

Seleção *in silico* de íntrons de cana-de-açúcar com potencial para melhorar o processo de edição de genoma.

Beatriz Rodrigues Estevam^a; Diego Maurício Riaño-Pachón^a

^aCentro de Energia Nuclear na Agricultura, Campus "Luiz de Queiroz", Universidade de São Paulo, Av Centenário, 303 - Piracicaba - SP
bia.estevam.25@usp.br e diego.riano@cena.usp.br

INTRODUÇÃO

ALTERNATIVA AOS COMBUSTÍVEIS FÓSSEIS

A cana-de-açúcar (*Saccharum spp.*) desempenha um papel crucial na produção global de energia de baixa emissão de carbono. Este cultivo é o **segundo mais utilizado para a produção de biocombustíveis no mundo**, e a sua maior produção destinada à geração de energia ocorre no **Brasil, que ocupa a posição de segundo maior produtor de biocombustíveis no mundo** [1].

A aplicação desse cultivo como **alternativa aos combustíveis não renováveis** pode auxiliar o desenvolvimento contínuo, sustentável e a mitigação das mudanças climáticas.

Contudo, para que se isso seja possível, ainda é **vital aumentar a eficiência do uso da energia proveniente do cultivo de cana**, visto que a venda da energia proveniente de biomassa é mais de 7 vezes menor em relação às fontes não renováveis no Brasil [2].

ALTA COMPLEXIDADE DO GENOMA DE CANA-DE-AÇÚCAR

No entanto, a alta complexidade do genoma de cana-de-açúcar [3] **apresenta desafios à edição genômica**. A cana-de-açúcar apresenta um genoma com:

- Aneuploidia
- Poliploidia
- Altas taxas de polimorfismos
- Mistura de cromossomos aneuploides e homo(eo)logos

Assim, **otimizar técnicas de edição nesse cultivo é imperativo. Para isso, propomos explorar o mecanismo IME (Intron-Mediated Enhancement)**. [4].

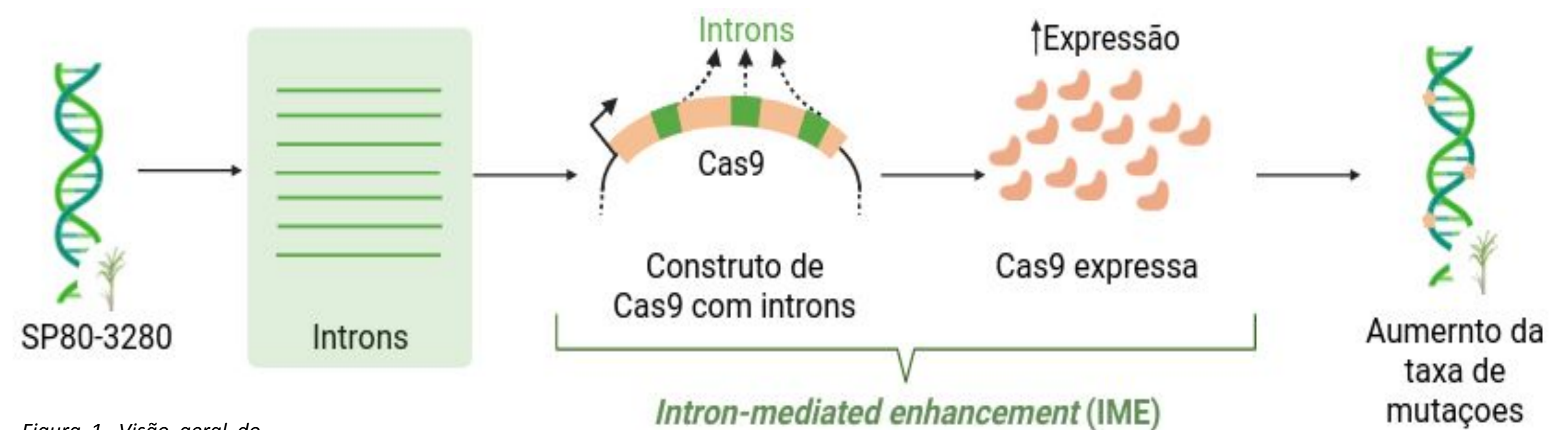


Figura 1. Visão geral do projeto.

