Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar, usando a abordagem de modelos mistos

Maria Marta Pastina

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2010

Maria Marta Pastina
Engenheira Agrônoma

Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar, usando a abordagem de modelos mistos

Orientador:
Prof. Dr. ANTONIO AUGUSTO FRANCO GARCIA

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2010

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

Pastina, Maria Marta
Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar, usando a abordagem de modelos mistos / Maria Marta Pastina. - - Piracicaba, 2010.
89 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010.

1. Cana-de-açúcar 2. Correlação genética e ambiental 3. Mapeamento genético 4. Marcador molecular 5. Melhoramento genético vegetal 6. Variação genética em plantas I. Título

CDD 633.61
P291m

Aos meus pais, **Affonsa** e **Marcus**,
pelo amor, apoio e grande contribuição
para a realização dos meus sonhos,
sem eles não chegaria tão longe. . .
Dedico.

“The best of all things is to learn. Money can be lost or stolen, health and strength may fail but what you have committed to your mind is yours forever.”
Louis L’Amour (22/03/1908 - 10/07/1988)

Agradecimentos

A Deus.

Ao Prof. Dr. Antonio Augusto Franco Garcia, pela confiança, paciência e dedicada orientação nesses dez anos de convívio. Além dos inúmeros conselhos e ensinamentos, essenciais para meu desenvolvimento científico e pessoal.

Ao CNPq, pelas bolsas de estudo concedidas (Doutorado e Doutorado Sanduíche no Exterior), que permitiram a realização deste trabalho através de uma colaboração entre ESALQ/USP e Universidade de Wageningen (Holanda).

Ao Prof. Dr. Fred van Eeuwijk da Universidade de Wageningen, por ter aceitado participar deste projeto.

Ao Prof. Dr. Marcos Malosetti da Universidade de Wageningen, pela valiosa contribuição no estudo de modelos mistos.

À FAPESP, pelo financiamento do Projeto de Iniciação Tecnológica (PITE) que permitiu a obtenção dos dados utilizados no presente trabalho, além de conceder-me a bolsa de Iniciação Científica durante a graduação. Em especial à coordenadora do projeto, Profa. Dra. Anete Pereira de Souza.

Ao Prof. Dr. Roland Vencovsky, pela orientação e ensinamentos durante meus dois primeiros anos de doutorado.

À Dra. Karine Miranda Oliveira e ao Centro de Tecnologia Canavieira (CTC), pelos dados fenotípicos concedidos para a realização deste trabalho.

À Dra. Luciana Rossini Pinto, pela grande ajuda na organização dos dados fenotípicos e pelas valiosas discussões sobre genética da cana-de-açúcar.

Ao Departamento de Genética da ESALQ/USP, pela oportunidade e excelência em ensino.

Ao Rodrigo, Marcelo e Gabriel, que muito contribuíram para o desenvolvimento da pesquisa, mesmo à distância.

Ao Marcelo, por já ter feito sua dissertação no LaTeX, o que muito me ajudou na escrita desta tese.

Aos amigos do laboratório de Genética-Estatística: Priscilla, Rodrigo, Marcelo, Gabriel, Graciela, Renato, Edjane e João Ricardo, pela grande amizade a agradável convivência.

Em especial ao Rodrigo, pela amizade e por sempre ter uma palavra de apoio nos momentos difíceis.

Aos colegas de pós-graduação, pela amizade e convívio durante o curso.

Às amigas e companheiras de moradia, Joze e Cecília, que muito me apoiaram nas últimas etapas do doutorado.

Aos meus queridos amigos da Holanda: Carol Castro, Sampaio, Carol Mosca, Taís Leite, Nicolau Bliska, Thais Córdova, Jullyana e Adriano Marques, Martha González, Carol Silva, Dennis Bijl, Camila Patreze, Clarissa Santos, Martin Boer, Dindo Tabanao, Patrícia Menendez e Sabine Schnabel, pela amizade e companheirismo.

Às minhas tias Anita, Natalina e Rita, pelas inúmeras orações em minha intenção.

Ao meu amado irmão Marquinhos, pelo carinho, amor e compreensão.

Em fim, a todos os amigos e familiares que de alguma forma fizeram parte de minha vida e contribuíram para a realização deste trabalho.

SUMMARY

RESUMO

Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar, usando a abordagem de modelos mistos

Os programas de melhoramento da cana-de-açúcar demandam aproximadamente 12 anos para a obtenção de um novo cultivar. Assim, os marcadores moleculares podem ser usados como uma ferramenta valiosa, uma vez que possibilitam o estudo da arquitetura genética de caracteres quantitativos, ajudando a reduzir este tempo. Embora a cana-de-açúcar seja uma cultura perene, para a qual o desempenho genotípico é avaliado através de ensaios estabelecidos ao longo de diferentes locais e cortes, a maior parte dos estudos de mapeamento de QTLs ignora a existência de interação entre QTLs, corte e local (QTL $\times$ H $\times$ L). Neste contexto, o presente trabalho apresenta uma estratégia que foi desenvolvida para a detecção de QTLs em cana-de-açúcar, com base em modelos mistos e mapeamento por intervalo, considerando diferentes estruturas de (co)variância que permitem supor heterogeneidade de variâncias genéticas e existência de correlações genéticas entre cortes e locais. A metodologia de modelos mistos foi aplicada aos dados de uma população segregante obtida a partir do cruzamento entre dois cultivares pré-comerciais de cana-de-açúcar, constituída por 100 indivíduos avaliados em dois locais (Piracicaba e Jaú, SP, Brasil) e em três cortes para produção (toneladas de cana por hectare, TCH), produção de açúcar (toneladas de Pol por hectare, TPH), porcentagem de fibra e Pol (teor de sacarose). A análise fenotípica resultou na seleção do modelo não-estruturado, que assume heterogeneidade de variâncias e existência de correlação genética específica para cada combinação de corte e local, para todos os caracteres avaliados. Na análise de mapeamento, foram detectados 50 QTLs, incluindo 14 QTLs para TCH, 15 para TSH, 10 para Pol e 11 para Fibra. Além disso, os resultados mostram que os efeitos das interações entre QTL e corte (QTL $\times$ H), QTL e local (QTL $\times$ L) e QTL, corte e local (QTL $\times$ H $\times$ L) foram importantes para todos os caracteres avaliados. Do total de QTLs identificados, 33 (66 %) apresentaram algum tipo de interação e apenas 17 (34 %) mostraram mesmo efeito entre as diferentes combinações de corte e local. Estes resultados fornecem informações importantes para o entendimento da base genética de caracteres quantitativos relacionados com produção e teor de sacarose em cana-de-açúcar.

Palavras-chave: Poliplóides; Progênie de irmãos completos; Mapa genético integrado; Análise multiponto; Mapeamento por intervalo; QTL $\times$ E

ABSTRACT

A mixed-model QTL analysis for sugarcane multiple-harvest-location trial data

Sugarcane breeding programs take at least twelve years to develop new commercial cultivars. Thus, molecular markers can be used as a valuable tool since they offer the possibility to study the genetic architecture of quantitative traits, helping to reduce this time. Although the performance of genotypes in sugarcane breeding programs has been evaluated across a range of locations and harvest years, since sugarcane is a perennial crop, many of the QTL detection methods ignore QTL by harvest by location interaction (QTL $\times$ H $\times$ L). In this work, a strategy for QTL detection in sugarcane was developed, based on mixed models and interval mapping, considering different (co)variance structures for the modeling of heterogeneous genetic variances and genetic correlations between harvests and locations. The mixed model approach was applied to a data set provided by a segregating population developed from a cross between two pre-commercial Brazilian cultivars, consisted of 100 individuals planted in two locations in 2003 (Piracicaba and Jaú, SP, Brazil) and evaluated in the first, second and third subsequent harvest years for cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), fiber percent and Pol (sucrose content). Phenotypic analysis provided the selection of the unstructured model, which allows the assumption of heterogeneity of variance and presence of a specific genetic correlation for each combination of harvest and location. In the QTL mapping procedure, 50 QTLs were detected, including 14 QTLs for TCH, 15 for TSH, 10 for Pol and 11 for Fiber. In addition, the results show that QTL by harvest (QTL $\times$ H), QTL by location (QTL $\times$ L) and QTL by harvest by location (QTL $\times$ H $\times$ L) interaction effects were important for all evaluated traits. From the total of QTLs identified, 33 (66%) had some interaction and only 17 (34%) showed stable effects across the different combinations of harvest and location. These results can provide useful information to understand the genetic control of complex traits related with sugarcane production and sucrose content.

Keywords: Polyploids; Full-sib progeny; Integrated linkage map; Multipoint analysis; Interval mapping; QTL $\times$ E

1 INTRODUCTION

Sugarcane (*Saccharum* spp.) is a clonally propagated outcrossing polyploid crop of great importance in tropical agriculture, as a source of sugar and ethanol. Modern commercial sugarcane cultivars are derived from interspecific crosses, followed by few cycles of intercrossing and selection, between *Saccharum officinarum* ($x = 10$, $2n = 8x = 80$) and its wild relative *S. spontaneum* ($x = 8$, $2n = 5 - 16x = 40 - 128$), with chromosome number in somatic cells ($2n$) ranging from 100 to 130 (D'HONT et al., 1998; IRVINE, 1999; GRIVET; ARRUDA, 2001).

Quantitative Trait Loci (QTL) mapping is a useful tool for having a better understanding of the genetic architecture of quantitative traits, which are difficult to handle and are of main importance for breeding. Several reasons make QTL mapping more complicated and challenging in sugarcane than in other species. First, the high level of polyploidy and aneuploidy results in a complex pattern of chromosomal segregation at meiosis (HEINZ; TEW, 1987). Second, linkage map construction and QTL mapping rely on segregating progenies derived from bi-parental crosses between highly heterozygous outbred parents, since inbred lines are not available. Therefore, each loci (marker or QTL) could have a different number of alleles, resulting in a mixture of different segregating patterns in the progeny (GARCIA et al., 2006; OLIVEIRA et al., 2007; LIN et al., 2003). Moreover, linkage phases between markers and QTL are also unknown.

The development of genetic linkage maps in sugarcane started with segregation analysis of single dose markers (SDMs) (WU et al, 1992). They correspond to alleles present at one copy (dose) in one of the parents or at one copy in both parents segregating, at 1:1 (presence : absence) or 3:1 (presence : absence) ratio, respectively, in the progeny. SDMs can be used to build genetic maps in any cross between heterozygous individuals with bivalent pairing at meiosis, commonly using the *double pseudo-testcross* strategy (GRATTAPAGLIA; SEDEROFF, 1994; PORCEDDU et al., 2002; SHEPHERD et al., 2003; CARLIER et al., 2004; CAVALCANTI; WILKINSON, 2007; CHEN et al., 2008), which provides two separated maps, one for each parent. In spite of the relative success obtained with the *double pseudo-testcross* strategy in sugarcane (for example, Al-Janabi et al., 1993; Ming et al., 1998; Hoarau et al., 2001; McIntyre et al., 2005b), the use of integrated maps combining 1:1 and 3:1 segregation (GARCIA et al, 2006; OLIVEIRA et al., 2007) presents several advantages, as they allow better saturation and characterization of the polymorphic variation in the

genome, which could provide a better framework for QTL mapping.

For QTL mapping in outcrossing species, a limited number of statistical methods was described (SONG; SOLLER; GENIZI, 1998; JOHNSON; JANSEN; ARENDONK, 1999; LIN et al., 2003). For sugarcane, single marker analysis (SM), interval mapping (IM) and/or composite interval mapping (CIM) are commonly used, considering the two maps obtained using the *double pseudo-testcross* strategy (SILLS et al., 1995; DAUGROIS et al., 1996; HOARAU et al., 2002; JORDAN et al., 2004; SILVA; BRESSIANI, 2005; McINTYRE et al., 2005a; REFFAY et al., 2005; AITKEN et al., 2008; RABOIN et al., 2006; AL-JANABI et al., 2007; PIPERIDIS et al., 2008; PINTO et al., 2009). In this approach, statistical analyses are carried out through well stablished models for backcrosses, using softwares developed for inbred-based populations. However, these models were not developed for outcrossing species with integrated maps.

In addition to its genetic complexity, sugarcane is a perennial crop, in which individual plants have several harvests. Thus, traits are repeatedly measured not only across different locations, but also along successive harvests, adding a time dimension to the phenotypic data. Varietal selection for quantitative traits in sugarcane is usually based on information from a series of field trials, considering different harvests and locations, called here multi-harvest-location trials (MHLT). QTL studies in sugarcane usually ignore QTL by harvest by location (QTL $\times$ H $\times$ L) interaction, considering data only from each harvest and location combination, one at a time in separated analysis (HOARAU et al., 2002; JORDAN et al., 2004; McINTYRE et al., 2005a; REFFAY et al., 2005; PINTO et al., 2009). The use of statistical models that allow the identification of consistent QTL across different environmental conditions, locations and years, can provide powerful and useful results for breeding purposes, with possibility of application in marker assisted selection (MAS).

Mixed models have been successfully employed in studies of genotype-by-environment (G $\times$ E) interaction (DENIS; PIEPHO; EEUWIJK, 1997; PIEPHO, 1997; CULLIS et al., 1998; CHAPMAN, 2008; SMITH; CULLIS; THOMPSON, 2001; SMITH et al., 2007), as well as for QTL by environment (QTL $\times$ E) interaction (PIEPHO, 2000, 2005; VERBYLA et al., 2003; MALOSETTI et al., 2004, 2008; EEUWIJK et al., 2005; BOER et al., 2007; MATHEWS et al., 2008). Since they provide great flexibility to model the complex variance-covariance structure in the data caused by genetic correlations between harvests and locations, they were used in this work to develop a strategy for QTL mapping in sugarcane using integrated maps. Different models were then adjusted to sugarcane data and the results are presented and discussed.

2 REVIEW

2.1 General Aspects of Sugarcane Breeding

Sugarcane (*Saccharum* spp.) is one of the most important industrial crop. It is used as source of sucrose and ethanol, in addition to several fiber products (for example: paper, cardboard and fiber board). Although sugarcane-growing areas are mostly located between the tropics, it is cultivated in more than 100 countries all over the world (FOOD AND AGRICULTURE ORGANIZATION - FAO, 2008). Brazil is the world's largest producer. For the 2009/2010 harvest year, it is estimated that 629 million tonnes of sugarcane will be produced in an area around 7.74 million hectares, representing an increase of 10% over the previous brazilian harvest (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2009). The development of flex-fuel vehicles (which can use either ethanol or gasoline), the reduction of the world petroleum reserves and the climate changes caused by the greenhouse effect, granted the expansion of the worldwide ethanol demand. Moreover, it is expected that the decline of the European Union (EU) sugar exports and the reduction of the sugar production in India, will provide new commercial opportunities for Brazil (FNP - CONSULTORIA & COMÉRCIO, 2008). In this respect, plant breeding has a significant role to increase productivity and allow the expansion of sugarcane production in marginal areas, due to its importance in the selection of more adapted cultivars with good agronomic performance.

The *Saccharum* complex includes the *Saccharum* genus and several related genera. Sugarcane is a member of Poaceae family and Andropogoneae tribe, like maize and sorghum. Many authors mention the occurrence of six polyploid outcrossing species in the genus *Saccharum*: two wild species, *S. spontaneum* Linnaeus ($2n = 40 - 128$) and *S. robustum* Brandes and Jeswiet ex Grassl ($2n = 60 - 205$), and four cultivated species, *S. officinarum* Linnaeus ($2n = 80$), *S. barberi* Jeswiet ($2n = 81 - 124$), *S. sinense* Roxb. ($2n = 111 - 120$) and *S. edule* Hassk. ($2n = 60 - 80$). However, morphological and molecular evidences suggest that these six species could be more properly represented by only two: one being *S. spontaneum*, and the other *S. officinarum* including the other four species and all interspecifc hybrids (IRVINE, 1999). *S. officinarum* is known as noble cane because of its splendid appearance, bright colors, broad leaves, thick stems, high sucrose and low fiber content. *S. spontaneum* is a wild and vigorous highly polyploid relative, characterized by the high capacity for adaptation to different environmental conditions. The latter species gave many

contributions for the development of the current commercial cultivars, due to its desirable agronomic characteristics, such as tillering, high capacity for regrowth, resistance to pests and diseases. In addition, other species such as *S. robustum*, *S. sinense* and *S. edule* were widely used as varieties in breeding programs, and nowadays, represent an important source of variability (MATSUOKA; GARCIA; ARIZONO, 1999).

Cultivated sugarcane had two geographic centers of origin, New Guinea and South East Asia. The history of domesticated sugarcane is not well known. *S. officinarum*, the tropical species, is probably originated from the wild species *S. robustum* in the New Guinea region. *S. spontaneum* has the widest geographic range, extending across three geographic zones: i) The East Zone, which includes South Pacific Islands, the Philippines, Taiwan, Japan, China, Vietnam, Thailand, Malaysia and Burma; ii) the Central Zone, which includes India, Nepal, Bangladesh, Sri Lanka, Pakistan, Turkmenistan, Afghanistan, Iran and Middle East; and iii) the West Zone, which includes Egypt, Sudan, Kenya, Uganda and Tanzania (CLAYTON; DANIELS, 1975 apud AITKEN; McNEIL, ca. 2010). Some authors suggested that *S. sinense* and *S. barberi* were originated through interspefic crosses between *S. officinarum* and *S. spontaneum*, followed by sucessive backcrosses. However, the current commercial cultivars are hybrids derived from different species (DANIELS; ROACH, 1987 apud MATSUOKA; GARCIA; ARIZONO, 1999).

Early sugarcane breeding programs started with the obtention of interspecific hybrids between *S. officinarum* and *S. spontaneum*, and then repeatedly backcrossing the hybrids to *S. officinarum*. This process is termed ‘nobilization’ and was performed mainly to recover the high sugar-producing ability of *S. officinarum* and to minimize the negative effects of *S. spontaneum*, resulting in improved cane yields, ratooning ability and increased resistance to biotic and abiotic stresses (ALWALA; KIMBENG, ca. 2010). However, the genetic contribution of each parent was not proportional, i.e., the maternal parent had a higher contribution to the number of chromosomes in these new materials, resulting in hybrids with high levels of ploidy and aneuploidy (presence of a distinct number of chromosomes among the different homology groups) (HEINZ; TEW, 1987). Thus, modern sugarcane is highly polyploid, in which the number of homologous and homoeologous chromosomes can vary among genotypes from the same cross. The genome composition of modern sugarcane commercial cultivars is about 70-80% *S. officinarum*, 10-20% *S. spontaneum* and 5-17% of recombinant chromosomes (D’HONT et al., 1996; GRIVET; ARRUDA, 2001). The main consequences of this phenomenon were the reduction of genetic diversity and increased complexity of the sug-

arcane genome. Particularly, this high genetic complexity makes plant breeding more challenging and complicated.

In general, sugarcane breeding is based on the selection and cloning of superior genotypes from a segregating population obtained by biparental or multiparental crosses, the latter also named polycrosses. Most of the agronomic important traits, such as sugarcane production, fiber content, stalk diameter, stalk length, stalk weight, among several others, are of quantitative nature, i.e., controlled by many genes whose expression is highly influenced by environmental action. In this case, selections made in the early stages can not be done at high intensity, because there are few seedlings of each genotype, making phenotypic measures susceptible to the environmental conditions and not consistent with the real genotypic value for the individual (characters of median and low heritability), affecting the selection of promising materials. Thus, the release of a new variety only occurs when there are many experimental results from different locations and years of cultivation. Therefore, a breeding program demands about 12 years to release new sugarcane cultivars (MATSUOKA; GARCIA; ARIZONO, 1999).

Sugarcane breeding programs can take much advantage by the development of methods that allow an early and efficient selection of superior genotypes. The recent development of techniques to detect and use molecular markers allowed a better understanding of the breeding process, since molecular markers provide information at the DNA level. Thus, several studies have been performed in sugarcane using molecular markers, including genetic diversity (LIMA et al., 2002), construction of genetic linkage maps (AITKEN; JACKSON; McINTYRE, 2005, 2007; RABOIN et al., 2006; GARCIA et al., 2006; OLIVEIRA et al., 2007), QTL mapping by linkage analysis (McINTYRE et al., 2006; AL-JANABI et al., 2007; PIPERIDIS et al., 2008; AITKEN et al ., 2008; PINTO et al., 2009), and also by association mapping (WEI; JACKSON; McINTYRE, 2006; RABOIN et al., 2008). QTL mapping can be an important tool as it allow a better knowledge about the genetic architecture of quantitative traits, contributing for MAS (DEKKERS; HOSPITAL, 2002).

2.2 Genetic Mapping

2.2.1 Molecular Markers and Linkage Maps

A genetic marker provides information about allelic variation at a given locus. In the first studies, genetic markers were based on the products of gene expression, being named morphological mark-

ers. Due to the fact that these markers can be influenced by the environment and/or by the action of other genes, they were of limited use. Then, the development of DNA-based molecular markers circumvented these problems, resulting in a wide use of these types of genetic markers to increase the knowledge about the genome constitution and genetic architecture of many plant species including several major crops (SCHLÖTTERER, 2004). Molecular markers, such as Restriction Fragment Length Polymorphism (RFLP), Expressed Sequence Tag RFLP (EST-RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeat (SSR) and Expressed Sequence Tag SSR (ESR-SSR), have been widely used in sugarcane to build genetic linkage maps (AL-JANABI et al., 1993; SILVA et al., 1993, 1995; D'HONT et al., 1994; GRIVET et al., 1996; MUDGE et al., 1996; MING et al., 1998, 2002b; GUIMARÃES; SILLS; SOBRAL, 1997; GUIMARÃES et al., 1999; HOARAU et al., 2001; AITKEN; JACKSON; McINTYRE, 2007; RABOIN et al., 2006; GARCIA et al., 2006; OLIVEIRA et al., 2005, 2007) and for QTL mapping (SILLS et al., 1995; DAUGROIS et al., 1996; MING et al., 2001, 2002a, 2002c; HOARAU et al., 2002; JORDAN et al., 2004; SILVA; BRESSIANI, 2005; McINTYRE et al., 2005a, 2005b, 2006; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2006; AITKEN et al., 2008; RABOIN et al., 2006, 2008; WEI; JACKSON; McINTYRE, 2006; AL-JANABI et al., 2007; PIPERIDIS et al., 2008; PINTO et al., 2009). Genetic linkage maps can be useful for gene tagging (SOBRAL; HONEYCUTT, 1993), map-based cloning (DIETRICH et al., 1996), QTL mapping (DOERGE; ZENG; WEIR, 1997) and supported MAS in plant breeding programs (TAKEDA; MATSUOKA, 2008).

Contrary to other crops, such as maize and soybean, in sugarcane it is not possible to generate inbred lines, mainly due to the high inbreeding depression that occurs when plants are selfed. However, as sugarcane is highly heterozygous and polyploid, segregation can be observed already at the first generation of a cross. Thus, conventional mapping in sugarcane generally relies on first generation progenies derived from biparental crosses (full-sib progenies). Several reasons make genetic mapping more laborious and difficult in polyploid species: i) high level of polyploidy and aneuploidy, resulting in different configurations of random pairing for homologous chromosomes at meiosis; ii) distinct patterns of marker segregation, not observable in diploid species, are expected in polyploids, such as 7:1, 7:2, 11:1, 11:3, 13:1, 15:1, 64:1 and 69:1, as a consequence of the different allele dosages (EDMÉ; GLYNN; COMSTOCK, 2006); iii) the genome constitution, for some polyploid species, is still unclear (in sugarcane, for example, the genomes of *S. officinarum*

and *S. spontaneum* occur in different proportions in the genomes of *S. barberi*, *S. sinense* and cultivated sugarcane); iv) statistical models are more complicated for polyploid than for diploid species (ALWALA; KIMBENG, ca. 2010).

