

UNIVERSIDADE DE SÃO PAULO
ESCOLA SUPERIOR DE AGRICULTURA “LUIZ DE QUEIROZ”
PROGRAMA DE PÓS-GRADUAÇÃO EM “GENÉTICA E MELHORAMENTO DE
PLANTAS”

TÓPICOS ESPECIAIS EM GENÉTICA E MELHORAMENTO DE PLANTAS
PROJETO DE PESQUISA

ANA LETYCIA BASSO GARCIA
JOSÉ HENRIQUE SOLER GUILHEN
DOCENTE: ROBERTO FRITSCHÉ NETO

“MODELOS DE SELEÇÃO GENÔMICA AMPLA EM CANA DE AÇÚCAR BASEADOS EM
FENOTIPAGEM E GENOTIPAGEM DE ALTO DESEMPENHO”

PIRACICABA
Estado de São Paulo – Brasil
Novembro / 2017

1 INTRODUÇÃO

A cana de açúcar é cultivada em regiões tropicais com a finalidade de industrial para produção de açúcar ou, mais recentemente, pelo seu potencial bioenergético. A biomassa é uma característica agronomicamente importante para a produção de etanol de segunda geração. Os programas de melhoramento em cana visam, nesse sentido, plantas com elevado teor de fibra e com paredes celulares fáceis de serem quebradas, além de alta produtividade e outras características correlacionadas. Para o desenvolvimento de cultivares superiores, esses programas têm trabalhado com métodos de seleção fenotípica tradicional, combinando seleção de famílias e seleção massal. Entre a identificação de matérias elite e as etapas finais do programa de melhoramento são necessários entre sete e dez anos, com grandes experimentos multi-locais, em algumas etapas. A cana de açúcar moderna cultivada é um híbrido entre *S. officinarum* e *S. spontaneum*, e possui genoma poliploide, com número de cromossomos variáveis e altamente heterozigoto. Essa complexidade dificulta a seleção por marcadores moleculares (MOORE; BOTHA, 2014).

A seleção genômica (SG) foi proposta por Meuwissen et al. (2001) para prever fenótipos de caracteres quantitativos em futuras gerações de uma população, baseada em marcadores moleculares distribuídos por todo o genoma. Esse método tem sido utilizado com eficiência no melhoramento de diploides, como soja, arroz e milho, e mais recentemente em poliploides, como alfafa e trigo. Em cana de açúcar, mesmo com a complexidade do genoma, Gouy et al. (2013) mostraram que o método é viável na seleção para açúcar e biomassa.

A SG baseia-se nos efeitos dos marcadores genéticos estimados em uma população de treinamento. Essa população é uma amostra da população de seleção. Os efeitos dos marcadores, por sua vez, consistem na variação fenotípica para o caráter de interesse a que cada um deles está associado. Os marcadores que explicam a maior parte da variância fenotípica dos caracteres de interesse são validados em uma população teste e então podem ser incorporados na etapa de seleção, para prever os valores genéticos genômicos (GBVs) das populações de melhoramento.

A acurácia da SG depende da herdabilidade do caráter, da densidade de marcadores em relação ao tamanho efetivo da população, da extensão e padrão do desequilíbrio de ligação (LD) existente na população, da precisão dos dados utilizados para medir os efeitos dos marcadores e da eficiência dos métodos estatísticos utilizados na predição. A precisão dos dados fenotípicos pode ser aumentada com o uso de equipamentos de fenotipagem de alto desempenho (HTP – *high throughput phenotyping*), em que é possível avaliar um grande número de caracteres em populações maiores do que vem sendo aplicado no melhoramento convencional (CROSSA et al., 2017).

Em geral, os caracteres primários, como a biomassa, possuem herança complexa. A HTP permite avaliação de um grande número de caracteres secundários que podem ser utilizados para prever os valores genotípicos de caracteres primários, bem como serem utilizados na seleção indireta desses caracteres. A partir de comprimentos de onda do espectro de infravermelho próximo (NIR), é possível prever a biomassa e caracterizar a porção da fibra do colmo da cana, bem como de mensurar o teor de sucrose, relacionado a produção de açúcar, que tradicionalmente são estimados a partir dos conteúdos de celulose, hemicelulose e lignina. Esse conhecimento permite que caracteres secundários correlacionados a marcadores em forte LD sejam utilizados na seleção (seleção *multi-trait*), em preferência aos caracteres primários, por vezes mais complexos e em LD mais fraco.

