



PROJETO DE PESQUISA

Modelos de seleção genômica ampla em cana de açúcar baseados em fenotipagem e genotipagem de alto desempenho

Ana Letycia Basso Garcia

José Henrique Soler Guilhen

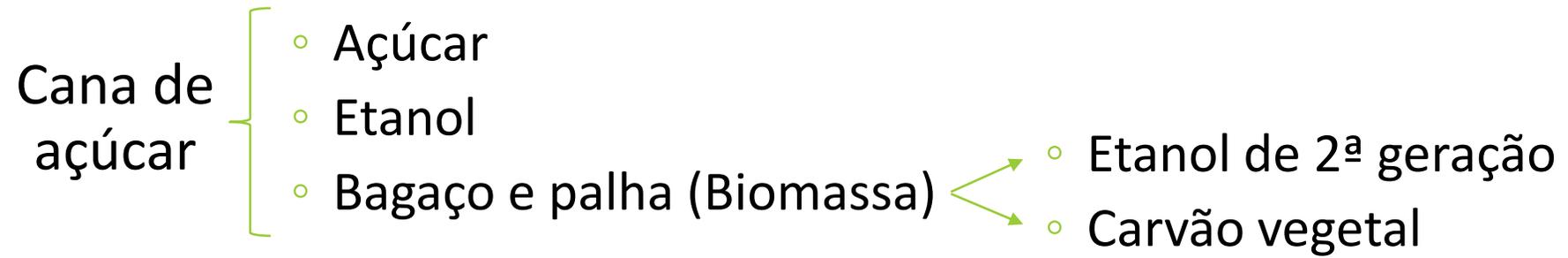
AGENDA

- Introdução
- Objetivos do trabalho
- Material e métodos
- Plano de trabalho
- Forma de avaliação dos resultados
- Considerações finais

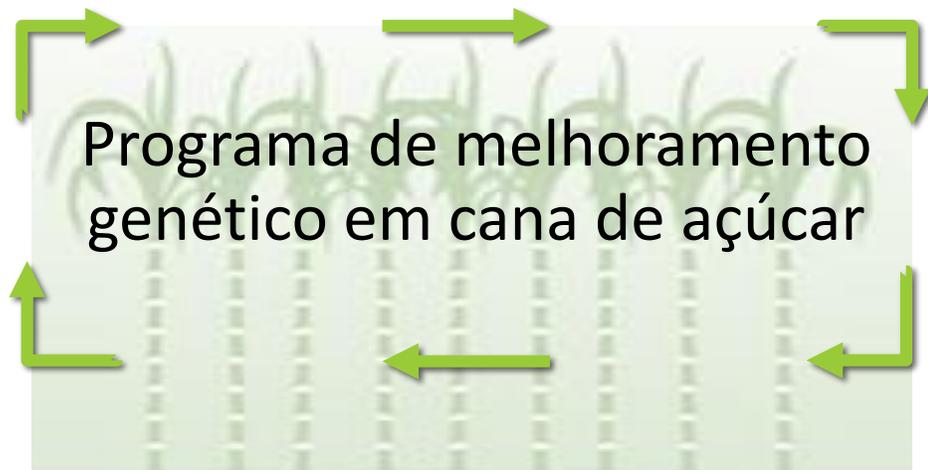
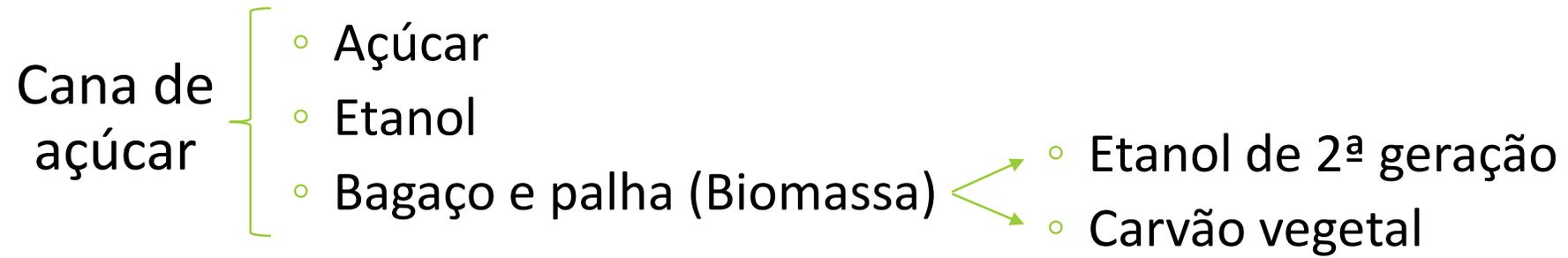
INTRODUÇÃO