Genetic mapping in sugarcane started after Wu et al. (1992) proposed a general strategy for mapping in highly polyploids with bivalent pairing, based on segregation of SDM. In this method, only markers present at one copy (dose) in each parent or at one copy in both parents, segregating into 1:1 (presence:absence) or 3:1 (presence:absence) ratio in the progeny, respectively, are able to be mapped. SDM opened the opportunity to start the construction of genetic linkage maps not only in sugarcane but also in several other crops, including diploids (PORCEDDU et al., 2002; SHEPHERD et al., 2003; CARLIER et al., 2004; CAVALCANTI; WILKINSON, 2007; CHEN et al., 2008). This approach is also known as *pseudo-testcross* or *double pseudo-testcross*, when involving one or both parents respectively (GRATTAPAGLIA; SEDEROFF, 1994), in which the testcross mating configuration is not known *a priori*, but inferred *a posteriori* from segregation analysis on the progeny. Two separated maps are obtained, one for each parent. Although SDM (or simplex markers) are present at high frequencies in the sugarcane genome (approximately 70%), duplex markers (present at two copies in only one parent) can also be used to map regions with low levels of simplex markers, and thus increase map coverage (AITKEN; JACKSON; McINTYRE, 2007).

Despite the many results obtained with the *double pseudo-testcross* strategy in sugarcane, the construction of integrated maps can be done using the information provided by SDM present in one copy in both parents, segregating into 3:1 ratio (GARCIA et al., 2006; OLIVEIRA et al., 2007), allowing better saturation and the characterization of the polymorphic variation in the whole genome, which could generate better results for QTL mapping. Several authors presented alternatives for the construction of integrated genetic maps in populations derived from crosses between non-inbred parents (RITTER; GEBHARDT; SALAMINI, 1990; RITTER; SALAMINI, 1996; MALIEPAARD; JANSEN; OOIJEN, 1997; RIDOUT et al., 1998; RIPOL et al., 1999; WU et al., 2002a, 2002b; GARCIA et al., 2006; MARGARIDO; SOUZA; GARCIA, 2007; OLIVEIRA et al., 2007). Wu et al. (2002a, 2002b) have developed approaches based on maximum likelihood to simultaneously estimate recombination fraction and linkage phase between markers. A method based on multipoint maximum likelihood, using Hidden Markov Models (HMM), was presented by Wu et al (2002b).

Based on the *double pseudo-testcross* strategy, the first sugarcane genetic maps were developed

in order to understand the organization of sugarcane genome and also to investigate on the best mapping population type needed to maximize the acquisition of SDM (SILVA et al., 1993; D'HONT et al., 1994; GRIVET et al., 1996) rather than to map QTLs. Silva et al. (1993) and Al-Janabi et al. (1993) published the first results for genetic mapping in sugarcane using, respectively, RFLP and RAPD markers, segregating in an 1:1 ratio. Both markers were detected in progenies from the cross between the doubled-haploid ADP85-0068 (female parent) derived through anther culture from the *S. spontaneum* clone SES208, which was also used as the male parent. Later, data from both marker types were joined into a single map, resulting in 64 linkage groups with a genome coverage of 3,300 cM (SILVA et al., 1995).

Grivet et al. (1996), considering isozyme and RFLP markers segregating in an 1:1 fashion, constructed a genetic map for a selfed progeny derived from the cultivar R570. This map comprised 408 linked markers placed onto 96 co-segregation groups, assembled into the 10 basic linkage groups. RFLP markers derived from maize probes allowed the investigation of synteny and colinearity between homo(eo)logous co-segregation groups and species origin (*S. officinarum* or *S. spontaneum*), provinding insights about the genome organization of a sugarcane cultivar. Later, as R570 cultivar is resistant to brown rust (*Puccinia melanocephala*), its selfed progeny was evaluated for rust resistance segregation (DAUGROIS et al., 1996) as also for tagging of a major gene responsible for resistance (ASNAGHI et al., 2001). Then, R570 selfed progeny was extended to 295 individuals and used to construct a reference genetic map based on AFLP markers. This map covered 5,849 cM, representing approximately 1/3rd of the sugarcane genome length, and was considered as the most saturated sugarcane map of that time (HOARAU et al., 2001). The variety R570 was also used in crosses with MQ76-53, an old Australian clone derived from a cross between Trojan and SES528 (RABOIN et al., 2006), and with the yellow spot (*Mycovellosiella koepkei*) resistant sugarcane variety M134/75 (AL-JANABI et al., 2007). The map obtained for MQ76-53 (R570 x MQ76-53) was also used to identify a gene controlling the red stalk color and a new brown rust resistance gene (RABOIN et al., 2006) while the M134/75 map (M134/75 x R570) was used to determine the number and location of QTL for resistance to yellow spot (AL-JABABI et al., 2007).

Interespecific crosses involving *S. officinarum* (La Purple) and its supposed progenitor species *S. robustum* (Mol 5829) (MUDJE et al., 1996; GUIMARÃES et al., 1999), *S. officinarum* (Green German) and *S. spontaneum* (IND 81-146), and between *S. spontaneum* (PIN84-1) and *S. officinarum* (Muntok Java) (MING et al., 2002b) allowed the construction of six genetic linkage maps,

one for each parent. These maps were used on comparative mapping studies among the Andropogonae tribe, mainly among sugarcane, maize and sorghum, and this was of great contribution to sugarcane, where sorghum linkage maps were used to understand the complex sugarcane genome (MING et al., 2002a). Another interspecific cross involving the progenitor species of cultivated sugarcane, *S. officinarum* (La Striped) and *S. spontaneum* (SES147B), were used to construct genetic linkage maps for each progenitor species using either simplex and duplex markers through *JoinMap* software (OOIJEN; VOORRIPS, 2001), considering AFLP, Sequence Related Amplified Polymorphism (SRAP) and Target Region Amplification Polymorphism (TRAP) (ALWALA et al., 2008).

Another linkage map was built by Aitken, Jackson and McIntyre (2005) in a population derived from the cross betweeen *S. officinarum* clone IJ76-514 (female parent) and variety Q165 (male parent). The main goal was to construct a high coverage genetic map of the variety Q165, based on information provided by AFLP and SSR markers, highlighting the utility of SSRs to allocate the linkage groups to homology groups and to compare linkage maps. Thus, genetic maps from different cultivars are useful to reveal different chromosome arrangements, having a great impact on the use of molecular markers for sugarcane breeding. Due to the important contribution of *S. officinarum* genome to the commercial sugarcane varieties, the cross between IJ76-514 and Q165 was also used to construct a map for IJ76-514, integrating simplex (1:1 and 3:1 segregation ratio) and duplex (11:3 segregation ratio) markers (AITKEN; JACKSON; McINTYRE, 2007).

Garcia et al. (2006), considering a full-sib family derived from the cross between the pre-commercial Brazilian varieties SP80-180 and SP80-4966, constructed a single (integrated) map based on the simultaneous maximum-likelihood estimation of linkage and linkage phases approach developed by Wu et al. (2002a). The integrated genetic map obtained with this approach gave rise to 357 linked markers (RFLP, SSP and AFLP) assigned to 131 co-segregation groups and had a total length of 2,602.4 cM. Later, expressed sequence tag (EST) derived markers obtained from the SUCEST database, EST-SSRs and EST-RFLPs, were added to the SP80-180 and SP80-4966 previous map. This genetic linkage map containing function-associated markers had 664 single dose markers distributed into 192 co-segregation groups and a total length of 6,261.1 cM (OLIVEIRA et al., 2007).

2.2.2 QTL Mapping

Molecular markers can provide useful information in several ways for plant breeding, especially through QTL mapping, which allow a better knowledge of the genetic architecture of quantitative traits. One of the most important applications of this type of study is the possibility to develop breeding methods that incorporate marker and phenotypic information. Moreover, it also allows a better understanding about the genetic correlation among traits (JIANG; ZENG, 1995; MACKAY, 2001; MALOSETTI et al., 2008), the interaction between genotypes and the environment (BOER et al., 2007; EEUWIJK et al., 2005; EEUWIJK; MALOSETTI; BOER, 2007; MALOSETTI et al., 2004), the genetic basis of heterosis (GARCIA et al., 2008), as well as to permit the determination of the genetic value of individuals in MAS (MOHAN et al., 1997; MORGANTE; SALAMINI, 2003; CHARCOSSET; MOREAU, 2004; TAKEDA; MATSUOKA, 2008).

Since the beginning of the twentieth century, a large number of researchers have found significant associations between mendelian markers (qualitative) and quantitative traits for different crop species (SAX, 1923; LINDSTROM, 1924, 1931; SMITH, 1937; RASMUSSON, 1927; WEXELESEN, 1933; GREEN, 1931, 1933). Such studies show that QTL can be mapped with some precision, since a population with genetic variability and highly linkage disequilibrium (LD), and suficient number of markers are available. Initially, a genetic map is built to serve as a basis to locate QTL. One of the major limitations of QTL mapping in outcrossing species, like sugarcane, is the construction of a saturated linkage map (LYNCH; WALSH, 1998).

QTL mapping means finding genomic regions that are associated with phenotypic expression, estimate their effects, gene action, incorporate number of loci and interaction among them and with environment (ZENG; KAO; BASTEN, 1999). It is based on the establishment of relations between the phenotype (expression of quantitative traits) and the genotype (evaluated with molecular markers). For doing this, sophisticated statistical methodologies are commonly used, relied on a strong computational support due to the complexity of the analysis. Such models include single marker (SM) analysis (SOLLER; BRODY; GENEZI, 1976; WELLER, 1986; EDWARDS; STUBER; WENDEL, 1987; STUBER; EDWARDS; WENDEL, 1987), interval mapping (IM - LANDER; BOTSTEIN, 1989), composite interval mapping (CIM - ZENG, 1993, 1994; JANSEN; STAM, 1994), Bayesian inference (SATAGOPAN et al., 1996; HEALTH, 1997; SILLANPAA; ARJAS, 1998; YI; XU, 2001; YI et al., 2005, 2007a, 2007b), multiple interval mapping (MIM - KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999) and the mixed models approach (MALOSETTI

et al., 2004, 2008; EEUWIJK et al. 2005; EEUWIJK; MALOSETTI; BOER, 2007; BOER et al., 2007). In the case of sugarcane, as will be presented, due to its polyploid nature and high genomic complexity, SM analysis is widely used, as well as, IM or CIM, based on genetic maps of each parent (*double pseudo-testcross* strategy).

The majority of the experimental crosses used for the construction of genetic maps and QTL mapping in plants are based on populations derived from crosses between inbreed lines, such as RILs, backcross and F_2. For these populations, statistical methods are well established and implemented in several softwares, such as MAPMAKER/EXP (LANDER et al., 1987) for genetic map construction and QTL-Cartographer (BASTEN; WEIR; ZENG, 2005) and WinQTL-Cartographer (WANG; BASTEN; ZENG, 2007) for QTL mapping. Thus, considering these methodologies, QTL mapping were performed for several plant species, such as maize (CARDINAL et al., 2001; SIBOV et al., 2003; MANGOLIN et al., 2004; ZHAO et al., 2006; SABADIN et al., 2008; WASSOM et al., 2008a,b), wheat (ABATE; LIU; McKENDRY, 2008; MACCAFERRI et al., 2008), rice (CHO et al., 2007), (SEMAGN et al., 2007) and soybean (LI; PFEIFFER; CORNELIUS, 2008).

However, for sugarcane, as already discussed in section 2.2.1, it is impractical or even impossible to obtain inbreed lines. In this case, none of the sophisticated models could be directly used, making difficult the obtention of good mapping results. A commonly used type of population in sugarcane is obtained from the crosses between non-inbred parents. Thus, a single locus could present several segregating alleles having different patterns of segregations for the markers (WU et al. 2002a) and QTLs (LIN et al. 2003). Moreover, the linkage phases between markers and QTLs are unknown. Specifically for sugarcane, the high level of polyploidy brings additional problems, due to the complex pattern of chromosomal segregation at meiosis (HEINZ; TEW, 1987). Therefore, an approach that has being used is based on the mapping of quantitative trait allele (QTA), in which significant associations between the phenotypic variation observed for the trait of interest and the different alleles that can be segregating for a specific locus are investigated. The effects of these QTAs in polyploid species might be smaller than those observed for diploid species, mainly due to the high number of segregating alleles per locus for the target trait (AITKEN et al., 2008).

2.2.2.1 Statistical Models

SM analysis, widely used for QTL mapping in sugarcane, is based on the comparison between trait means of different marker genotype classes. This can be easily implemented through t-tests,

analysis of variance, simple and multiple linear regression, maximum likelihood analysis, among others. Such analyses test the null hypothesis (H_0) that the observed phenotypic values are independent of the genotype at a particular marker (if the marker is unlinked to the putative QTL). If the null hypothesis is rejected, it is assumed that there is a putative QTL associated with the marker for the quantitative trait (DOERGE; ZENG; WEIR, 1997; LIU, 1998; LYNCH; WALSH, 1997; DOERGE, 2002). The main advantages of this method are: simplicity and fast speed of execution; it can be carried out through widely used statistical softwares, such as SAS (SAS INSTITUTE, 1989) and R (R DEVELOPMENT CORE TEAM, 2009). In addition, it can be easily extended to multiple loci models and does not need a genetic linkage map established for the population, enabling the inclusion of unlinked markers (which is common in the sugarcane linkage maps). However, the statistical tests are biased, because there is a bias due to the recombination fraction between the QTL effect and its distance from the marker, resulting in a low power to detect QTL when the markers do not completely cover the genome and/or when a small sample size is considered (DOERGE, 2002). Furthermore, it is not possible to infer about the location of the QTL.

In order to avoid some of these limitations, Lander and Botstein (1989) proposed the IM method, which uses information from a pair of adjacent molecular markers to make inferences about the existence of a putative QTL at each position within the interval between these markers. This approach combines SM analysis with statistical methods based on maximum likelihood. As the genotype of the QTL is not observable directly, appropriate mixture models are used, based on conditional probabilities, allowing to estimate separately the parameters involved in the analysis (effect and QTL position), increasing the power of the tests (LANDER; BOTSTEIN, 1989; LYNCH; WALSH, 1998). IM requires an estimated genetic map as a framework for the localization of QTL. Although the IM allows estimating the effects and positions of putative QTLs, it also has some limitations, since this method uses information only from two markers at a time, not considering the interference of other QTLs located outside the mapping interval, which can result in the detection of false positives (DOERGE, 2002).

Zeng (1993, 1994) and Jansen and Stam (1994) extended IM using an approach named CIM, which is a model that also includes markers outside the mapping interval as cofactors. This makes possible to control the effects of other QTL(s) outside the mapping interval, combining IM with a multiple regression model. The inclusion of markers with a significant effect in the quantitative trait decrease the residue of the model, increasing the statistical power. In comparison with SM

analysis, IM and CIM have several advantages, such as the lack of confounding between effect and position of the putative QTL, and higher statistical power (DOERGE; ZENG; WEIR, 1997). However, IM and CIM also have limitations, overcome by other methods, such as MIM (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999), which in addition to more accuracy to estimate genetic parameters, allows to map simultaneously all QTL controlling the quantitative trait and the identification of interactions between QTLs (epistasis). To our knowledge, MIM have not yet been used in sugarcane.

Considering that for sugarcane in general a genetic map for each parent is estimated based on markers segregating in an 1:1 ratio, QTL mapping is usually done through statistical methods developed for backcross, such as SM analysis and IM or CIM. Such analyses are implemented in softwares developed for experimental populations, such as MAPMAKER/QTL (PATERSON et al., 1988), QTLCartographer (BASTEN; WEIR; ZENG, 2005), WinQTL-Cartographer (WANG; BASTEN; ZENG, 2007), R/qtl (BROMAN et al., 2003), QTL Express (SEATON et al., 2002), PLABQTL (UTZ; MELCHINGER, 2003), QTX (MANLY; CUDMORE; MEER, 2001) among others. Some authors were able to map QTLs using these approaches for different traits, such as sugarcane brown rust resistance (DAUGROIS et al., 1996; McINTYRE et al., 2005b; RABOIN et al., 2006), flourish, plant height (MING et al., 2002a), sugar production, weight and stalk number, fiber content (MING et al. 2001, 2002c; HOARAU et al. 2002; JORDAN et al. 2004; da SILVA; BRESSIANI, 2005; AITKEN; JACKSON; McINTYRE, 2006). However, it is important to mention that the *double pseudo-testcross* strategy, has some disadvantages including: i) reduction of the genome coverage, because normally only markers segregating in an 1:1 ratio are used to build the map (other segregation types such as 3:1 are sometimes included, but not making usage of modern multipoint features for map construction); ii) as a consequence, low-density genetic linkage maps are obtained; iii) reduced statistical power; iv) difficulty to interprete the results, since the mapping should refer to the mapping population, rather than for each parent; v) use of non-appropriate statistical models for QTL mapping.

In terms of statistical analysis, a limited number of methods are available for QTL mapping in full-sib progenies, obtained from a cross between two heterozygous parents (SONG; SOLLER; GENIZI, 1999; JOHNSON; JANSEN; ARENDONK, 1999; LIN et al., 2003). Based on the IM approach, Lin et al. (2003) presented a statistical method that allows QTL mapping, considering information from molecular markers with different segregation types and an integrated genetic

linkage map. However, this method has some timitations: i) the conditional probabilities for the genotypes of the QTL are not based on multipoint estimates; ii) computational difficulties to estimate the linkage phase between QTL and markers using the EM algorithm based on the mixture model. Moreover, this model and other similar ones are not able to provide useful QTL mapping results, since they do not allow the study of QTL by environment (QTL $\times$ E) interaction which is of core importance for breeding purposes.

2.2.2.2 QTL Mapping in Sugarcane

A comprehensive review of QTL mapping in sugarcane is presented by Pastina et al. (ca. 2010). Some of their results will be discussed here.

QTL mapping in sugarcane is usually done considering a segregating population obtained from biparental crosses, such as those populations used for genetic mapping, for example: interspecific crosses between *S. officinarum* and *S. robustum*, *S. officinarum* and *S. spontaneum*, self-fertilization and biparental crosses between commercial materials. The most studied traits are: brown rust (*Puccinia melanocephala*) resistance; smut (*Ustilago scitaminea*) resistance; yellow spot (*Mycovellosiella koepkey*) resistance; flowering time; sugar yield; stalk length; stalk weight; stalk diameter; stalk number; and fiber content. Early studies, in general, considered only markers segregating in an 1:1 ratio. More recently, there was the inclusion of markers segregating in a 3:1 ratio or even multiplexes (markers present in more than one copy in one or both parents). However, the mapping strategy remains basically the same for the majority of the studies: *double pseudo-testcross* for the construction of genetic maps, taking into account only information from markers segregating in an 1:1 ratio, and SM analysis to detect QTLs associated with different traits of agronomic importance in sugarcane.

The first results of QTL mapping in sugarcane were presented by Sills et al. (1995), considering RAPD markers and a progeny obtained from the cross between *S. officinarum* (LA Purple) and *S. robustum* (Mol 5829). SM analysis resulted in the identification of 24 significant associations for stalk number, stalk diameter, tasseled stalks, Pol (sucrose content), fiber content and smut resistance. One significant epistatic interactions was identified for stalk diameter, considering multiple regression analysis. Later, for the same progeny, Mudge et al. (1996) and Guimarães, Sills and Sobral (1997) cited the association of molecular markers with eyespot disease and flowering time, respectively.

Daugrois et al. (1996), through the study of the self progeny of the cultivar R570 (resistant to rust) and using SM analysis, identified an RFLP marker linked at 10 cM with a possible dominant resistance gene to rust, showing a 3 (resistant): 1 (susceptible) segregation ratio. Hoarau et al. (2002), using the same progeny raised to 295 individuals with AFLP markers in single dose (segregating in 1:1 and 3:1 ratio) and multiple dose, detected 73 putative QTAs, consistent across both years of evaluation. SM analysis was carried out for each year separately. In addition, 41 epistatic interactions were identified for the different years and traits (stalk number, stalk diameter, stalk length and brix).

Two different full-sib progenies derived from heterozygous parents, one obtained from a cross between *S. officinarum* (Green German) and *S. spontaneum* (IND 81-146), and another from the cross between *S. spontaneum* (PIN 84-1) and *S.officinarunm* (Muntok Java), were evaluated by Ming et al. (2001), resulting in the identification of 36 associations for sugar content, of which 14 were detected for the first progeny and 22 for the second. Later, Ming et al. (2002b), using the same full-sib progenies, identified 102 significant associations for sugar yield, Pol (sucrose content), stalk weight, stalk number, fiber content and ash content. Of these 102 marker-QTL associations, 61 were identified for markers placed on the linkage map and 41 for unlinked markers. Still using the same full-sib progenies, Ming et al. (2002c) detected 65 significant associations for plant height and flowering time, of which 30 were identified for markers placed on linkage groups and 35 for unlinked ones. For the three studies reported, SM (for unlinked markers) and IM (for markers placed on individual maps obtained for each parent, considering the *double pseudo-testcross* strategy) approaches were implemented, using RFLP markers.

A population obtained from a cross between two Australian elite clones, Q117 (female) and 74C42 (male), evaluated in two different sites and years, was considered for QTL mapping by Jordan et al. (2004). This analysis resulted in the identification of seven and six RFLP markers associated with stalk number and sucker number, respectively. These associations were consistents across sites and years, and three of these markers were identified to be related to the two traits simultaneously. The tests were carried out based on SM analysis. Comparative mapping with sorghum suggested that there are two unlinked regions of the genome associated with stalk number and sucker number, suggesting the possibility to select simultaneously for high stalk number and low sucker types.

Based on the same progeny and combining information from different marker types (AFLP,

RFLP and SSR), McIntyre et al. (2005b) identified in addition to an unlinked marker segregating in a 3:1 ratio, two genomic regions, one in each parent, associated with Pachymetra root rot (*Pachymetra chaunorhiza*) resistance. For brown rust resistance, significant associations were detected in two genomic regions, one in each parent. Moreover, association analysis carried out for 154 elite clones found that some of these markers remained associated with these diseases (WEI; JACKSON; McINTYRE, 2006). Such results suggest that these markers can be used for MAS in breeding programs looking for resistance to both diseases. Later, McIntyre et al. (2005a) used similar sequences of genes for brown rust resistance to identify QTLs associated with RFLP, AFLP and SSR markers in the same progeny. Through the comparative mapping with sorghum, it was found that markers obtained from one of these genes were associated with a chromosomal region of sorghum, in which a major QTL had already been identified for brown rust resistance. In these studies, QTL analyses were performed based on SM approach.

Considering a progeny derived from the cross between two *Saccharum* spp. pre-commercial Brazilian cultivars, SP80-180 (female) and SP80-4966 (male), Silva and Bressiani (2005) identified an EST-RFLP marker inversely associated with sucrose content (Pol) through SM analysis. Thus, plants without this marker would have a higher Pol value than plants with the marker. However, this marker was present in the parent SP80-4966 (high sugar yield) and absent in the parent SP80-180 (low sugar yield), indicating a transgressive segregation of such marker in the progeny. Moreover, this suggests that parent SP80-4966 has other alleles that contribute to Pol increase, compensating the negative effect of the EST-RFLP marker.