OTIMIZAÇÃO DAS TÉCNICAS DE EDIÇÃO

Quais íntrons de Cana-de-açúcar promovem o efeito desejado?

METODOLOGIA

EFEITO DESEJADO

FERRAMENTAS

SELEÇÃO

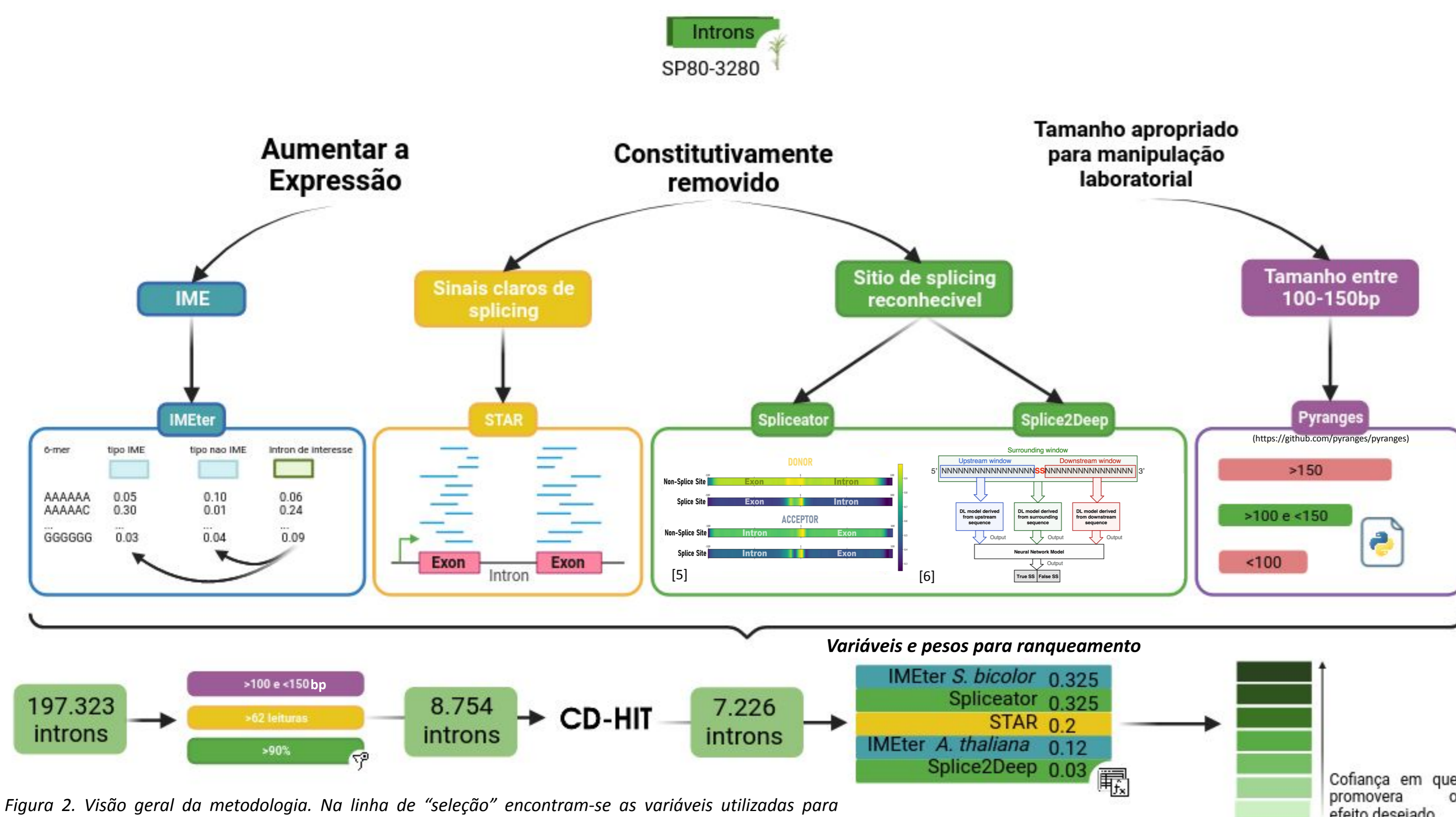


Figura 2. Visão geral da metodologia. Na linha de "seleção" encontram-se as variáveis utilizadas para ranqueamento dos íntrons e seus respectivos pesos para o ranqueamento.