- Cana de
açúcar
- Açúcar
 - Etanol
 - Bagaço e palha (Biomassa)

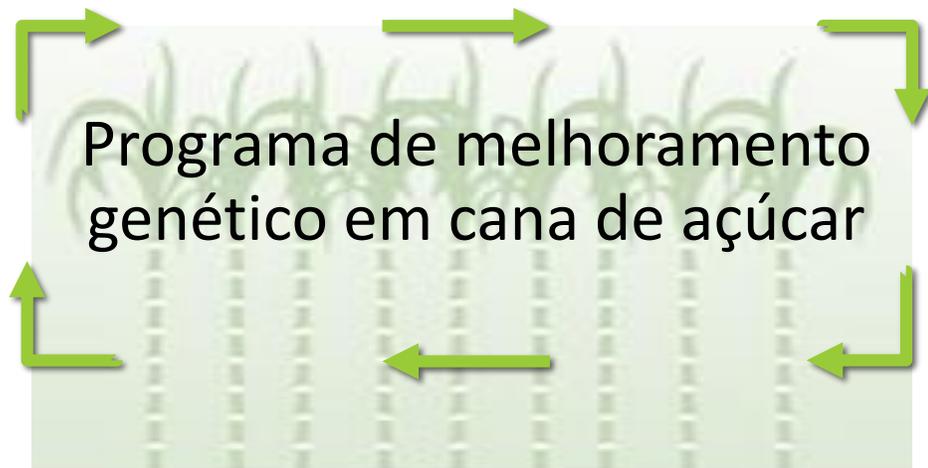
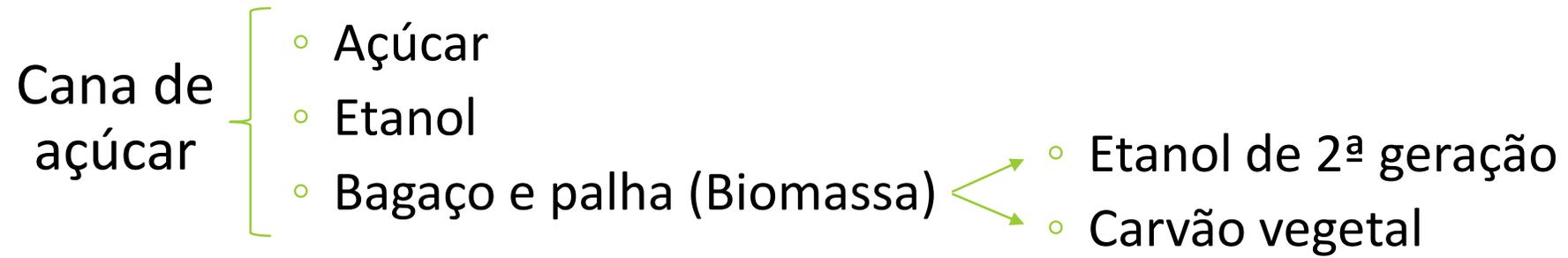
INTRODUÇÃO