Reffay et al. (2005), to study the genetic contribution of the Mandalay (*S. spontaneum*) clone for Australian elite varieties and parents, performed QTL mapping through SM analysis in a progeny derived from the cross between the clones Q117 and MQ77-340, the latter being a direct descendant of Mandalay (*S. officinarum* Korpi clone $\times$ Mandalay). From a total of 101 linkage groups, 65 had markers originated from the parent Mandalay and/or Korpi. Markers from both parents of MQ77-340 (Mandalay and Korpi) were identified to be associated with different traits (Pol, brix, Commercial Cane Sugar, fiber content, stalk number, tonnes of cane per hectare - TCH and tonnes of sugar per hectare - TSH), expressing both positive and negative phenotypic effects. Recently, Piperidis et al. (2008), through the comparative mapping between individual maps obtained for each parent of the same progeny (Q117 $\times$ MQ77-340) and maps of cultivars R570 (French origin) and Q165A (Australian origin), identified significant marker-QTL associations for brix, in linkage

groups consistent across two or three maps of each parent. The analyses were performed using SM approach, for unlinked markers, and IM approach, for markers placed on individual maps for each parent.

Aitken, Jackson and McIntyre (2006) identified 37 marker-QTL associations for brix and Pol in a progeny derived from the cross between the clone *S. officinarum* IJ76-514 and the cultivar Q165A. In addition, 97 epistatic interactions were identified, of which nine were consistent across two years (eight affecting more than one trait simultaneously). QTL detection were performed for each harvest year separately, based on SM approach, considering data from AFLP and SSR markers. McIntyre et al. (2006), considering the same progeny and SNP (polymorphisms in the sucrose phosphate synthase - SPS - gene family III) markers, identified by SM analysis that there is no significant difference in the SNP frequency among individuals with low and high sucrose content in the progeny. However, using an ecotilling approach, two of the SPS gene family III haplotypes were mapped to two different linkage groups from homology group 1 in Q165A. Both haplotypes mapped near QTLs for increased sucrose content, but none of them were associated with any sugar related traits. Recently, Aitken et al. (2008), using AFLP and SSR markers, identified, for the same progeny, 27 genomic regions significantly associated with at least one of the traits including: cane yield (TCH), stalk weight, stalk number, stalk length, stalk diameter. About 46% of the marker-QTL associations were consistent across different years of evaluation. The QTL analyses were performed for each year and the results were compared. In addition, using SNPs, two alleles of the TEOSINTE BRANCHED 1 gene (TB1, a major gene controlling branching in maize) showed some association with stalk number in two years of evaluation, but with a small effect in sugarcane. Moreover, 195 epistatic interactions were identified considering all the traits and years under study.

The progeny obtained from the cross between the modern cultivar R570 and the clone MQ76-53 was used for QTL mapping by Raboin et al. (2006), allowing the identification of an AFLP marker linked at 6.5 cM to a gene controlling the red stalk color (segregating in an 1:1 ratio) and another AFLP marker linked at 23 cM to a new brown rust resistance gene (segregating in an 1:1 ratio), both genes placed on the MQ76-53 genetic linkage map.

Although the yellow spot disease, caused by the imperfect fungus *Mycovellosiella koepkey*, is not of great importance for many sugarcane producing countries, it can be severe and can cause low juice purity, high reducing sugar/sucrose ratio, sucrose losses at early harvest and it can also affect cane yield at late harvest, when grown under high relative humidity. Al-Janabi et al. (2007)

based on a progeny derived from a biparental cross between M134/75 (resistant cultivar) and R570 (susceptible cultivar), performed QTL mapping through IM and CIM approach (for markers segregating in an 1:1 ratio and placed on the individual maps of each parent) and SM analysis (for unlinked markers). A major QTL was found to be associated with a resistance gene (segregating in a 3 resistant :1 susceptible ratio in the progeny).

Through the several results reported for QTL mapping in sugarcane, it is possible to conclude that most of the analyses were performed based on SM and IM approaches, considering two maps, one for each parent, obtained unsing the *double pseudo-testcross* strategy and data from SDM segregating in an 1:1 and 3:1 ratio. Moreover, some authors describe the identification of consistent QTLs across harvests. However, the statistical models implemented in these analyses were not appropriate for this conclusion, since it is based on the information provided by analysis performed separately for each harvest year and location. Thus, it highlights the need for the development of more powerful statistical methods that combine information from multiple harvests and locations, allowing the obtention of more reliable estimates about the QTL effects across environments, producing better results for possible applications in MAS.

2.3 Mixed Models

Mixed models correspond to linear models that incorporate both fixed effects, which are parameters associated with an entire population or with certain specific repeatable levels of experimental factors, and random effects, which are associated with individual experimental units drawn at random from a population, varying from experiment to experiment (PINHEIRO; BATES, 2000). This definitions do not consider the mean (intercept) and error terms, which are respectively fixed and random effects in all standard linear models. The traditional fixed linear models are, in general, too restrictive to perform satisfactory data analysis for the typical data structure of most breeding programs because of the independence assumption. Actually, error structure in breeding experiments is much more complex than that considered in standard linear models for conventional data analysis (BALZARINI, 2001). In contrast, linear mixed models can take into account covariances among observations, dealing with correlated data by incorporating random effects and estimating their associated variance components (SMITH; CULLIS; THOMPSON, 2005).

The choice of fixed and random terms is not always determined by the structure of the experiment, since it may depend on the information required (SEARLE, 1971; SMITH; CULLIS;

THOMPSON, 2005). For example, in plant breeding, variety trials are often carried out over different locations and several years, through the multi-environment trials (MET). If a general assessment of varieties over time is required, then the years present in the trial are considered as a random selection of years, and year would be defined as a random term in the model. On the other hand, if the effect of the specific years present in the trial was to be assessed, year would be defined as a fixed term (PAYNE et al., 2009).

A general form of a linear mixed model is (BALZARINI, 2001):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the response vector (data), $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known design matrices, $\boldsymbol{\beta}$ is a vector of fixed parameters, $\boldsymbol{u}$ and $\boldsymbol{e}$ are unobservable random vectors of random effects and errors terms, respectively. For these random terms, it is generally assumed normal distribution, with $E(\boldsymbol{u})$ and $E(\boldsymbol{e})$ equal to zero, and variance-covariance matrices $\boldsymbol{G}$ and $\boldsymbol{R}$. Different models for the VCOV matrix of the data, $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$, can be considered in the mixed models approach, specifying the structure of $\boldsymbol{Z}$, $\boldsymbol{G}$ and $\boldsymbol{R}$. Note that when $\boldsymbol{Z} = 0$ and $\boldsymbol{R} = \sigma^2\boldsymbol{I}$, mixed models reduces to the standard linear model.

For plant breeding experiments, genetic correlations may be included into the mixed model through the $\boldsymbol{G}$ matrix, and experimental correlations among observations may be fitted by the off-diagonal elements of $\boldsymbol{R}$. Thus, several models for the (co)variance structure of $\boldsymbol{G}$ and $\boldsymbol{R}$ can be considered (Table 1) and will be discussed in details in the next section.

Mixed model solutions can be obtained by (HENDERSON, 1990):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} + \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

$$\hat{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

In this case, if $\boldsymbol{G}$, $\boldsymbol{R}$, $\boldsymbol{Z}$ and, therefore, $\boldsymbol{V}$ are known, $\hat{\boldsymbol{\beta}}$ is the Best Linear Unbiased Estimator (BLUE) for $\boldsymbol{\beta}$, and $\hat{\boldsymbol{u}}$ is the solution for the prediction of random effects, i.e., the Best Linear Unbiased Predictor (BLUP). Theoretically, BLUPs have the smallest mean squared error of prediction among all linear unbiased predictors, since the parameters of the model are known (BALZARINI, 2001). However, $\boldsymbol{V}$ is usually unknown in practical situations. Thus, the best approach is to use

likelihood-based methods for the estimation of covariance parameters prior to the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{u}$. Assuming that $\boldsymbol{u}$ and $\boldsymbol{e}$ are normally distributed, restricted maximum likelihood (REML) method (PATTERSON; THOMPSON, 1971), a variant of the maximum likelihood (ML) method, is usually preferred for estimating variance components in mixed models. Given estimates of the variance components, in the REML approach, the fixed effects may be estimated using Empirical Best Linear Unbiased Estimation (E-BLUE) and the random effects predicted using Empirical Best Linear Unbiased Prediction (E-BLUP). To indicate that $\boldsymbol{G}$ and $\boldsymbol{R}$ have been estimated prior to getting the BLUPs, the term E-BLUP is frequently used to refer to $\hat{\boldsymbol{u}}$ (BALZARINI, 2001; PAYNE et al., 2009).

Consider a simple model with one radom effects vector ($\boldsymbol{u}$) representing genotypic effects and a response vector ($\boldsymbol{y}$) containing the phenotypic data for $i = 1, \ldots, I$ genotypes. Thus, under a mixed model assumption, the equation $\hat{\mu}_i = \hat{\mu} + w(y_i - \mu)$ provides the prediction for the mean performance of genotype i, where μ is the population mean and w is a weighting or shrinkage factor. If $G = \sigma_u^2 \boldsymbol{I}$ and $R = \sigma_e^2 \boldsymbol{I}$, the simplest structures, where $\boldsymbol{I}$ is the identity $I \times I$ matrix, the elements of $\boldsymbol{GZ'V}^{-1}$, which defines the weights w, are function of $\sigma_u^2/(\sigma_u^2 + \sigma_e^2)$. In this aspect, this weights may represent the broad sense heritability of the trait under evaluation. BLUP estimates for random effects are smaller than if the effects had been estimated as fixed, with more shrinkage taking place for smaller value ratios of the estimated variance components in w. For this reason, the BLUP random effects estimates are often called 'shrunken' parameter estimates (BALZARINI, 2001).

REML was developed in order to avoid the biased variance component estimates that are produced by the ordinary maximum ML method, because it takes into account the degrees of freedom used to estimate treatment effects. Thus, ML methods have a downwards bias which increases with the number of fixed effects in the model, leading to underestimates of standard errors for fixed effects, resulting in incorrect inferences being made from the data (PAYNE et al., 2009).

For model selection, there are several strategies, such as graphical methods and diagnostics (CHRISTENSEN; PEARSON; JOHNSON, 1992) and likelihood-based methods (DIGGLE, 1988; OMAN, 1991; WOLFINGER, 1993). The likelihood ratio (LR) test can be used to verify nested mixed models. Thus, the test is:

$$LR = -2 \times \log\left(\frac{L_R}{L_F}\right)$$

where L_R is the residual likelihood of the reduced model and L_F is the residual likelihood of the full

model. The null hypothesis tested (H_0) is that the reduced model is not different from the full model. The LR test has a χ^2 distribution, with degrees of freedom equal to the difference in the number of parameters between the two models. Since the fixed part is the same for all models, only the number of parameters in the variance-covariance structure needs to be considered. An alternative likelihood-based mixed model selection strategy was proposed by Wolfinger (1996) who suggested to use the Akaike Information Criterion (AIC - AKAIKE, 1974) and the Bayesian Information Criterion (BIC - SCHWARZ, 1978) to compare the non-nested mixed models, as following:

$$AIC = -2\log(L) + 2 \times n_{PAR}$$

$$BIC = -2\log(L) + log(N) \times n_{PAR}$$

where $\log(L)$ is the residual loglikelihood, N is the total number of observations and n_{PAR} is the number of parameters in the VCOV matrix (PIEPHO, 2000). BIC is a Bayesian model selection based on Bayes factors, and involves a penalty for the number of parameters, which tends to favor parsimonious models. AIC is an estimator of expected relative Kullback-Leibler (K-L) information (frequently conceptualized as a 'distance' between full reality and a model) based on the maximized log-likelihood function, corrected for asymptotic bias (BURNHAM; ANDERSON, 2004). The smaller the value of the criterion, the more preferable is the model (WOLFINGER, 1993). For the fixed terms in mixed models, Wald tests is commonly used (VERBEKE; MOLENBERGHS, 2000).

2.3.1 Mixed Models for Multi-Environment Trials

The main goal in many advanced plant breeding programs involve the evaluation of a set of genotypes in designed experiments performed at a range of environmental conditions. To this purposes, MET are considered, in which the genotypes are evaluated across several locations and years. Typically, cultivars are tested for global performance across a series of geographically and temporally varying conditions, or for local performance under temporally varying conditions at specific sites. In addition, cultivars are also often evaluated for adaptability and stability in relation to changing environmental conditions, taking into account the genotype by environment (G $\times$ E) interaction (EEUWIJK et al., 2005).

MET analysis has been made through the Additive Main Effects and Multiplicative Interaction

(AMMI) models, well described by Gauch (1988, 1992) and attributed to Fisher and Mackenzie (1923) and Gollob (1968), in addition to biplots for visual representation (GABRIEL, 1971) and traditional methods, such as joint analysis of variance (ANOVA) and linear regression. However, these techniques have some limitations: i) consider a fixed-model framework for genotype and G $\times$ E; ii) not consider spatial variation within trials; and iii) not consider heterogeneity of variance between trials. In this aspect, mixed model approach can be a useful tool for dealing with MET, since different models for $\boldsymbol{G}$ and $\boldsymbol{R}$ can be used for genotype mean predictions and G $\times$ E interaction studies, taking into account also the spatial variance observed among trials. Moreover, the likelihood-based methodologies involved in mixed models estimation provide a more flexible analytical approach for MET data, since balanced data are not required (BALZARINI, 2001).

MET data may be summarized in two-way tables, indexed by genotypes and environments. Thus, a typical model is (RESENDE; THOMPSON, 2004):

$$y_{ij} = \mu + e_j + g_i + ge_{ij} + \varepsilon_{ij} \tag{1}$$

where y_{ij} is the phenotypic response for genotype i at environment j; μ is the general mean; g_i is genetic effect of genotype i; e_j is the environment effect; ge_{ij} is the G $\times$ E interaction; and ε_{ij} is the error term. The μ and e_j effects can be regarded as fixed and the others as random. In the context of mixed models, the following model can be considered:

$$y_{ij} = \mu + e_j + g_{ij} + \varepsilon_{ij} \tag{2}$$

where g_{ij} is the random effect for genotype i in environment j. Thus, different classes of structures can be considered for the VCOV matrix of g_{ij} (G matrix, Table 1). In this matrix, the diagonal elements are the genetic variances for individual environments and the off-diagonal elements are the genetic covariances between pairs of environments.

The simplest model for the G matrix is one that considers independence and homogeneous variation, i.e., there are no genetic correlations between environments and the genetic variances are homogeneous across the environments. These assumptions are rarely realistic, and this model is named independent model (ID). Following the same idea of independence, the heterogeneous genetic variation (DIAG) model allows for heterogeneous genetic variances, reflecting the magnitude of variation between genotypes in individual environments, but assumes that there are no genetic

Table 1 – Example of different models for the (co)variance structure of $\boldsymbol{G}$ matrix that can be considered for MET data

Model	Description	$\boldsymbol{G}$ matrix
ID	Identical variation	$\begin{bmatrix} \sigma_g^2+\sigma_{ge}^2 & 0 & \dots & 0 \\ 0 & \sigma_g^2+\sigma_{ge}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_g^2+\sigma_{ge}^2 \end{bmatrix}$
DIAG	Heterogeneous variation	$\begin{bmatrix} \sigma_{g_1}^2+\sigma_{ge_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{g_2}^2+\sigma_{ge_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{g_J}^2+\sigma_{ge_J}^2 \end{bmatrix}$
CS	Compound symmetry with homogeneous variation	$\begin{bmatrix} \sigma_g^2+\sigma_{ge}^2 & \sigma_g^2 & \dots & \sigma_g^2 \\ \sigma_g^2 & \sigma_g^2+\sigma_{ge}^2 & \dots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \dots & \sigma_g^2+\sigma_{ge}^2 \end{bmatrix}$
CS$_{Het}$	Compound symmetry with heterogeneous variation	$\begin{bmatrix} \sigma_{g_1}^2+\sigma_{ge_1}^2 & \sigma_g^2 & \dots & \sigma_g^2 \\ \sigma_g^2 & \sigma_{g_2}^2+\sigma_{ge_2}^2 & \dots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \dots & \sigma_{g_J}^2+\sigma_{ge_J}^2 \end{bmatrix}$
AR1	First-order autoregressive model with homogeneous variation	$\begin{bmatrix} \sigma_g^2+\sigma_{ge}^2 & \sigma_g^2\rho_g & \dots & \sigma_g^2\rho_g^{d(1,J)} \\ \sigma_g^2\rho_g & \sigma_g^2+\sigma_{ge}^2 & \dots & \sigma_g^2\rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2\rho_g^{d(J,1)} & \sigma_g^2\rho_g^{d(J,2)} & \dots & \sigma_g^2+\sigma_{ge}^2 \end{bmatrix}$
AR1$_{Het}$	First-order autoregressive model with heterogeneous variation	$\begin{bmatrix} \sigma_{g_1}^2+\sigma_{ge_1}^2 & \sigma_g^2\rho_g & \dots & \sigma_g^2\rho_g^{d(1,J)} \\ \sigma_g^2\rho_g & \sigma_{g_2}^2+\sigma_{ge_2}^2 & \dots & \sigma_g^2\rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2\rho_g^{d(J,1)} & \sigma_g^2\rho_g^{d(J,2)} & \dots & \sigma_{g_J}^2+\sigma_{ge_J}^2 \end{bmatrix}$
FA1	First-order factor analytic model	$\begin{bmatrix} \lambda_1^2+\Psi_1 & \lambda_1\lambda_2 & \dots & \lambda_1\lambda_J \\ \lambda_2\lambda_1 & \lambda_2^2+\Psi_2 & \dots & \lambda_2\lambda_J \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_J\lambda_1 & \lambda_J\lambda_2 & \dots & \lambda_J^2+\Psi_J \end{bmatrix}$
Unst	Unstructured model	$\begin{bmatrix} \sigma_{g_1}^2+\sigma_{ge_1}^2 & \sigma_{g_{12}}^2 & \dots & \sigma_{g_{1J}}^2 \\ \sigma_{g_{21}}^2 & \sigma_{g_2}^2+\sigma_{ge_2}^2 & \dots & \sigma_{g_{2J}}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g_{J1}}^2 & \sigma_{g_{J2}}^2 & \dots & \sigma_{g_J}^2+\sigma_{ge_J}^2 \end{bmatrix}$

σ_g^2 and σ_{ge}^2: variance components for genotype main effect and G $\times$ E interaction, respectively; $\rho_g^{d(j,j')}$: genetic correlation between environments, where $d(j,j')$ correspond to the distance in time between them; $\sigma_{g_j}^2$ and $\sigma_{ge_j}^2$: environment-specific genetic variance for genotype main effect and G $\times$ E interaction; $\sigma_{g_{jj'}}^2$: genetic covariance between environments j and j'; Ψ_j: environment-specific residual variance; λ_j and $\lambda_{j'}$: loadings of the factor in environment j and j'.

correlations between environments.

Compound symmetry (CS) and compound symmetry heterogeneous (CS$_{Het}$) assume the exis-

tence of genetic correlations, reflecting the agreement in genotype rankings between environments, but homogeneous and heterogeneous genetic variances, respectively. The First-order autoregressive models also take into account the assumption of genetic correlation between environments, considering homogeneous (AR1) or heterogeneous ($AR1_{Het}$) variances. These models are especially important for perennial crops, such as sugarcane, where data may relate to multiple harvests (repeated measures over time). Note that, for such crop species, the genetic correlation decrease with the distance in time between harvests (environments), since the genes expressed in the first harvest year could may not have the same expression in the subsequent years.

The unstructure model for the (co)variance matrix considers the general case, in which all elements of the matrix are allowed to be different. However, in some cases, estimation of such structure may be inefficient or unstable for even moderately values of J (number of environmentes), since the number of parameters is equal to $J(J+1)/2$. First-order factor analytic (FA1) model, considering $k = 1$ factor, can be regarded as an approximation to the completely unstructured VCOV and can provide parsimonious models. The factor analysis technique can be considered as an extension of the principal component analysis (GAUCH, 1988; 1992; RESENDE; THOMPSON, 2004). In a general case, a factor analytic model of order k, denoted as FAk, is postulated in terms of unobservable genotype effects in different environments (RESENDE; THOMPSON, 2004):

$$g_{ij} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij}$$

where g_{ij} is the effect of genotype i in environment j, $\sum_{r=1}^{k} \lambda_{jr} f_{ir}$ is the sum of multiplicative terms used to explain G $\times$ E interaction, in which λ_{jr} is the loading (slope regression) for factor (latent variable) r in environment j, f_{ir} is the score for genotype i in factor r, and δ_{ij} is the error representing the lack of fit of the model. This model leads to a VCOV structure for G in which: $\sigma^2_{g_j} + \sigma^2_{ge_j} = \sum_{r=1}^{k} \lambda^2_{jr} + \Psi_j$ is the genotype variance in environment j, where Ψ_j is the specific residual variance for δ_{ij}; and $\sigma^2_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr} \lambda_{j'r}$ is the covariance between environments j and j'.

A model analogous to the Eberhart and Russel (1966) regression model can be obtained in this mixed model framework by using one multiplicative term:

$$g_{ij} = \lambda_j f_i + \delta_{ij}$$

where λ_j is the loading (slope regression) for the factor in environment j, f_i is the score for genotype i (factor). Thus, $\sigma^2_{g_j} + \sigma^2_{ge_j} = \lambda^2_j + \Psi_j$ is the genotype variance in environment j, and $\sigma^2_{g_{jj'}} = \lambda_j \lambda_{j'}$ is the covariance between environments j and j' that compound the VCOV structure of the G matrix. This is the FA1 model, as presented in Table 1.

Several authors described the use of mixed models to analyze MET data in different plant species (DENIS; PIEPHO; EEUWIJK, 1997; PIEPHO, 1997; CULLIS et al., 1998; RESENDE; THOMPSON, 2004; CHAPMAN, 2008; SMITH; CULLIS; THOMPSON, 2001, 2005; SMITH et al., 2007; EEUWIJK et al., 2007). Smith et al. (2007) presented an alternative method to model the VCOV structure of the G matrix for MET data obtained from perennial crops. Sugarcane is an important example of perennial species, where MET data may relate to multiple harvests (repeated measures along time), in addition to multiple trials (locations). The authors proposed an approach for the analysis of yield data combined not only across trials but also harvests. The method is an extension of Smith et al. (2001) and allow to regard a VCOV matrix for harvest ($\boldsymbol{G}^H_{K\times K}$) and other one for locations ($\boldsymbol{G}^L_{J\times J}$), instead to consider an individual VCOV matrix ($\boldsymbol{G}$) for the factorial combination of harvests and locations (in this case, each harvest-location combination is considered as a single environment). In this context, the final $\boldsymbol{G}$ matrix is:

$$\boldsymbol{G} = \boldsymbol{G}^L_{J\times J} \otimes \boldsymbol{G}^H_{K\times K}$$

where $j = 1, \ldots, J$ is the number of locations, $k = 1, \ldots, K$ is the number of harvests, and $\otimes$ denotes the Kronecker product. The different VCOV structures presented in Table 1 can be regarded for each individual matrix, depending of the data and the objectives of the analysis. This may provide models with a reduced number of parameters when compared with the conventional approach, allowing, in some situations, the selection of parsimonious models for the VCOV structure.