A associação entre SG e HTP, permite aumento do ganho de seleção, uma vez que tende a aumentar a acurácia da predição, aumenta a intensidade de seleção e reduz o tempo do ciclo de melhoramento. O grande desafio da associação entre SG e HTP é incorporar o grande volume de dados fenotípicos gerados por HTP aos modelos de predição. Os métodos *Genomic Best Linear Unbiased Prediction* (GBLUP), *Random Regression Best Linear Unbiased Predictor* (RR-BLUP), o modelo de predição bivariada (BV), e alguns métodos bayesianos, como o BayesA, tem sido eficiente na predição de genótipos superiores em diversas culturas. É preciso, portanto, compreender qual dos métodos melhor se ajusta ao novo modelo de SG que está sendo proposto.

Esse trabalho justifica-se pela necessidade de implementar a seleção genômica em cana de açúcar, a fim de aumentar a eficácia dos programas de melhoramento e reduzir custos. Dada a complexidade do genoma da cana, a associação entre marcas de grande efeito ou muitas marcas de pequeno efeito para caracteres secundários relacionados à biomassa, avaliados durante vários pontos do ciclo de desenvolvimento da planta, podem auxiliar a construção de modelos preditivos de elevada acurácia, viabilizando a seleção genômica em cana de açúcar.

2 OBJETIVOS

- Prever os valores genotípicos para a população a partir de dados obtidos por HTP;
- Avaliar os padrões de desequilíbrio de ligação e haplótipos dentro da população;
- Definir modelos estatísticos de elevada acurácia para predição de fenótipos para seleção genômica em cana de açúcar.

3 MATERIAL E MÉTODOS

A população de cana de açúcar a ser utilizada neste trabalho foi desenvolvida na Austrália pelo laboratório da *Queensland Alliance for Agriculture and Food Innovation*, na Universidade de Queensland. A população é composta por 186 genótipos, com duas ou três repetições e três testemunhas. O DNA genômico desses indivíduos foi obtido segundo protocolo de (ALJANABI;

FORGET; DOOKUN, 1999). O material foi sequenciado e genotipado por GBS (*genotype-by-sequencing*) pela empresa *Diversity Arrays Technology Pty Ltd*, resultando um conjunto de marcadores SNP e DArT. A fenotipagem, bem como as características da população, são descritas por Hoang et al. (2017). Em síntese, na fenotipagem de alto de desempenho, os colmos das amostras foram coletados e submetidos ao sistema de análises NIR denominado SpectraCane (Biolab, Australia), sob diferentes espectros de NIR para estimação do conteúdo de biomassa e caracterização de fibra e teor de açúcar. Essa avaliação foi feita por três anos.

A análise de qualidade e análise descritiva dos dados genotípicos será realizada para verificar a quantidade de marcadores identificados. Além disso, será avaliado o padrão de distribuição dos marcadores na população, a partir de uma análise de divergência genética e o padrão de equilíbrio de ligação existente. Será processada também uma análise da segregação mendeliana para quantificação da dosagem alélica. A partir disso, serão construídas as matrizes de relacionamento (GRM – *genomic relationship matrix*).

Os dados fenotípicos obtidos serão avaliados em relação a qualidade, sendo submetidos também a uma avaliação descritiva. A partir desse conjunto de dados, composto por medidas NIR e outros caracteres secundários, serão estimados índices vegetativos, como NDVI e preditos os valores genéticos para biomassa e teor de açúcar.