INTRODUÇÃO

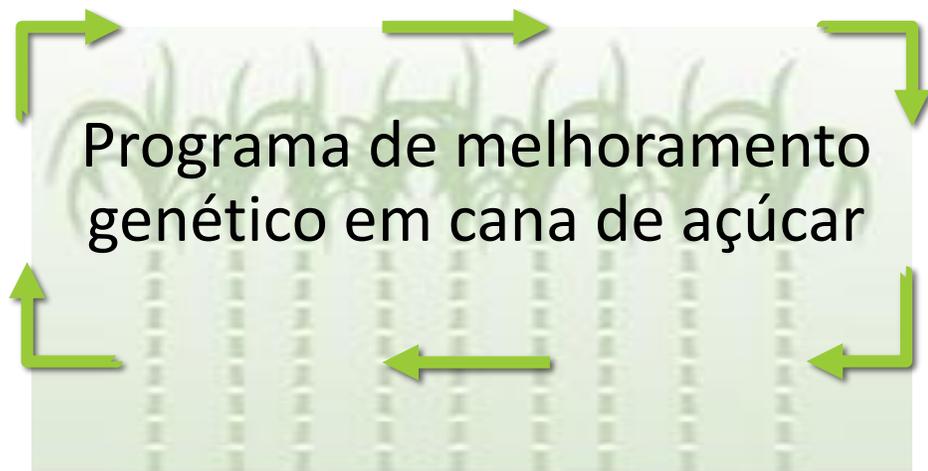
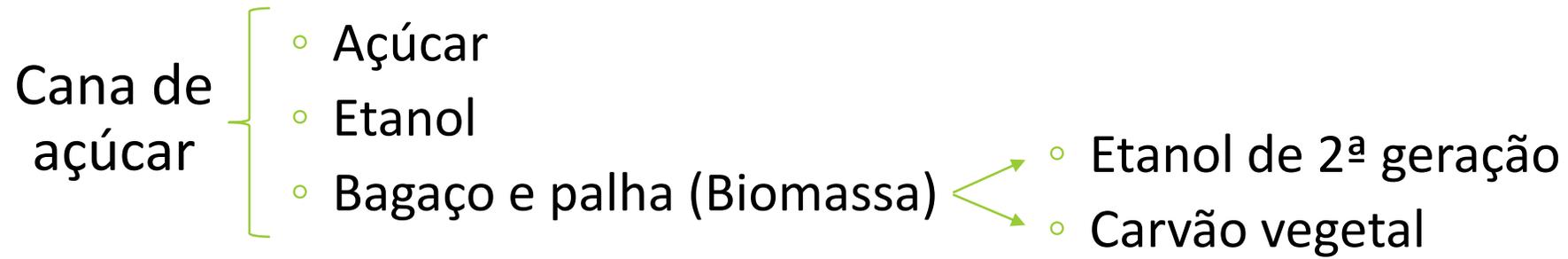


