

UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE BOTUCATU

**Estudo genético da característica fibra em Cana-de-
Açúcar**

Karine Kettener

Tese apresentada para obtenção do título de
Doutor pelo programa de Pós-Graduação em
Ciências Biológicas-Genética.

Botucatu

2016

Karine Kettener

Licenciada e Bacharel em Ciências Biológicas

**Estudo genético da característica fibra em Cana-de-
Açúcar**

Orientador: Prof. Dr. Celso Luis Marino

Co-orientador: Prof. Dr. Jorge A. Da Silva

Tese apresentada para obtenção do título de
Doutor pelo programa de Pós-Graduação em
Ciências Biológicas-Genética.

Botucatu

2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Kettener, Karine.

Estudo genético da característica fibra em
cana-de-açúcar / Karine Kettener. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: Celso Luis Marino

Coorientador: Jorge A. da Silva

Capes: 20203004

1. Cana-de-açúcar. 2. Plantas - Melhoramento genético.
3. Mapeamento cromossômico. 4. Biomassa vegetal.

Palavras-chave: Biomassa; Cana-de-açúcar; Mapeamento
genético; Parâmetros genéticos.

À minha mãe, **Maria José**, que me ensinou a simplicidade, o amor e a alegria em tudo o que faz. Ao meu pai, **Felício**, por me ensinar o valor do trabalho honesto. Ao meu marido, **Tiago**, por acreditar em meus sonhos e por participar deles de corpo e alma.
Dedico.

“No matter how you feel, get up,
dress up, show up and never give
up”.

“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

Agradecimentos

Ao meu orientador, amigo e padrinho Celso Marino, por todos os ensinamentos, longas conversas e momentos de muito aprendizado. Foram 8 anos trabalhando juntos e que me mudaram para uma pessoa melhor, obrigada por tudo!

Ao meu co-orientador Jorge A. Da Silva, que, juntamente com sua esposa Denise, sempre estiveram dispostos a ajudar mesmo nos momentos mais difíceis dessa tese. Eu e o Tiago seremos eternamente gratos e saibam que vocês foram como nossos “pais americanos”.

Ao Programa de Pós-graduação em Ciências Biológicas (Genética), funcionários da pós graduação e do departamento de genética, que sempre me auxiliaram em minha caminhada com muita alegria e disposição em ajudar. Aos amigos do Programa que conheci em disciplinas, organizações de eventos e afins, obrigada pelos ensinamentos e momentos prazerosos que passamos juntos. Agradeço especialmente minhas amigas Juliana Giusti, Juliana Marino, Ana Carolina Tomaz Tavares e Maria Cecília Perantoni Fuchs, as quais em um simples curso de férias, se tornaram pessoas para a vida toda e que mesmo longe, sempre mostraram-se presentes, Vocês são demais!

A CAPES e ao programa Ciência sem Fronteiras pela bolsa concedida.

À empresa Centro de Tecnologia Canavieira-CTC, pela proposta de parceria, em especial a Ex Gerente de Biotecnologia Sabrina Moutinho Chabregas pela oportunidade, amizade e contribuição. A pesquisadora Karine Miranda Oliveira, Thiago Benatti, Itaraju Brum, Francisco Claudio Lopes pelas discussões

científicas e apoio. As técnicas Márcia Patrícia Moreno, Natália Spagnoll, Daiane Silva, Ana Paula Almeida, Daniele Truffi pela amizade, incentivo e auxílio nas etapas iniciais de minha tese e pela fiel amizade.

Aos amigos da Texas A&M AgriLife Research, em especial a Nora Solis, Victoria Mora, Jong-won Park, Everton Barreto por todo auxílio no projeto e no dia a dia texano, meus sinceros agradecimentos. Vocês são muito especiais e fizeram toda a diferença no período que morei no Texas.

Aos amigos que fiz na Texas A&M Agrilife Genomics and Bioinformatics Service em College Station, Charlie Jonshon, Joshua, Sabrina e em especial ao Rick, que me acompanhou e me ensinou com toda a paciência a arte da genotipagem.

Aos amigos do Centro de Análises Genômicas-CAGEN da Unesp de Botucatu, Lídia Carolina Arneiro, Paola Mazza Revolti, Maria Cecília Perantoni Fuchs, Leonardo Curi, Hernando, Bruno, Izabel Cristina e em especial, Vanusa Socorro Leite, Júlio César Otto, Tânia Mara Bortoloto que me ensinaram a arte da ciência quando eu ainda era uma aluna de graduação. Agradeço especialmente a Lídia Carolina Arneiro que sempre torceu por mim e me incentivou até o fim. Você foi como a irmã que não tive, obrigada por tudo!

A ESALQ-USP na pessoa do professor Antonio Augusto Franco Garcia, pela oportunidade de cursar disciplinas e de aprender a arte do mapeamento genético. A aluna de doutorado Marianella Quezada por ter me auxiliado nas análises de mapeamento e por ter se tornado uma amiga para a vida toda.

A minha querida amia Carolina Grandó, por sempre me acolher com muita alegria todas as vezes que precisei ficar em Piracicaba tanto por conta da tese,

como por outros motivos, meus sinceros agradecimentos.

Aos queridos amigos do grupo Nárnia pelo apoio e incentivo aos estudos e pela compreensão devida às ausências. Obrigada por tornarem esse período muito mais leve e divertido! Desculpem-me por não me arriscar a citar nomes, mas felizmente somos muitos e estamos sempre aumentando.

Enfim, a todos os amigos e familiares que, direta ou indiretamente, contribuíram para minha formação não só como cientista, mas principalmente como cidadão consciente, meus sinceros agradecimentos.

Sumário

Prefácio	12
Capítulo I	14
1 Introdução	14
2 Material e métodos	18
2.1 População de estudo e delineamento experimental	18
2.2 Caracteres avaliados	19
2.2.1 Análises tecnológicas	19
2.2.2 Componentes lignocelulósicos	20
2.3 Análises de variância	20
2.4 Herdabilidade	23
2.5 Análises de associação dos marcadores microsatélites	24
2.6 Análise por componentes principais	25
3 Resultados e discussão	26
3.1 Análise dos componentes lignocelulósicos	26
3.2 Análise por componentes principais	27
3.3 Análises de variância	28
3.4 Herdabilidade	39
3.5 Análises de associação dos marcadores microsatélites	40
4 Conclusão	44
Referências	45
Capítulo II	47
Abstract	48
Introduction	49
Materials and methods	51
Plant material and nucleic acid extraction	51
RAD-Seq library construction and sequencing	51
SNP discovery and genotyping	52
SNPs filtering and classifying by segregation pattern	53
Construction of genetic maps	54
Results	55
Marker development	55
SNP calling	56
Segregation analysis	58
Sugarcane genetic map	58
Discussion	63
Conclusion	65
References	66

Lista de tabelas

Capítulo I

Tabela 1: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico PBU (peso do bagaço úmido) das três testemunhas dos 8 grupos que compõem o experimento. **29**

Tabela 2: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico PBS (peso do bagaço seco) das três testemunhas dos 8 grupos que compõem o experimento. **30**

Tabela 3: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Fibra industrial das três testemunhas dos 8 grupos que compõem o experimento. **31**

Tabela 4: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Celulose das três testemunhas dos 8 grupos que compõem o experimento. **32**

Tabela 5: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Lignina das três testemunhas dos 8 grupos que compõem o experimento. **33**

Tabela 6: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Lignina nos clones da população. **35**

Tabela 7: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Celulose nos clones da população. **36**

Tabela 8: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Fibra nos clones da população. **37**

Tabela 9: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico PBS (peso seco do bagaço) nos clones da população. ... **38**

Tabela 10: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico PBU (peso úmido do bagaço) nos clones da população. . **39**

Capítulo II

Table 1: Alignment by Bowtie2 of 19,013,358 tags of RadSeq in a population of sugarcane mapping. The total aligned tags are the sum of the unique and non-unique alignments. **56**

Table 2: Number of SNP markers generated after analyzing the RadSeq data from a sugarcane population by Tassel-GBS pipeline. **57**

Table 3: Total markers in each reference genome to be used for genetic mapping..... **59**

Lista de figuras

Capítulo I

- Figura 1:** Delineamento experimental em blocos incompletos balanceados da população segregante para produção de fibra cedida para estudo pelo programa de melhoramento da empresa Centro de Tecnologia Canavieira-CTC..... **19**
- Figura 2:** Porcentagem de lignina na cana planta e cana soca na população após a análise de variância com efeitos de estrutura fixa. **26**
- Figura 3:** Porcentagem de lignina e celulose na cana planta e cana soca na população após a análise de variância com efeitos de estrutura fixa..... **27**
- Figura 4:** Análise de componentes principais com os genitores e a progênie da população segregante para fibra através de marcadores do tipo microssatélites. Os percentuais nos eixos referem-se à quantidade de variação nos dois primeiros componentes. **28**
- Figura 5:** Estimativas do coeficiente de herdabilidade para os caracteres PBU, PBS, Fibra, Lignina e Celulose..... **40**
- Figura 6:** *Single marker analysis* para detectar os marcadores de microssatélites relacionados com as características PBU, PBS, Fibra, Lignina e Celulose. **41**
- Figura 7:** Regressão linear múltipla com os marcadores SSRs nas características PBU, PBS, Fibra, Lignina e Celulose. **43**

Capítulo II

- Figure 1:** Circular graph showing redundancy between single-dose markers obtained from four genome references used to map Rad tag in a sugarcane mapping population. Red regions represent a redundancy within each reference, while regions in green, orange and blue respectively represent a redundancy of four, three and two references..... **58**
- Figure 2:** Genetic map of sugarcane displaying the distribution of linkage groups (LGs). Genetic distances between adjacent markers are shown on the left of each co-segregation group (CG). In red we have the representation of the type B3.7 markers, in green the type D1.12 and D2.15 in blue. **60**

PREFÁCIO

Os resultados obtidos durante o desenvolvimento deste trabalho estão apresentados na forma de capítulos. O primeiro capítulo apresenta uma análise dos parâmetros genéticos de uma população F1 obtida a partir do cruzamento de duas variedades comerciais de cana-de-açúcar oriundas do programa de melhoramento do Centro de Tecnologia Canavieira, Piracicaba/SP. Também avaliamos marcadores moleculares microsatélites buscando uma relação entre os mesmos com as características agrônômicas aqui estudadas (Peso do bolo úmido, Peso do bolo seco, Fibra, Lignina e Celulose). No segundo capítulo, o qual está em formato de artigo científico que será submetido para a revista *Genome*, com o título “*A SNP genetic map constructed using restriction site-associated DNA sequencing approach for sugarcane*”, apresenta o primeiro mapa genético com marcadores SNPs (*Single nucleotide polymorphisms*) obtidos via RADSeq utilizando a população F1 de mapeamento obtida a partir do cruzamento de uma variedade comercial dos Estados Unidos com *S. spontaneum*, oriundos do programa de melhoramento da *Texas A&M AgriLife Research*, Weslaco, TX, USA.

Durante o doutorado realizei estágio na *Texas A&M AgriLife Research*, Weslaco, TX, USA, sob orientação do Dr. Jorge A. G. da Silva, onde foram realizadas as análises dos componentes lignocelulósicos e o screening de marcadores SSR. Também houve o estágio na *Texas A&M AgriLife Genomics and Bioinformatics Service* (Texas A&M University, College Station, TX, EUA), sob orientação do Dr. Charles Johnson, onde a genotipagem para a descoberta de SNPs foi realizada através do protocolo RadSeq.

As análises de bioinformática para o mapeamento genético foi realizado em parceria com o Prof. Dr. Antonio Augusto Franco Garcia da ESALQ-USP, Piracicaba.

Capítulo I

Análise de parâmetros genéticos para biomassa em Cana-de-açúcar

1 INTRODUÇÃO

A Cana de açúcar (*Saccharum spp.*) pertence a família Poaceae com a propriedade de acumular até 50-60% de seu peso seco em sacarose (Bonnett G.D. et al., 2004), o que a torna economicamente importante para a produção de açúcar e bioetanol. Em moinhos de cana, o sumo é extraído a partir de caules e utilizado tanto para a cristalização do açúcar e fermentação como para a produção do chamado etanol de primeira geração (Arruda P., 2012). A Cana está entre as plantas produtoras de biomassa mais eficientes (Vermerris W., 2011), de modo que o setor pode igualmente beneficiar da produção de etanol a partir da matéria-prima lignocelulósica representada pelo bagaço, que é atualmente utilizado para a geração de calor e eletricidade para a usina (Maciel M.R.W. et al., 2009). A Lignocelulose é a biomassa vegetal composta por polímeros de celulose, hemicelulose e lignina, sendo um recurso renovável utilizado para a produção de biocombustíveis (Stewart C.N.Jr. et al., 2014). No entanto, possui uma natureza altamente complexa, na qual as microfibrilas de celulose são incorporados numa matriz de polissacarídeos reticulado, constituindo uma barreira física que limita a liberação dos açúcares fermentáveis, o que demanda tratamentos químicos relativamente caros para a remoção da lignina, permitindo o acesso aos polissacarídeos para a sacarificação (Vega-Sánchez M.E., Ronald P.C., 2010; Zhao X., Zhang L., Liu D., 2012). A engenharia genética de plantas para bioenergia surgiu como uma alternativa para otimizar este processo, tendo

como objetivo produzir plantas que acumulem menos lignina no intuito de serem mais passíveis de degradação química.

Saccharum X officinarum designa cientificamente a cana-de-açúcar autoploiploide cultivada (Cheavegatti-Gianotto et al., 2011), com tamanho de genoma monoplóide estimado em ~1 Gb (Souza et al., 2011), o que tem sido apontado como desafiador, pois dificulta os estudos genéticos e a aplicação de metodologias mais modernas no melhoramento da espécie.