For the analysis of MET data using mixed models, taking genotype effects as random or fixed depends on the aim of the analysis and considerations about the properties of the two types of estimation procedures, the E-BLUP for random effects, and the E-BLUE for fixed effects. In this context, if the objective of the analysis is to identify the best genotypes for selection purposes, the ranking of the estimated genotype effects should be as close as possible to the rankings of the true genotype effects. This implies the use of BLUP (random genotype effects). However, the optimality properties of BLUP are based on the assumption that the variance parameters in the model are known, but in practice, this is not the case and the parameters are estimated from the data

(E-BLUP). It is not possible to state that the estimates of the variance parameters are sufficiently precise to ensure that the optimality of BLUP is maintained with E-BLUP. If the aim of the analysis is to determine the difference between specific pairs of genotypes, then the use of BLUP as an estimation method is inappropriate, since the BLUP of a specific difference is biased. In this case, genotype effects should be regarded as fixed (SMITH; CULLIS; THOMPSON, 2005).

2.3.2 Mixed Models and QTL Mapping

QTL mapping studies frequently use phenotypic data obtained from multiple environments. The detection of QTL main effects and QTL $\times$ E interaction in such studies requires the use of appropriate statistical tests. However, most part of the methodologies available to study QTL $\times$ E interaction does not account for the fact that the same genotypes are grown in each environment, which introduces genetic correlation among phenotypic observations on the same genotype (PIEPHO, 2005). Mixed models are a natural framework for the analysis of such complex data sets, allowing to model the genetic (co)variances between environments in combination with the heterogeneous residuals, achieving more realistic and reliable conclusions about G $\times$ E interaction.

Based on regression methods (HALEY; KNOTT, 1992; MARTÍNEZ; CURNOW, 1992), QTL mapping can integrate molecular marker information into mixed models to test not only for the effect of DNA polymorphisms on phenotypic traits, but also to identify regions (QTLs) with effects on multiple environments and, as consequence, the occurrence of QTL $\times$ E interaction, which is caused by changes in QTL expression across environments. The main advantage of the regression-based QTL mapping methods is its computational simplicity, when compared with mixture models.

A first step in a mixed-model QTL mapping analysis is to select a phenotypic model for the MET data, with the aim of identifying a variance-covariance model, considering the possibility of heterogeneity of genetic variances across individual environments and heterogeneity of genetic correlations between pairs of environments. Then, the next step is to include marker information in this mixed model. In this context, model (2) can be expanded:

$$y_{ij} = \mu + e_j + g_{ij} + x_i a_j + \varepsilon_{ij} \tag{3}$$

where g_{ij} is the random effect for genotype i in environment j; e_j is the environment effect; x_i contains genetic information on QTL genotypes at a particular genome position, also named as genetic predictor; a_j is the additive effect of the QTL in environment j; and ε_{ij} is the error term.

In an F_2 population, if the QTL is positioned over markers, x_i assumes the values -1, 0 and 1 for individual's marker having the genotypes aa, Aa and AA (interpreted as the QTL genotypes qq, Qq, and QQ respectively). QTL genotypes between markers are not directly observable, but can be calculated based on conditional probabilities obtained from flanking marker genotypes (JIANG; ZENG, 1997). Thus, the value of x_i for positions between markers needs to be adjusted and the most used approach is based on conditional expectations of QTL genotypes, given marker phenotypes (HALEY; KNOTT, 1992; MARTÍNEZ; CURNOW, 1992). Through this model, QTL environment-specific effects can be easily considered. A test for QTL main effects along the genome can also be performed.

Mixed models have been widely used to investigate the causes of G $\times$ E interaction, through the identification of QTLs with consistent expression under different environmental conditions (years and locations), as well as to investigate the genetic base of correlated traits (pleiotropy and/or linked QTLs). Several authors reported the use of mixed models applied to QTL mapping in different crop species, such as maize (CROSSA et al., 1999; VARGAS et al., 2006; BOER et al., 2007; EEUWIJK et al., 2010), barley (PIEPHO, 2000; VERBYLA et al., 2003; MALOSETTI et al., 2004), rice (EMRICH; PRICE; PIEPHO, 2008) and wheat (MATHEWS et al., 2008). To our knowledgement, none of the current QTL mapping results published for sugarcane were provided by mixed models analysis.

3 MATERIAL AND METHODS

3.1 Material

3.1.1 Plant Material

Phenotypic and molecular data were obtained for a population based on 100 individuals derived from a cross between two pre-commercial Brazilian cultivars, SP80-180 (B3337 x polycross) and SP80-4966 (SP71-1406 x polycross). SP80-180 was the female parent and has lower sucrose content and high stalk production, whereas SP80-4966 (male parent) has higher sucrose and lower stalk production. Both parents and population were developed at the Experimental Station of the Centro de Tecnologia Canavieira (CTC) in Camamu-BA, Brazil.

3.1.2 Molecular Data

Restriction Fragment Length Polymorphism (RFLP), Microsatellite or Simple Sequence Repeat (SSR), EST (Expressed Sequence Tag) RFLP and SSR derived markers were used to genotype parents and progeny. All these markers had already been generated and coded, as described in detail by Garcia et al. (2006) and Oliveira et al. (2007). For marker scoring, each segregating allele was scored based on the presence (1) or absence (0) in the progeny, therefore behaving as a dominant marker. Only single-dose markers were considered. Marker segregation patterns were tested for the expected ratios using chi-square test (χ^2), considering single dose markers in only one parent (1:1 fashion) and in both parents (3:1 fashion). All loci with strong deviation from expected proportions were discarded after controlling type I error for multiple tests using Bonferroni's procedure.

3.1.3 Phenotypic Data

The mapping population was planted in two locations in 2003 (Piracicaba and Jaú, both in the State of São Paulo, Brazil) and evaluated in the first, second and third harvest years for cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), Fiber percent and Pol (sucrose content). In each location the experimental design consisted of a randomized complete block design with two replicates. However, the 100 clones were not fully randomized. The clones were randomly split in three groups of 36, 38, and 26 clones, respectively. The clones were only

randomized within those groups, while the groups were not randomized. In the experiments, each of the groups of clones (genotypes) was augmented by four checks (commercial cultivars SP80-1842, SP81-3250, SP80-1816 e RB72454). Both parents were also included in the first group, but not considered in the statistical analysis. A layout of the field experiments is presented in Figure 1.

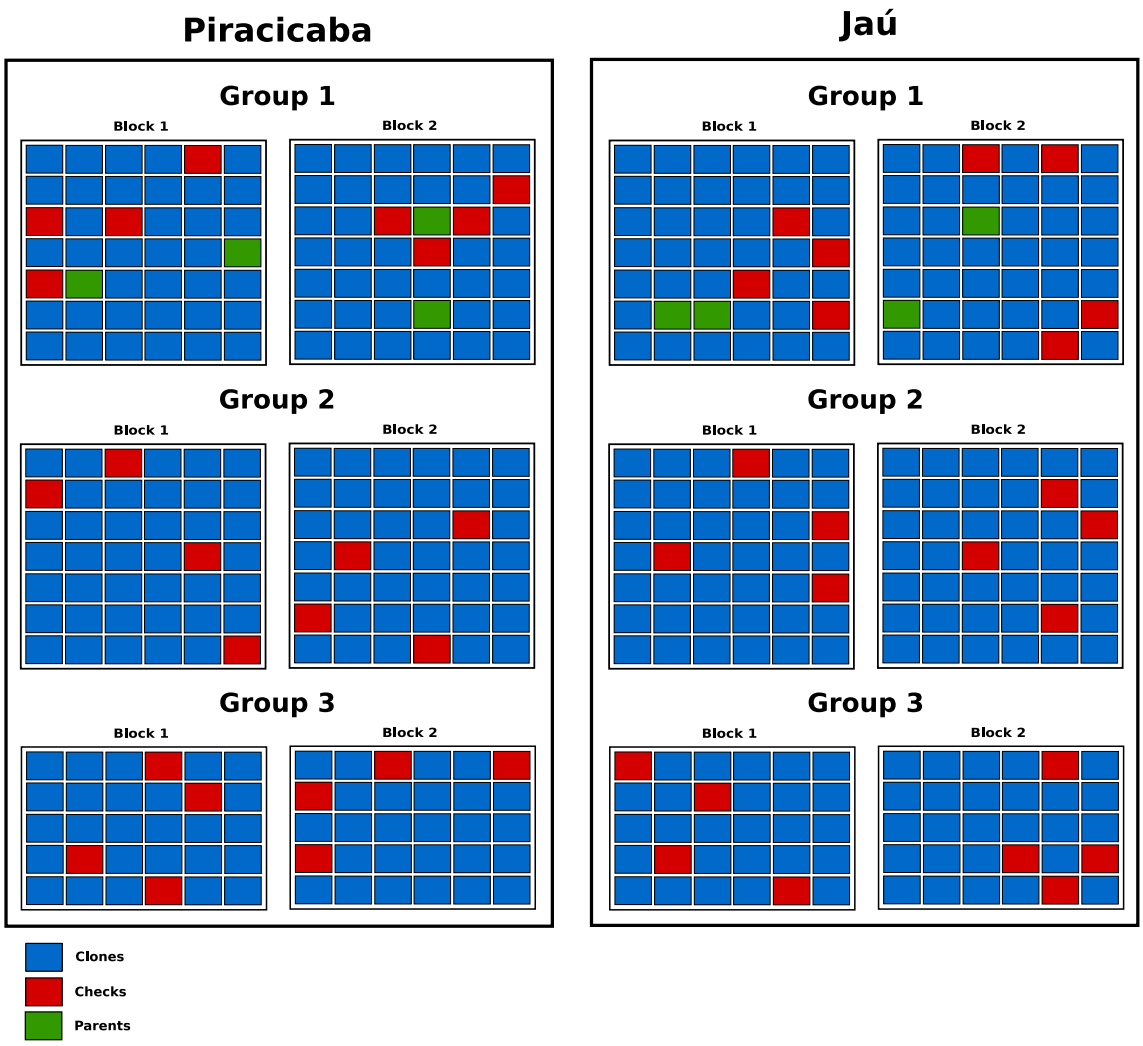

Figure 1 – Location 1: Piracicaba, SP, Brazil; and Location 2: Jaú, SP, Brazil

3.2 Methods

3.2.1 Linkage Map

Based on a multipoint approach (WU et al., 2002a; WU et al., 2002b), map construction was

carried out using the package *OneMap* (MARGARIDO; SOUZA; GARCIA, 2007). For this purpose, 741 molecular markers were used, including 459 loci displaying an 1:1 segregation ratio (100 RFLP, 27 EST-RFLP, 332 EST-SSR) and 282 loci segregating in a 3:1 ratio (88 RFLP, 10 EST-RFLP, 184 EST-SSR). Following the notation presented by Wu et al. (2002a), markers segregating for the parent SP80-180 (P_1) were denoted as D_1, considering the configuration '$ao \times oo$', in which the a allele is dominant to the o (null) allele. Informative loci for the parent SP80-4966 (P_2) were denoted as D_2, with the configuration '$oo \times ao$', and markers segregating for both parents were denoted as C, with configuration '$ao \times ao$'. For the determination of the linkage groups, two point analysis was performed considering a minimum LOD Score threshold of 6 and 0.5 for the recombination fraction. Linkage groups with a maximum of five loci were ordered through the comparison of all possible orders, in a procedure analogous to the comand `compare` implemented in the MAPMAKER/EXP software (LANDER et al., 1987). For linkage groups with more than 5 markers, the `order` algorithm took five adjacent markers, which were ordered through the comparison of all possible orders, and then the other markers were sequentially placed on the linkage group based on the initial order, in a similar way to that performed by the comand `try` in the MAPMAKER/EXP software and validated by Mollinari et al. (2009). After, the `ripple` comand was applied to verify if local inversions had occured. Map distances were expressed in centiMorgans based on Kosambi's function (KOSAMBI, 1944).

3.2.2 Genetic Predictors

For notation purposes, in a similar way to that proposed by Lin et al. (2003), consider a full-sib progeny obtained from the cross between two outbred diploid parental individuals, denoted as P and Q (Figure 2). They could be seen as a general case when compared with the loci configuration observed for sugarcane, where only SDM were considered. For an interval flanked by two markers, m and $m+1$, each one with alleles 1 and 2, the genotypes for these loci can be represented by $P_m^{\{1,2\}}$, $Q_m^{\{1,2\}}$, $P_{m+1}^{\{1,2\}}$ and $Q_{m+1}^{\{1,2\}}$, in which $\{1,2\}$ indicates the allelic possibilities for each locus. Suppose that there is a QTL between these two markers, with alleles P^1 and P^2 for parent P, Q^1 and Q^2 for parent Q. Thus, QTL segregation in the progeny will fit into four genotypic classes (P^1Q^1, P^1Q^2, P^2Q^1 and P^2Q^2), in a 1:1:1:1 ratio. Therefore, it is possible to define three orthogonal contrasts involving these four genotypic classes:

$$
\begin{array}{cccccccc}
+ & P^1Q^1 & + & P^1Q^2 & - & P^2Q^1 & - & P^2Q^2 \\
+ & P^1Q^1 & - & P^1Q^2 & + & P^2Q^1 & - & P^2Q^2 \\
+ & P^1Q^1 & - & P^1Q^2 & - & P^2Q^1 & + & P^2Q^2
\end{array}
$$

The first and second contrasts relate to QTL additive effects in parent P and Q, denoted as α_p and α_q, respectively, while the third one refers to QTL dominance effect (intra-locus interaction) between the additive effects on each parent, denoted as δ_{pq}. Genetic predictors were constructed for a grid of evaluation points, w, along the genome ($w = 1, ..., W$). These genetic predictors were introduced as explanatory variables in the mixed models. For individual i and evaluation point w, the genetic predictors are:

$$
\begin{aligned}
x_{p_{iw}} &= p(P^1Q^1|\mathbf{M}_i) + p(P^1Q^2|\mathbf{M}_i) - p(P^2Q^1|\mathbf{M}_i) - p(P^2Q^2|\mathbf{M}_i) \\
x_{q_{iw}} &= p(P^1Q^1|\mathbf{M}_i) - p(P^1Q^2|\mathbf{M}_i) + p(P^2Q^1|\mathbf{M}_i) - p(P^2Q^2|\mathbf{M}_i) \\
x_{pq_{iw}} &= p(P^1Q^1|\mathbf{M}_i) - p(P^1Q^2|\mathbf{M}_i) - p(P^2Q^1|\mathbf{M}_i) + p(P^2Q^2|\mathbf{M}_i)
\end{aligned}
$$

where $x_{p_{iw}}$, $x_{q_{iw}}$ and $x_{pq_{iw}}$ are the expected values of the explanatory variables for the additive QTL effects in parents P and Q, and dominance effect, respectively, at position w, given all the marker information $\mathbf{M}_i$ in a particular linkage group for individual i (HALEY; KNOTT, 1992; MARTÍNEZ; CURNOW, 1992; LYNCH;WALSH, 1998).The conditional multipoint probabilities $p(P^1Q^1|\mathbf{M}_i)$, $p(P^1Q^2|\mathbf{M}_i)$, $p(P^2Q^1|\mathbf{M}_i)$ and $p(P^2Q^2|\mathbf{M}_i)$ were calculated by a hidden Markov chain model implemented in a new version of the *OneMap* package (MARGARIDO; SOUZA; GARCIA, 2007), at all marker positions and at an additional grid of points with a step size of 1 cM along the genome.

Due to the lack of information provided by SDMs, since only 1:1 and 3:1 segregation patterns could be obtained, in some genomic positions the matrix of genetic predictors could be singular, i.e. some genetic predictors could be linear combinations of others. Since collinearity could cause serious problems with the estimation and interpretation of the parameters, its presence was investigated by examining the singular values and by calculating the condition number of the genetic predictors matrix in all genomic positions. Only informative contrasts (without collinearity), were then considered. For example, linkage groups with only markers of type D_1 have enough information solely for the estimation of one contrast for the additive effect in parent P, $x_{p_{iw}}$. The same principle was applied for all linkage groups and genomic positions.

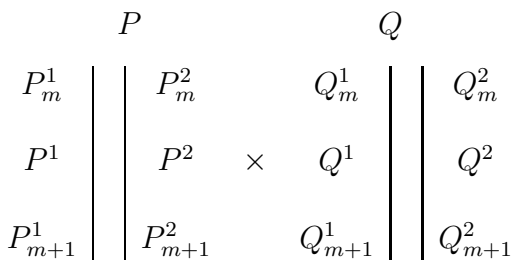

Figure 2 – Graphical representation of a bi-parental cross between outbred parents P and Q. $P_m^{\{1,2\}}$, $Q_m^{\{1,2\}}$, $P_{m+1}^{\{1,2\}}$ and $Q_{m+1}^{\{1,2\}}$ are the marker alleles for loci m and $m+1$; P^1, P^2, Q^1 and Q^2 are the QTL alleles

3.2.3 Multi-Harvest-Location Phenotypic Analysis

Prior to QTL detection, the identification of an appropriate mixed model for phenotypic data was done by comparing different variance-covariance (VCOV) structures for the genetic effect within location and harvest (Table 2). For mathematical description of the model, a notation similar to that presented by Eckermann et al. (2001), Verbyla et al. (2003) and Boer et al. (2007) was used. The statistical model, in which the underline indicates a random variable, is:

$$\underline{y}_{isjkr} = \mu + L_j + H_k + LH_{jk} + \underline{G}_{ijk} + \underline{\varepsilon}_{isjkr} \tag{4}$$

$\underline{y}_{isjkr}$ is the phenotype of the r^{th} replicate of the i^{th} genotype in group s, location j and harvest k; μ is the general mean; L_j is the location effect; H_k is the harvest effect; LH_{jk} is the location by harvest interaction effect; $\underline{G}_{ijk}$ is the genetic effect of genotype i at location j and harvest k; and $\underline{\varepsilon}_{isjkr}$ is a nongenetic effect. The genotypes can be separated into two groups, $n = n_g + n_c$, where n_g is the number of genotypes (clones) in the progeny ($i = 1, ..., n_g$), and n_c is the number of check entries ($i = n_g + 1, ..., n_g + n_c$). The model for $\underline{G}_{ijk}$ is given by:

$$\underline{G}_{ijk} = \begin{cases} \underline{g}_{ijk} & i = 1, ..., n_g \\ c_{ijk} & i = n_g + 1, ..., n_g + n_c \end{cases} \tag{5}$$

where $\underline{g}_{ijk}$ is a random variable for the genetic effect of genotype i in location j and harvest k, and c_{ijk} represents a fixed effect for check i in location j and harvest k. Although check entries are not relevant to the detection of QTL, they are important in providing information on the nongenetic variation that may be present (VERBYLA et al., 2003; BOER et al., 2007). It was assumed that vector $\underline{\mathbf{g}} = (g_{111}, ..., g_{IJK})$ have a multivariate normal distribution with zero mean and VCOV

matrix $\mathbf{G}$ (Table 2), $\underline{g} \sim N(0, \mathbf{G})$. For the nongenetic term ($\underline{\varepsilon}_{isjkr}$), the model was:

$$\underline{\varepsilon}_{isjkr} = t_s + t_{sjk} + b_{sjkr} + \underline{\eta}_{isjkr} \tag{6}$$

where t_s is the group effect, t_{sjk} is the effect of group s, appropriate for location j and harvest k; b_{sjkr} is the effect of block r within group s, location j and harvest k; $\underline{\eta}_{isjkr} \sim N(0, \sigma^2)$ represents a residual error term. All two-way and three-way interactions between fixed effects were also included, but for clarity purposes, not showed here. Seven different models for the VCOV matrix were analysed and compared based on AIC and BIC.

Table 2 – Different models for the genetic (co)variance matrix ($\mathbf{G}$) analysed

$\mathbf{G}$ matrix	Model	$n_{PAR}{}^a$	Description
$\mathbf{G} = \mathbf{G}^{L-H}_{M\times M}$	a) ID	1	Identical genetic variation
	b) DIAG	M	Heterogeneous genetic variation
	c) CS$_{Het}$	$M+1$	Compound symmetry with heterogeneous genetic variation
	d) FA1	$2M$	First-order factor analytic model
	e) Unst	$\frac{M(M+1)}{2}$	Unstructured model
$\mathbf{G} = \mathbf{G}^{L}_{J\times J} \otimes \mathbf{G}^{H}_{K\times K}$	f) Unst $\otimes$ AR1$_{Het}$	$\frac{J(J+1)+2(H+1)}{2} - 1$	Unstructured and first-order autoregressive models for the genetic variance within location and within harvest, respectively
	g) Unst $\otimes$ Unst	$\frac{J(J+1)+K(K+1)}{2} - 1$	Unstructured models for the genetic variance within location and harvest

Models (a-e) use the factorial combination of locations and harvests as different environments. Models (f-g) use the direct product of (co)variance matrices for locations and harvests. ^{a} The number of parameters for the models (f-g) follows from the sum of the parameters for the component matrices minus the number of identification constraints. $M = J \times K$, where J is the number of locations and K is the number of harvests.

3.2.4 QTL Analysis

Based on the interval mapping approach (LANDER; BOTSTEIN, 1989), the presence of a putative QTL was tested along the genome. In this context, the model was expanded to include marker information:

$$\underline{y}_{isjkr} = \mu + L_j + H_k + LH_{jk} + x_{p_{iw}}\alpha_{p_{jkw}} + x_{q_{iw}}\alpha_{q_{jkw}} + x_{pq_{iw}}\delta_{pq_{jkw}} + \underline{G}_{ijk} + \underline{\varepsilon}_{isjkr} \tag{7}$$

where $\alpha_{p_{jkw}}$, $\alpha_{q_{jkw}}$ and $\delta_{pq_{jkw}}$ are the harvest-location-specific effects of the additive genetic predictor for parent P and Q, and dominance genetic predictor, respectively, at evaluation point w. The VCOV matrix used for $\underline{G}_{ijk}$ was that selected in the previous phenotypic analyses. Assuming that the putative QTL has no effect across locations and harvests, the null hypothesis tested using a Wald test (VERBEKE; MOLENBERGHS, 2000) was:

$$H_0 : \begin{cases} \alpha_{p_{11w}} = \alpha_{p_{12w}} = ... = \alpha_{p_{JKw}} = 0 \\ \alpha_{q_{11w}} = \alpha_{q_{12w}} = ... = \alpha_{q_{JKw}} = 0 \\ \delta_{pq_{11w}} = \delta_{pq_{12w}} = ... = \delta_{pq_{JKw}} = 0 \end{cases}$$

A test for QTL main effects were also performed along the genome, using:

$$\underline{y}_{isjkr} = \mu + L_j + H_k + LH_{jk} + x_{p_{iw}}\alpha_{p_w} + x_{q_{iw}}\alpha_{q_w} + x_{pq_{iw}}\delta_{pq_w} + \underline{G}_{ijk} + \underline{\varepsilon}_{isjkr} \quad (8)$$

Genome positions with P-values ≤ 0.01 in the QTL profile produced by models (7) and (8) were selected to build a multi-QTL model.

Models (7) and (8) were also applied for the analysis of all unlinked markers (424), considering that for them the genetic predictor could only have the values -1 (allele *o*) and 1 (allele *a*). The Wald test was also used to identify putative QTL effects associated with individual markers.