INTRODUÇÃO



7 – 10 anos

INTRODUÇÃO



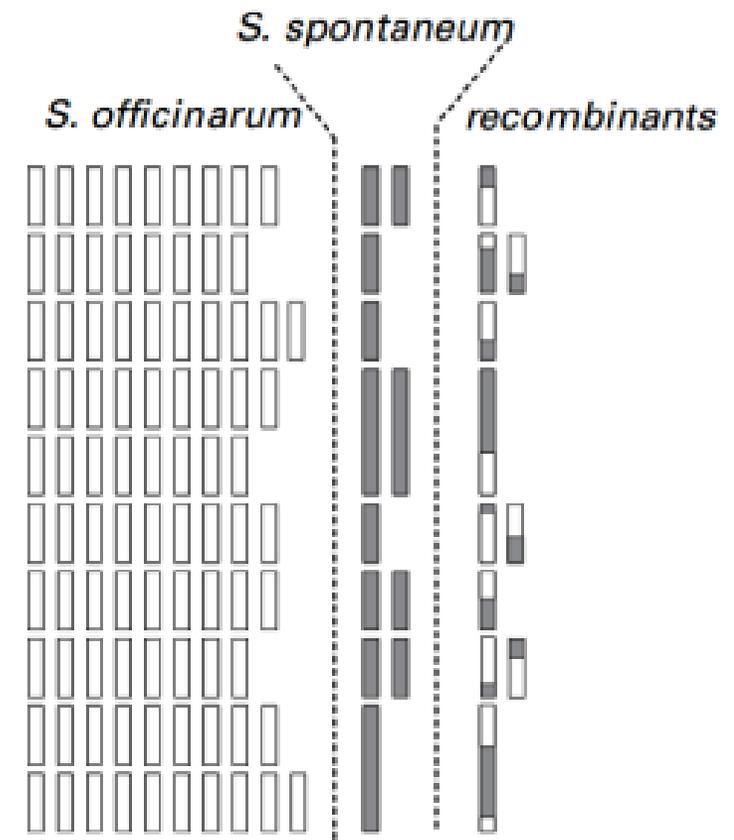
7 – 10 anos



- Alto teor de fibra
- Alto teor de celulose e hemicelulose
- Baixo teor de lignina

Genoma da Cana de açúcar

- Genoma poliploide
- Número de cromossomos é variável
- Alta heterozigozidade



(MOORE; BOTHA, 2014).

Seleção genômica



Seleção genômica



Seleção genômica



Seleção genômica



**População de
treinamento**

Estimativa dos efeitos
dos marcadores
genéticos

Seleção genômica



População de treinamento

Estimativa dos efeitos dos marcadores genéticos



População de validação

Predição dos GBVs

Seleção genômica



População de treinamento

Estimativa dos efeitos dos marcadores genéticos



População de validação

Predição dos GBVs



População de melhoramento

Seleção Genômica

Seleção genômica em
poliploides

Seleção genômica em poliploides

ORIGINAL ARTICLE

Increased genomic prediction accuracy in wheat breeding using a large Australian panel

Adam Norman^{1,3}  · Julian Taylor¹ · Emi Tanaka² · Paul Telfer^{1,3} · James Edwards^{1,3} · Jean-Pierre Martinant⁴ · Haydn Kuchel^{1,3}

Seleção genômica em poliploides

Increased genomic prediction accuracy in wheat breeding using a large Australian panel

Adam Norman^{1,3}  · Julian Taylor¹ · Emi Tanaka² · Paul Telfer^{1,3} · James Edwards^{1,3} · Jean-Pierre Martinant⁴ · Haydn Kuchel^{1,3}

Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population

Xuehui Li, Yanling Wei, Ananta Acharya, Julie L. Hansen, Jamie L. Crawford, Donald R. Viands, Réal Michaud, Annie Claessens, and E. Charles Brummer*

Seleção genômica em poliploides

ORIGINAL ARTICLE

Increased genomic prediction accuracy in wheat breeding using a large Australian panel

Adam Norman^{1,3}  · Julian Taylor¹ · Emi Tanaka² · Paul Telfer^{1,3} · James Edwards^{1,3} · Jean-Pierre Martinant⁴ · Haydn Kuchel^{1,3}

Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population

Xuehui Li, Yanling Wei, Ananta Acharya, Julie L. Hansen, Jamie L. Crawford, Donald R. Viands, Réal Michaud, Annie Claessens, and E. Charles Brummer*

Theor Appl Genet (2013) 126:2575–2586
DOI 10.1007/s00122-013-2156-z

ORIGINAL PAPER

Experimental assessment of the accuracy of genomic selection in sugarcane

M. Gouy · Y. Rousselle · D. Bastianelli · P. Lecomte · L. Bonnal · D. Roques · J.-C. Efile · S. Rocher · J. Daugrois · L. Toubi · S. Nabeneza · C. Hervouet · H. Telismart · M. Denis · A. Thong-Chane · J. C. Glaszmann · J.-Y. Hoarau · S. Nibouche · L. Costet