A estimação de parâmetros genéticos tem fundamental importância para os programas de melhoramento, pois permitem identificar a natureza da ação dos genes envolvidos no controle dos caracteres quantitativos e, assim, avaliar a eficiência das diferentes estratégias de melhoramento, pela obtenção de ganhos genéticos preditos e manutenção de uma base genética adequada. As variâncias genéticas aditivas e não-aditivas, as correlações e as herdabilidades são os parâmetros genéticos de maior importância (Cruz e Carneiro, 2003).

O termo parâmetro é utilizado para designar as constantes características de uma população, particularmente média e variância. No caso de populações utilizadas em programas de melhoramento, os parâmetros de interesse são de duas naturezas: genética e não-genética. A estimação dos parâmetros genéticos é necessária para: (a) obter informações sobre a natureza da ação dos genes envolvidos na herança dos caracteres sob investigação; e (b) estabelecer a base para a escolha dos métodos de melhoramento aplicáveis à população.

A herdabilidade (h^2) é um dos parâmetros genéticos mais informativos para o trabalho do melhorista. Este parâmetro fornece a proporção da variância

genética presente na variância fenotípica total. Dessa forma, ela infere sobre a confiabilidade do valor fenotípico como indicador do valor genético para espécies que se reproduzem sexuadamente e são propagadas pelo cultivo de sementes.

A herdabilidade pode ser estimada em sentido amplo (h^2_g) e sentido restrito (h^2_a). No sentido amplo, considera toda a variação genética aditiva e não-aditiva (Ramalho et al., 1993) e, no sentido restrito, é determinada pela relação entre variância genética aditiva e a variância fenotípica (Resende, 2001).

O coeficiente de herdabilidade, tanto no sentido restrito como no amplo, pode variar de zero a um. No caso de $h^2 = 1$, as diferenças fenotípicas entre os indivíduos são causadas unicamente por diferenças genéticas entre os mesmos; quando $h^2 = 0$ significa que a variabilidade do caráter não tem origem genética. Neste caso, não existe correlação alguma entre valor genético e valor fenotípico da unidade de seleção (Allard, 1964).

A variância aditiva é a variância dos valores genéticos aditivos, um dos fatores determinantes da covariância ou semelhança entre parentes e, por conseguinte, o principal determinante das propriedades genéticas da população e de sua resposta (Falconer, 1981).

Segundo Cruz e Regazzi (2001), a existência da variância aditiva indica maior facilidade na identificação de genótipos superiores com maior concentração de alelos favoráveis. Por sua vez, a variância atribuída à dominância é indicativa de dificuldades no processo de seleção, mas é importante quando se deseja explorar o vigor híbrido.

No intuito de se obter informações sobre os componentes genéticos relativos à biomassa em Cana de açúcar, este trabalho objetiva estudar os parâmetros genéticos de cana-planta e cana-soca para verificar se há diferença entre cortes, gerando informações para os programas de melhoramento para cana energia.

2 MATERIAL E MÉTODOS

2.1 População de estudo e delineamento experimental

A população de estudo consiste em um cruzamento entre o clone CT fibra alta e o clone CT fibra baixa, cedido pelo Centro de Tecnologia Canavieira –CTC, Piracicaba, SP. O ensaio é formado por 250 clones F1 e os genitores. Os clones foram plantados em Piracicaba-SP em uma área aproximada de 3,0 ha em delineamento de 2x5 (dois sulcos de cinco metros). O experimento foi dividido em dois conjuntos, cada conjunto formado por 4 grupos e 16 blocos. Os indivíduos se repetem em um mesmo grupo e os controles (C1, C2 e C3) se repetem em todos os blocos, como demonstrado na figura 1.

O período da condução do experimento foi de dois anos, que correspondem aos estágios de cana-planta (colheita em julho de 2012) e cana-soca (colheita em julho de 2013).

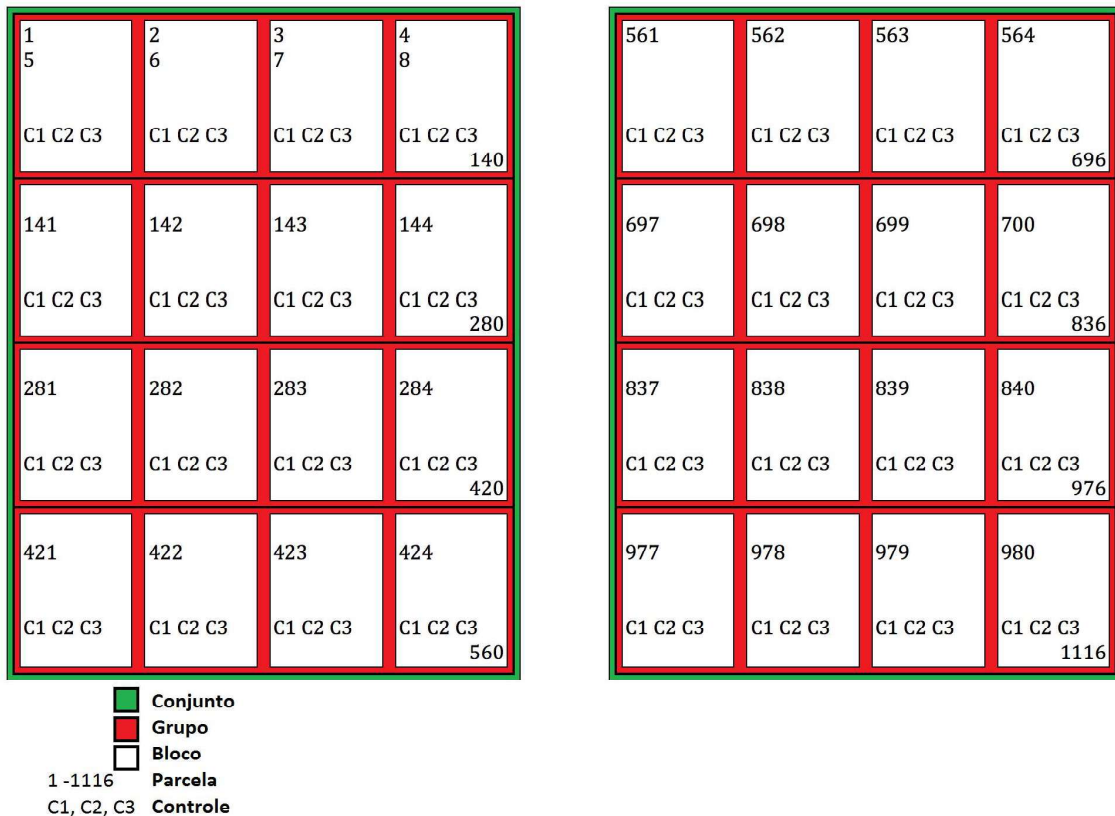


Figura 1: Delineamento experimental em blocos incompletos balanceados da população segregante para produção de fibra cecida para estudo pelo programa de melhoramento da empresa Centro de Tecnologia Canaveira-CTC.

2.2 Caracteres Avaliados

2.2.1 Análises tecnológicas

Os seguintes caracteres foram avaliados:

- a) Peso do bolo úmido (PBU): Obtido através da prensagem da parte fibrosa e subsequente peso do bolo (bagaço) úmido.
- b) Peso do bolo seco (PBS): Obtido a partir da secagem em estufa do bolo úmido e subsequente pesagem.

c) Fibra: A Fibra da cana foi estimada através do método de Tanimoto (1964), o qual utiliza os valores de PBU e PBS para a obtenção da porcentagem de fibra industrial.

2.2.2 Componentes lignocelulósicos

Cinquenta clones foram submetidos a análises de *wet chemistry* para a construção da curva padrão que foi utilizada para a medição dos demais clones via *Near Infrared Spectroscopy (NIR)*. O protocolo utilizado foi o proposto por Rodríguez-Zúñiga et al. (2014) com algumas alterações. Os componentes medidos foram lignina e celulose para cana planta e cana soca.

2.3 Análises de variância

As análises de Variância e coeficiente de variação ambiental para os cinco caracteres (PBU, PBS, Fibra, Lignina e Celulose) foi realizada com base em 8 experimentos (que está sendo chamado, nesse caso, de "Grupo"), dessa forma, a indexação original de dois conjuntos de 4 experimentos (1,2,3,4), foi modificada para apenas um (1,2,3,4,5,6,7,8). As análises foram realizadas corte a corte e também em ambos os cortes conjuntamente, como demonstrado nas tabelas seguintes:

Análise Individual – Apenas um corte

Considerando:

- $E = \text{N}^\circ \text{ de Experimentos} = 8$
- $B / E = \text{Blocos dentro de experimentos} = 4$
- $T = \text{Testemunhas} = 3$

FV	GL	QM	F
Experimentos	$E-1 = 7$	QM E	QM E / QM B
Blocos / Exp	$E(B-1) = 24$	QM B	QM B / QM R
Tratam	$T-1 = 2$	QM T	QM T / QM E
Trat x Exp	$(T-1)(E-1) = 14$	QM TxE	QM TxE / QM R
Resíduo	$E(B-1)(T-1) = 48$	QM R	
Total	$TBE - 1 = 95$		

Análise Conjunta – Dois cortes

Considerando:

- E = N° de Experimentos = 8
- B / E = Blocos dentro de experimentos = 4
- T = Testemunhas = 3
- C = Cortes = 2

FV	GL	QM	F
Exp	E-1	7	QM 1 QM1/QM2
Blocos Exp	/ E(B-1)	24	QM 2 (QM2+QM11)/(QM10+QM5)
Trat	T-1	2	QM 3 QM3/QM5
Trat x Exp	(T-1) (E-1)	14	QM 4 QM4/QM5
Erro A (T x B/E)	(T-1) [(B-1)*E]	48	QM 5 QM5/QM11
Cortes	(C-1)	1	QM 6 QM6/QM10
Cortes Trat	x (C-1) (T-1)	2	QM 7 QM7/QM11
Cortes Exp	x (C-1) (E-1)	7	QM 8 QM7/QM10
Cortes Trat x Exp	x (C-1) (T-1) (E-1)	14	QM 9 QM9/QM11
Erro B (C x B/E)	(C-1) [(B-1)*E]	24	QM 10 QM10/QM11
Erro C (C x T x B/E)	(C-1)(T-1)	48	QM 11

	1)(B- 1)*E]	
Total	ETCB- 1	191

FV	E(QM)
Exp	$\sigma_{TCB}^2 + i\sigma_{CB}^2 + l\sigma_{TB}^2 + il\sigma_B^2 + ilj\varphi_E$
Blocos / Exp	$\sigma_{TCB}^2 + i\sigma_{CB}^2 + l\sigma_{TB}^2 + il\sigma_B^2$
Trat	$\sigma_{TCB}^2 + l\sigma_{TB}^2 + klj\varphi_T$
Trat x Exp	$\sigma_{TCB}^2 + l\sigma_{TB}^2 + lj\varphi_{TE}$
Erro A (T x B/E)	$\sigma_{TCB}^2 + l\sigma_{TB}^2$
Cortes	$\sigma_{TCB}^2 + i\sigma_{CB}^2 + kij\varphi_C$
Cortes x Trat	$\sigma_{TCB}^2 + kj\varphi_{TC}$
Cortes x Exp	$\sigma_{TCB}^2 + i\sigma_{CB}^2 + ij\varphi_{EC}$
Cortes x Trat x Exp	$\sigma_{TCB}^2 + j\varphi_{ETC}$
Erro B (C x B/E)	$\sigma_{TCB}^2 + i\sigma_{CB}^2$
Erro C (C x T x B/E)	σ_{TCB}^2
Total	

2.4 Herdabilidade

Trabalhou-se com um modelo multivariado com a matriz de covariância não estruturada, assumindo cada colheita como um caráter diferente, mas correlacionados (Resende, 2007). A implementação da modelagem foi realizada com o pacote LME4 em linguagem R.

Esse modelo é o mais completo, pois utiliza toda a informação simultaneamente e trata idades ou safras como sendo caracteres diferentes e correlacionados, considerando suas diferentes herdabilidades e correlações genéticas.

2.5 Análises de associação dos marcadores microssatélites

Os tipos de segregação do marcador foram identificados através de em um teste de qui-quadrado para desvios das proporções de segregação esperadas de 1:1 e 3:1 (marcadores em dose única em apenas um dos pais e marcadores em dose única em ambos os pais, respectivamente) com a correção de Bonferoni para controlar o erro tipo I para múltiplos testes.

A *Single Marker Analysis (SMA)* calcula se os valores fenotípicos diferem entre os genótipos para um determinado marcador molecular, ou seja, pode detectar associações entre marcadores moleculares e características agrônômicas de interesse para os melhoristas. A SMA pode ser realizada utilizando uma variedade de análises estatísticas, incluindo testes t, ANOVA, regressão, *maximum likelihood estimations* e *log likelihood ratios*. No caso, a ANOVA foi realizada baseada na equação linear que segue:

$$Y = \mu + f(\text{marcador}) + \text{erro}$$

Onde:

Y: valor da característica agronômica

μ : média populacional

F (marcador): função do marcador molecular

Os critérios adotados para declarar associações do marcador com a característica agronômica foram p-valores menores que 0,05 e 0,01.

Usou-se a regressão linear múltipla para determinar a proporção de variação fenotípica explicada pela associação entre marcadores e características agronômicas.

2.6 Análise por Componentes Principais

A análise de componentes principais (ACP) é um método estatístico essencialmente descritivo. A técnica foi inicialmente proposta por K. Pearson em 1901 (Morrison, 1976), e o objetivo é apresentar em uma forma gráfica o máximo de informação contida em uma matriz de dados, com o objetivo de visualizar as proximidades entre os indivíduos e os vínculos entre as variáveis. Com os marcadores microssatélites, foi utilizado o comando *prcomp* para o cálculo da PCA e para a plotagem dos resultados, foi usado o pacote *ggplot2*, todos na linguagem R.