Genomic positions with evidence of putative QTLs were included in a multi-QTL model. To determine which QTL were significant in this model, the Wald statistic was calculated after dropping each individual QTL separately from the full model. Non-significant QTL with P-value greater than 0.05 were then excluded. Finally, each of the remaining QTL were tested to determine significance of QTL $\times$ Location (QTL $\times$ L), QTL $\times$ Harvest (QTL $\times$ H) and QTL $\times$ Harvest $\times$ Location (QTL $\times$ H $\times$ L), also using Wald test. Only significant QTL effects were kept in the model and their effects were estimated. All the statistical analysis involving mixed models were performed through the Genstat software (PAYNE et al., 2009).

4 RESULTS

4.1 Linkage Map

From a total of 741 molecular markers, 317 (42.8%) were mapped to 96 linkage groups (LGs) with a total map length of 2468.14 cM and average distance between markers (marker density) of 7.5 cM. Most of the LGs (42, or 43.7%) were consisted of only two linked markers; 27 had three markers; 9 had four; 9 had five; 6 had six markers. The largest linkage groups had 10, 11 and 14 markers. The marker loci were substantially clustered along the LGs, while a minority were sparsely distributed with gaps larger than 20 cM, being observed on 11.8% of the intervals between two adjacent marker loci (Figure 3).

4.2 Multi-Harvest-Location Phenotypic Analysis

For each trait, different VCOV structures for the modeling of genetic correlations between locations and harvests were evaluated (Table 3). Models (a-e) consider each harvest-location combination as a single environment, while models (f-g) use direct products of (co)variance matrices for locations and harvests (SMITH et al., 2007; MALOSETTI et al., 2008). Model (a) considers homogeneous variation (ID), i.e. there are no genetic correlations between environments, and genetic variances are homogeneous across environments. Model (b) allows for heterogeneous genetic variances but assumes there are no genetic correlations between environments. Model (c) considers heterogeneous genetic variance and common genetic covariance between environments. Model (d) uses a multiplicative model called factor analytic model of order 1, to approximate a fully unstructured (co)variance matrix (OMAN, 1991; GOGEL; CULLIS; VERBYLA, 1995). Model (e) allows the VCOV matrix $\mathbf{G}$ to contain unique genetic variances and covariances. The other models combine in two different ways structures that make sense for the current data: autoregressive of order 1, in which the correlation between harvests decay with distance in time, with heterogeneous genetic variances (AR1_{Het}) and unstructured (UN) models, as proposed by Smith et al. (2007).

AIC and BIC provided different results in certain cases, for example, model (e) and (g) must be selected using AIC and BIC for TCH, respectively. However, the first and second best models for Fiber, for example, presented a difference smaller than 1 for the BIC values. Thus, for this reason we

decided to use AIC for the model selection in the phenotypic analyses. Although model (e) requires the estimation of a higher number of parameters, it showed the smallest AIC for all evaluated traits (Table 3). Based on these results model (e) was selected to be used in the QTL mapping procedure.

Table 3 – AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) for the mixed models, considering different VCOV structure for the genetic effect within location and harvest ($M = J \times K$, where J is the number of locations and K is the number of harvests; TCH: tonnes of cane per hectare; TSH: tonnes of sugar per hectare; Pol: sucrose content; and Fiber percent)

Trait	$\mathbf{G}$ matrix	Model	n_{PAR}	AIC	BIC
	$\mathbf{G} = \mathbf{G}^{L-H}_{M\times M}$	a) ID	1	7831.4	7834.0
		b) DIAG	6	7801.6	7817.2
		c) CS$_{Het}$	7	7083.0	7101.2
TCH		d) FA1	12	7039.4	7070.7
		e) Unst	21	**6909.3**	6964.0
	$\mathbf{G} = \mathbf{G}^{L}_{J\times J} \otimes \mathbf{G}^{H}_{K\times K}$	f) Unst $\otimes$ AR1$_{Het}$	$(3+4)-1=6$	6970.9	6986.5
		g) Unst $\otimes$ Unst	$(3+6)-1=8$	6934.0	**6954.8**
	$\mathbf{G} = \mathbf{G}^{L-H}_{M\times M}$	a) ID	1	3331.8	3334.4
		b) DIAG	6	3282.2	3297.8
		c) CS$_{Het}$	7	2693.8	2712.0
TSH		d) FA1	12	2646.1	2677.2
		e) Unst	21	**2560.8**	**2615.3**
	$\mathbf{G} = \mathbf{G}^{L}_{J\times J} \otimes \mathbf{G}^{H}_{K\times K}$	f) Unst $\otimes$ AR1$_{Het}$	$(3+4)-1=6$	2631.5	2647.1
		g) Unst $\otimes$ Unst	$(3+6)-1=8$	2601.6	2622.4
	$\mathbf{G} = \mathbf{G}^{L-H}_{M\times M}$	a) ID	1	1428.8	1431.4
		b) DIAG	6	1391.5	1407.1
		c) CS$_{Het}$	7	1026.9	1045.1
Pol		d) FA1	12	974.0	1005.1
		e) Unst	21	**944.7**	**999.2**
	$\mathbf{G} = \mathbf{G}^{L}_{J\times J} \otimes \mathbf{G}^{H}_{K\times K}$	f) Unst $\otimes$ AR1$_{Het}$	$(3+4)-1=6$	1091.1	1106.7
		g) Unst $\otimes$ Unst	$(3+6)-1=8$	1076.4	1097.2
	$\mathbf{G} = \mathbf{G}^{L-H}_{M\times M}$	a) ID	1	1072.4	1075.0
		b) DIAG	6	1075.8	1091.4
		c) CS$_{Het}$	7	254.9	273.1
Fiber		d) FA1	12	241.3	**272.4**
		e) Unst	21	**218.8**	273.3
	$\mathbf{G} = \mathbf{G}^{L}_{J\times J} \otimes \mathbf{G}^{H}_{K\times K}$	f) Unst $\otimes$ AR1$_{Het}$	$(3+4)-1=6$	271.3	286.9
		g) Unst $\otimes$ Unst	$(3+6)-1=8$	273.5	294.3

$\mathbf{G}$: genetic (co)variance matrix; ID: independent; DIAG: diagonal; CS$_{Het}$: compound symmetry (heterogeneous); FA1: factor analytic of order 1; AR1$_{Het}$: autoregressive of order 1 (heterogeneous); and Unst: unstructured; bold values: the smallest AIC or BIC value, indicating the best model.

4.3 QTL Analysis

The results of the QTL mapping through the interval mapping (IM) approach are summarised in Figure 4. These provided the identification of 28 putative QTLs, 9 for TCH, 8 for TSH, 4 for Pol and 7 for Fiber. Single marker (SM) analysis resulted in the detection of 22 marker-QTL associations: 5 for TCH, 8 for TSH, 8 for Pol and 5 for Fiber (Figure 5).

Significant positions or markers identified associated with a putative QTL by the IM and SM approaches were included in the multi-QTL model for the estimation of QTL main effects and QTL harvest-location-specific effects. Several QTLs (66%) showed a significant QTL $\times$ H (24%), QTL $\times$ L (14%) and QTL $\times$ H $\times$ L (28%) interaction, and 17 QTLs (34%) had the same effect across harvests and locations. The final multi-QTL model for TCH had 14 QTLs identified by the previous analyses and that remained significant in the multi-QTL analysis (Table 4). For this trait, the QTLs positioned on LG9 and LG19 had only significant positive additive main effects, which means that there are no QTL $\times$ H, QTL $\times$ L or QTL $\times$ H $\times$ L interaction effects. However, QTL $\times$ H interaction was detected for QTLs placed on LG25, LG32, LG72, LG92 and for markers EST3EC and ESTC81m3C, indicating that these QTLs showed the same behavior along the two locations, but not along harvests. For QTLs identified in LG66 and associated with marker ESTB64m3C, QTL $\times$ L interaction was detected, i.e. the effects of these QTLs are significantly different between locations, but keep the same effect along harvests. Moreover, the QTLs identified associated in LG8 and LG28, and with markers SG61BD1 and ESTC03m2D2, presented QTL $\times$ H $\times$ L interactions, wich means that the effects of these QTLs are significantly different along the combinations of harvest and location.

For TSH, 15 QTLs remained significant in the multi-QTL model (Table 5). QTLs placed on LG19 and LG21, and associated with markers SG105AD1, SG140CC and EST9BD2, presented significant additive main effects. The QTLs detected in LG6 and associated with markers EST3EC and ESTC03m2D2 had interaction with harvest and location, which means that each QTL had a different expression across the combinations of location and harvest. The other QTLs detected for this trait had QTL $\times$ L interaction (LG9 and associated with marker SG61BD1) or a QTL $\times$ H interaction (LG25, LG32, LG72, LG92 and associated with marker ESTC02m1D2).

From a total of 12 QTLs identified by IM and SM analyses, 10 QTLs remained significant in the multi-QTL model for Pol (Table 5). QTLs detected in LG6, LG35 and associated with markers SG06AD1, ESTA6AD1, ESTC15m2D2 and ESTA03m4C showed significant additive main effects

Table 4 – QTL effects estimated with the multi-QTL mixed model and their average standard error of difference (avsed) or standard error (TCH: tonnes of cane per hectare; α_p, α_q and δ_{pq} are the additive main effects on parents P and Q and dominance interaction, respectively; $\alpha_{p_{jk}}$, $\alpha_{q_{jk}}$ and $\delta_{pq_{jk}}$ are the harvest-location-specific effects)

Trait	LG (effect)	Markers	Position	Location-Harvest						(avsed)
			(cM)	1-1	1-2	1-3	2-1	2-2	2-3	
	8 ($\alpha_{q_{jk}}$)	EST2DD2/SG04AD1	0.0	1.43	0.60	0.66	-2.21	-0.42	0.44	(1.27)
	9 (α_p)	ESTB27m2D1/ESTC123m4D1	42.0	3.79	3.79	3.79	3.79	3.79	3.79	(1.27)[a]
	19 (α_q)	ESTB157m4D2/ESTB157m1D2	13.0	4.12	4.12	4.12	4.12	4.12	4.12	(1.27)[a]
	25 (α_{q_k})	EST1CC/ESTC47m3D1	2.0	-1.68	-4.15	-5.99	-1.68	-4.15	-5.99	(1.17)
	28 ($\alpha_{p_{jk}}$)	SG11FC/ESTA15m3C	13.0	2.34	-0.63	2.05	0.19	0.52	-1.01	(2.03)
	32 (α_{q_k})	ESTA63m3D2/ESTA48m2D2	22.0	-2.20	-3.07	-1.91	-2.20	-3.07	-1.91	(0.68)
TCH	66 (α_{p_j})	ESTA68m1C/ESTC129m5C	12.7	-7.08	-7.08	-7.08	-1.29	-1.29	-1.29	(1.66)
	72 (α_{p_k})	ESTA54m3D1/ESTB94m6D1	3.0	1.57	4.25	4.04	1.57	4.25	4.04	(0.73)
	92 (α_{q_k})	ESTB65m1D2/ESTC44m1D2	4.0	1.84	1.65	0.14	1.84	1.65	0.14	(0.68)
	NL ($\alpha_{p_{jk}}$)	SG61BD1		5.03	3.57	2.77	2.28	3.53	3.47	(1.25)
	NL (α_{p_k})	EST3EC		1.81	0.72	-0.93	1.81	0.72	-0.93	(0.85)
	NL (α_{p_j})	ESTB64m3C		6.97	6.97	6.97	1.28	1.28	1.28	(1.64)
	NL (α_{p_k})	ESTC81m3C		1.85	5.06	5.78	1.85	5.06	5.78	(0.83)
	NL ($\alpha_{q_{jk}}$)	ESTC03m2D2		0.57	1.91	2.08	2.65	0.89	-0.40	(1.16)

[a] Standard error; NL: not-linked.

across all combinations of location and harvest. QTL $\times$ H $\times$ L interactions were detected for QTLs associated with markers ESTB122m8D2 and ESTA03m5D1. The QTLs placed on LG81 and associated with marker ESTC49m3D1 presented QTL $\times$ L interaction and QTL $\times$ H interaction, respectively.

The multi-QTL model resulted in the identification of 11 QTLs for Fiber (Table 5). One had significant dominance effect in LG3. QTLs placed on LG35 and associated with markers SG25BC and ESTC110m2C presented significant main effects across all the combinations of locations and harvests. QTL $\times$ L interaction was detected on LG44 and associated with marker SG99DC, which means that these putative QTLs had a different expression across locations, but not across harvests. The other QTLs, positioned in LG37, LG55 and LG83, and associated with markers SG105AD1 and ESTB153m1D2, showed QTL $\times$ H $\times$ L interaction, changing the behavior not only across locations, but also along harvests.

Table 5 – QTL effects estimated with the multi-QTL mixed model and their average standard error of difference (avsed) or standard error (TSH: tonnes of sugar per hectare; Pol: sucrose content; Fiber percent; α_p, α_q and δ_{pq} are the additive main effects on parents P and Q and dominance interaction, respectively; $\alpha_{p_{jk}}$, $\alpha_{q_{jk}}$ and $\delta_{pq_{jk}}$ are the harvest-location-specific effects)

Trait	Linkage Group (type of effect)	Markers	Position (cM)	Location-Harvest 1-1	1-2	1-3	2-1	2-2	2-3	(avsed)
	6 ($\alpha_{p_{jk}}$)	ESTB07m1C/ESTA68m2C	15.1	-0.18	-0.79	-0.90	-0.20	-0.16	-0.83	(0.35)
	9 (α_{p_j})	ESTC123m4D1/ESTC48m5D1	56.8	-0.29	-0.29	-0.29	0.17	0.17	0.17	(0.13)
	19 (α_q)	ESTB157m4D2/ESTB157m1D2	16.3	0.43	0.43	0.43	0.43	0.43	0.43	(0.14)[a]
	21 (α_p)	SG26DD1/SG23BD1	0.0	-0.51	-0.51	-0.51	-0.51	-0.51	-0.51	(0.14)[a]
	25 (α_{q_k})	EST1CC/ESTC47m3D1	3.0	-0.46	-1.10	-1.47	-0.46	-1.10	-1.47	(0.21)
	32 (α_{q_k})	ESTA63m3D2/ESTA48m2D2	24.0	-0.35	-0.62	-0.46	-0.35	-0.62	-0.46	(0.13)
	72 (α_{p_k})	ESTA54m3D1/ESTB94m6D1	6.0	0.13	0.68	0.78	0.13	0.68	0.78	(0.13)
TSH	92 (α_{q_k})	ESTB65m1D2/ESTC44m1D2	4.0	0.31	0.38	0.15	0.31	0.38	0.15	(0.12)
	NL (α_{p_j})	SG61BD1		0.80	0.80	0.80	0.40	0.40	0.40	(0.15)
	NL (α_p)	SG105AD1		0.40	0.40	0.40	0.40	0.40	0.40	(0.14)[a]
	NL (α_p)	SG140CC		0.39	0.39	0.39	0.39	0.39	0.39	(0.16)[a]
	NL (α_q)	EST9BD2		-0.52	-0.52	-0.52	-0.52	-0.52	-0.52	(0.14)[a]
	NL ($\alpha_{p_{jk}}$)	EST3EC		-0.19	-0.09	-0.49	0.25	-0.06	-0.24	(0.24)
	NL (α_{q_k})	ESTC02m1D2		0.22	0.67	0.82	0.22	0.67	0.82	(0.13)
	NL ($\alpha_{q_{jk}}$)	ESTC03m2D2		0.26	0.72	0.85	0.53	0.23	0.00	(0.21)
	6 (α_p)	ESTA63m1C/ESTB111m2C	69.0	0.37	0.37	0.37	0.37	0.37	0.37	(0.13)[a]
	35 (α_p)	ESTB69m2D1/ESTB65m3D1	25.2	-0.16	-0.16	-0.16	-0.16	-0.16	-0.16	(0.07)[a]
	81 (α_{q_j})	ESTC113mD2/ESTC24m1D2	7.1	-0.12	-0.12	-0.12	-0.25	-0.25	-0.25	(0.05)
	NL (α_p)	SG06AD1		0.27	0.27	0.27	0.27	0.27	0.27	(0.07)[a]
Pol	NL (α_p)	ESTA6AD1		-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	(0.08)[a]
	NL (α_q)	ESTC15m2D2		-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	(0.07)[a]
	NL ($\alpha_{q_{jk}}$)	ESTB122m8D2		-0.16	0.01	-0.02	0.20	-0.13	0.11	(0.11)
	NL ($\alpha_{p_{jk}}$)	ESTA03m5D1		0.27	0.10	0.11	0.13	0.35	0.65	(0.10)
	NL (α_{p_k})	ESTC49m3D1		-0.37	-0.15	-0.09	-0.37	-0.15	-0.09	(0.06)
	NL (α_p)	ESTA03m4C		-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	(0.08)[a]
	3 (δ_{pq})	ESTA10m2D1/SG08A	69.0	-0.47	-0.47	-0.47	-0.47	-0.47	-0.47	(0.22)[a]
	35 (α_p)	ESTB69m2D1/ESTB65m3D1	20.0	0.26	0.26	0.26	0.26	0.26	0.26	(0.11)[a]
	37 ($\alpha_{q_{jk}}$)	ESTA61m3D2/ESTB75m1D2	7.0	-0.09	-0.18	-0.23	-0.21	-0.03	-0.11	(0.07)
	44 (α_{p_j})	ESTC123m3D1/ESTA06m4D1	20.0	-0.02	-0.02	-0.02	-0.16	-0.16	-0.16	(0.04)
	55 ($\alpha_{p_{jk}}$)	SG41FC/SG94EC	0.0	-0.34	-0.67	-0.54	-0.64	-0.46	-0.46	(0.13)
Fiber	83 ($\alpha_{p_{jk}}$)	ESTC129m1D1/ESTC119m1D1	6.3	0.19	0.28	0.10	0.23	0.08	0.18	(0.07)
	NL ($\alpha_{p_{jk}}$)	SG105AD1		-0.15	-0.03	-0.12	-0.03	-0.07	0.10	(0.07)
	NL (α_p)	SG25BC		-0.29	-0.29	-0.29	-0.29	-0.29	-0.29	(0.11)[a]
	NL (α_{p_j})	SG99DC		0.25	0.25	0.25	0.03	0.03	0.03	(0.05)
	NL (α_p)	ESTC110m2C		-0.41	-0.41	-0.41	-0.41	-0.41	-0.41	(0.15)[a]
	NL ($\alpha_{q_{jk}}$)	ESTB153m1D2		0.10	0.11	0.06	-0.03	0.18	0.10	(0.07)

[a] Standard error; NL: not-linked.

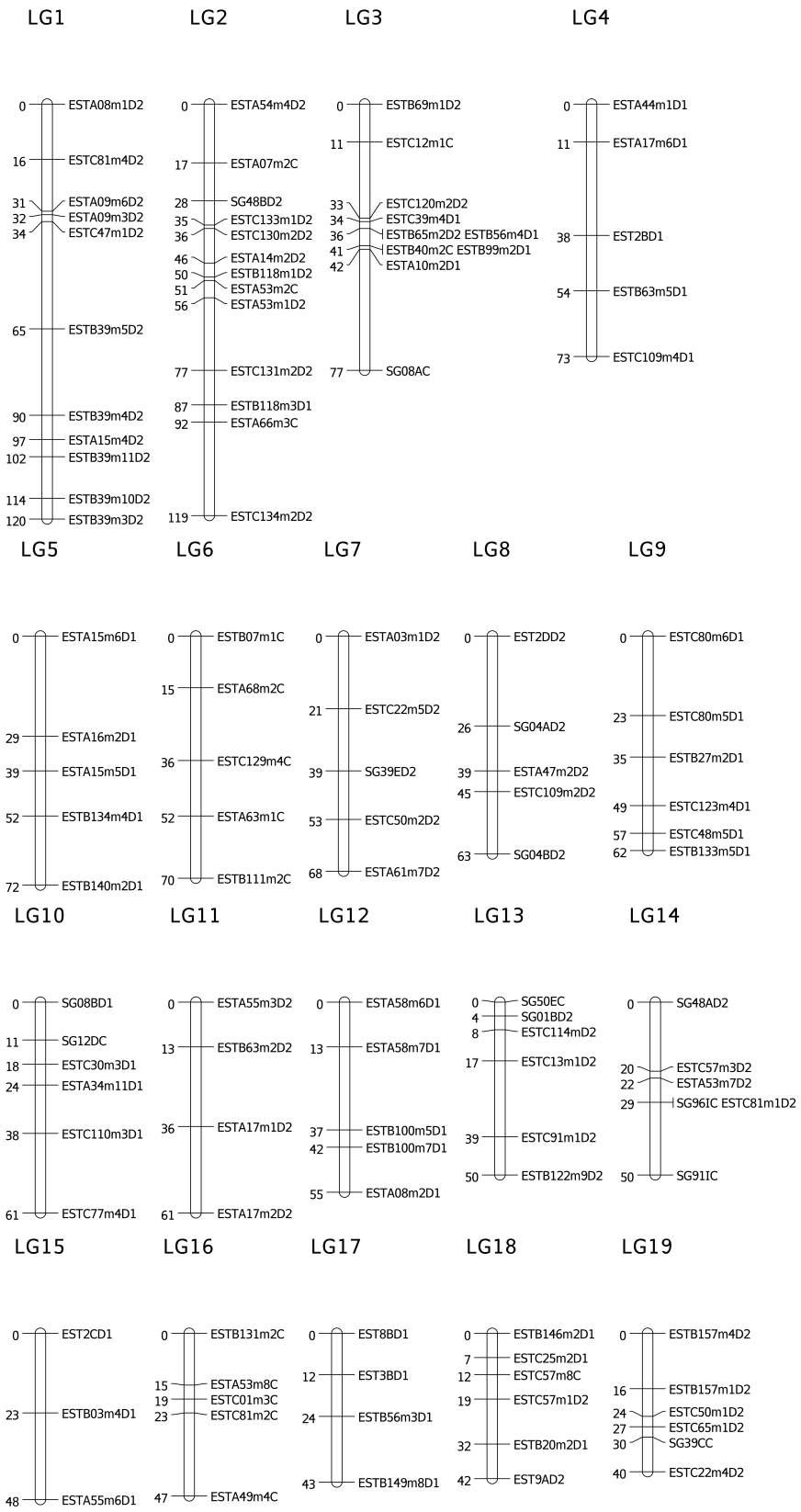

Figure 3 – Integrated genetic map of a sugarcane commercial cross (SP80-180 × SP80-4966) based on 100 individuals. Map distances are given in centi-Morgans (Kosambi). LG: linkage group

(Continues)