Para promover o efeito desejado os íntrons selecionados devem 1) Ter **sinais claros de splicing**; 2) Ter indícios de **aumentar a expressão por IME**; 3) Serem **removidos de forma constitutiva** pelo splicing evitando prejuízos à sequência da nuclease e 4) Ter **tamanho entre 100 e 150 bp** facilitando processos de **manipulação laboratorial**.

Para isso, extraímos **197.323 íntrons** de tamanho apropriado da **variedade SP80-3280**. Então, empregamos um conjunto de ferramentas para avaliar a aptidão desses íntrons em relação às características de interesse previamente definidas.

- Capacidade de aumentar a expressão:** **IMEter** (<https://bio.tools/imeter>) nas versões V1 e V2 usando como modelo *Sorghum bicolor*, *Arabidopsis thaliana* e *Oryza sativa* (Figura 3).
- Íntrons com sítio de splicing reconhecível:** **Splice2Deep** e **Spliceator** [5, 6] (Figura 4).
- Sinais claros de splicing:** número de leituras de RNASeq mapeadas nas bordas exon-exon adjacentes igual ou superior ao valor do primeiro quartil (Figura 5) - sendo no mínimo **62 leituras**.

RESULTADOS

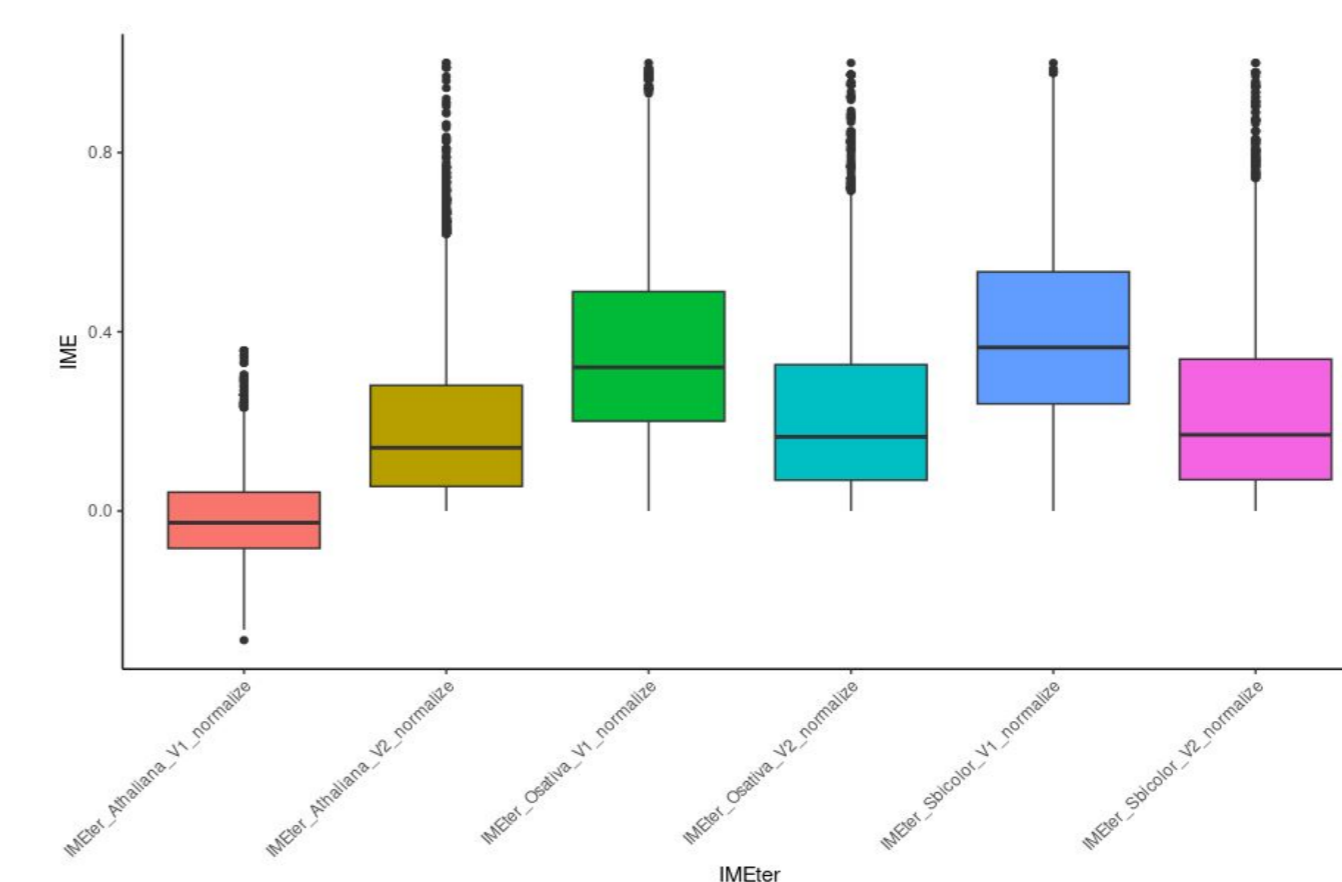


Figura 3. Distribuição da predição de promover IME dos íntrons selecionados pelas ferramentas Spliceator e Splice2Deep.

segundo as versões V1 e V2 feitas para os modelos *A. thaliana*, *S. bicolor* e *O. sativa*.

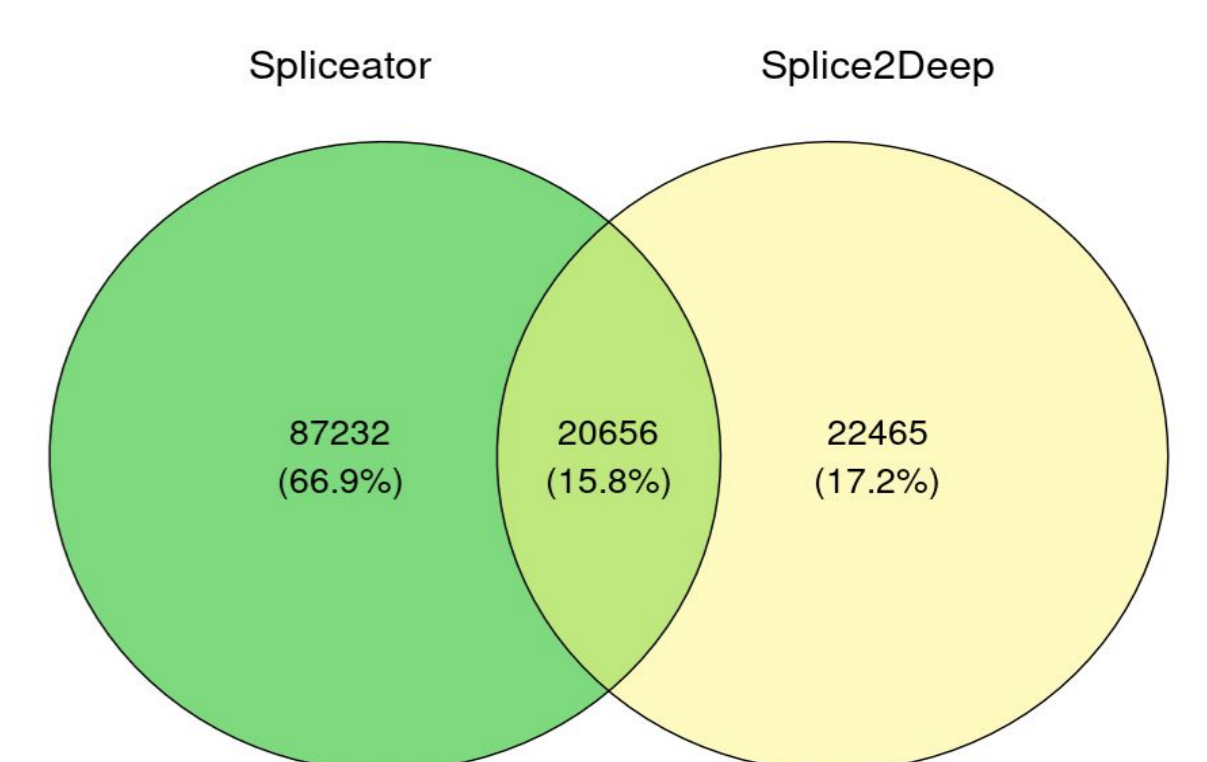


Figura 4. Sobreposição dos íntrons selecionados pelas ferramentas Spliceator e Splice2Deep.