Acurácia da seleção genômica

h^2

Acurácia da seleção genômica

h^2

N

Acurácia da seleção genômica

h^2

LD

N

Acurácia da seleção genômica

h^2

LD

N

Densidade de
marcadores

Acurácia da seleção genômica

h^2

LD

N

Densidade de
marcadores

Precisão dos
dados utilizados

Acurácia da seleção genômica

h^2

LD

N

Densidade de
marcadores

Precisão dos
dados utilizados

Fenotipagem de alto
desempenho (HTP)

Acurácia da seleção genômica

h^2

LD

N

Densidade de
marcadores

Precisão dos
dados utilizados

Fenotipagem de alto
desempenho (HTP)

Eficiência dos
modelos estatísticos

HTP para biomassa e açúcar + SG

$$RS = \frac{i \cdot r \cdot Vg}{T}$$

HTP para biomassa e açúcar + SG

$$RS = \frac{\uparrow i \cdot r \cdot Vg}{T}$$

HTP para biomassa e açúcar + SG

$$RS = \frac{\uparrow i \cdot r \cdot Vg}{T \downarrow}$$

HTP para biomassa e açúcar + SG

$$RS = \frac{\uparrow i \cdot \uparrow r \cdot Vg}{T \downarrow}$$

Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield

Jin Sun, Jessica E. Rutkoski, Jesse A. Poland, José Crossa, Jean-Luc Jannink, and Mark E. Sorrells*

Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat

Jessica Rutkoski,^{*,†,‡,§} Jesse Poland,[§] Suchismita Mondal,[‡] Enrique Autrique,[‡] Lorena González Pérez,[‡] José Crossa,[‡] Matthew Reynolds,[‡] and Ravi Singh[‡]

^{*}International Programs, College of Agriculture and Life Sciences, and [†]Plant Breeding and Genetics Section, School of Integrated Plant Sciences, Cornell University, Ithaca, New York 14853, [‡]Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Ciudad de Mexico, 06600, Mexico, and [§]Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506

High-Throughput Phenotyping of Sorghum Plant Height Using an Unmanned Aerial Vehicle and Its Application to Genomic Prediction Modeling

Kakeru Watanabe¹, Wei Guo², Keigo Arai³, Hideki Takanashi⁴, Hiromi Kajiya-Kanegae¹, Masaaki Kobayashi⁵, Kentaro Yano⁵, Tsuyoshi Tokunaga⁶, Toru Fujiwara⁷, Nobuhiro Tsutsumi⁴ and Hiroyoshi Iwata^{1*}

Métodos de modelagem para SG

- G-BLUP
- RR-BLUP
- Modelo de predição bivariada (BV)
- Métodos bayesianos (BayesA, BayesB, BLASSO...)
- Outros...

Métodos de modelagem para SG

- G-BLUP
- RR-BLUP
- Modelo de predição bivariada (BV)
- Métodos bayesianos (BayesA, BayesB, BLASSO...)
- Outros...

Como incorporar as informações de HTP para estimar os efeitos dos marcadores?

OBJETIVOS

- Predizer 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.

OBJETIVOS

- Predizer 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.

OBJETIVOS

- Predizer 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.

MATERIAL E MÉTODOS

- População desenvolvida: Universidade de Queensland (Austrália)
- 186 genótipos, com duas ou três repetições
- 3 testemunhas
- 3 anos

Bioenerg. Res. (2017) 10:400–416
DOI [10.1007/s12155-016-9801-8](https://doi.org/10.1007/s12155-016-9801-8)



High-Throughput Profiling of the Fiber and Sugar Composition of Sugarcane Biomass

Nam V. Hoang^{1,2} · Agnelo Furtado¹ · Laura Donnan³ · Eloise C. Keeffe³ ·
Frederik C. Botha^{1,3} · Robert J. Henry^{1,4}