3 RESULTADOS E DISCUSSÃO

3.1 Análise dos componentes lignocelulósicos

A composição química do bagaço de cana pré-tratado (celulose e lignina) foi analisada utilizando a técnica de espectroscopia infravermelha (NIR). O spectra gerado foi correlacionado com resultados das análises químicas clássicas usando uma regressão dos mínimos quadrados parciais. Observou-se que a porcentagem de lignina aumentou consideravelmente na cana soca e a porcentagem de celulose teve um aumento menos significativo da cana planta para a cana soca (Figura 2 e 3).

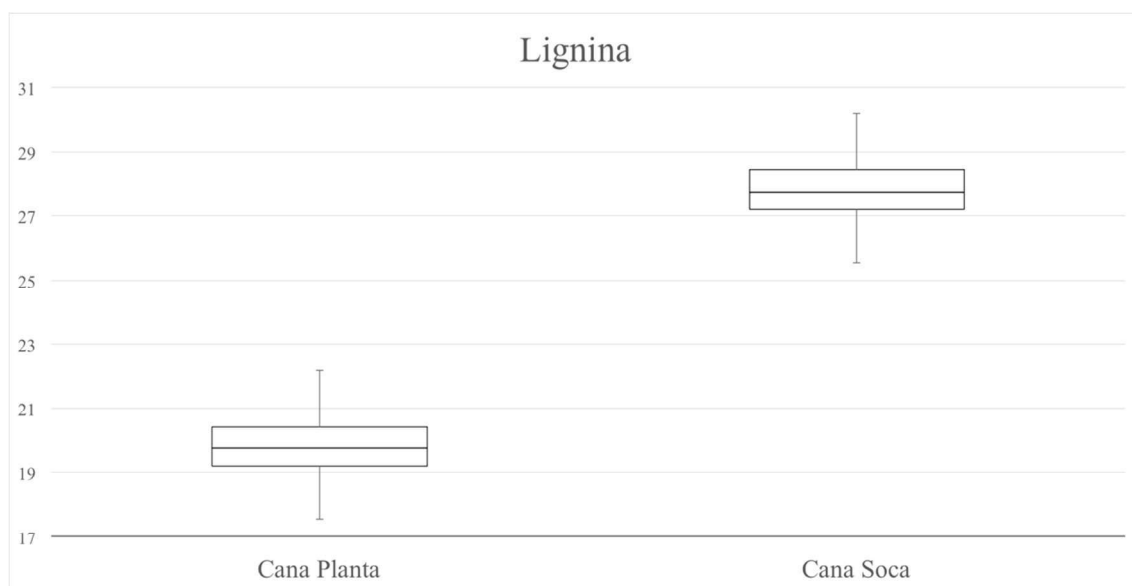


Figura 2: Porcentagem de lignina na cana planta e cana soca na população após a análise de variância com efeitos de estrutura fixa.

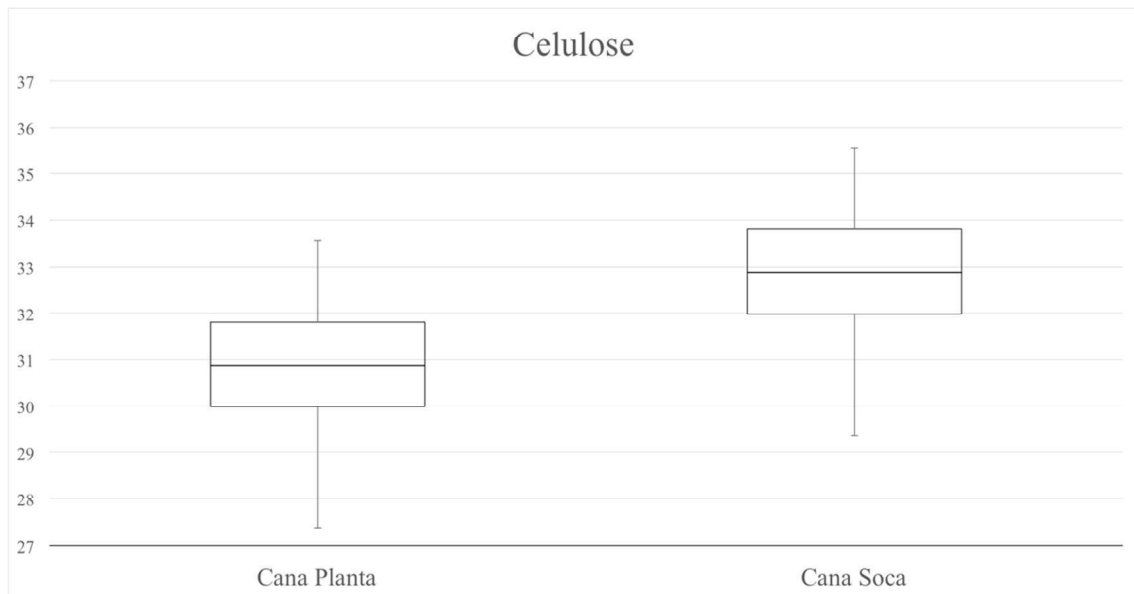


Figura 3: Porcentagem de lignina e celulose na cana planta e cana soca na população após a análise de variância com efeitos de estrutura fixa.

3.2 Análise por Componentes Principais

Este tipo de análise apenas mostra como a progênie se distribui em relação aos genitores, eventualmente refletindo aqueles indivíduos mais próximos de um ou outro genitor em termos genéticos. Observa-se que a população de estudo tem ampla distribuição, com diferença significativa da progênie e dos genitores, o que era esperado tendo em vista que o cruzamento gerador da população foi de uma planta considerada com alto teor de fibra com uma outra planta com baixo teor de fibra (Figura 4).

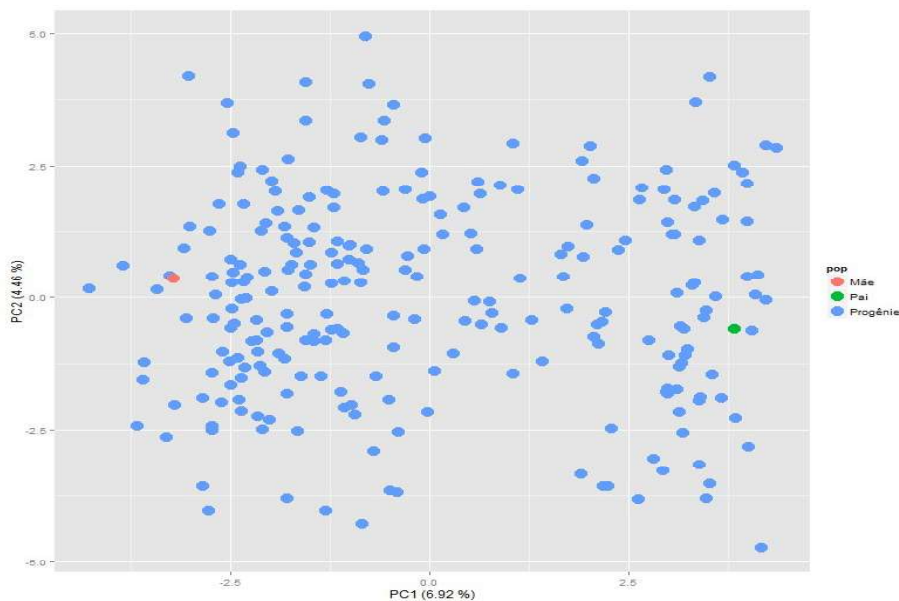


Figura 4: Análise de componentes principais com os genitores e a progênie da população segregante para fibra através de marcadores do tipo microssatélites. Os percentuais nos eixos referem-se à quantidade de variação nos dois primeiros componentes.

3.3 Análises de variância

Como a ligação entre os grupos do experimento em campo são as testemunhas, uma análise de variância entre elas foi realizada afim de verificar se há efeito do ambiente sobre os genótipos. Constatou-se efeito significativo do ambiente sobre os genótipos, o que torna o experimento não homogêneo (Tabela 1-5). A fonte de variação “cortes” também mostrou-se altamente significativa, mostrando que há diferenças entre os cortes.

Tabela 1: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônomo PBU (peso do bagaço úmido) das três testemunhas dos 8 grupos que compõem o experimento.

PBU CV% 3.754064

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	532.8	76.11	3.21682164	0.014993864
Variedade	2	164.82	82.41	3.201631702	0.049521165
Corte	1	660.08	660.08	27.40058115	2.30E-05
Grupo:Bloco	24	567.93	23.66	0.979931768	
Grupo:Variedade	14	405.65	28.98	1.125874126	0.361191392
Variedade:Corte	2	70.05	35.03	1.3914	0.25858
Grupo:Corte	7	1363.03	194.72	8.083022001	4.60E-05
Grupo:Bloco:Variedade	48	1235.41	25.74	1.0224	0.46955
Grupo:Bloco:Corte	24	578.11	24.09	0.9569	0.53346
Grupo:Variedade:Corte	14	271.08	19.36	0.7692	0.6955
Resíduo	48	1208.32	25.17		

Tabela 2: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônomo PBS (peso do bagaço seco) das três testemunhas dos 8 grupos que compõem o experimento.

PBS CV% 4.330169

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	282.94	40.42	8.962306	2.02961E-05
Variedade	2	64.63	32.31	3.3070624	0.045130506
Corte	1	1156.89	1156.89	113.53189	1.40E-10
Grupo:Bloco	24	108.24	4.51	0.7154309	
Grupo:Variedade	14	163.35	11.67	1.1944729	0.310205332
Variedade:Corte	2	73.83	36.92	3.7777	0.029947
Grupo:Corte	7	269.31	38.47	3.7752699	0.006734398
Grupo:Bloco:Variedade	48	468.95	9.77	0.9997	0.500362
Grupo:Bloco:Corte	24	244.66	10.19	1.0431	0.437101
Grupo:Variedade:Corte	14	119.34	8.52	0.8723	0.591799
Resíduo	48	469.08	9.77		

Tabela 3: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônomo Fibra industrial das três testemunhas dos 8 grupos que compõem o experimento.

Fibra CV% 4.707286

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	11.039	1.577	11.38628	2.72487E-06
Variedade	2	8.3489	4.1744	13.70903	1.95E-05
Corte	1	27.1502	27.1502	77.973	5.26E-09
Grupo:Bloco	24	3.3242	0.1385	0.652827	
Grupo:Variedade	14	3.7395	0.2671	0.877176	0.586893227
Variedade:Corte	2	4.1457	2.0729	7.2074	0.0018319
Grupo:Corte	7	5.1656	0.7379	2.119184	0.080544654
Grupo:Bloco:Variedade	48	14.6183	0.3045	1.0589	0.4218116
Grupo:Bloco:Corte	24	8.3575	0.3482	1.2108	0.2803198
Grupo:Variedade:Corte	14	3.6759	0.2626	0.9129	0.5514329
Resíduo	48	13.805	0.2876		

Tabela 4: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Celulose das três testemunhas dos 8 grupos que compõem o experimento.

Celulose CV% 12.32367

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	190.15	27.16	3.001105	0.0206448
Variedade	2	42.9	21.45	1.82398	0.172390657
Corte	1	954.47	954.47	112.5554	1.53E-10
Grupo:Bloco	24	217.26	9.05	1.060771	
Grupo:Variedade	14	350.78	25.06	2.130952	0.026549545
Variedade:Corte	2	16.28	8.14	0.6552	0.52392
Grupo:Corte	7	220.8	31.54	3.71934	0.007283279
Grupo:Bloco:Variedade	48	564.4	11.76	0.9466	0.575
Grupo:Bloco:Corte	24	203.57	8.48	0.6828	0.84301
Grupo:Variedade:Corte	14	369.78	26.41	2.1264	0.02689
Resíduo	48	596.24	12.42		

Tabela 5: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Lignina das três testemunhas dos 8 grupos que compõem o experimento.

Lignina CV% 16.3675

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	175.31	25.04	3.009615	0.020383721
Variedade	2	51.65	25.82	2.364469	0.1048608
Corte	1	342.53	342.53	40.20305	1.48E-06
Grupo:Bloco	24	199.62	8.32	0.966564	
Grupo:Variedade	14	248.15	17.72	1.622711	0.107073028
Variedade:Corte	2	2.14	1.07	0.1021	0.90312
Grupo:Corte	7	157.98	22.57	2.649061	0.035213746
Grupo:Bloco:Variedade	48	524.2	10.92	1.0435	0.44166
Grupo:Bloco:Corte	24	204.53	8.52	0.8143	0.70207
Grupo:Variedade:Corte	14	234.84	16.77	1.6028	0.11284
Resíduo	48	502.34	10.47		

A análise de variância conjunta foi realizada para todos os clones. A fonte de variação “cortes” também se mostrou altamente significativa, mostrando que há diferenças entre os cortes. As fontes de variação (FV) “Genótipos” e “Cortes” mostraram-se significativas para todos os caracteres. Esses resultados mostram que o comportamento dos genótipos, para todos os caracteres, são influenciados pelo ano agrícola e que os genótipos não apresentam um comportamento

consistente ao longo dos cortes.

O coeficiente de variação experimental (CV%), que dá uma idéia da precisão experimental, foi satisfatório, pois apresentou-se abaixo dos níveis normalmente encontrados em experimentos com cana-de-açúcar (CV% = 20 a 22).

O menor valor de CV% do caráter lignina e celulose é justificável, pois, segundo Pimentel Gomes (1982), variáveis medidas em laboratórios sofrem menores variações do que variáveis medidas no campo, como é o caso de PBU e PBS (Tabela 6-10).

Tabela 6: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agronômico Lignina nos clones da população.