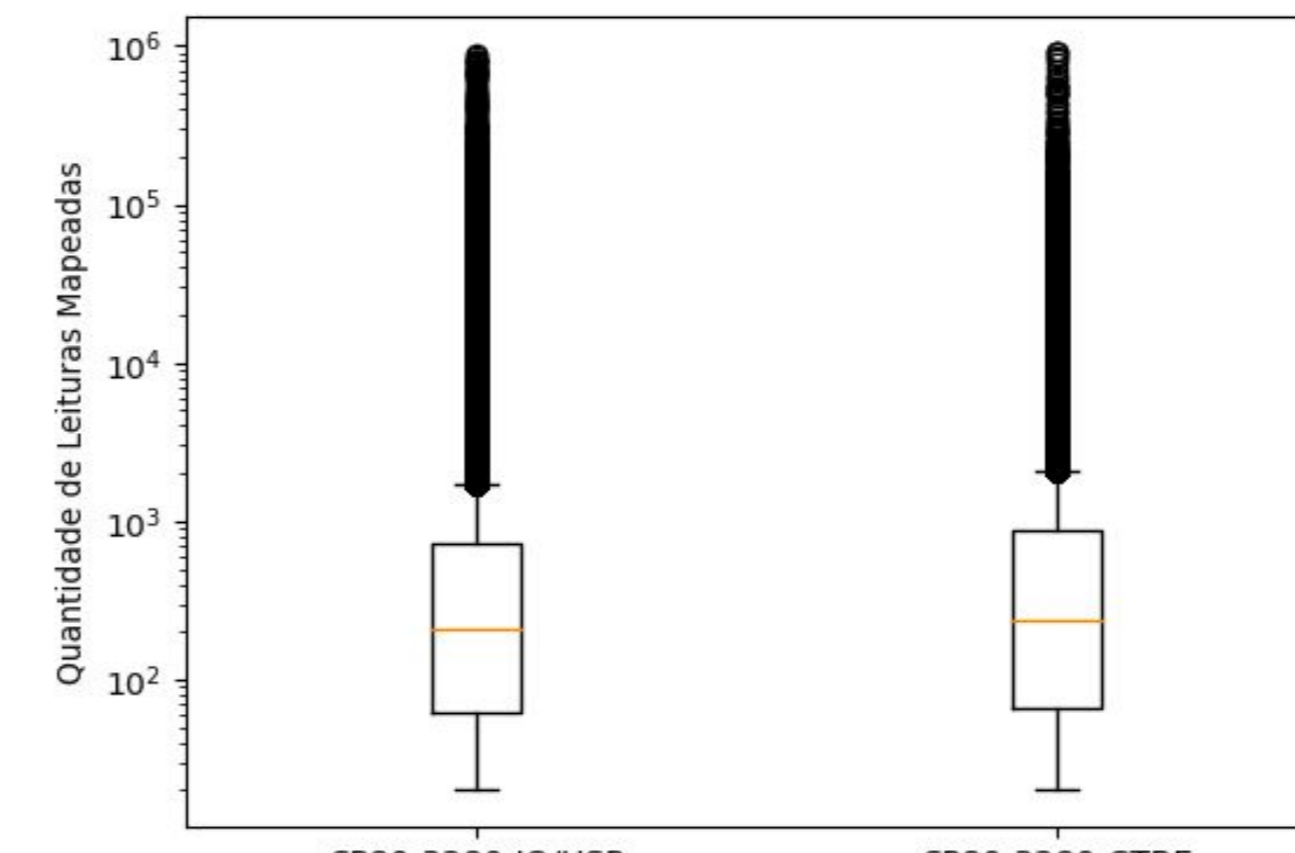


Figura 5. Gráfico de caixa e bigodes representando a quantidade de leituras mapeadas nas junções de splicing 1ª quartil para variedade CTBE igual a 62 e para variedade IQ/USP igual a 65.