Os BLUPs de cada genótipo serão preditos segundo o modelo misto a seguir, descrito em notação matricial como $y = Xb + Zg + Wt + Qp + \varepsilon$, em que: y é o vetor de observações para biomassa; b é o efeito de bloco (fixo); g é o efeito genético (aleatório); t é o efeito de anos (aleatório); p é o efeito de bloco ano (efeito aleatório); ε é o erro residual e X , Z , W e Q são as matrizes de incidência de efeito. As características secundárias, também conhecidas como dados longitudinais, serão avaliadas por três modelos estatísticos: repetibilidade simples (SR), *multi-trait* (MT) e regressão aleatória (RR).

Para predição genômica, os BLUPs das características secundárias ou da biomassa serão utilizados como variáveis dependentes na modelagem. Serão testados quatro métodos de predição, com base nos valores BLUPs obtidos e na matriz de relacionamento genômico. Os modelos a serem testados serão fundamentados nos métodos *Genomic Best Linear Unbiased Prediction* (GBLUP), *Random Regression Best Linear Unbiased Predictor* (RR-BLUP), BayesA e o modelo de predição bivariada (BV).

A validação e confirmação dos melhores resultados entre as diferentes metodologias empregadas, tanto em relação aos modelos de predição quanto ao tamanho e composição da população de treinamento, será feita pela técnica de validação cruzada. Essa técnica consiste na divisão da população de plantas em dois subconjuntos. A primeira parte do subconjunto (80%) é

utilizada para as estimativas e a segunda parte do subconjunto (20%) para validação. A acurácia da SG predita será medida pelo coeficiente de correlação de Pearson, calculado a partir da correlação entre os valores genéticos preditos (GBVs) e os valores genéticos observados (GVs).

4 PLANO DE TRABALHO

O cronograma de execução do plano de trabalho e os recursos e orçamento estão listados nas tabelas 1 e 2, respectivamente.

Tabela 1. Cronograma de execução do plano de trabalho

Atividade	2018				2019			
	1º	2º	3º	4º	1º	2º	3º	4º
Controle de qualidade dos dados fenotípicos	X							
Análise descritiva e ajuste dos dados fenotípicos	X							
Predição dos BLUPs para os caracteres avaliados		X						
Controle de qualidade dos dados genotípicos			X					
Modelagem do LD ¹ e análise da diversidade genética			X					
Estimação das matrizes de relacionamento genômico				X				
Estimação dos parâmetros genéticos e correlações				X				
Teste dos modelos estatísticos de predição				X	X	X		
Validação dos modelos de predição						X	X	
Redação do artigo científico e submissão								X

* Trimestres; ¹ Desequilíbrio de ligação

Tabela 2. Recursos e orçamento necessário para realização do plano de trabalho

Item	Valor
Rendimentos para recursos humanos (bolsas)	R\$ 65.000,00
Materiais de escritório (folhas, canetas, tinta para impressora, etc.)	R\$ 800,00
Aluguel de servidor para análise de dados	R\$ 750,00
Total	R\$ 66.500,00

5 FORMA DE ANÁLISE DOS RESULTADOS E RESULTADOS ESPERADOS

Com os valores preditos na SG será calculada a acurácia e herdabilidade, e com os valores da validação cruzada a correlação Pearson.

Espera-se que o presente trabalho tenha a capacidade de informar para os melhoristas a metodologia com melhor predição e resultará em otimização de mão de obra, tempo, custos, velocidade e uma maior eficiência na disponibilidade e lançamento de novos materiais genéticos.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALJANABI, S.; FORGET, L.; DOOKUN, A. An improved and rapid protocol for the isolation of polysaccharide-and polyphenol-free sugarcane DNA. **Plant Molecular Biology Reporter**, v. 17, p. 1–8, 1999.

CROSSA, J. et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961–975, 2017.

GOUY, M. et al. Experimental assessment of the accuracy of genomic selection in sugarcane. **Theoretical and Applied Genetics**, v. 126, n. 10, p. 2575–2586, 2 out. 2013.

HOANG, N. V. et al. High-Throughput Profiling of the Fiber and Sugar Composition of Sugarcane Biomass. **BioEnergy Research**, v. 10, n. 2, p. 400–416, 30 jun. 2017.

MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819–29, abr. 2001.

MOORE, P. H.; BOTHA, F. C. **Sugarcane Physiology Biochemistry & Functional Biology**. [s.l: s.n.].