Lignina CV% 8.401061

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	7392.0	1056.0	333.3099	2.2e-16
Variedade	256	7956.9	31.1	9.8104	2.2e-16
Corte	1	4855.9	4855.9	1532.6964	2.2e-16
Grupo:Bloco	24	32.3	1.3	0.4247	0.9934
Grupo:Variedade	14	248.1	17.7	5.5945	2.073e-10
Variedade:Corte	256	13842.3	54.1	17.0668	2.2e-16
Grupo:Corte	7	158.0	22.6	7.1335	2.79e-08
Grupo:Bloco:Variedade	810	2575.7	3.2	1.0037	0.4791
Grupo:Bloco:Corte	24	35.5	1.5	0.4670	0.9868
Grupo:Variedade:Corte	14	234.8	16.8	5.2946	1.054e-09
Resíduo	810	2566.3	3.2		

Tabela 7: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Celulose nos clones da população.

Celulose CV% 9.308936

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	12784.4	1826.34	469.4987	2.2e-16
Variedade	256	7740.1	30.23	7.7724	2.2e-16
Corte	1	3030.9	3030.93	779.1642	2.2e-16
Grupo:Bloco	24	54.1	2.26	0.5799	0.9469
Grupo:Variedade	14	350.8	25.06	6.4412	2.034e-12
Variedade:Corte	256	16327.1	63.78	16.3954	2.2e-16
Grupo:Corte	7	220.8	31.54	8.1089	1.471e-09
Grupo:Bloco:Variedade	810	3144.7	3.88	0.9981	0.5111
Grupo:Bloco:Corte	24	56.8	2.37	0.6086	0.9300
Grupo:Variedade:Corte	14	369.8	26.41	6.7900	2.999e-13
Resíduo	810	3150.9	3.89		

Tabela 8: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico Fibra nos clones da população.

Fibra CV% 2.659996

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	130.6	18.658	58.7433	2.2e-16
Variedade	256	5261.1	20.551	64.7032	2.2e-16
Corte	1	165.8	165.833	522.1113	2.2e-16
Grupo:Bloco	24	9.8	0.410	1.2896	0.16011
Grupo:Variedade	14	3.7	0.267	0.8410	0.62437
Variedade:Corte	256	288.4	1.127	3.5473	2.2e-16
Grupo:Corte	7	5.2	0.738	2.3234	0.02376
Grupo:Bloco:Variedade	810	358.9	0.443	1.3951	1.144e-06
Grupo:Bloco:Corte	24	8.1	0.338	1.0627	0.38158
Grupo:Variedade:Corte	14	3.7	0.263	0.8267	0.64028
Resíduo	810	257.3	0.318		

Tabela 9: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônomo PBS (peso seco do bagaço) nos clones da população.

PBS CV% 13.95805

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	2656	379.4	43.3783	2.2e-16
Variedade	256	129611	506.3	57.8903	2.2e-16
Corte	1	8798	8797.9	1005.9618	2.2e-16
Grupo:Bloco	24	428	17.8	2.0396	0.002406
Grupo:Variedade	14	163	11.7	1.3341	0.180733
Variedade:Corte	256	9181	35.9	4.1005	2.2e-16
Grupo:Corte	7	269	38.5	4.3991	8.310e-05
Grupo:Bloco:Variedade	810	10464	12.9	1.4771	1.585e-08
Grupo:Bloco:Corte	24	252	10.5	1.1986	0.233654
Grupo:Variedade:Corte	14	119	8.5	0.9747	0.477500
Resíduo	810	7084	8.7		

Tabela 10: Análises de Variância e Coeficiente de Variação Ambiental para o caráter agrônômico PBU (peso úmido do bagaço) nos clones da população.

PBU CV% 21.24138

	GI	SQ	QM	Fvalue	Pr(>F)
Grupo	7	10281	1468.7	72.5125	2.2e-16
Variedade	256	220473	861.2	42.5211	2.2e-16
Corte	1	5579	5579.5	275.4736	2.2e-16
Grupo:Bloco	24	1381	57.5	2.8413	7.688e-06
Grupo:Variedade	14	406	29.0	1.4306	0.13235
Variedade:Corte	256	28654	111.9	5.5262	2.2e-16
Grupo:Corte	7	1363	194.7	9.6138	1.616e-11
Grupo:Bloco:Variedade	810	24419	30.1	1.4884	8.535e-09
Grupo:Bloco:Corte	24	871	36.3	1.7913	0.01147
Grupo:Variedade:Corte	14	271	19.4	0.9560	0.49743
Resíduo	810	16406	20.3		

3.4 Herdabilidade

Os caracteres PBU, PBS e fibra apresentaram os maiores valores de herdabilidade (Figura 5), o que demonstra que são altamente determinados pelo genótipo, sofrendo pouca alteração ambiental. Já lignina e celulose, mostraram-se com maior influência ambiental, tendo a celulose destaque com a menor herdabilidade observada (Figura 5).

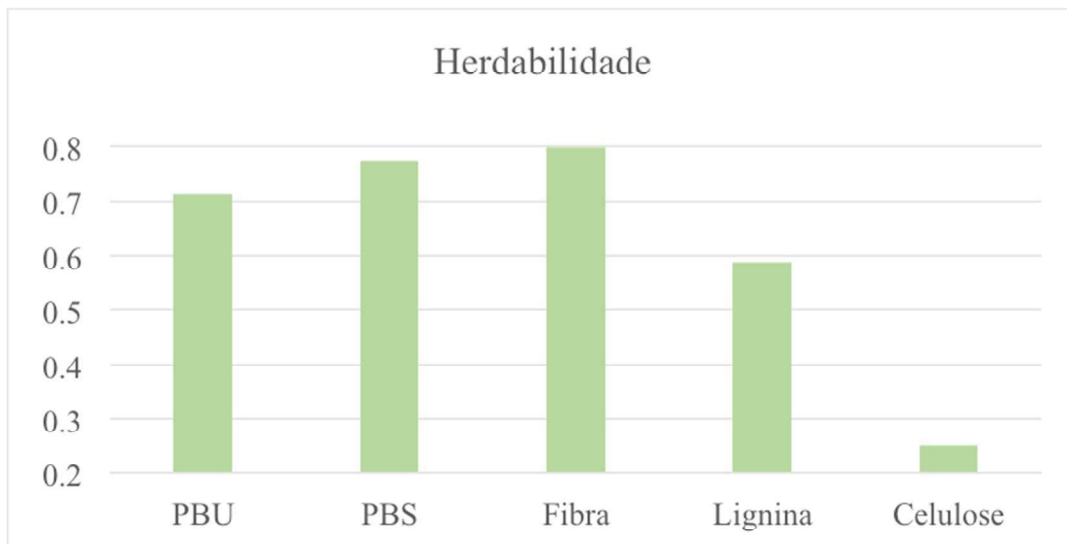


Figura 5: Estimativas do coeficiente de herdabilidade para os caracteres PBU, PBS, Fibra, Lignina e Celulose.

3.5 Análises de associação dos marcadores microssatélites

Oitenta e nove marcadores microssatélites oriundos de sequências expressas do genoma de cana-de-açúcar (EST-SSR) obtidos a partir de rotas metabólicas relacionadas a lignina e celulose e 15 marcadores originados de sequências genômicas (gSSR) foram genotipados na população, produzindo 166 marcadores polimórficos em dose única (Figura 6). Os marcadores **SCB381D** e o **SCESSR0148A** mostraram-se significativos para todas as características em questão, podendo ambos serem usados em programas de melhoramento que visem detectar plantas com alto rendimento de fibra. Alguns marcadores mostraram-se específicos para apenas um caráter, como é o caso do **SCB184A** para PBU.

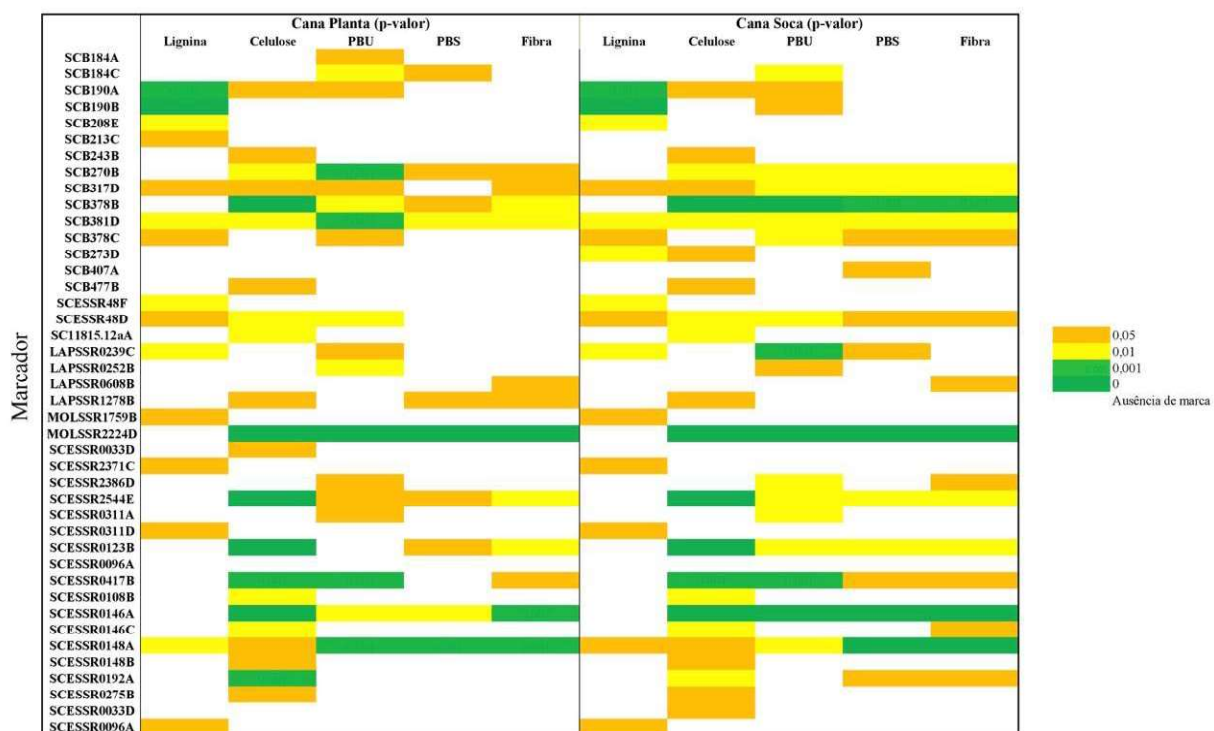


Figura 6: Single marker analysis para detectar os marcadores de microssatélites relacionados com as características PBU, PBS, Fibra, Lignina e Celulose.

A regressão linear múltipla foi utilizada para determinar a proporção de variação fenotípica explicada pela associação entre os marcadores e os SSRs (Figura 7).

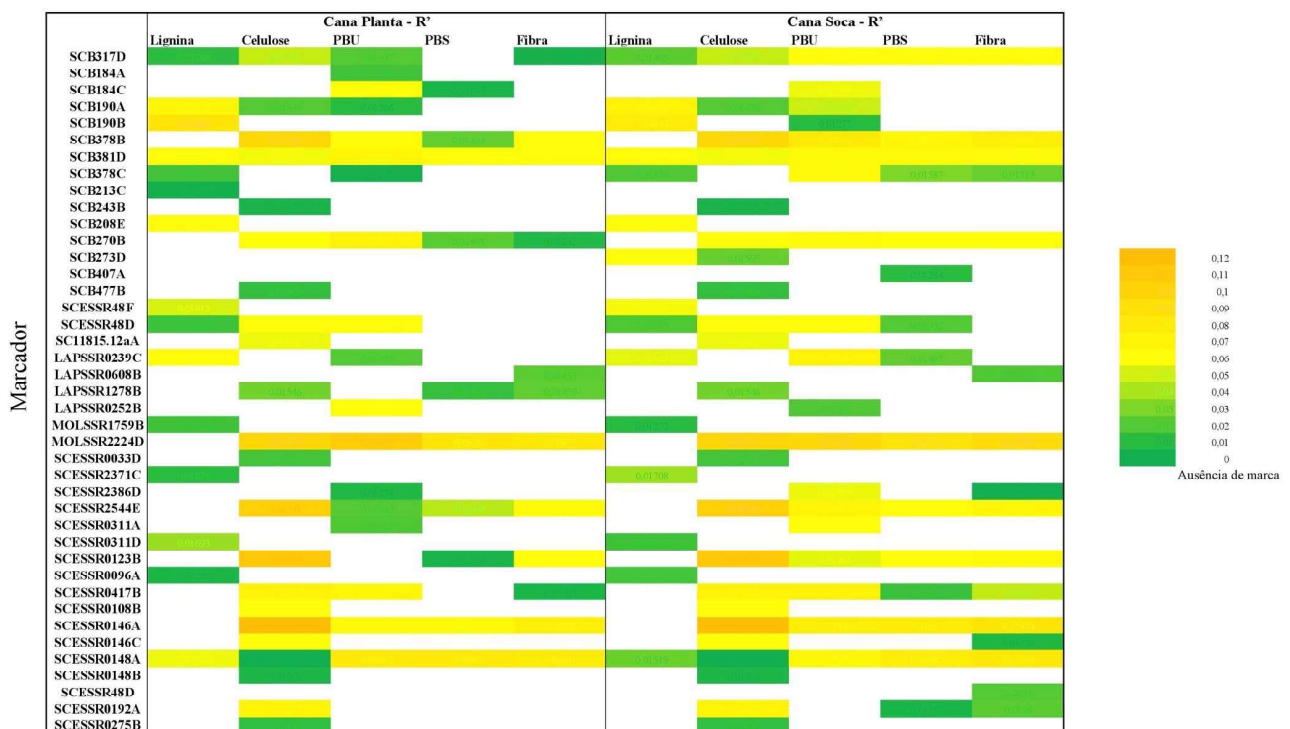


Figura 7: Regressão linear múltipla com os marcadores SSRs nas características PBU, PBS, Fibra, Lignina e Celulose.