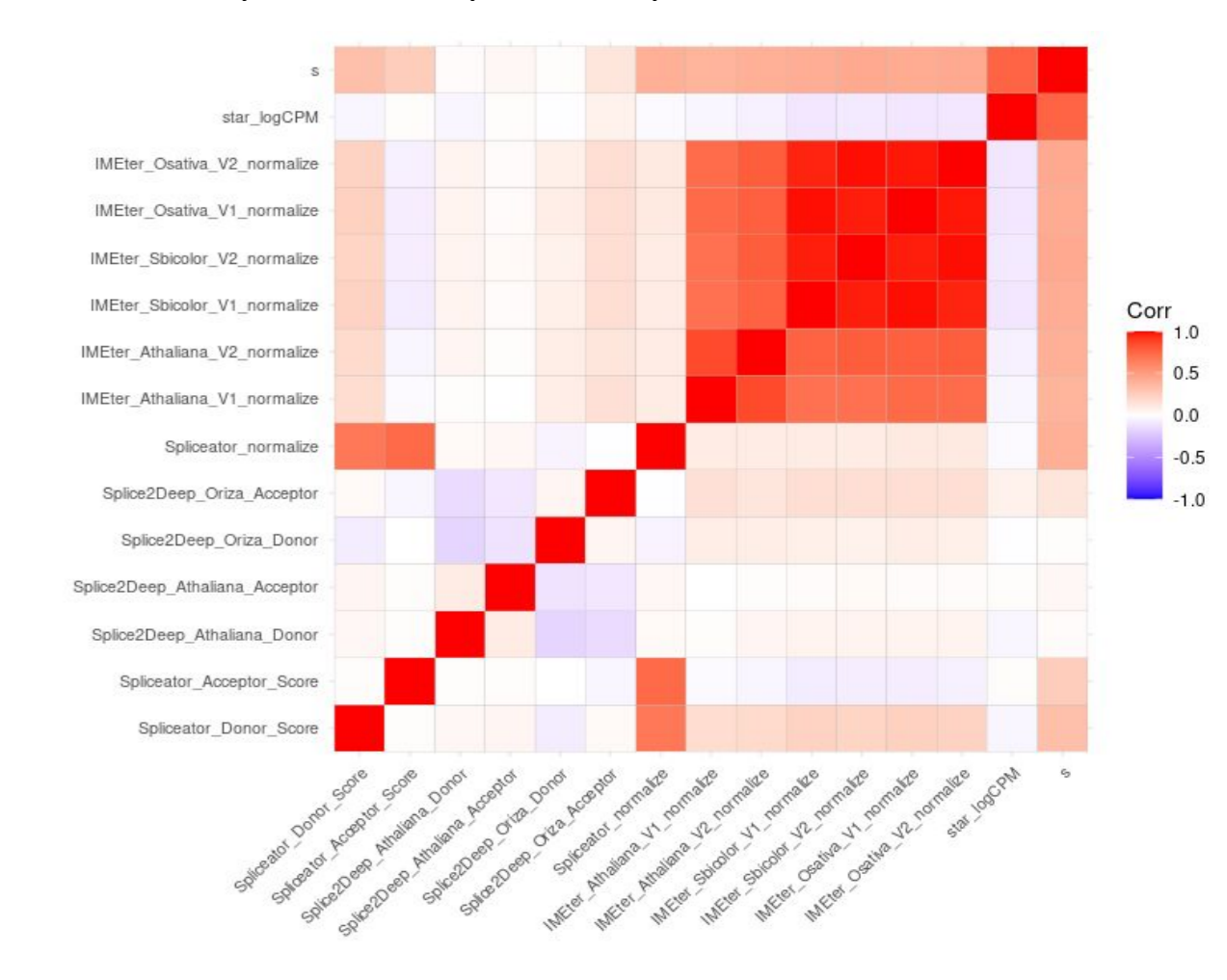


Figura 6. Correlação entre todas as variáveis calculadas e a métrica para ranqueamento dos íntrons (s). Nota-se a alta correlação entre os resultados de IMEter entre versões e organismos modelos, razão pela qual são removidos da fórmula para ranqueamento.

