1, out.cim2) if
(is.null(threshold)) ENVpeaks <- rbind(ENVpeaks, as.data.frame(summary(out, thr =
3, format = "allpeaks"))) if (dim(summary(out, thr = 3, format = "allpeaks"))[1] > 0)
row.names(ENVpeaks)[zz] <- colnames(twoWay)[i] zz <- zz + dim(summary(out, thr = 3,
format = "allpeaks"))[1] else ENVpeaks <- rbind(ENVpeaks, as.data.frame(summary(out,
thr = threshold, format = "allpeaks"))) if (dim(summary(out, thr = threshold, format =
"allpeaks"))[1] > 0) row.names(ENVpeaks)[zz] <- colnames(twoWay)[i]

zz <- zz + dim(summary(out, thr = threshold, format = "allpeaks"))[1]
write.csv(ENVpeaks, file = "ENVpeaks2.csv")
```