4 CONCLUSÃO

Todas as características avaliadas (PBU, PBS, Fibra, Lignina e Celulose) foram significativas para genótipos pelo teste F ($P < 0,01$), demonstrando haver diferenças significativas entre os genótipos;

Os parâmetros genéticos referentes ao presente estudo indicam grandes possibilidades de sucesso neste programa de melhoramento visando à seleção das características avaliadas;

A análise de variabilidade genética via marcadores moleculares, demonstrou uma significativa variabilidade genética na população, devido à segregação dos indivíduos;

A análise de componentes principais revelou presença de estrutura de população.

REFERÊNCIAS

ALLARD, R.W.; BRADSHAW, A.D. Implications of genotype-environmental interactions in applied plant breeding. **Crop Science**, Madison, v. 4, n. 5, p. 503-508, 1964.

ARRUDA P. Genetically modified sugarcane for bioenergy generation. **Curr Opin Biotechnol**. 2012.

CASU RE, DIMMOCK CM, CHAPMAN SC, GROF CP, MCINTYRE CL, BONNETT GD, ET AL. Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. **Plant Mol Biol**. CSIRO Plant Industry, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, Queensland 4067, Australia, 2004.

CHEAVEGATTI-GIANOTTO, A.; ABREU, H. M. C.; ARRUDA, P.; BESPALHOK-FILHO, J. C.; BURNQUIST, W. L.; CRESTE, S.; DI-CIERO, L.; FERRO, J. A.; FIGUEIRA, A. V. O.; FILGUEIRAS, T. S.; GROSSI-DE-SA, M. F.; GUZZO, E. C.; HOFFMANN, H. P.; LANDELL, M. G. A.; MACEDO, N.; MATSUOKA, S.; REINACH, F. C.; ROMANO, E.; SILVA, W. J.; SILVA-FILHO, M. C.; ULIAN, E. C. Sugarcane (*Saccharum → o cinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. **Tropical Plant Biology**, Berlin, v. 4, n. 1, p. 62–89, 2011.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Ed. UFV, 2003.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed., Viçosa: Ed. UFV, 2001.

DIAS MOS, ENSINAS A V., NEBRA SA, MACIEL FILHO R, ROSSELL CEV, MACIEL MRW. Production of bioethanol and other bio-based materials from sugarcane bagasse: Integration to conventional bioethanol production process. **Chem Eng Res Des**. 2009.

FALCONER, D. S., Introduction to Quantitative Genetics, Ed. 2. **Longmans Green**, London/New York, ed. 2, 1981.

MORRISSON, D.F. Multivariate Statistical Methods, 2nd ed.; **McGraw-Hill: NewYork**, 1976.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 10. ed. Piracicaba: Nobel, 430p. 1982.

RAMALHO, M.A.P.; SANTOS, J.; ZIMMERMANN, M.J.O. **Genética quantitativa em plantas autógamias**. Goiânia: UFG, 272 p. 1993.

RESENDE, M.D.V.; STURION, J.A. Análise genética de dados com dependência espacial e temporal no melhoramento de plantas perenes via modelos

geoestatísticos e de series temporais empregando REML/BLUP ao nível individual. **Colombo: Embrapa Florestas**, 2001.

RESENDE, M.D.V. Matemática e estatística na análise de experimentos e no melhoramento genético. **Colombo: Embrapa Florestas**, 2007.

RIBEIRO, M.P.A. Fast Determination of the Composition of Pretreated Sugarcane Bagasse Using Near-Infrared Spectroscopy, **Bioenerg. Res.** 2014.

SOUZA, G. M.; BERGES, H.; BOCS, S.; CASU, R.; D'HONT, A.; FERREIRA, J. E.; HENRY, R.; MING, R.; POTIER, B.; VAN SLUYS, M.-A.; VINCENTZ, M.; PATERSON, A. H. The sugarcane genome challenge: strategies for sequencing a highly complex genome. **Tropical Plant Biology**, Berlin, v. 4, p. 145–156, 2011.

TANIMOTO, T. The press method of cane analysis. **Hawaiian Planter's Record**, Honolulu, v. 57, n. 2, p. 133-150, 1964.

VEGA-SÁNCHEZ ME, RONALD PC. Genetic and biotechnological approaches for biofuel crop improvement. **Curr Opin Biotechnol.** 2010.

VERMERRIS W. Survey of Genomics Approaches to Improve Bioenergy Traits in Maize, Sorghum and Sugarcane. **J Integr Plant Biol.** Blackwell Publishing Asia; 2011.

ZHAO X, ZHANG L, LIU D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. **Biofuels, Bioprod Biorefining.** John Wiley & Sons, Ltd.; 2012.

YUAN JS, TILLER KH, AL-AHMAD H, STEWART NR, STEWART CN JR. Plants to power: bioenergy to fuel the future. **Trends Plant Sci.** Elsevier; 2014.

Capítulo II

A SNP genetic map constructed using restriction site-associated DNA sequencing approach for sugarcane.

K. Kettener¹, M. Quezada², A.A.F. Garcia², J.A.G. da Silva³, C.L. Marino¹

¹Departamento de Genética, Universidade Estadual Paulista (UNESP), CEP 18609000, Botucatu-SP, Brasil.

² Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-SP, Brasil.

³Texas A&M AgriLife Research, Weslaco-TX, USA.

*To whom correspondence should be addressed: UNESP-

Campus de Botucatu Instituto de Biociências

Rua Prof. Dr. Antonio Celso Wagner Zanin, s/nº

CEP: 18618-689

E-mail: karinekettener@gmail.com

Fax: (55-14) 3815-3131 ☐ Phone: (55-14) 3815-3131 or 3880-0363

Abstract

Advancements in next-generation sequencing technology have enabled whole genome re-sequencing in many species, providing new discoveries and characterization of molecular polymorphisms. There are limitations, however, to next-generation sequencing approaches for species with large complex genomes such as sugarcane. Restriction site-associated DNA sequencing (RADSeq) has been developed as a tool for association studies and genomics-assisted breeding in a range of species including those with complex genomes. Double digest RAD sequencing (ddRADSeq) with PstI and MluCI restriction enzymes for targeted complexity reduction followed by multiplex sequencing to produce high-quality polymorphism data at a relatively low per sample cost. Here we report the construction of a genetic map for sugarcane using ddRADseq markers in a segregating F1 population bi-parental cross. Sequences were analyzed by a modified version of Tassel-GBS software using four pseudo-references as reference genome. 6,188 SNP markers were scored in the mapping. Analysis of SNP marker data produced 191 linkage groups (LGs) with a total map length of 2487,12 cM. The genetic map will facilitate the identification of genes controlling important agronomic traits, and provide a reference for large-scale re-sequencing projects.

Keywords: Genetic maps; ddRADseq; SNP; Sugarcane.

Introduction

Sugarcane is a major crop of tropical and sub-tropical regions. This fast growing plant is a major source of sugar (sucrose). The high productivity of the sugarcane plant makes it a key target for use as an energy crop. The fiber of the plant is often used to generate electricity. The sugarcane plant is also used to produce ethanol as a fuel.

Sugarcane is a hybrid from an inter-specific hybridization between two wild polyploid relatives, *Saccharum officinarum* L. ($2n = 80$) and *Saccharum spontaneum* L. ($2n = 40-128$), which is genetically complex and the genome size (around 10,000Mbp) is likely to vary between genotypes especially those with different chromosome numbers, creating a challenge in the application of both conventional and molecular breeding to the genetic improvement of sugarcane as a sugar and energy crop. Genetic linkage maps have been difficult to construct in sugarcane due to the double genome structure, homologous and homoeologous chromosomes and aneuploid inheritance. However, with the advent and wide adoption of single dose markers, linkage mapping in sugarcane has witnessed several achievements, indicating the potential to greatly increase our understanding about this plant by application of emerging genomic technologies.

The evolution of next-generation sequencing (NGS) platforms able of producing millions of short (50-100 bp) DNA sequence reads has reduced the costs of DNA sequencing and offers the possibility of making direct, genotyping-by-sequencing (GBS) practical (Metzker L., 2010). Recent initiatives are seeking to sequence the sugarcane genome (Souza et al, 2011.), which already has a large number

of expressed sequence tags (ESTs) at project SUCEST (Vettore et al., 2003), from which they derived important markers to mapping studies. Lately, genotyping SNPs in species of complex genomes, such as sugarcane, has allowed advanced genetic studies (Garcia, 2013). Even with the promise of NGS approach and availability of next-generation sequencing platforms, by now, GBS methods retain significant limitations. The most protocols for the synthesis of DNA fragment libraries compatible with high-throughput sequencing platforms are onerous, costly and would be unreal for production efforts involving hundreds of samples. Sequence-based genotyping is restricted to those species with available genome assemblies (Han B et al., 2009). We are facing a lack of marker platforms that can provide GBS independent of the status of an assembled genome.

The restriction-site associated DNA sequencing (RAD-seq) was one of the NGS methods that sequencing only the DNA flanking specific restriction enzyme sites to produce a reduced representation of the genome, which ligated an adapter containing multiplex identifiers (MIDs) in the reduced-representation libraries (RRLs) (Guo ZH, 2013; Rotino GL, 2011; Johnson EA, 2007; Schultz Q, 2013; Li S, 2012). In these ways, several high-density genetic maps have been constructed in eggplant (Rotino GL, 2012), ryegrass (Slabaugh MB , 2011), barley (Hayes PM , 2011), grape (Li S, 2012) and even sesame (Zhang X, 2013).

Here we report on the construction of a genetic map in sugarcane using the restriction-site associated DNA sequencing (RAD-seq), which, from our knowledge, is the first application of RADseq in sugarcane.

Materials and Methods

Plant material and nucleic acid extraction

Young leaf tissue was collected from a progeny of 88 individuals selected randomly from a sugarcane F1 biparental derived from the cross between two varieties from Texas A&M AgriLife Research Breeding Program, 'H94-5753' (female, *S. officinarum*) X 'US56-14-4' (male, *S. spontaneum*) and the two parents. DNA was extracted using the Qiagen 96-plex DNeasy kit as per the Qiagen fresh leaf tissue 96-plex protocol (<http://www.qiagen.com/HB/DNeasy96Plant>).

RAD-Seq library preparation and sequencing

The procedure was performed as described by Peterson et al. (2012) with some modifications. First, genomic DNA was double digested separately with restriction enzymes. The double digest reactions were carried out in a volume of 25 μ l containing approximately 150 ng of genomic DNA, 5 U of *Pst*I and *MLu*CI (Fermentas), and 1 \times buffer. The reaction mixture was incubated at 37°C for 6 hours and 65°C for 90 min. Second, the fragments were ligated with adaptors. The ligation reaction was conducted in a reaction volume of 50 μ l at 16°C overnight, containing 10 pmol of *Pst*I and *MLu*CI adaptors, and 1,000 U of T4 DNA Ligase (New England Biolabs [NEB]). To ensure that the digestion was complete, the digestions were performed again with the same enzymes. Each sample was then amplified via PCR in a 50 μ l reaction volume, containing 50–100 ng of adaptor ligated DNA fragments as a template, 1 \times HF buffer, 3.5 mM MgCl₂, 0.4 mM dNTPs, 0.5 U of iProof polymerase (Bio-Rad), and 4 pmol of two

overhang primers. PCR amplification was performed according to the following program: 98°C for 2 min, followed by 13 cycles at 98°C for 30 s, 60°C for 30 s, and 72°C for 15 s, and a final extension at 72°C for 5 min. The PCR products were run on a 2% agarose gel, and fragments of 300–500 bp were recovered from the gel. The samples from 12 individuals were pooled together, and DNA was isolated using a Gel Extraction Kit (Qiagen). The libraries were quantified using Qubit fluorometer (Invitrogen), Agilent 2100 (Agilent Technologies) and real-time quantitative PCR, and then submitted for sequencing on the Illumina HiSeq2000 platform.

SNP discovery and genotyping

We used TASSEL-GBS v.5.2.11, discovery pipeline, to carry out the SNP discovery and indels. As this pipeline requires a reference genome and the sugarcane genome is still being sequenced (Souza et al., 2011), the following pseudo-references were used: methyl-filtrate genome of sugarcane (~674 Mb arranged in 1.109.444 scaffolds) (Grativol et al., 2014), sorghum genome (v. 2.1; ~ 726 Mb arranged in 10 chromosomes and 1.600 scaffolds remaining) (Paterson et al., 2009), transcriptome by RNA-Seq of sugarcane (~780 Mb arranged in 119.768 scaffolds) (Cardoso-Silva et al., 2014), and sequences of the SUCEST project (~152 Mb arranged in 237.954 sequences) (Vettore et al., 2003). First, all the reads were filtered using the default Illumina cluster QC pre filtering and separated reads with short inserts produce chimeric sequence including adapter sequence. The reads that pass this first filter were merged to generate a unique file suitable for the Tassel pipeline. A list of RadSeq tags was created using only the reads that were presented at least 5 times across all the individuals and begin

with a known barcode sequence follow for the expected cut restriction site. Using the Bowtie 2 algorithm (v. 2.2.1) (Langmead; Salzberg, 2013) the tags were aligned against each reference with default parameters. After the alignment, we modified the Java type variable from the “FastToTBTPugin” of Tassel-GBS from byte (-y argument) to short in order to enable the storage of a greater amount of reads per tag per individuals. This alteration improves the quantitative score of heterozygotes and homozygotes genotypes. Adjustments to work with short variables were also required for the subsequent pipeline plugins (Margarido, G. R. A., unpublished). Exact counts of allele-specific readings (depths) were stored in the variant call format type file (vcf).