MATERIAL E MÉTODOS

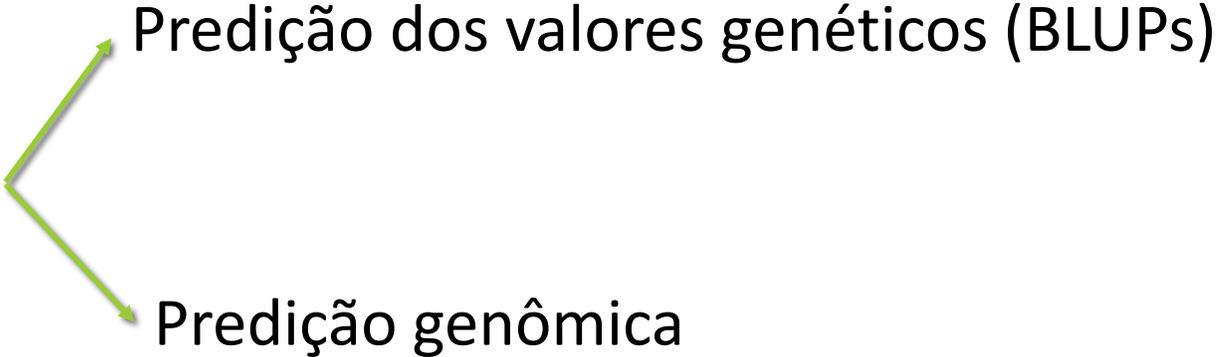
- Fenotipagem por HTP → SpectraCane (Biolab, Australia)



NIR: construção de índices vegetativos (NDVI) para estimação do conteúdo de biomassa, caracterização de fibra e teor de açúcar.

- Sequenciamento e genotipagem por GBS → marcadores SNP e DArT.

MATERIAL E MÉTODOS

- SG em duas etapas
 - Predição dos valores genéticos (BLUPs)
 - Predição genômica
- 
- ```
graph LR; A[SG em duas etapas] --> B[Predição dos valores genéticos (BLUPs)]; A --> C[Predição genômica];
```

# Predição dos valores genéticos (BLUPs)

---

- Biomassa:

$$y = Xb + Zg + Wt + Qp + \varepsilon$$

**y**: vetor de observações para biomassa;

**b**: vetor dos efeitos de bloco (fixo);

**g**: vetor de efeitos genéticos (aleatório);

**t**: vetor dos efeitos de anos (aleatório);

**p**: vetor dos efeitos da interação bloco x ano (aleatório);

**$\varepsilon$** : vetor de resíduos (aleatório)

**X**, **Z**, **W** e **Q**: matrizes de incidência dos efeitos.

## Predição dos valores genéticos (BLUPs) - Outros caracteres

---

- Modelos para HTP: **Repetibilidade Simples (SR)**, *multi-trait* (MT) e Regressão Aleatória (RR)

$$y = Xb + Zg + Wt + Sr + Qp + \varepsilon$$

em que:  $\mathbf{y}$  é o vetor de observações para as características secundárias;  $\mathbf{b}$  é o efeito para horário (fixo);  $\mathbf{g}$  é o efeito genético (aleatório);  $\mathbf{t}$  é o efeito de ano (aleatório);  $\mathbf{r}$  é o efeito de ano para repetição (aleatório);  $\mathbf{p}$  é o efeito de ano para bloco (efeito aleatório);  $\varepsilon$  é o erro residual e  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{W}$ ,  $\mathbf{S}$  e  $\mathbf{Q}$  são as matrizes de incidência de efeito.

## Predição dos valores genéticos (BLUPs) - Outros caracteres

---

- Modelos para HTP: Repetibilidade Simples (SR), **multi-trait (MT)** e Regressão Aleatória (RR)

$$y = Xb + Zg + Wt + Sr + Qp + \varepsilon$$

onde  $\mathbf{y}_i$  é o vetor de observações para a característica secundária no ano  $i$ ;  $\mathbf{X}_i$ ,  $\mathbf{Z}_i$ ,  $\mathbf{W}_i$ ,  $\mathbf{S}_i$  e  $\mathbf{Q}$  são matrizes de incidência relacionando as observações em  $\mathbf{y}_i$  com os efeitos fixos em  $\mathbf{b}_i$ , efeitos genéticos aleatórios em  $\mathbf{g}_i$ , efeitos ambientais aleatórios (teste, repetição e blocos) em  $\mathbf{t}_i$ ,  $\mathbf{r}_i$  e  $\mathbf{p}_i$ , respectivamente.