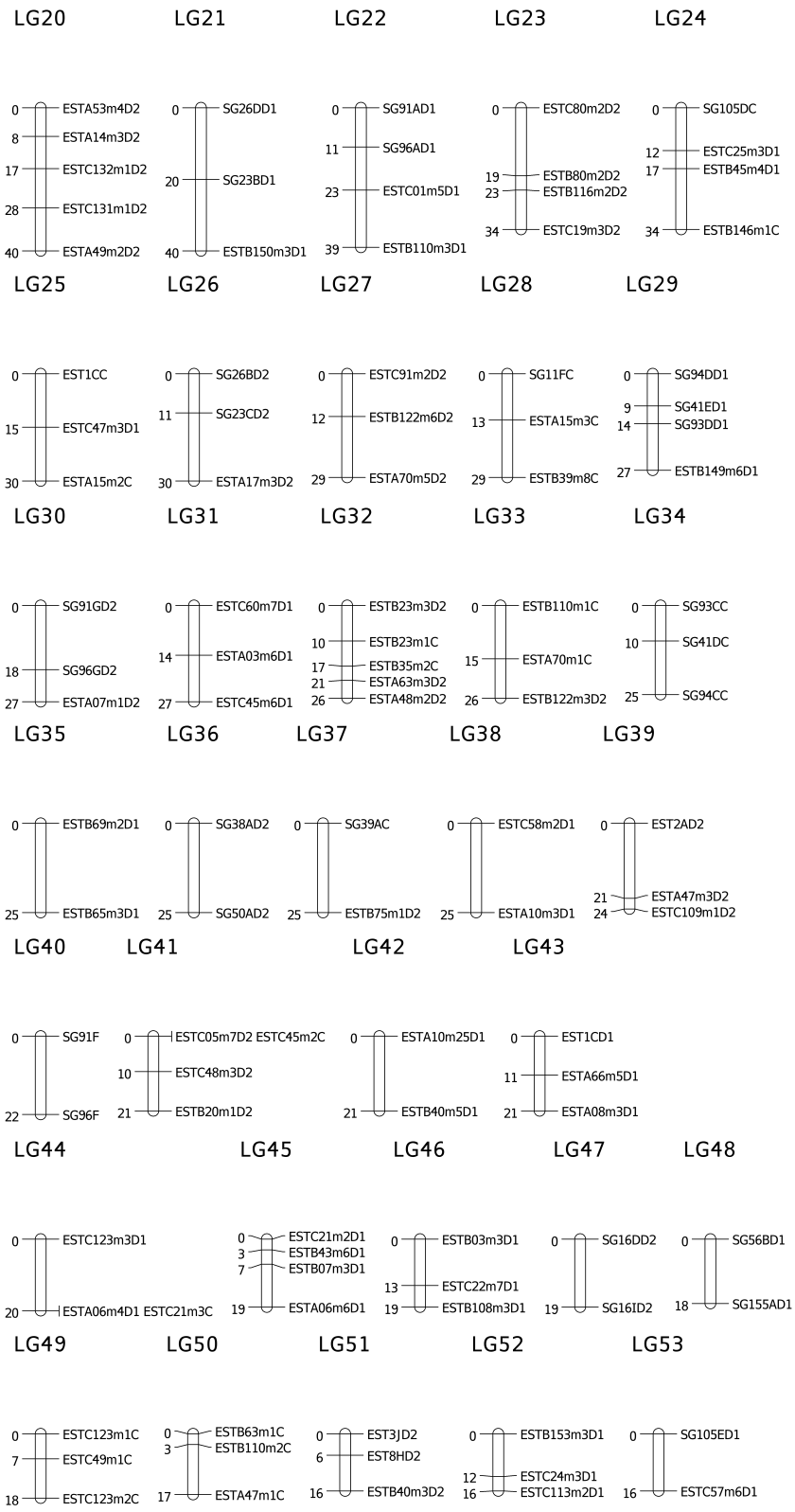

Figure 3 – Integrated genetic map of a sugarcane commercial cross (SP80-180 × SP80-4966) based on 100 individuals. Map distances are given in centi-Morgans (Kosambi). LG: linkage group

(Continued)

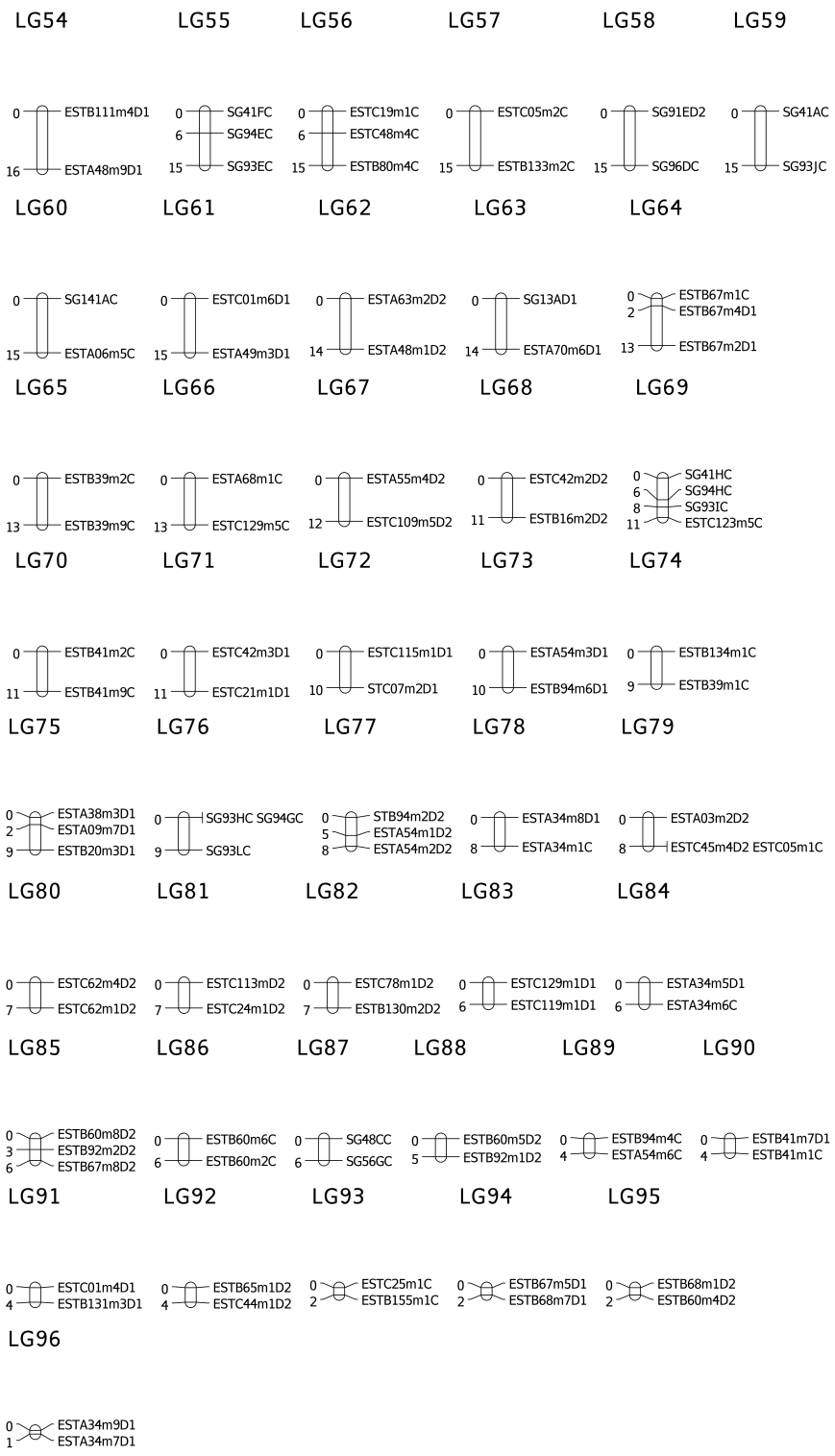

Figure 3 – Integrated genetic map of a sugarcane commercial cross (SP80-180 × SP80-4966) based on 100 individuals. Map distances are given in centi-Morgans (Kosambi). LG: linkage group

(Conclusion)

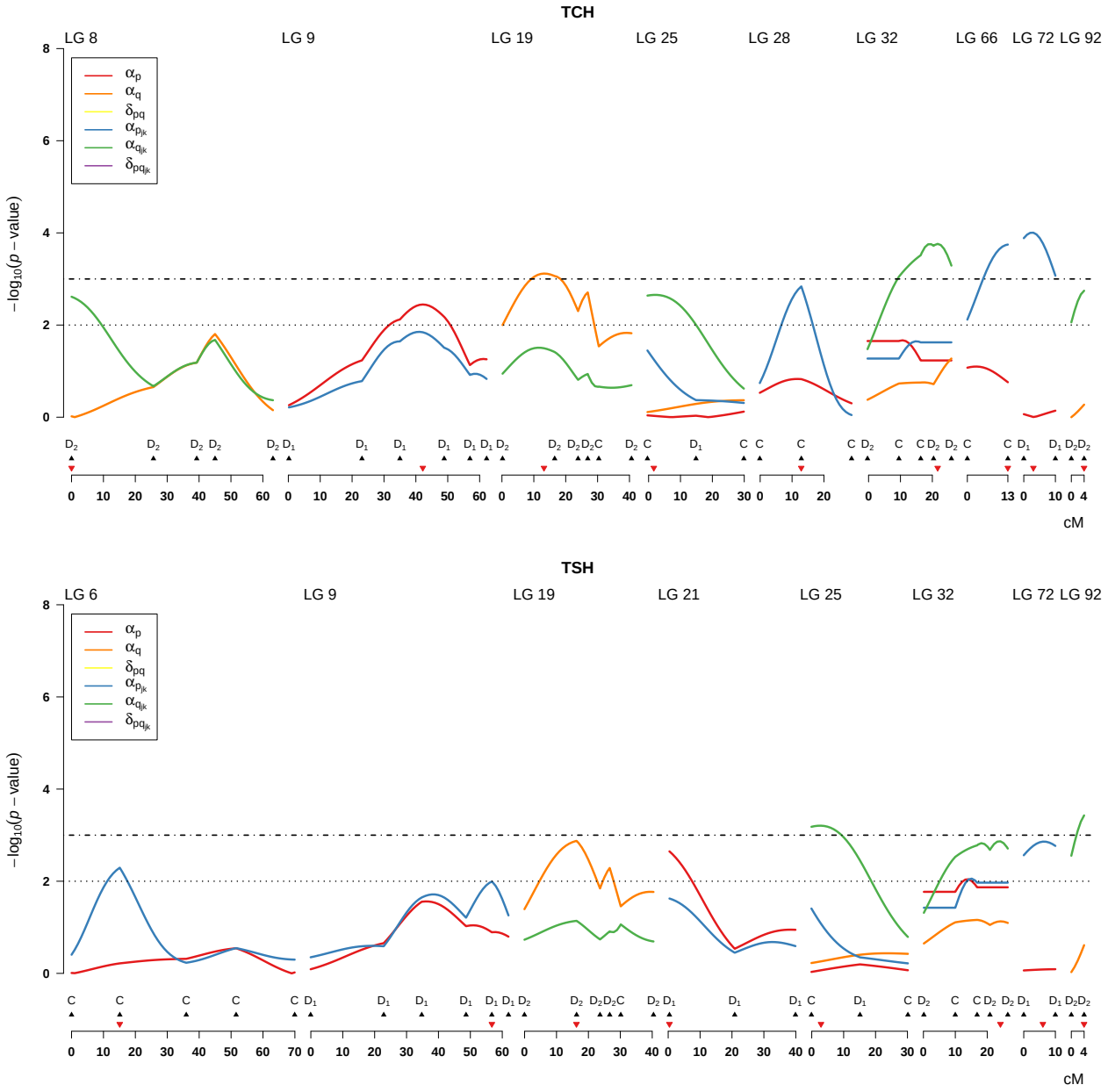

Figure 4 – Putative QTL (red triangle) associated with cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), Fiber percent and sucrose content (Pol) using the selected mixed model with QTL effects. Two different situations were considered: 1) using only main effects, model 8 (α_p, α_q and δ_{pq} for additive effects on parents P and Q and dominance interaction, respectively); 2) using genetic effects specifically for each harvest-location combination, through model 7 ($\alpha_{p_{jk}}$, $\alpha_{q_{jk}}$ and $\delta_{pq_{jk}}$). Not all effects were adjusted in all LG due to lack of information from SDM; black triangles: marker position; dotdashed line: $-\log_{10}(0.001)$ and doted line: $-\log_{10}(0.01)$

(Continues)

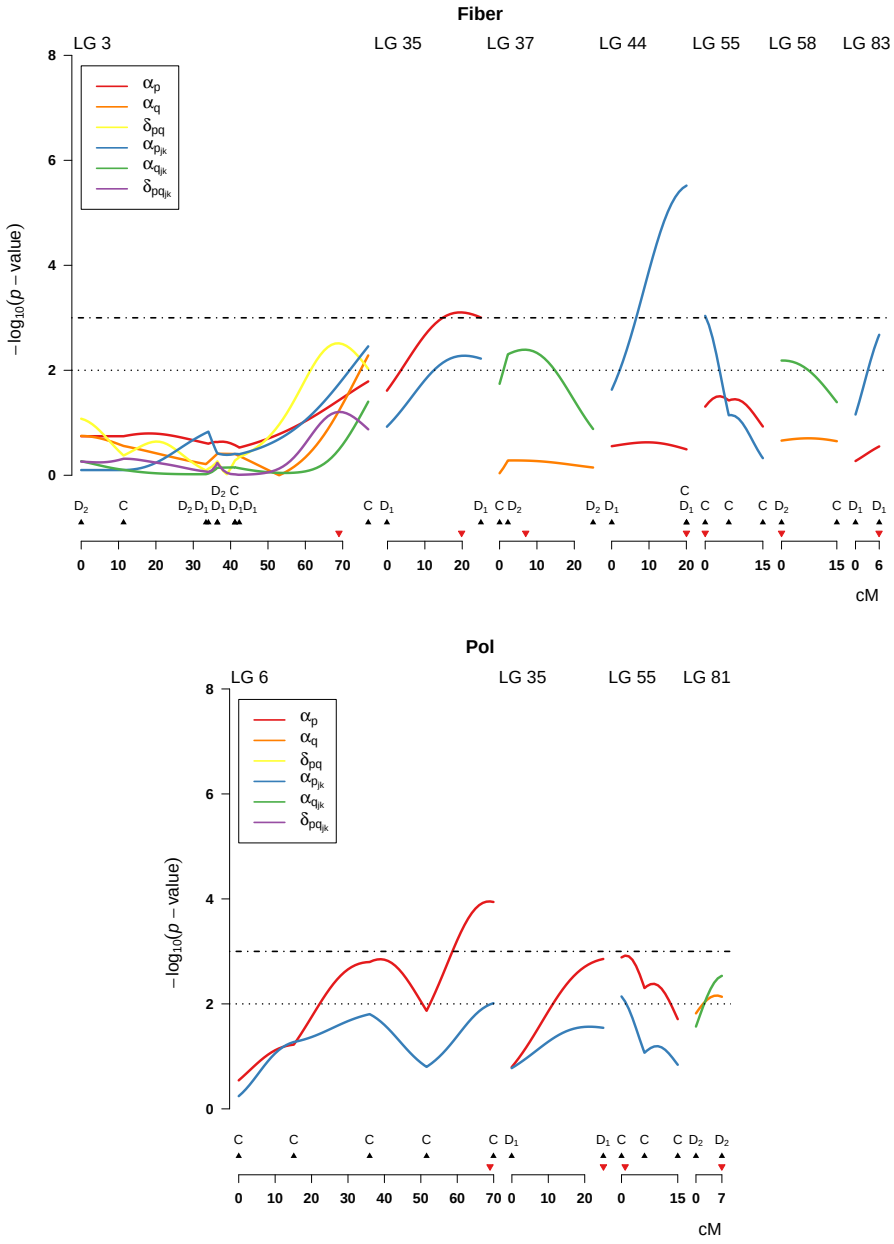

Figure 4 – Putative QTL (red triangle) associated with cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), Fiber percent and sucrose content (Pol) using the selected mixed model with QTL effects. Two different situations were considered: 1) using only main effects, model 8 (α_p, α_q and δ_{pq} for additive effects on parents P and Q and dominance interaction, respectively); 2) using genetic effects specifically for each harvest-location combination, through model 7 ($\alpha_{p_{jk}}$, $\alpha_{q_{jk}}$ and $\delta_{pq_{jk}}$). Not all effects were adjusted in all LG due to lack of information from SDM; black triangles: marker position; dotdashed line: $-\log_{10}(0.001)$ and doted line: $-\log_{10}(0.01)$
(Conclusion)

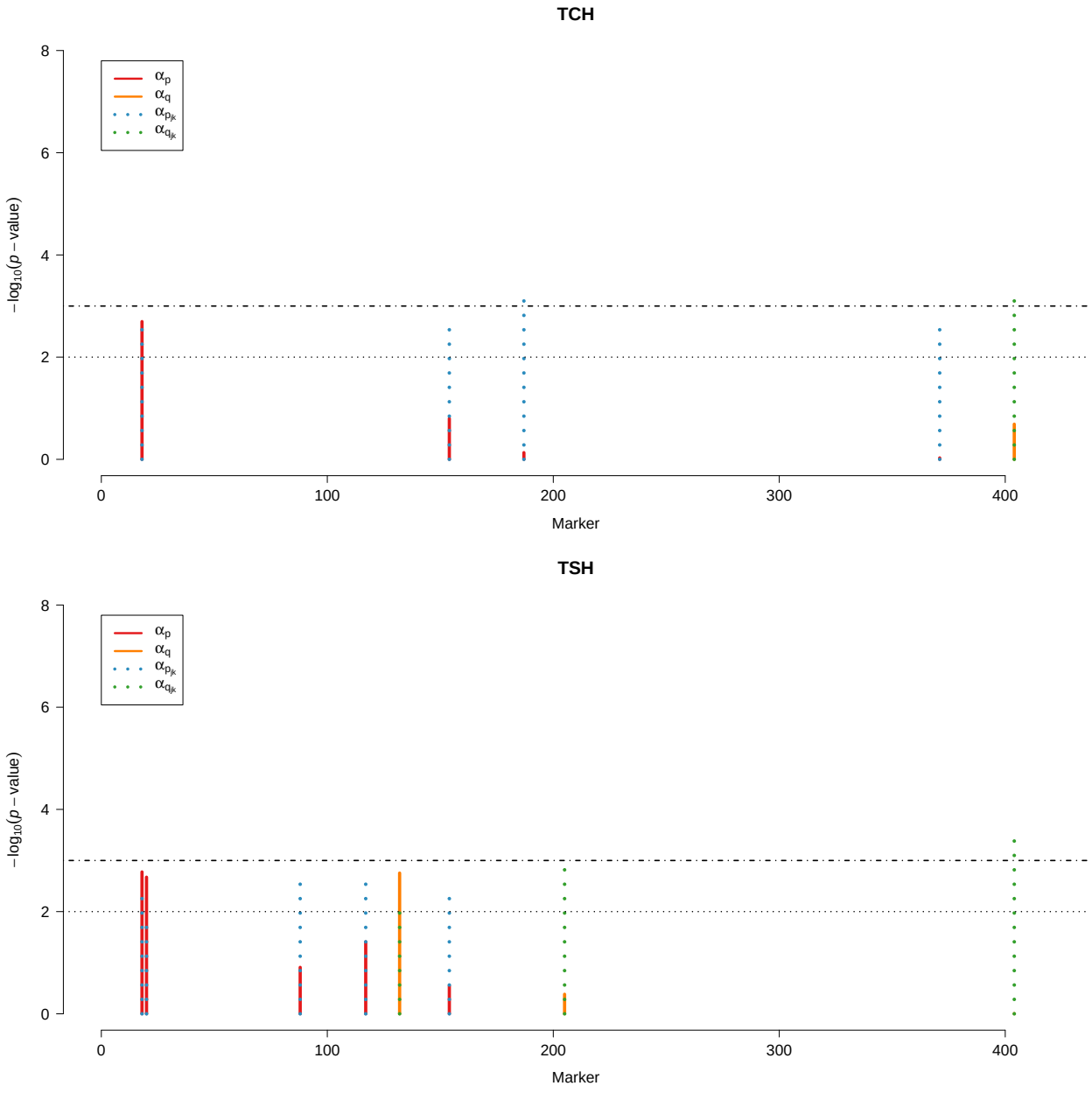

Figure 5 – Putative QTL associated with cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), Fiber percent and sucrose content (Pol) using the selected mixed model with QTL effects through SM analysis. Two different situations were considered: 1) using only main effects model (α_p and α_q for additive effects on parents P and Q, respectively); 2) using genetic effects specifically for each harvest-location combination ($\alpha_{p_{jk}}$ and $\alpha_{q_{jk}}$); dotdashed line: $-\log_{10}(0.001)$ and doted line: $-\log_{10}(0.01)$
(Continues)

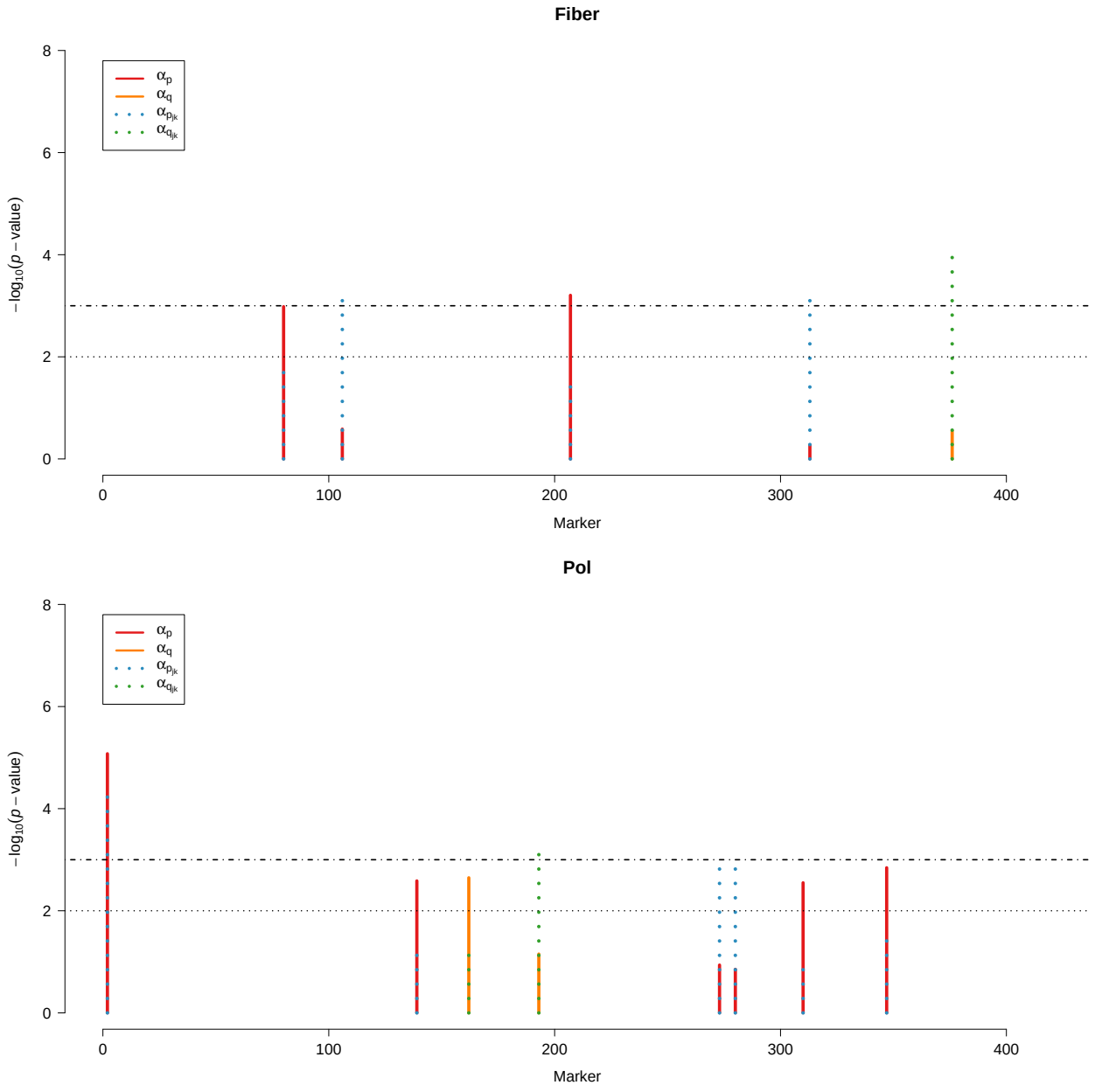

Figure 5 – Putative QTL associated with cane yield (tonnes of cane per hectare, TCH), sugar yield (tonnes of sugar per hectare, TSH), Fiber percent and sucrose content (Pol) using the selected mixed model with QTL effects through SM analysis. Two different situations were considered: 1) using only main effects model (α_p and α_q for additive effects on parents P and Q, respectively); 2) using genetic effects specifically for each harvest-location combination ($\alpha_{p_{jk}}$ and $\alpha_{q_{jk}}$); dotdashed line: $-\log_{10}(0.001)$ and doted line: $-\log_{10}(0.01)$ (Conclusion)

5 DISCUSSION

The number of LGs achieved here (96) is similar to $2n = 100 - 130$ chromosomes, expected for modern sugarcane cultivars (GRIVET; ARRUDA, 2001; HOARAU et al., 2001), although many LG could not be integrated (only D_1 or D_2 markers). However, the high number of unlinked markers (424) allied to the small length of most LGs and the reduced number of markers (loci) per LG indicates that the map is still not saturated. Probably, most of the small LGs represent unconnected parts of other groups. Usually, only single-dose polymorphisms have been selected for mapping (MING et al., 1998), thus gaps in sugarcane maps are commonly expected, due the exclusion of markers in multiples doses, i.e., duplex of monoparental origin, triplex or higher multiplex markers. As a consequence, the advantages of interval mapping could not be fully used for QTL mapping and therefore lower statistical power could be expected. Since multipoint estimates were used with the new version of *OneMap*, the map used here had higher likelihoods than the previous published ones, obtained for the same population (GARCIA et al., 2006; OLIVEIRA et al., 2007). The use of integrated maps, when compared with the *double pseudo-testcross* strategy, provides the obtention of linkage maps with increased marker saturation and higher representation of the genetic polymorphism generated by the cross, since markers with 3:1 and 1:1 segregation patterns may be used together, resulting in better maps and higher statistical power to detect QTL. Moreover, this approach allows the estimation of additive effects in each parent (α_p and α_q), which is presented for the first time for sugarcane in this article.