SNPs filtering and classifying by segregation pattern

Since sugarcane is highly polyploidy, only single-dose markers (Wu et al. 1992) were used for map analysis. According to Wu, et. al., 2002, SNPs in this F1 population can be classified into three groups, when both parents are heterozygous (B3.7: $ab \times ab$); when one parent is heterozygous and the other is homozygous (D1.10: $ab \times aa$) and the opposite (D2.15; $aa \times ab$). The SNPs were stringent filtered retaining only biallelic SNPs, without distortion of the expected mendelian segregation ratio and with less than 20% missing data. Marker segregation distortion was determined by calculating Chi-square test (χ^2) with a Bonferroni correction for multiple test. The threshold consider to discard SNPs with strong deviation from expected segregation was a global significance level of 0.05, implemented in *p.adjusted* function (R Development Core Team, 2012). Duplicate SNPs between references were identified and renamed, considering that they have the same information to construct the map.

Construction of genetic map

Markers used for map construction were filtered using the following criteria: i) for each marker, individuals that lacked genotyping data were less than 20% of the total; and ii) all markers were tested by Chi-square test (χ^2) with a Bonferroni correction for multiple tests ($\alpha=0.05$). Loci with more than 20% of missing data or distorted were excluded from the linkage analysis. Linkage analysis with the markers in single-dose, non-duplicated, and filtered to lost data and distorted segregation was performed by the R package OneMap (2.0-4 v.) (Margarido; Souza; Garcia, 2007). This package, based on the methods of Wu et al. (2002a, 2002b), allows simultaneous multipoint estimation of the recombination fraction and the linkage phases between markers. The co-segregation groups or linkage groups (LGs) were obtained considering LOD score > 9.0 and recombination fraction < 0.1 calculated from two-point test. For LGs up to five markers, an exhaustive search command was performed for order markers (*'compare' function*), choosing the order with the highest multipoint likelihood. For the larger groups, an informational subset of five markers was ordered as an initial framework upon which new markers were sequentially positioned (*'order.seq' function*), as recommended by Mollinari et al. (2009). Heatmaps of LOD scores with recombination fractions between each pair of markers were inspected visually and manual correction was made when necessary along the map's construction. The miss-alignments or originally unlinked markers were again tested on the ordered LGs by *'try.seq'* function. The order of groups with more than five markers was fine considering the output of *ripple* function, which applies the same name algorithm (Lander, Green, 1987) within a sliding window of five

markers. Finally, the multipoint estimates of recombination fractions were converted to map distances using the Kosambi function. A graphic representation of the map was generated using Mapchart 2.2 software (Voorrips, 2002).

Results

Marker development

There were 19,013,358 million readings taken to align against four different pseudo-reference genomes. Methyl-filtrate sugarcane genome (Grativol et al., 2014) resulted in greater alignment rate, with 82.46% (312,9) lined tags. In relation to references based on RNA, 31.53% (119,646) and 17.30% (65,627) the tags were aligned against RNA-Seq transcriptome (Cardoso- Silva et al., 2014) and sequences of SUCEST project (Vettore et al., 2003), respectively. However, their non-single alignment rates differed greatly (Table 1), in agreement with the findings of Vicentini et al. (2012) over the existing redundancy SUCEST database. Finally, the sorghum reference genome had a total of 35.01% (132,845) aligned tags (Table 1).

Table 1: Alignment by Bowtie2 of 19,013,358 tags of RadSeq in a population of sugarcane mapping. The total aligned tags are the sum of the unique and non-unique alignments.

Pseudo-reference genome	Tags aligned	non-	Tags aligned		
			Total	Single alignment	Non-single alignment
Methyl-filtrate genome	66,553 (17.54%)		82.46%	178,052 (46.92%)	134,848 (35.54%)
Sorghum genome	246,608 (64.99%)		35.01%	95,076 (25.06%)	37,769 (9.95%)
RNA-Seq transcriptome	259,807 (68.47%)		31.53%	100,250 (26.42%)	19,396 (5.11%)
Sequences of SUCEST project	313,826 (82.70%)		17.30%	23,754 (6.26%)	41,873 (11.04%)

SNP calling

Aligning the reads to all the four pseudo-reference genomes performed SNP calling. We chose uniquely mapped reads for SNP discovery and after applying the filtering procedure, a total of 62,960 SNPs was found in Sorghum genome, 173,274 SNPs for Methyl-filtrate genome, 74,253 SNPs for transcriptome and 27,594 SNPs for Sucest (Table 2). After the SNP calling procedure, two of the 88 individual progeny were found to have a high quantity of missing data and were subsequently dropped for further analysis.

Table 2: Number of SNP markers generated after analyzing the RadSeq data from a sugarcane population by Tassel-GBS pipeline.

Reference	Chromosome	SNPs
Sorghum genome	1	9.718
	2	7.786
	3	8.095
	4	7.180
	5	4.496
	6	5.939
	7	4.690
	8	3.835
	9	5.322
	10	5.202
	Total	62.960
Methyl-filtrate genome	1	173.274
RNA-Seq transcriptome	1	74.253
Sequences of SUCEST project	1	27.594

The redundancy of the selected single dose markers was inspected after filtering ploidy and quality and the genotypes already encoded as a, ab and b. For example, comparably, all references show very similar levels of redundancy within and between them (Figure 1). Only 70 markers were equally assigned to all four references. Among Sorghum, Methyl filtered and RNA-seq, 252 markers were given as ambiguous. Among SUCEST, Methyl filtered and RNA-seq, 58 ambiguous markers (Figure 1).

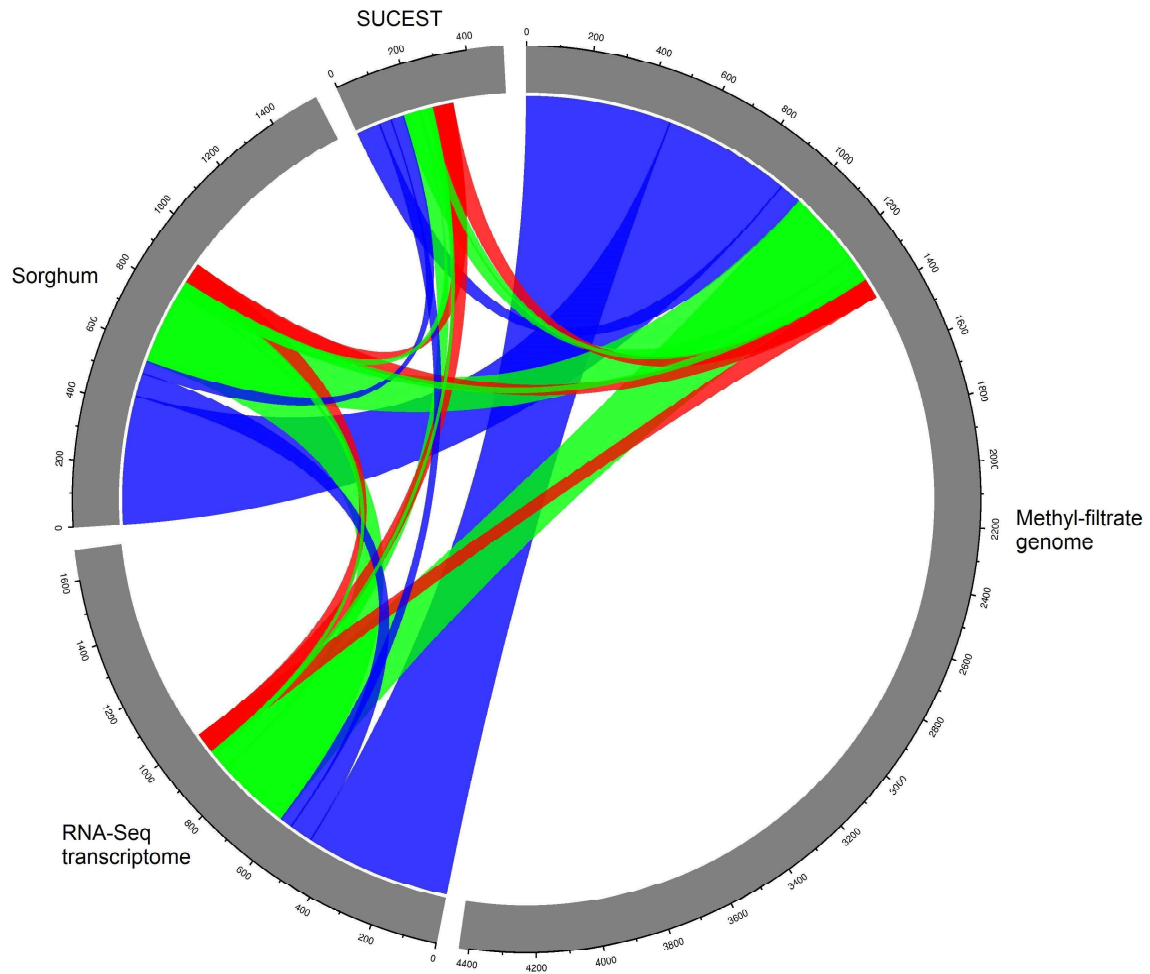


Figure 1 - Circular graph showing redundancy between single-dose markers obtained from four genome references used to map Rad tag in a sugarcane mapping population. Red regions represent a redundancy within each reference, while regions in green, orange and blue respectively represent a redundancy of four, three and two references.

Segregation analysis

SNP markers developed from RADseq cannot all be used in map construction. High missing values affected the map orders and reduced the map accuracy (Hackett, C.A., Broadfoot, L.B., 2003). In the cross type B3.7 (abXab), we found 262 markers using sorghum as a reference genome, 806 markers in methyl-filtrate genome, 277 in the transcriptome genome and 99 in the SUCEST genome. In the cross type D1.12 (boXaa), we found 625 markers using sorghum as a reference genome, 1741 markers in methyl-filtrate genome, 735 in the

transcriptome genome and 226 in the SUCEST genome. D2.15 (aaXab), we found 674 markers in the sorghum genome, 1953 in the methyl-filtrate genome, 703 in the transcriptome genome and 195 in the SUCEST genome (Table 4).

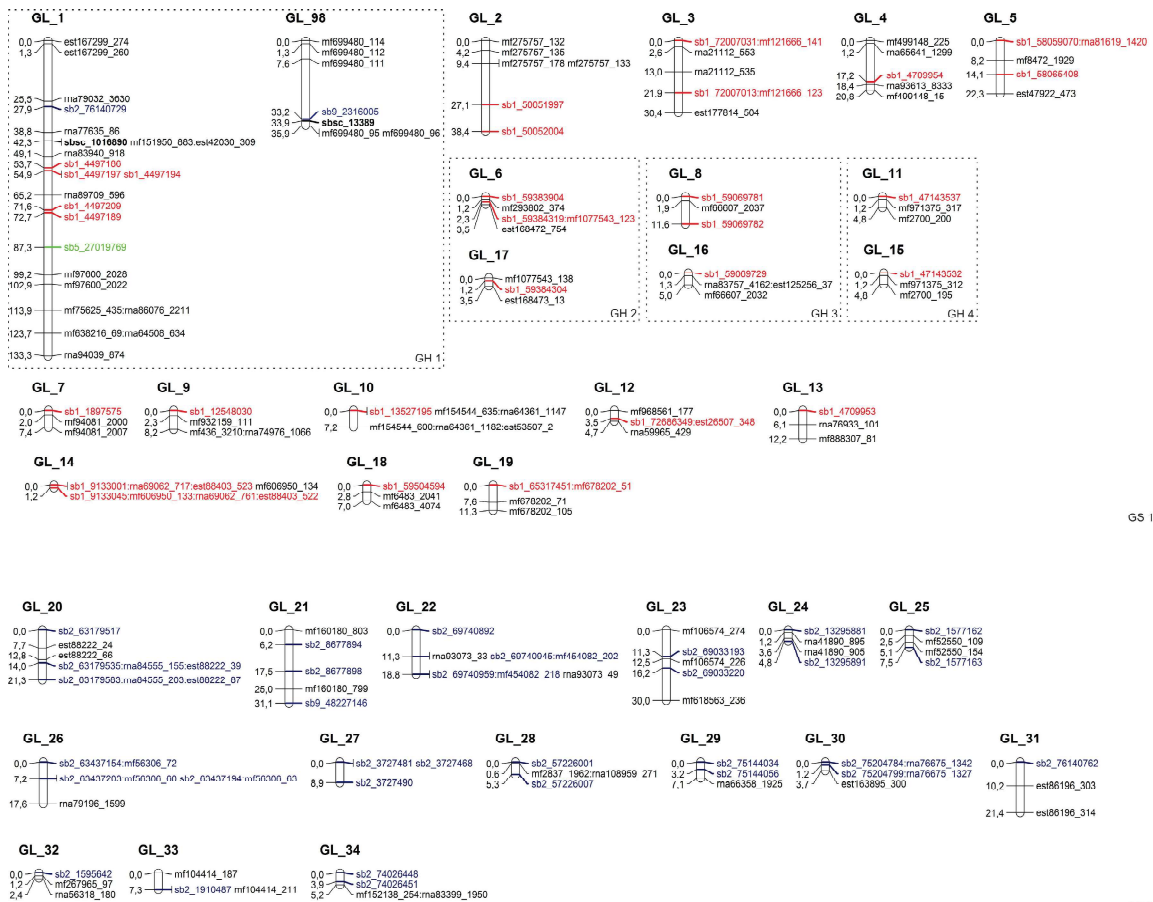
Table 3: Total markers in each reference genome to be used for genetic mapping.