## Predição dos valores genéticos (BLUPs) - Outros caracteres

---

- Modelos para HTP: Repetibilidade Simples (SR), *multi-trait* (MT) e **Regressão Aleatória (RR)**

$$y = Xb + Zg + Wt + Sr + Qp + \varepsilon$$

em que  $\mathbf{y}$  é o vetor de observações para caracteres secundários,  $\mathbf{X}$  é a matriz de incidência e aos coeficientes de regressão,  $\mathbf{b}$  é o vetor da regressão (efeito fixo). As matrizes  $\mathbf{Z}$ ,  $\mathbf{W}$ ,  $\mathbf{S}$  e  $\mathbf{Q}$  são as matrizes das covariáveis dos efeitos e é o efeito da regressão (aleatório);  $\mathbf{g}$  é o efeito genético (aleatório);  $\mathbf{t}$  é o efeito de ano (aleatório);  $\mathbf{r}$  é o efeito de ano para repetição (aleatório);  $\mathbf{p}$  é o efeito de ano para bloco (efeito aleatório);  $\varepsilon$  é o efeito residual.

# Predição genômica

---

1. *Genomic Best Linear Unbiased Prediction (GBLUP)*
2. *Random Regression Best Linear Unbiased Predictor (RR-BLUP)*
3. BayesA
4. Modelo de Predição Bivariada (BV)

# Predição genômica

---

## 1. *Genomic Best Linear Unbiased Prediction (GBLUP)*

- Capturar e considerar a heterogeneidade de variância genética da população
- Assume variância igual entre os locos

# Predição genômica

---

## 2. *Random Regression Best Linear Unbiased Predictor (RR-BLUP)*

- Arquitetura genética homogênea<sub>ε</sub>
- Seleção indireta de covariáveis

$$\boxed{y = Xb + Wm + \varepsilon} \quad \longrightarrow \quad \begin{bmatrix} X'X & X'W \\ W'X & W'W + I \frac{\sigma_\varepsilon^2}{(\sigma_g^2/n_Q)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \end{bmatrix}$$

**y**: vetor de informações fenotípicas; **b**: vetor de efeitos fixos; **m**: vetor de efeitos aleatórios dos marcadores.

# Predição genômica

---

## 3. BayesA

- Arquitetura genética flexível
- Seleção indireta de covariáveis
- Existência de muitos marcadores de pouco efeito e poucos marcadores de grande efeito  $\rightarrow \sigma^2$  de cada marcador assume distribuição  $\chi^2$  inversa
- Os efeitos genéticos tem distribuição *a priori*  $t \rightarrow$  *shrinkage* específicos, de acordo com o tamanho do efeito do marcador

# Predição genômica

---

## 4. Modelo de Predição Bivariada (BV)

- Incluiu um traço secundário único adicional em populações de treinamento e teste.

# Validação dos métodos

---

- Validação Cruzada:

1. Estimação: 80%

2. Validação: 20%

# PLANO DE TRABALHO

## ○ Cronograma de execução

| 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 <sup>1</sup> 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; <sup>1</sup> Desequilíbrio de ligação

# PLANO DE TRABALHO

---

- **Recursos e orçamento**

---

| 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 |

---

# FORMA DE AVALIAÇÃO DOS RESULTADOS

---

- Correlação de Pearson e regressão entre os valores genéticos preditos e observados (fenotípicos) na população de validação (a partir da validação cruzada)
- Herdabilidade a ser empregada ( $h_c^2$ )
- Acurácia

$$r_{\hat{g}g} = r_{gf} / h_c$$

# CONSIDERAÇÕES FINAIS

---

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

# 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.].

The background of the image is a dense, vertical pattern of green bamboo stalks, creating a textured, natural look. The stalks are arranged in a regular, repeating pattern, filling the entire frame.

OBRIGADO!