Mixed models were used here because of their flexibility and the possibility of modeling complex (co)variance structures resulting from repeatedly measures across locations and harvests. Although varietal selection for quantitative traits in sugarcane is usually based on information from series of field trials, with data for multiple harvests, the fitting of different VCOV structures for the genetic effect within location and harvest is rarely done (SMITH et al., 2007). For this mixed model analysis, genotypes were assumed to be random, since the main interest is in the genetic variation within the progeny rather than the genotypes themselves. The terms location (L) and harvest (H) were taken as fixed. Due to the reduced number of parameters, it was expected that the group of models that exploits the direct product of (co)variance matrices for location and harvest (models f-g) had the smallest values of AIC for the evaluated traits. However, the unstructured model con-

sidering the factorial combination of location and harvest (model e) was the best one for all traits, despite the large number of parameters. This could be a consequence of using AIC to compare models, since it tends to select model with more parameters. However, the differences with BIC values are, in general, small for this data set. Through this model it was possible to calculate the genetic variance intrinsic to each combination of location and harvest, showing the occurance of correlation and heterogeneity of variances among the different harvest-location combinations.

For QTL mapping, the model selected in the previous step was successfully applied, including fixed QTL main effects and harvest-location-specific QTL effects. In this case, QTL effects were tested taking into account the genetic correlation structure in the data. Piepho (2005), through a simulation study, showed that ignoring genetic correlations in multi-environment data leads to a substantial increase of the type I error rate when testing for QTL effects. Thus, it is expected that the current multiple-harvest-location mixed model approach will reduce the risk of over-optimistic conclusions, since an unstructured genetic (co)variance matrix was considered. Moreover, another important feature of the current approach is that all information was produced within the same model framework, avoiding to combine results from different analyses, what is commonly done in two-stage analyses, one for the BLUPs obtention and the other for QTL detection. Thus, the results show that QTL $\times$ H, QTL $\times$ L and QTL $\times$ H $\times$ L interaction effects were important for all evaluated traits, providing valuable information to understand the genetic control of complex traits related with sugarcane production and sucrose content.

From the total of QTLs identified, 17 (34%) showed stable effects across the diferent combinations of harvest and location, and 33 (66%) had some interaction. For the evaluated traits, TCH, TSH, Pol and Fiber, 12 (85.7%), 10 (66.7%), 4 (40%) and 7 (63.6%) of the QTLs identified, respectively, had some interaction with the environment. Most part of the interactions detected was QTL $\times$ Harvest. Probably, it can be explained by the fact that, for all evaluated traits, genotype by harvest (G $\times$ H) interaction compounded great part of the G $\times$ E interaction. QTLs with the same effect across harvests and locations (for example: QTLs detected in LG9 and LG19, for TCH; LG19 and LG21, for TSH; LG6 and LG35, and associated with markers SG06AD1, ESTA6AD1, ESTC15m2D2 and ESTA03m4C, for Pol; LG3 and LG35, and associated with markers SG25BC and ESTC110m2C, for Fiber) are important for studies looking for major genes controlling agronomic traits. However, if only stable QTLs across harvests are the goal of the research, the ones identified in LG66 and associated with marker ESTB64m3C, for TCH; LG9 and marker SG61BD1,

for TSH; LG81 for Pol; LG44 and marker SG99DC, are of main interest. In addition, for marker assisted seletion, it is important to consider the signals of the QTL efffects. For TCH, QTLs located in LG8 and LG28 showed effects changing in signal for the same harvest along the different locations, where QTL in LG28 had changes in signal from one harvest to another in the same location (harvest 1 and harvest 2, location 1). Moreover, the QTL identified in LG25 showed a negative effect increasing intensity with harvest time, which is particularly interesting in sugarcane, since yield decrease with harvest time.

In some cases, QTLs were identified in common linkage groups or associated with common markers for different traits. For example, TCH and TSH had 9 QTLs associated with the same linkage groups and with markers in common, including 6 QTLs in LG9, LG19, LG25, LG32, LG72 and LG92, and 3 QTLs related with markers SG61BD1, EST3EC and ESTC03m2D2. As all the common QTLs were close, it is possible to infer that they can be QTLs with pleiotropic effect on the two traits, TCH and TSH, i.e. a single QTL controlling two different traits. Common QTLs were also detected for Pol and Fiber in LG35, possibly with a pleiotropic effect, i.e. the same QTL controlling simultaneously the two different traits, as they were positioned in the same interval between markers. A special attention should be given to this QTL in breeding programs when multi-trait selection is involved, since it had opposite effects controlling Pol and Fiber.

It is difficult to compare the results presented here with other QTL studies in sugarcane, since this is the first to perform a study of G $\times$ H $\times$ L using mixed models. Hoarau et al. (2002), Jordan et al. (2004), McIntyre et al. (2005a, 2005b), Reffay et al. (2005), Aitken et al. (2008), Al-Janabi et al. (2007) and Piperidis et al. (2008) carried out SM analysis separately for each harvest or harvest-location combination (when available), considering maps obtained for each parent through the *double pseudo-testcross* strategy, and verified which QTLs were present or absent comparing the results for the different environments, concluding about stable QTLs across environments and QTL by environment interaction based on separate analyses.

QTL mapping in sugarcane still has several difficulties, such as the use of only SDM, low saturated linkage maps, reduced sample size (N), the occurrence of collinearity between the additive genetic predictors estimated for each parent (as a consequence of the lack of information provided by the markers). However, the present study provides many contributions, allowing the identification of a considerable number of QTLs for the evaluated traits, with information about effects, position, stable QTLs, QTL $\times$ H, QTL $\times$ L and QTL $\times$ H $\times$ L interaction. In addition, the statis-

cal models used to perform the analyses presented here can be used in future studies about QTL mapping involving different marker doses.

6 CONCLUSION

i.) The mixed models approach was succesfuly applied to the analysis of sugarcane multi-harvest-location trials (MHLT) data. It was possible to model complex (co)variance structures, taking into account heterogeneous genetic variance and the existence of genetic correlation between harvest-location combinations. The unstructured VCOV model showed the smallest values of AIC for all evaluated traits and was selected for the QTL analysis.

ii.) QTL $\times$ H, QTL $\times$ L and QTL $\times$ H $\times$ L interaction effects were important for all evaluated traits. From the total of QTLs identified, 33 (66%) had some interaction and only 17 (34%) showed stable effects across the different combinations of harvest and location. Most of the interactions were due to QTL $\times$ H.

iii.) After the final model was adjusted, several QTLs for the evaluated traits were identified, providing information about genetic effects, positions, stable QTLs and QTL $\times$ E interactions (QTL $\times$ H, QTL $\times$ L and QTL $\times$ H $\times$ L). These results could provide useful information in order to have a better understanding of the genetic control of complex traits, such as biomass production and sucrose content.

REFERENCES

ABATE, Z.A.; LIU, S.; McKENDRY, A.L. Quantitative trait loci associated with deoxynivalenol content and kernel quality in the soft red winter wheat 'Ernie'. **Crop Science**, Madison, v. 48, p. 1408-1418, 2008.

AITKEN, K.S.; HERMANN, S.; KARNO, K.; BONNETT, G.D.; McINTYRE, C.L.; JACKSON, P.A. Genetic control of yield related stalk traits in sugarcane. **Theoretical and Applied Genetics**, New York, v. 117, p. 1191-1203, 2008.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. **Theoretical and Applied Genetics**, New York, v. 110, p. 789-801, 2005.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. **Genome**, Ottawa, v. 50, p. 742-756, 2007.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar x *Saccharum officinarum* population. **Theoretical and Applied Genetics**, New York, v. 112, p. 1306-1317, 2006.

AITKEN, K.S.; McNEIL, M. Diversity analysis. In: HENRY, R. (Ed.) **Genetics, Genomics and Breeding of Sugarcane**. New Hampshire: Science Publishers. ca. 2010. 1v.

AKAIKE, H. A new look at the statistical model identification. **The Institute of Electrical and Electronics Engineers Transaction and Automatic Control**, Notre Dame, v. 19, p. 716-723, 1974.

AL-JANABI, S.M.; HONEYCUTT, R.J.; McCLELLAND, M.; SOBRAL, B.W.S. A genetic linkage map of *Saccharum spontaneum* (L.) 'SES 208'. **Genetics**, Bethesda, v. 134, p. 1249-1260, 1993.

AL-JANABI, S.M.; PARMESSUR, Y.; KROSS, H.; DHAYAN, S.; SAUMTALLY, S.; RAMDOYAL, K.; AUTREY, L.J.C.; DOOKUN-SAUMTALLY, A. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. **Molecular Breeding**, Berlin, v. 19, p. 1-14, 2007.

ALWALA, S.; KIMBENG, C.A. Molecular genetic linkage mapping in *Saccharum*: strategies, resources and achievements. In: HENRY, R. (Ed.) **Genetics, genomics and breeding of sugarcane**. New Hampshire: Science Publishers., ca. 2010. 1v.

ALWALA, S.; KIMBENG, C.A.; VEREMIS, J.C.; GRAVOIS, K.A. Linkage mapping and genome analysis in a *Saccharum* interespecific cross using AFLP, SRAP and TRAP markers. **Euphytica**, Wageningen, v. 164, p. 37-51, 2008.

ASNAGHI, C.; D'HONT, A.; GLASZMANN, J.C.; ROTT, P. Resistance of sugarcane cultivar R 570 to *Puccinia melanocephala* isolates from different geographic locations. **Plant Disease**, Sant Paul, v. 85, p. 282-286, 2001.

BALZARINI, M. Applications of mixed models in plant breeding. In: KANG, M.S. (Ed.) **Quantitative genetics, genomics and plant bredding**. New York: CABI Publishing. 2001. p. 353-363.

BASTEN, C.J.; WEIR, B.S.; ZENG, Z.B. QTL-Cartographer: a reference manual and tutorial for QTL mapping. Haley, 2005. Disponível em: <http://statgen.ncsu.edu/qtlcart>. Acesso em: 20 jan. 2010.

BOER, M.P.; WRIGHT, D.; FENG, L.; PODLICH, D.W.; LUO; L.; COOPER, M.; EEUWIJK, F.A. van. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. **Genetics**, Bethesda, v. 177, p. 1801-1813, 2007.

BROMAN, K.W.; WU, H.; SEN, S.; CHURCHILL, G.A. R/qtl: QTL mapping in experimental crosses. **Bioinformatics**, Oxford, v. 19, p. 889-890, 2003.

BURNHAM, K.P.; ANDERSON, D.R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods & Research**, London, v. 33, n. 2, p. 261-304, 2004.

CARDINAL, A.J.; LEE, M.; SHAROPOVA, N. WOODMAN-CLIKEMAN, W.L.; LONG, M.L. Genetic mapping and analysis of quantitative Trait Loci for resistance to stalk tunneling by the european corn borer in maize. **Crop Science**, Madison, v. 41, p. 835-845, 2001.

CARLIER, J.D.; REIS, A.; DUVAL, M.F.; D'EECKENBRUGGE, G.C.; LEITÃO, M. Genetic maps of RAPD, AFLP and ISSR markers in *Ananas bracteatus* and *A. comosus* using the pseudotestcross strategy. **Plant Breeding**, Berlin, v. 123, p. 186-192, 2004.

CAVALCANTI, J.J.V.; WILKINSON, M.J. The first genetic maps of cashew (*Anacardium occidentale* L.). **Euphytica**, Wageningen, v. 157, p. 131-143, 2007.

CHAPMAN, S.C. Use of crop models to understand genotype by environment interactions for drought in realworld and simulated plant breeding trials. **Euphytica**, Wageningen, v. 161, p. 195-208, 2008.

CHARCOSSET, A.; MOREAU, L. Use of molecular markers for the development of new cultivars and the evaluation of genetic diversity. **Euphytica**, Wageningen, v. 137, p. 81-94, 2004.

CHEN, C.; KIM D. BOWMAN, K.D.; CHOI, Y.A.; DANG, P.M.; RAO, M.N.; HUANG, S., SONEJI, J.R.; McCOLLUM, T.G.; GMITTER, F.G.Jr. EST-SSR genetic maps for *Citrus sinensis* and *Poncirus trifoliata*. **Tree Genetics & Genomes**, Berlin, v. 4, p. 1-10, 2008.

CHO, Y.G.; KANG, H.J.; LEE, J.S.; LEE, Y.T.; LIM, S.J.; GAUCH, H.; EUN, M.Y.; McCouch, S.R. Identification of quantitative trait loci in rice for yield, yield components, and agronomic traits across years and locations. **Crop Science**, Madison, v. 47 p. 2403-2417, 2007.

CHRISTENSEN, R.; JOHNSON, W.; PEARSON, L.M. Prediction diagnostics for spatial linear models. **Biometrika**, Cambridge, v. 79, n. 3, p. 583-591, 1992.

CONAB - Companhia Nacional de Abastecimento. 3^{o} levantamento de cana-de-açúcar - Dezembro de 2009. Disponível em: <http://www.conab.gov.br/conabweb>. Acesso em: 20 jan. 2010.

CULLIS, B.; GOGEL, B.; VERBYLA, A.; THOMPSON, R. Spatial analysis of multi-environment early generation variety trials. **Biometrics**, Washington, v. 54, n. 1, p. 1-18, 1998.

CROSSA, J.; VARGAS, M.; van EEUWIJK, F.A.; JIANG, C.; EDMEADES, G.O.; HOISINGTON, D. Interpreting genotype × environment interaction in tropical maize using linked molecular markers and environmental covariables. **Theoretical and Applied Genetics**, New York, v. 99, p. 611-625, 1999.

DAUGROIS, J.H.; GRIVET, L.; ROQUES, D.; HOARAU, J.Y.; LOMBARDI, H.; GLASZMANN, J.C.; D'HONT, A. A putative major gene for rust resistence linked with a RFLP marker in sugarcane cultivar 'R570'. **Theoretical and Applied Genetics**, New York, v. 92, p. 1059-1064, 1996.

DEKKERS, J.C.M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, New York, v.3, p.22-32, 2002.

DENIS, J.B.; PIEPHO, H.P.; van EEUWIJK, F.A. Modelling expectation and variance for genotype by environment data. **Heredity**, London, v. 79, p. 162-171, 1997.

D'HONT, A.; GRIVET, L.; FELDMANN, P.; Rao, P.; BERDING, N.; GLASZMANN, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular Genetics and Genomics**, Berlin, v. 250, p. 405-413, 1996.

D'HONT, A.; ISON, D.; ALIX, K.; GLASZMANN, J.C. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. **Genome**, Ottawa, v. 41, p. 221-225, 1998.

D'HONT, A.; LU, Y.H.; GONZÁLES de LEON, D.; GRIVET, L.; FLEDMANN, P.; LANAUD, C.; GLASZMANN, J.C. A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the andropogoneae. **Genome**, Ottawa, v. 37, p. 222-230, 1994.

DIETRICH, W.F.; MILLER, J.; STEEN, R.; MERCHANT, M.A.; DAMRON-BOLES, D.; HUSAIN, Z.; DREDGE, R.; DALY, M.J.; INGALLS, K.A.; O'CONNOR, T.J.; EVANS, C.A.; DeANGELIS, M.M.; LEVINSON, D.M.; KRUGLYAK, L.; GOODMAN, N.; COPELAND, N.G.; JENKINS, N.A.; HAWKINS, T.L.; STEIN, L.; PAGE, D.C.; LANDER, E.S. A comprehensive genetic map of the mouse genome. **Nature**, London, v. 380, p. 149-152, 1996.

DIGGLE, P.J. An Approach to the Analysis of Repeated Measurements. **Biometrics**, Washington, v. 44, n. 4, p. 959-971, 1988.

DOERGE, R.W. Mapping and analysis of quantitative trait loci in experimental populations. **Nature Reviews Genetics**, New York, v. 3, p. 43-52, 2002.

DOERGE, R.W.; ZENG, Z.B.; WEIR, B.S. Statistical issues in the search for genes affecting quantitative traits in experimental populations. **Statistical Science**, Hayward, v. 12, p. 195-219, 1997.

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

ECKERMANN, P.J.; VERBYLA, A.P.; CULLIS, B.R.; THOMPSON, R. The abalysis of quantitative traits in wheat mapping populations. **Australian Journal of Agricultural Research**, Melbourne, v. 52, p. 1195-1206, 2001.

EEUWIJK, F.A. van; BOER, M.; TOTIR, L.R.; BINK, M.; WRIGHT, D.; WINKLER, C.R.; PODLICH, D.; BOLDMAN, K.; BAUMGARTEN, A.; SMALLEY, M.; ARBELBIDE, M.; ter BRAAK, C.J.F.; COOPER, M. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. **Theoretical and Applied Genetics**, New York, v. 120, 429-440, 2010.

EEUWIJK, F.A. van; MALOSETTI, M.; BOER, M.P. Modelling the genetic basis of response curves underlying genotype × environment interaction. In: SPIERTZ, J.H.J.; STRUIK, P.C.; van LAAR, H.H. (Ed.) **Scale and complexity in plant systems research:** gene-plant-crop relations. Dordrech: Springer, 2007, p. 115-126.

EEUWIJK, F.A. van; MALOSETTI, M.; YIN, X.; STRUIK, P.C.; STAM, P. Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. **Australian Journal of Agricultural Research**, Melbourne, v. 56, p. 883-894, 2005

EDMÉ, S.J.; GLYNN, N.G.; COMSTOCK, J.C. Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum*. **Heredity**, London, v. 97, p. 366-375, 2006.

EDWARDS, M.D.; STUBER, C.W.; WENDEL, J.F. Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. **Genetics**, New York, v. 116, p. 113-125, 1987.

EMRICH, K.; PRICE, A.; PIEPHO, H.P. Assessing the importance of genotype × environment interaction for root traits in rice using a mapping population. III: QTL analysis by mixed models. **Euphytica**, Wageningen, v. 161, n. 1-2, p. 229-240, 2008.

FAO. **FAOSTAT - Agriculture**. Disponível em: <http://faostat.fao.org/site/339/default.aspx>. Acesso em: 20 jan. 2010.

FISHER, R. A.; MACKENZIE, W. A. Studies in crop variation. II. The manurial response of different potato varieties. **The Journal of Agriculture Science**, Cambridge, v. 13, p. 311-20, 1923.

FNP - CONSULTORIA & COMÉRCIO. **AGRIANUAL 2009:** Anuário da Agricultura Brasileira, São Paulo, 2008. 497 p.

GABRIEL, K.R. Biplot display of multivariate matrices with application to principal components analysis. **Biometrika**, Cambridge, v. 58, p. 453-467, 1971.

GARCIA, A.A.F.; KIDO, E.A.; MEZA, A.N.; SOUZA, H.M.B.; PINTO, L.R.; PASTINA, M.M.; LEITE, C.S.; SILVA, J.A.G. DA; ULIAN, E.C.; FIGUEIRA, A.; SOUZA, A.P. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, New York, v. 112, p. 298-314, 2006.

GARCIA, A.A.F.; WANG, S.; MELCHINGER, A.E.; ZENG, Z.B. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Bethesda, v. 180, p. 1707-1724, 2008.

GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v. 44, p. 705-15, 1988.

GAUCH, H.G. **Statistical analysis of regional yield trials:** AMMI analysis of factorial designs, Amsterdam: Elsevier, 1992. 172 p.

GOGEL, B.J.; CULLIS, B.R.; VERBYLA, A.P. REML estimation of multiplicative effects in multi-environment variety trials. **Biometrics**, Washington, v. 51, n. 2, p. 744-749, 1995.

GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance technique. **Psychometrika**, Baltimore, v. 33, n. 1, p. 73-115, 1968.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategyand RAPD markers. **Genetics**, Bethesda, v. 137, p. 1121-1137, 1994.

GREEN, C.V. Further evidence of linkage in size inheritance. **American Naturalist**, Chicago, v. 67, p. 377-380, 1933.

GREEN, C.V. Linkage in size inheritance. **American Naturalist**, Chicago, v. 65, p. 502-511, 1931.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, London, v. 5, p. 122-127, 2001.

GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.E.; GLASZMANN, J.C. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interespecific hybrid. **Genetics**, Bethesda, v. 142, p. 987-1000, 1996.

GUIMARÃES, C.T.,; HONEYCUTT, R.J.; SILLS, G.R.; SOBRAL, B.W.S. Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes and Jew. Ex. Grassl. **Genetics and Molecular Biology**, Ribeirão Preto, v. 22, p. 125-132, 1999.

GUIMARÃES, C.T.; SILLS, G.R.; SOBRAL, B.W.S. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Sciences**, Washington, v. 94, p. 14261-14266, 1997.

HALEY, C.S.; KNOTT, S.A.; ELSEN J.M. Mapping quantitative trait loci in crosses between outbred lines using least squares. **Genetics**, Bethesda, v.136, p. 195-207, 1994.

HEALTH, S.C. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. **American Journal of Human Genetics**, Baltimore, v. 61, p. 748-760, 1997.

HEINZ, D.J.; TEW, T.L. Hybridization procedures. In: HEINZ, D.J. (Ed.) **Sugarcane improvement through breeding**. Amsterdam: Elsevier. 1987. p. 313-342.

HENDERSON, C.R. Statistical methods in animal improvement: historical overview. In: GIANOLA, D.; HAMMOND, K. (Ed.) **Advances in Statistical Methods for Genetic Improvement of Livestock**. New York: Springer-Verlag, 1990, p. 1-14.