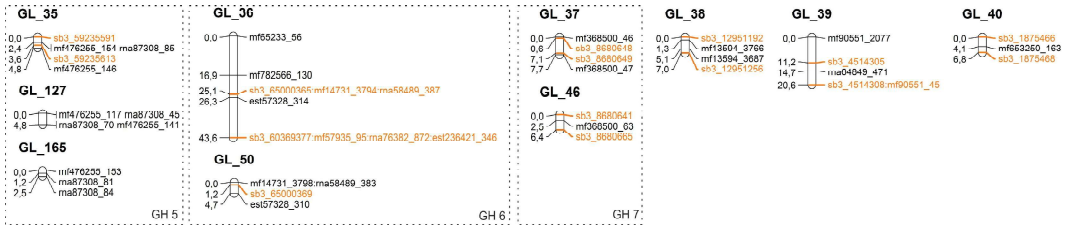
Marker type	Sorghum genome	Methyl-filtrate genome	Transcriptome genome	SUCEST genome
B 3.7	262	806	277	99
D 1.12	625	1741	735	226
D 2.15	674	1953	703	195
Total	1561	4500	1715	520

Sugarcane genetic map

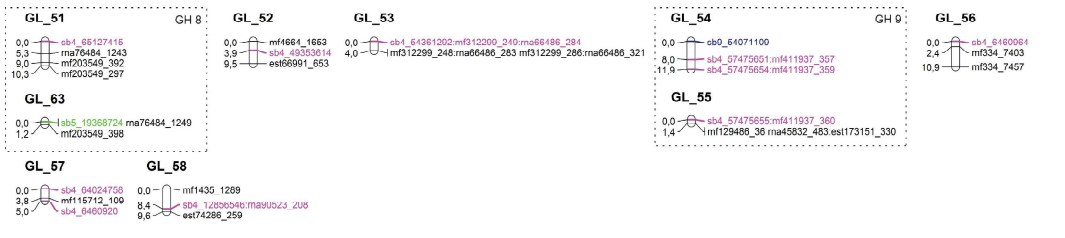
Of the 8,296 developed SNP markers, 720 (11.63%) were included in the map, which were combined into 191 linkage groups (Figure 2). The total map length was 2,487.12 cM, with a marker density of 3.45 cM. We obtained 124 linkage groups (64.21%) with 3 markers, 37 linkage groups with 4 markers (19.47%) 17 linkage groups with 5 markers (8.94%) 3 linkage groups with 6 markers (10.2%) 5 linkage groups with 7 markers (2.63%) and 5 linkage groups with more than 7 markers (2.63%). GLs sizes ranged from 1.2 cM (GL 14; 41; 48; 63; 104 and 126) to 133.3 cM (GL 1), with an average of 12.72 cM; 9 GLs had sizes smaller than 2 cM, 112 GLs exhibited greater than or equal 2 cM sizes and smaller than 10 cM, and the remaining 79 GLs have sizes greater than or equal to 10 cM. Distances greater than 15 cM between pairs of consecutive markers were observed in slightly more than 10 % of cases representing very few gaps.

Figure 2 shows the distribution of GLs grouped into 23 groups of hom (e) ology (GHs), obtained from common sources of genomic loci mapped in different LGs. The number of GLs allocated in GHs ranged from two to four, the highest GH containing 18 markers (GH 12), and the lowest with five markers (GH 9; 11). GLs also were gathered in synteny groups (GSs) with sorghum, based on the distribution of locos coming from RadSeq with reference to the genome of this species. Here, the GSs numbering refers to sorghum chromosome loci which they originated. The number of GLs in each GS varied between seven (GS 5) to 20 (GS 1). In fact, the greatest number of loci (53) reed alignment originating from the genome of sorghum was obtained for chromosome 1, while the lowest number of loci (18) was available for chromosome 5.

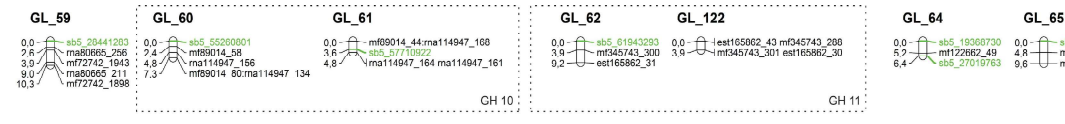




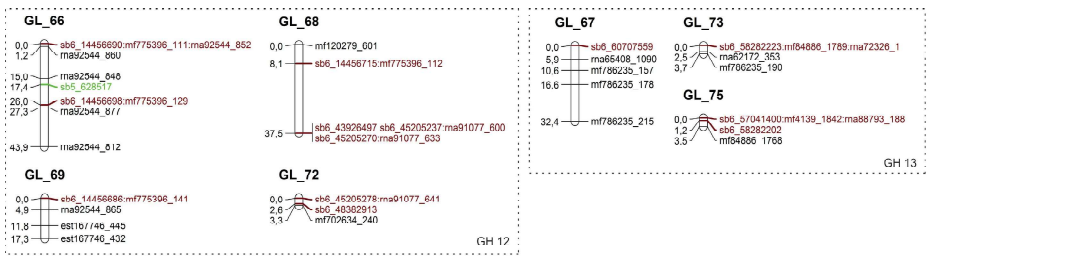
GS 3



GS 4



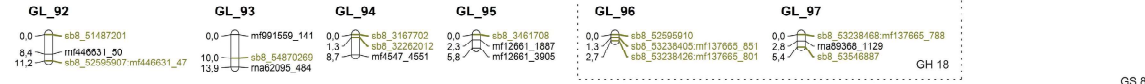
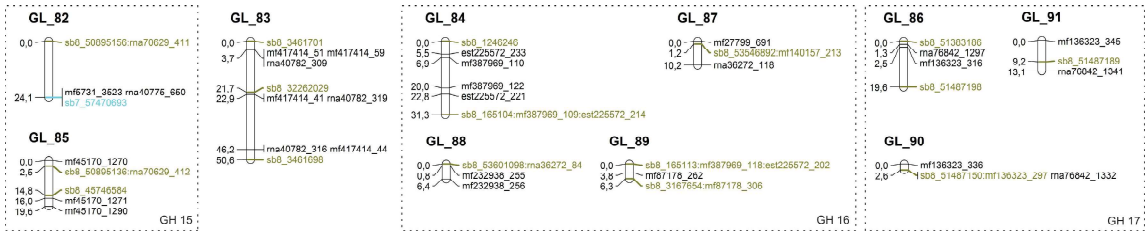
GS 5



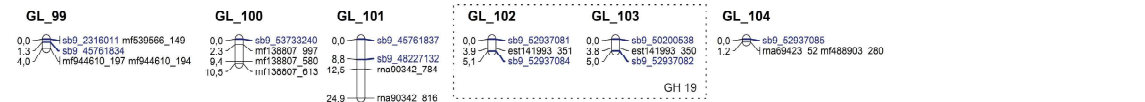
GS 6



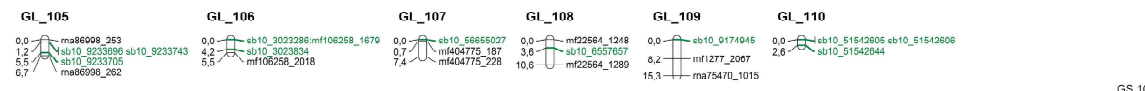
GS 7



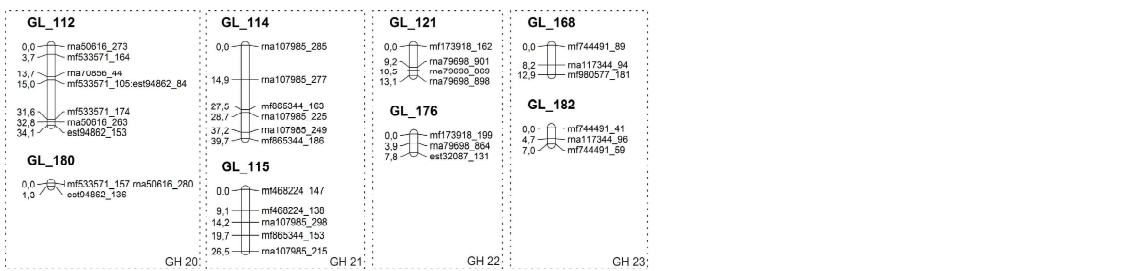
GS 8



GS 9



GS 10



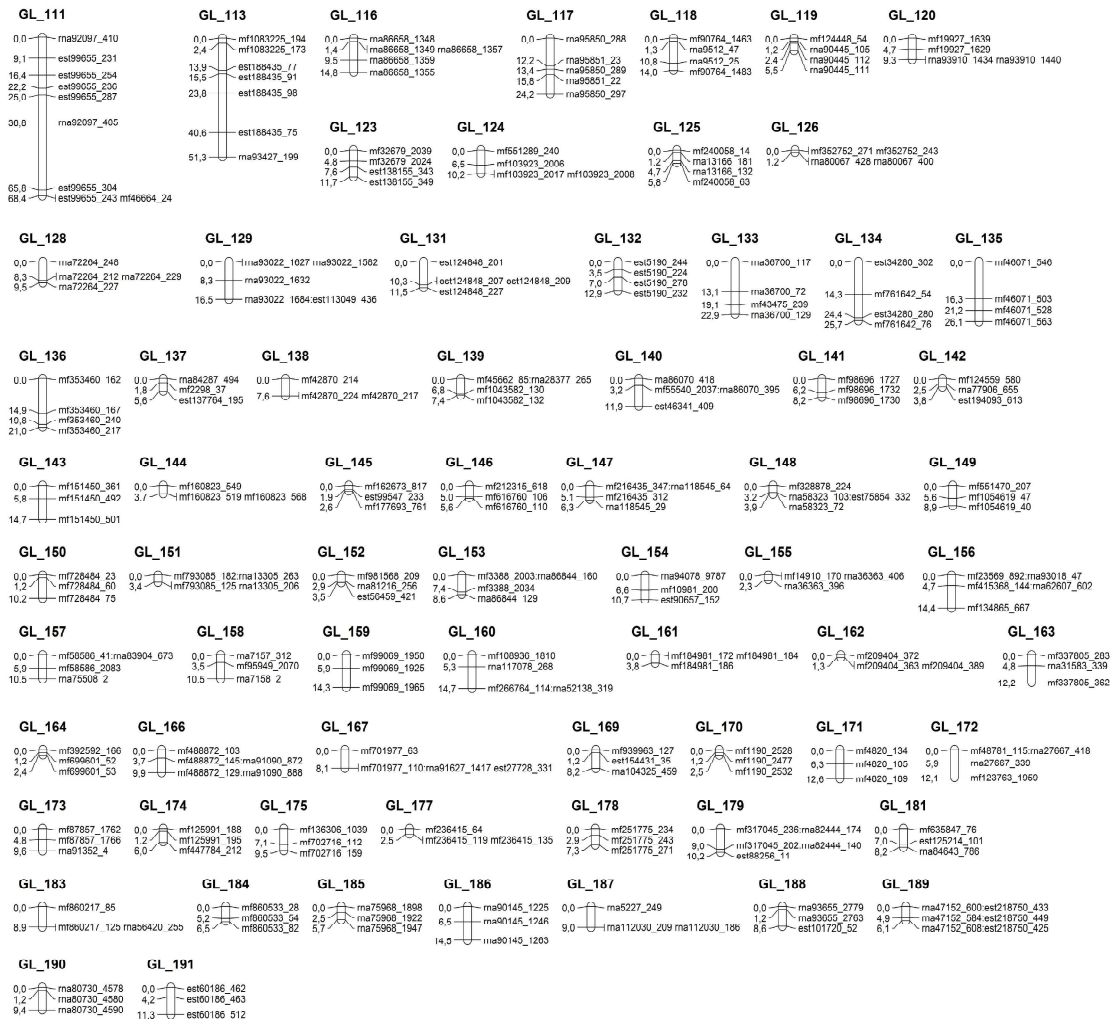


Figure 2: Genetic map of sugarcane displaying the distribution of linkage groups (LGs) along the hom(e)ology (GHs) and synteny with sorghum (GSs). Genetic distances between adjacent markers are shown on the left of each co-segregation group (CG). In red we have the representation of the type B3.7 markers, in green the type D1.12 and D2.15 in blue.

Discussion

This was the first time that large numbers of SNP markers were developed using RADseq in sugarcane. In this work, a total of 6,188 valid SNP markers was generated through *Pst*I and *MLu*CI restriction enzymes, which is a satisfactory number by comparison with EST-based SNPs that have low coverage of the genome, but useful for bridging functional and structural genomics (Cordeiro GM, et al., 2006). SNP arrays and chips needed genomic information in advance and

were always custom designed or commercialized, which means it was costly for screening a big population (Lawley C, et al, 2012; Korban SS. Et al, 2012; Guardo MD, et al., 2014). For a species with high heterozygosity, more than one restriction enzyme may be required to reveal more molecular diversity (Zerbini, F.M. et al, 2014). Altogether, it is suggested that RADseq was a rapid, highly efficient and less cost methods for molecular breeding using larger population and larger amount of markers.

In total, 720 markers (11.63%) were mapped, which shows that we lost a large amount of markers, probably due to the crop ploidy. The markers were distributed over 191 linkage groups (GLs) and covered a total of 2,487.12 cM (Figure 1), determining a high average density of 3.45 cM compared to other studies (Pastina et al. 2010). GLs sizes ranging from 1.17 cM to 133.3 cM, with an average of 12.98 cM. The number of gaps greater than 20 cM between pairs of consecutive markers was observed in 9 cases, representing very few gaps. Although the use of the marks has been relatively low, these numbers show that for sugarcane, this map can be considered dense and saturated. The 11.63% utilization rate refers mainly to the careful disposal of redundant markers and exclusion of loci distantly allocated at the first ordination. Aitken et al. (2014) developed a sugarcane map based on Diversity Array Technology (DArT) markers, with an approximate density (4.3 cM), but with a higher number of markers (2,267); In this case, no treatment was conducted to avoid redundant mapping markers, well known for genotyping platforms on a large scale.