HOARAU, J.Y.; GRIVET, L.; OFFMAN, B.; RABOIN, L.M.; DIORFLAR, J.P.; PAYET, J.; HELLMAN, M.; D'HONT, A.; GLASZMANN, J.C. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTL's for yield components. **Theoretical and Applied Genetics**, New York, v. 105, p. 1027-1037, 2002.

HOARAU, J.Y.; OFFMAN, B.; D'HONT, A.; RISTERUCCI0, A.M.; ROQUES, D.; GLASZMANN, J.C.; GRIVET, L. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. **Theoretical and Applied Genetics**, New York, v. 103, p. 84-97, 2001.

IRVINE, J.E. *Saccharum* species as horticultural classes. **Theoretical and Applied Genetics**, New York, v. 98, p. 186-194, 1999.

JANSEN, R.C.; STAM, P. Resolution of quantitative traits into multiple loci via interval mapping. **Genetics**, Bethesda, v. 136, p. 1447-1455, 1994.

JIANG, C.; ZENG, Z.B. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Bethesda, v. 140, p. 1111-1127, 1995.

JOHNSON, D.L.; JANSEN, R.C.; van ARENDONK, J.A.M. Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. **Genetical Research**, London, v. 73, p. 75-83, 1999.

JORDAN, D.R.; CASU, R.E.; BESSE, P.; CARROLL, B.C.; BERDING, N.; McINTYRE, C.L. Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. **Genome**, Ottawa, v. 47, p. 988-993, 2004.

KAO, C.H.; ZENG, Z.B. General formulae for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. **Biometrics**, Washington, v. 53, p. 653-665, 1997.

KAO, C.H., ZENG, Z.B.; TEASDALE, R D. Multiple interval mapping for quantitative trait loci. **Genetics**, Bethesda, v. 152, p. 1203-1216, 1999.

KOSAMBI, D.D. The estimation of map distances from recombination values. **Annual of Eugene**, London, v. 12, p. 172-175, 1944.

LANDER, E.S.; GREEN, P.; ABRAHANSON, J.; BARLOW, A.; DALY, M.J.; LINCON S.E.; NEWBURG, L. MAPMAKER: An interactive computing package for constructing primary genetic linkages of experimental and natural populations. **Genomics**, Orlando, v. 1, p. 174-181, 1987.

LANDER; E.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Bethesda, v. 121, p. 185-199, 1989.

LI, D.; PFEIFFER, T.W.; CORNELIUS, P.L. Soybean QTL for yield and yield components associated with Glycine soja alleles. **Crop Science**, Madison, v. 48, p. 571-581, 2008.

LIMA, M.L.A.; GARCIA, A.A.F.; OLIVEIRA, K.M.; MATSUOKA, S.; ARIZONO, H.; SOUZA JR., C.L.; SOUZA, A.P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). **Theoretical and Applied Genetics**, New York, v. 104, p. 30-38, 2002.

LIN, M.; LOU, X.Y.; CHANG, M.; WU, R. A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. **Genetics**, Bethesda, v. 165, p. 901-913, 2003.

LINDSTROM, E.W. A genetic linkage between size and color factors in the tomato. **Science**, Washington, v. 60, p. 182-183, 1924.

LINDSTROM, E.W. Genetic tests for linkage between row and certain qualitative genes in maize. **Research Bulletin**, Ames, v. 142, p. 250-288, 1931.

LIU, B.H. **Statistical genomics:** linkage, mapping, and QTL analysis. Boca Raton: CRC Press, 1998. 611p.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980p.

MACCAFERRI, M.; SANGUINETI, M.C.; CORNETI, S.; ORTEGA, J.L.A.; SALEM, M.B.; BORT, J.; DeAMBROGIO, E.; MORAL, L.F.G. del; DEMONTIS, A.; EL-AHMED, A.; MAALOUF, F.; MACHLAB, H.; MARTOS, V.; MORAGUES, M.; MOTAWAJ, J.; NACHIT, M.; NSERALLAH, N.; OUABBOU, H.; ROYO, C.; SLAMA, A.; TUBEROSA, R. Quantitative trait loci for grain yield and adaptation of Durum Wheat (*Triticum durum* Desf.) across a wide range of water availability. **Genetics**, Bethesda, v. 178, p. 489-511, 2008.

MACKAY, T.F.C. Quantitative trait loci in Drosophila. **Nature Reviews Genetics**, New York, v. 2, p. 11-20, 2001.

MALIEPAARD, C.; JANSEN, J.; van OOIJEN, J.W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, Cambridge, v. 70, p. 237-250, 1997.

MALOSETTI, M.; RIBAUT, J.M.; VARGAS, M.; CROSSA, J.; van EEUWIJK, F.A. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Wageningen, v. 161, p. 241-257, 2008.

MALOSETTI, M.; VOLTAS, J.; ROMAGOSA, I.; ULLRICH, S.E.; van EEUWIJK, F.A. Mixed models including environmental covariables for studying QTL by environment interaction. **Euphytica**, Wageningen, v. 137, p. 139-145, 2004.

MANGOLIN, C.A.; SOUZA JR., C.L. de; GARCIA, A.A.F.; GARCIA, A.F.; SIBOV, S.T.; SOUZA, A.P. de. Mapping QTLs for kernel oil content in a tropical maize population. **Euphytica**, Wageningen, v. 137, p. 251-259, 2004.

MANLY, K.F.; CUDMORE JR, R.H.; MEER, J.M. Map Manager QTX, cross-platform software for genetic mapping. **Mammalian Genome**, New York, v. 12, p. 930-932, 2001.

MARGARIDO, G.R.A.; SOUZA, A.P.; GARCIA, A.A.F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, Lund, v. 144, p. 78-79, 2007.

MARTÍNEZ, O.; CURNOW, R.N. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. **Theoretical and Applied Genetics**, New York, v. 85, p. 480-488, 1992.

MATHEWS, K.L.; MALOSETTI, M.; CHAPMAN, S.; McINTYRE, L.; REYNOLDS, M.; SHORTER, R.; van EEUWIJK, F. MultienvironmentQTL mixed models for drought stress adaptation in wheat. **Theoretical and Applied Genetics**, New York, v. 117, p. 1077-1091, 2008.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A (Ed.) **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999. p. 205-252.

McINTYRE, C.L.; CASU, R.E.; DRENTH, J.; KNIGHT, D.; WHAM, V.A.; CROFT, B.J.; JORDAN, D.R.; MANNERS. Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance. **Genome**, Ottawa, v. 48, p. 391-400, 2005a.

McINTYRE, C.L.; JACKSON, P.A.; CORDEIRO, G.M.; AMOUYAL, O.; HERMANN, S.; AITKEN, K.S.; ELIOTT, F.; HENRY, R.J.; CASU, R.E.; BONNETT, G.D. The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. **Molecular Breeding**, Berlin, v. 18, p. 39-50, 2006.

McINTYRE, C.L.; WHAN, V.A.; CROFT, B.; MAGAREY, R.; SMITH, G.R. Identification and validation of molecular markers associated with Pachymetra Root Rot and brown rust resitance in sugarcane using map- and association-based approachs. **Molecular Breeding**, Berlin, v. 16, p. 151-161, 2005b.

MING, R.; DEL MONTE, T.A.; HERNANDEZ, E.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. **Genome**, Ottawa, v. 45, p. 794-803, 2002a.

MING, R.; LIU, S.C.; BOWERS, J.E.; IRVINE, J.E.; PATERSON, A.H. Construction of a *Saccharum* consensus genetic map from two interspecific crosses. **Crop Science**, Madison, v. 42, p. 570-583, 2002b.

MING, R.; LIU, S.C.; LIN, Y.R.; da SILVA, J.; WILSON, W.; BRAGA, D.; van DEYNZE, A.; WENSLAFF, T.F.; WU, K.K.; MOORE, P.H.; BURNQUIST, W.; SORRELLS, M.E.; IRVINE, J.E.; PATERSON, A.H. Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. **Genetics**, Bethesda, v. 150, p. 1663-1682, 1998.

MING, R.; LIU, S.C.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. QTL Analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. **Genome Research**, London, v. 11, p. 2075-2084, 2001.

MING, R.; WANG, Y.W.; DRAYE, X.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. Molecular dissection of complex traits in autopolyploids: mapping QTL's affecting sugar yield and related traits in sugarcane. **Theoretical and Applied Genetics**, New York, v. 105, p. 332-345, 2002c.

MOHAN, M.; NAIR, S.; BHAGWAT, A.; KRISHNA, T.G.; YANO, M.; BHATIA, C.R.; SASAKI, T. Genome mapping, molecular markers and markers-assisted selection in crop plants. **Molecular Breeding**, Berlin, v. 3, p. 87-103, 1997.

MOLLINARI, M.; MARGARIDO, G.R.A.; VENCOVSKY, R.; GARCIA, A.A.F. Evaluation of algorithms used to order markers on genetic maps. **Heredity**, London, v. 103, p. 494-502, 2009.

MORGANTE, M.; SALAMINI, F. From plant genomics to breeding practice. **Current Opinion in Plant Biotechnology**, London, v. 14, p. 214-219, 2003.

MUDGE, J.; ANDERSEN, W.R.; KEHRER, R.; FAIRBANKS, D.J. A RAPD genetic map of *Saccharum officinarum*. **Crop Science**, Madison, v. 36, p. 1362-1366, 1996.

OLIVEIRA, K.M.; PINTO, L.R.; MARCONI, T.G.; MARGARIDO, G.R.A.; PASTINA, M.M.; TEIXEIRA, L.H.M.; FIGUEIRA, A.M.; ULIAN, E.C.; GARCIA, A.A.F.; SOUZA, A.P. Functional genetic linkage map on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, Berlin, v. 20, p. 189-208, 2007.

OMAN, S.D. Multiplicative Effects in Mixed Model Analysis of Variance. **Biometrika**, Cambridge, v. 78, n. 4, p. 729-739, 1991.

OOIJEN, J.W. van; VOORRIPS, R.E. **JoinMap 3.0, Software for the Calculation of Genetic Linkage Maps**. Plant Research International. Kyazma BV, Wageningen, The Netherlands, 2001.

PASTINA, M.M.; PINTO, L.R.; OLIVEIRA, K.M.; SOUZA, A.P.; GARCIA, A.A.F. Molecular mapping of complex traits. In: HENRY, R. (Ed.) **Genetics, Genomics and Breeding of Sugarcane**. New Hampshire: Science Publishers., ca. 2010. 1V.

PATERSON, A.; LANDER, E.; LINCOLN, S.; HEWITT, J.; PETERSON, S.; TANKSLEY, S. Resolution of quantitative traits into mendelian factors using a complete RFLP linkage map. **Nature**, London, v. 225, p. 721-726, 1988.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v. 58, p. 545-54, 1971.

PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. GenStat for Windows (12th Edition) Introduction. **VSN International**, Hemel Hempstead, 2009.

PIEPHO, H.P. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. **Genetics**, Bethesda, v. 156, p. 2043-2050, 2000.

PIEPHO, H.P. Analyzing Genotype-Environment Data by mixed Models with Multiplicative Terms. **Biometrics**, Washington, v. 53, n. 2, p. 761-766, 1997.

PIPEHO, H.P. Statistical tests for QTL and QTL-byenvironment effects in segregating populations derived from line crosses. **Theoretical and Applied Genetics**, New York, v. 110, p. 561-566, 2005.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-Plus**. New York: Springer-Verlag. 2000. 528 p.

PINTO, L.R.; GARCIA, A.A.F.; PASTINA, M.M.; TEIXEIRA, L.H.M.; BRESSIANI, J.A.; ULIAN, E.C.; BIDOIA, M.A.P.; SOUZA, A.P. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. **Euphytica**, Wageningen, 2009. Disponível em: <http://www.springerlink.com>.

PIPERIDIS, N.; JACKSON, P.A.; D'HONT, A.; BESSE, P.; HOARAU, J.Y.; COURTOIS, B.; AITKEN, K.S.; McINTYRE, C.L. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. **Molecular Breeding**, Berlin, v. 21, p. 233-247, 2008.

PORCEDDU, A.; ALBERTINI, E.; BARCACCIA, G.; FALISTOCCO, E.; FALCINELLI, M. Linkage mapping in apomictic and sexual Kentucky bluegrass (*Poa pratensis* L.) genotypes using a two way pseudo-testcross strategy based on AFLP and SAMPL markers. **Theoretical and Applied Genetics**, New York, v. 104, p. 273-280, 2002.

RABOIN, L.M.; OLIVEIRA, K.M.; LECUNFF, L.; TELISMART, H.; ROQUES, D.; BUTTERFIELD, M.; HOARAU, J.Y.; D'HONT, A. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. **Theoretical and Applied Genetics**, New York, v. 112, p. 1382-1391, 2006.

RABOIN, L.M.; PAUQUET, J.; BUTTERFIELD, M.; D'HONT, A.; GLASZMANN, J.C. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. **Theoretical and Applied Genetics**, New York, v. 116, p. 701-714, 2008.

RASMUSSON, J. Genetically changed linkage values in Pisum. **Hereditas**, Lund, v. 10, p. 1-152, 1927.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2009. Disponível em: <http://www.R-project.org>. Acesso em: 20 jan. 2010.

REFFAY, N.; JACKSON, P.A.; AITKEN, K.S.; HOARAU, J.Y.; D'HONT, A.; BESSE, P.; McINTYRE, C.L. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. **Molecular Breeding**, Berlin, v. 15, p. 367-381, 2005.

RESENDE, M.D.V; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Brazilian Journal of Mathematics and Statistics**, São Paulo, v. 22, n. 2, p. 31-52, 2004.

RIDOUT, M.S., TONG, S.; VOWDEN, C.J.; TOBUTT, R.K. Three-point linkage analysis in crosses of allogamous plant species. **Genetical Research**, London, v. 72, p. 111-121, 1998.

RIPOL, M.I.; CHURCHILL, G.A.; da SILVA, J.A.G; SORRELLS, M. Statistical aspects of genetic mapping in autopolyploids. **Gene**, Amsterdam, v. 235, p. 31-41, 1999.

RITTER, E.; GEBHARDT, C.; SALAMINI, F. Estimation of recombination frequencies and construction RFLP linkage maps in plants from crosses between heterozygous parents. **Genetics**, Bethesda, v. 125, p. 645-654, 1990.

RITTER, E.; SALAMINI, F. The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. **Genetical Research**, London, v.67, p.55-65, 1996.

SABADIN, P.K.; SOUZA JR., C.L.; SOUZA, A.P.; GARCIA, A.A.F. QTL mapping for yield components in a tropical maize population using microsatellite markers. **Hereditas**, Lund, v. 145, p. 194-203, 2008.

SAS Institute, SAS/STAT User's Guide. Cary, 1989. Disponível em: <http://www.d.umn.edu/math/docs/saspdf/stat/pdfidx.htm>. Acesso em: 20 jan. 2010.

SATAGOPAN, J.M.; YANDELL, B.S.; NEWTON, M.A.; OSBORN, T.C. A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. **Genetics**, Bethesda, v. 144, p. 805-816, 1996.

SAX, K. The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. **Genetics**, Bethesda, v. 8, p. 552-560, 1923.

SCHLÖTTERER, C. The evolution of molecular markers - just a matter of fashion?. **Nature Reviews Genetics**, New York, v. 5, p. 63-69, 2004.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Philadelphia, v. 6, p. 461-464, 1978.

SEARLE, S. R. **Linear models**. New York: J. Wiley & Sons, 1971. 532 p.

SEATON, G.; HALEY, C.S.; KNOTT, S.A.; KEARSEY, M.; VISSCHER, P.M. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. **Bioinformatics**, Oxford, v. 18, p. 339-340, 2002.

SEMAGN, K.; SKINNES, H.; BJØRNSTAD, A.; MARØOY, A.G.; TARKEGNE, Y. Quantitative trait loci controlling fusarium head blight resistance and low deoxynivalenol content in hexaploid wheat population from ‘Arina’ and NK93604. **Crop Science**, Madison, v. 47, p. 294-303, 2007.

SHEPHERD, M.; CROSS, M.; DIETERS, M.J.; HENRY, R. Genetic maps for *Pinus elliottii* var hondurensis using AFLP and microsatellite markers. **Theoretical and Applied Genetics**, New York, v. 106, p. 1409-1419, 2003.

SIBOV, S.T.; SOUZA JR, C.L.; GARCIA, A.A.F.; SILVA, A.R.; GARCIA, A.F.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (Zea mays L.) using microsatellite markers. 2. Quantitative trait loci (QTL) for grain yield, plant heigth, ear height and grain moisture. **Hereditas**, Lund, v. 139, p. 107-115, 2003.

SILLANPAA, M.J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. **Genetics**, Bethesda, v. 148, p. 1373-1388, 1998.

SILLS, G.R.; BRIDGES, W.; AL-JANABI, S.M.; SOBRAL, B.W.S. Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. ex Grassl). **Molecular Breeding**, Berlin, v. 1, p. 355-363, 1995.

SILVA, J.A.G. da; BRESSIANI, J.A. Sucrose synthasemolecular marker associated with sugar content in elite sugarcane progeny. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, n. 2, p. 294-298, 2005.

SILVA, J.A.G. da; HONEYCUTT, R.J.; BURNQUIST, W.; AL-JANABI, S.M.; SORRELLS, M.E.; TANKSLEY, S.D.; SOBRAL, W.S. *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP and PCR based markers. **Molecular Breeding**, Berlin, v. 1, p. 165-179, 1995.

SILVA, J.A.G. da; SORRELLS, M.E.; BURNQUIST, W.; TANKSLEY, S.D. RFLP linkage map of *Saccharum spontaneum*. **Genome**, Ottawa, v. 36, p. 782-791, 1993.

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. **Biometrics**, Washington, v. 57, n. 4, p. 1138-1147, 2001.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, Cambridge, v. 143, p. 449-462, 2005.

SMITH, A.B.; STRINGER, J.K.; WEI, X.; CULLIS, B.R. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, Wageningen, v. 157, p. 253-266, 2007.

SMITH, H.H. The relation between genes affecting size and color in certain species of Nicotiana. **Genetics**, Bethesda, v. 22, p. 361-375, 1937.

SOBRAL, B.W.S.; HONEYCUTT, R.J. High output genetic mapping in polyploids using PCR-generated markers. **Theoretical and Applied Genetics**, New York, v. 86, p. 105-112, 1993.

SOLLER, M.; BRODY, T.; GENIZI, A. On the power of experimental design for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. **Theoretical and Applied Genetics**, New York, v. 47, p. 35-39, 1976.

SONG, J.Z.; SOLLER, M.; GENIZI, A. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. **Genetical Research**, London, v. 73, p. 61-73, 1998.

STUBER, C.W.; EDWARDS, M.D.; WENDEL, J.F. Molecular-marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. **Crop Science**, Madison, v. 27, p. 639-648, 1987.

TAKEDA, S.; MATSUOKA, M. Genetic approaches to crop improvement: responding to environmental and population changes. **Nature Reviews Genetics**, New York, v. 9, p. 444-457, 2008.

UTZ, H.F.; MELCHINGER, A.E. PLABQTL: a Program for Composite Interval Mapping of QTL. Stuttgart, 2003. Disponível em: <http://www.uni-hohenheim.de/~zipspwww/soft.html>. Acesso em: 20 jan. 2010.

VARGAS, M.; van EEUWIJK, F.A.; CROSSA, J.; RIBAUT, J.M. Mapping QTLs and QTL × environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. **Theoretical and Applied Genetics**, New York, v. 112, p. 1009-1023, 2006.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Spinger-Verlag, 2000, 568 p.

VERBYLA, A.; ECKERMAN, P.J.; THOMPSON, R.; CULLIS, B. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. **Australian Journal of Agricultural Research**, Melbourne, v. 54, p. 1395-1408, 2003.

WANG, S.; BASTEN, C.J.; ZENG, Z.B. Windows QTL-Cartographer 2.5. Raleigh, 2007. Disponível em: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Acesso em: 20 jan. 2010.

WASSOM, J.J.; MIKKELINENI, V.; BOHN, M.O.; ROCHEFORD, T.R. QTL for Fatty Acid Composition of Maize Kernel Oil in Illinois High Oil × B73 Backcross-Derived Lines. **Crop Science**, Madison, v. 48, p, 69-78, 2008a

WASSOM, J.J.; JeFFREY, C.W.; MARTINEZ, E.; KING, J.J.; DeBAENE, J. QTL Associated with Maize Kernel Oil, Protein, and Starch Concentrations; Kernel Mass; and Grain Yield in Illinois High Oil × B73 Backcross-Derived Lines. **Crop Science**, Madison, v. 48, p. 243-252, 2008b.

WEI, X.; JACKSON, P.A.; McINTYRE, C.L. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. **Theoretical and Applied Genetics**, New York, v. 114, p. 155-164, 2006.

WELLER, J.I. Maximum likelihood techiques for the mapping and analyses of quantitative trait loci with the aid of genetic markers. **Biometrics**, Washington, v. 42, p. 627-640, 1986.

WEXELSEN, H. Quantitative inheritance and linkage in barley. **Hereditas**, Lund, v. 18, p. 307-348, 1933.

WOLFINGER, R.D. Covariance structure selection in general mixed linear models. **Communications in Statistics A, Theory and Methods**, Zug, v. 22, p. 1079-1106, 1993.

WU, R.; MA, C.X.; PAINTER, I.; ZENG, Z.B. Simultaneous maximum likelihood estimation of linkage and linkage-phases in outcrossing species. **Theoretical Population Biology**, New York, v. 61, p. 349-363, 2002a.

WU, R.; MA, C.X.; WU, S.S.; ZENG, Z.B. Linkage mapping of sex-specific differences. **Genetical Research**, London, v. 79, p. 85-96, 2002b.

WU, K.K.; BURNQUIST, W.; SORRELLS, M.E.; TEW, T.L.; MOORE, P.H.; TANKSLEY, S.D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, New York, v. 83, p. 294-300, 1992.

YI, N.; XU, S. Bayesian mapping of quantitative trait loci under complicated mating designs. **Genetics**, Bethesda, v. 157, p. 1759-1771, 2001.

YI, N.; YANDELL, B.S.; CHURCHILL, G.A.; ALLISON, D.B.; EISEN, E.J.; POMP, D. Bayesian model selection for genome-wide epistatic QTL analysis. **Genetics**, Bethesda, v. 170, p. 1333-1344, 2005.

YI, N.; BANERJEE, S.; POMP, D.; YANDELL, B.S. Bayesian mapping of genome-wide interacting quantitative trait loci for ordinal traits. **Genetics**, Bethesda, v. 176, p. 1855-1864, 2007a.

YI, N.; SHRINER, D.; BANERJEE, S.; MEHTA, T.; POMP, D.; YANDELL, B.S. An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. **Genetics**, Bethesda, v. 176, p. 1865-1877, 2007b.

ZENG, Z.B. Theoretical basis of precision mapping of quantitative trait loci. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 90, p. 10972-10976, 1993.

ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, Bethesda, v. 136, p. 1457-1468, 1994.

ZENG, Z.B., KAO, C.H.; BASTEN, C.J. Estimating the genetic architecture of quantitative traits. **Genetical Research**, London, v. 74, p. 279-289, 1999.

ZHAO, M.; ZHANG, Z.; ZHANG, S.; LI, W.; JEFFERS, D.P.; RONG, T.; PAN, G. Quantitative trait loci for resistance to banded leaf and sheath blight in maize. **Crop Science**, Madison, v. 46, p. 1039-1045, 2006.