Methyl-filtrate sugarcane genome (Grativol et al., 2014) resulted in a higher rate alignment with 82.46% (312.9) aligned tags, in relation to transcriptome obtained

via RNA-Seq (Cardoso- Silva et al., 2014) that was 31.53% (119 646) and 17.30% (65 627) the tags were aligned against sequences of SUCEST project (Vettore et al., 2003), respectively. However, their non-single alignment rates differed greatly (Table 1), in agreement with the findings of Vicentini et al. (2012) over the existing redundancy SUCEST database. Finally, the sorghum reference genome had a total of 35.01% (132,845) aligned tags (Table 1). These finds suggests that our tags are in regulatory regions of the genome, which explains the high rate alignment in the methyl-filtrate pseudo-reference genome, that is a quality group of sequences that well represents the sugarcane gene content.

Conclusion

This study provides insights into map construction in sugarcane. RADseq, first time performed on sugarcane, was a rapid and efficient strategy for SNP maker development and genotyping in mapping populations. A main contribution was that we constructed a new standard genetic map of sugarcane with 6,188 SNP markers and 88 individuals derived from a cross between 'H94-5753' (female) X 'US56-14-4' (male, *S. spontaneum*). This may improves the accuracy of genetic maps greatly. The availability of such maps may facilitate a variety of genomic studies in sugarcane, including QTL analysis, candidate genes searching and chromosomal variation.

REFERENCES

AITKEN KS, MCNEIL MD, BERKMAN PJ, HERMANN S, KILIAN A, BUNDOCK PC, ET AL. Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. **BMC Plant Biol.** 2014.

BARCHI L, LANTERI S, PORTIS E, VALE G, VOLANTE A, PULCINI L, CIRIACI T, ACCIARRI N, BARBIERATO V, TOPPINO L, ROTINO GL: A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. **PLoS One** 2012.

BARCHI L, LANTERI S, PORTIS E, ACQUADRO A, VALE G, TOPPINO L, ROTINO GL: Identification of SNP and SSR markers in eggplant using RAD tag sequencing. **BMC Genomics** 2011.

BENJAMINI, Y.; YEKUTIELI, D. The control of the false discovery rate in multiple testing under dependency. **The Annals of Statistics**, Hayward, v. 29, n. 4, p. 1165–1188, 2001.

BIANCO L, CESTARO A, SARGENT DJ, BANCHI E, DERDAK S, GUARDO MD, ET AL. Development and validation of a 20 K Single Nucleotide Polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica* Borkh). **PLoS One.** 2014.

CARDOSO-SILVA, C. B.; COSTA, E. A.; MANCINI, M. C.; BALSALOBRE, T. W. A.; CANESIN, L. E. C.; PINTO, L. R.; CARNEIRO, M. S.; GARCIA, A. A. F.; SOUZA, A. P.; VICENTINI, R. *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. **PLoS ONE**, San Francisco, v. 9, n. 2, p. 1–10, 2014.

CELTON JM, TUSTIN DS, CHAGNÉ D, GARDINER SE. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. **Tree Genet Genomes.** 2009.

CORDEIRO GM, ELIOTT F, MCINTYRE CL, CASU RE, HENRY RJ. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. **Theor Appl Genet.** 2006.

CHAGNÉ D, CROWHURST RN, TROGGIO M, DAVEY MW, GILMORE B, LAWLEY C, ET AL. Genome-wide SNP detection, validation, and development of an 8 K SNP array for apple. **PLoS One.** 2012.

D'HONT A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. **Cytogenet Genome Res.** 2005.

GAO SJ, DAMAJ MB, PARK JW, BEYENE G, BUENROSTRO-NAVA MT, MOLINA J, ET AL. Enhanced transgene expression in sugarcane by co-expression of virus- encoded rna silencing suppressors. **Plos One.** 2013.

GRATIVOL, C.; REGULSKI, M.; BERTALAN, M.; MCCOMBIE, W. R.; SILVA, F.

R.; ZERLOTINI-NETO, A.; VICENTINI, R.; FARINELLI, L.; HEMERLY, A. S.; MARTIENSSEN, R. A.; FERREIRA, P. C. G. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. **The Plant Journal**, Oxford, v. 79, p. 162–172, 2014.

GRIVET L, ARRUDA P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Curr Opin Plant Biol**. 2002.

HACKETT CA, BROADFOOT LB. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. **Heredity**. 2003.

HAN YP, CHAGNÉ D, GASIC K, RIKKERINK EHA, BEEVER JE, GARDINER SE, ET AL. BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. **Genomics**. 2009.

HUANG X., FENG Q., QIAN Q., ZHAO Q., WANG L., WANG A., ET AL..High-throughput genotyping by whole genome resequencing. **Genome Res**. 2009.

JOYCE P, HERMANN S, O'CONNELL A, DINH Q, SHUMBE L, LAKSHMANAN P. Field performance of transgenic sugarcane produced using Agrobacterium and biolistics methods. **Plant Biotech J**. 2014.

KHAN MA, KORBAN SS. Association mapping in forest trees and fruit crops. **J Exp Bot**. 2012.

LANDER, E. S.; GREEN, P. Construction of multilocus genetic linkage maps in humans. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 84, p. 2363–2367, 1987.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v. 9, n. 4, p. 357–359, 2013.

MANICKAVASAGAM M, GANAPATHI A, ANBAZHAGAN VR, SUDHAKAR B, SELVARAJ N, VASUDEVAN A, ET AL. Agrobacterium-mediated genetic transformation and development of herbicide-resistant sugarcane (*Saccharum* species hybrids) using axillary buds. **Plant Cell Rep**. 2004.

MARGARIDO, G. R. A.; SOUZA, A. P.; GARCIA, A. A. F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, Lund, v. 144, n. 3, p. 78–79, 2007.

METZKER M. L. Sequencing technologies – the next generation. **Nat. Rev. Genet**. 2010.

MILLER MR, DUNHAM JP, AMORES A, CRESKO WA, JOHNSON EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. **Genome Res**. 2007.

MORRIS, G. P.; RAMU, P.; DESHPANDE, S. P.; HASH, C. T.; SHAH, T.; UPADHYAYA, H. D.; RIERA-LIZARAZU, O.; BROWN, P. J.; ACHARYA, C. B.;

MITCHELL, S. E.; HARRIMAN, J.; GLAUBITZ, J. C.; BUCKLER, E. S.; KRESOVICH, S. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 110, n. 2, p. 453–458, 2013.

MOLINARI HBC, MARUR CJ, DAROS E, DE CAMPOS MKF, DE CARVALHO J, BESPALHOK JC, ET AL. Evaluation of the stress-inducible production of proline in transgenic sugarcane (*Saccharum* spp.): osmotic adjustment, chlorophyll fluorescence and oxidative stress. **Physiol Plant**. 2007.

MOLLINARI, M.; MARGARIDO, G. R. A.; VENCOVSKY, R.; GARCIA, A. A. F. Evaluation of algorithms used to order markers on genetic maps. **Heredity**, v. 103, n. 6, p. 494–502, 2009.

GARCIA, A.A.F. et al. SNP genotyping allows an in-depth characterization of the genome of sugarcane and other complex autopolyploid. **Sci. Rep.** 3,3399; DOI:10.1038/srep03399, 2013.

MURRAY, S. C.; ROONEY, W. L.; HAMBLIN, M. T.; MITCHELL, S. E.; KRESOVICH, S. Sweet sorghum genetic diversity and association mapping for brix and height. **The Plant Genome**, Madison, v. 2, n. 1, p. 48, 2009.

PASTINA, M. M.; PINTO, L. R.; OLIVEIRA, K. M.; SOUZA, A. P.; GARCIA, A. A. F. Molecular mapping of complex traits. In: HENRY, R. J.; KOLE, C. (Ed.). **Genetics, Genomics and Breeding of Sugarcane**. Enfield: Science Publishers, cap. 7, p. 117–148, 2010.

PATERSON, A. H.; BOWERS, J. E.; BRUGGMANN, R.; DUBCHAK, I.; GRIMWOOD, J.; GUNDLACH, H.; HABERER, G.; HELLSTEN, U.; MITROS, T.; POLIAKOV, A.; SCHMUTZ, J.; SPANNAGL, M.; TANG, H.; WANG, X.; WICKER, T.; BHARTI, A. K.; CHAPMAN, J.; FELTUS, F. A.; GOWIK, U.; GRIGORIEV, I. V.; LYONS, E.; MAHER, C. A.; MARTIS, M.; NARECHANIA, A.; OTILLAR, R. P.; PENNING, B. W.; SALAMOV, A. A.; WANG, Y.; ZHANG, L.; CARPITA, N. C.; FREELING, M.; GINGLE, A. R.; HASH, C. T.; KELLER, B.; KLEIN, P.; KRESOVICH, S.; MCCANN, M. C.; MING, R.; PETERSON, D. G.; RAHMAN, M.; WARE, D.; WESTHOFF, P.; MAYER, K. F. X.; MESSING, J.; ROKHSAR, D. S. The *Sorghum bicolor* genome and the diversification of grasses. **Nature**, London, v. 457, n. 7229, p. 551–556, 2009.

PEGADARAJU V, NIPPER R, HULKE B, QI L, SCHULTZ Q: De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. **BMC Genomics** 2013.

PETERSON BK, WEBER JN, KAY EH, FISHER HS, HOEKSTRA HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PLoS One**. 2012.

PFENDER WF, SAHA MC, JOHNSON EA, SLABAUGH MB. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust

resistance in *Lolium perenne*. **Theor Appl Genet**. 2011.

QUINONES PM, URQUIZA GC, ZERBINI PA, ZERBINI FM. Molecular diversity of begomoviruses in the pepper crop in Cuba. **Rev Prot Veget**. 2014.

SOUZA, G. M.; BERGES, H.; BOCS, S.; CASU, R.; D'HONT, A.; FERREIRA, J. E.; HENRY, R.; MING, R.; POTIER, B.; VAN SLUYS, M.A.; VINCENTZ, M.; PATERSON, A. H. The sugarcane genome challenge: strategies for sequencing a highly complex genome. **Tropical Plant Biology**, Berlin, v. 4, p. 145–156, 2011.

TAMMISOLA J. Towards much more efficient biofuel crops - can sugarcane pave the way? **GM Crops**. 2010.

VON ZITZEWITZ J, CUESTA-MARCOS A, CONDON F, CASTRO AJ, CHAO S, COREY A, FILICHKIN T, FISK SP, GUTIERREZ L, HAGGARD K, KARSAI I, MUEHLBAUER GJ, SMITH KP, VEISZ O, HAYES PM: The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. **Plant Genome** 2011.

VETTORE, A. L.; SILVA, F. R.; KEMPER, E. L.; SOUZA, G. M.; SILVA, A. M.; FERRO, M. I. T.; HENRIQUE-SILVA, F.; GIGLIOTI, E. A.; LEMOS, M. V. F.; COUTINHO, L. L.; NOBREGA, M. P.; CARRER, H.; FRANCA, S. C.; BACCI-JUNIOR, M.; GOLDMAN, M. H. S.; GOMES, S. L.; NUNES, L. R.; CAMARGO, L. E. A.; SIQUEIRA, W. J.; VAN SLUYS, M.-A.; THIEMANN, O. H.; KURAMAE, E. E.; SANTELLI, R. V.; MARINO, C. L.; TARGON, M. L. P. N.; FERRO, J. A.; SILVEIRA, H. C. S.; MARINI, D. C.; LEMOS, E. G. M.; MONTEIRO-VITORELLO, C. B.; TAMBOR, J. H. M.; CARRARO, D. M.; ROBERTO, P. G.; MARTINS, V. G.; GOLDMAN, G. H.; OLIVEIRA, R. C.; TRUFFI, D.; COLOMBO, C. A.; ROSSI, M.; ARAUJO, P. G.; SCULACCIO, S. A.; ANGELLA, A.; LIMA, M. M. A.; ROSA-JUNIOR, V. E.; SIVIERO, F.; COSCRATO, V. E.; MACHADO, M. A.; GRIVET, L.; DI-MAURO, S. M. Z.; NOBREGA, F. G.; MENCK, C. F. M.; BRAGA, M. D. V.; TELLES, G. P.; CARA, F. A. A.; PEDROSA, G.; MEIDANIS, J.; ARRUDA, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, Cold Spring Harbor, v. 13, n. 12, p. 2725–2735, 2003.

VICENTINI, R.; DEL BEM, L. E. V.; VAN SLUYS, M. A.; NOGUEIRA, F. T. S.; VINCENTZ, M. Gene content analysis of sugarcane public ESTs reveals thousands of missing coding-genes and an unexpected pool of grasses conserved ncRNAs. **Tropical Plant Biology**, Berlin, v. 5, n. 2, p. 199–205, 2012.

VOORRIPS, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. **The Journal of Heredity**, Washington, v. 93, n. 1, p. 77–78, 2002.

ZALE J, JUNG, J.H., KIM, J.Y., PATHAK, B., KARAN, R.LIU, H.CHEN, X.WU, H. CANDREVA, J.ZHAI Z.SHANKLIN, J. ALTPETER, F.: Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. **Plant Biotech J**. 2015.

ZHANG H, MIAO H, WEI L, LI C, ZHAO R, WANG C: Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). **PLoS One** 2013.

WANG N, FANG L, XIN H, WANG L, LI S: Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. **BMC Plant Biol** 2012.

WANG XQ, ZHAO L, EATON DA, LI DZ, GUO ZH. Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. **Mol Ecol Resour.** 2013.

WU K. K., BURNQUIST W., SORRELLS M. E., TEW T. L., MOORE P. H., ET AL. , 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theor. Appl. Genet.** 1992.

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**, Berlin, v. 61, n. 3, p. 349–363, 2002a.

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