Posteriormente excluímos as duplicatas entre os genomas usando a ferramenta CD-HIT (<https://sites.google.com/view/cd-hit>), **somando 7.226 íntrons finais**.

Para concluir, fizemos o **ranqueamento dos íntrons usando o valor de S**, calculado com base nos valores normalizados das variáveis que possuíam menor correlação entre si (Figura 2 e 6) e que, portanto, oferecem mais informações **para diferenciar os íntrons**. Os valores de S variam de 0.22 a 1.30 (Figura 7).

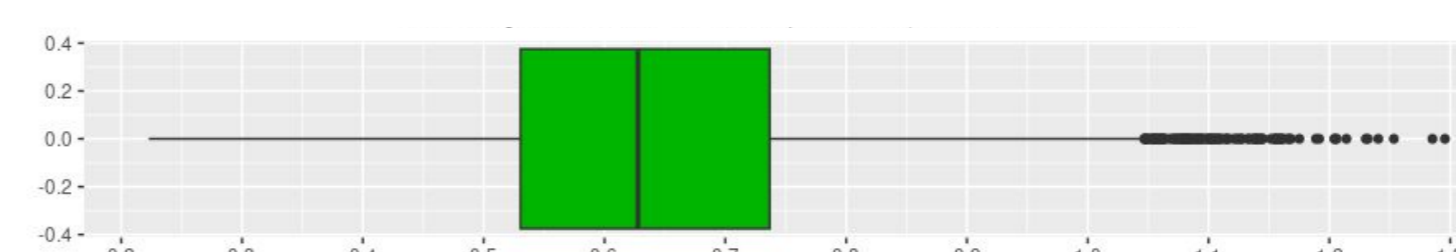


Figura 7. Distribuição do valor S - usado para ranquear os íntrons.

CONCLUSÃO E PRÓXIMOS PASSOS

Assim, obtivemos **7.226 íntrons únicos** ranqueados pela confiança de que promoverão a ampliação da expressão gênica e o splicing constitutivo simultaneamente. Com essa lista obtida, nosso próximo desafio é responder:

Onde inserir esses íntrons na sequência de Cas9?

ACESSO



labbcsc/
caneintrons

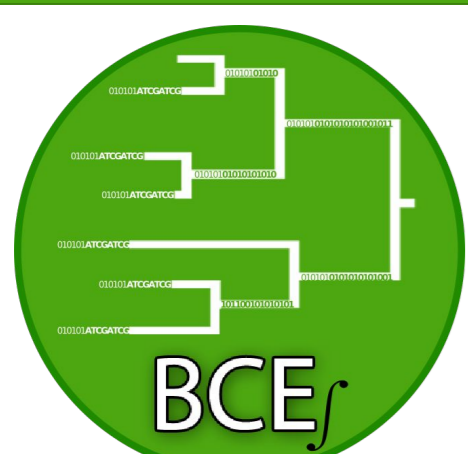
Sugarcane introns for CRISPR/Cas9

2 Contributors 0 Issues 2 Stars 0 Forks

REFERÊNCIAS

- "Biofuel energy production". Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/grapher/biofuel-production?tab=table> [Online Resource]. Acesso em: 19 julho, 2022.
- ANEEL [Agência Nacional de Energia Elétrica]. **Resultados dos Leilões de Geração no Ambiente Regulado**. 2022. Disponível em: <https://app.powerbi.com/view?e=eyJrIjoiaWYmMzN2Y0NGMYjEjYjYzZjU0ZTUwMDpSODE3IiwidCI6IjE6fjEwZDZmOWI4LWVjYjYtNDZlMjM0Mm00LWVhNGU5ZyAxNzBIMSIsImMiOiR9>
- THIRUGNANASAMBANDAM, P. P.; HOANG, N. V.; HENRY, R. J. The Challenge of Analyzing the Sugarcane Genome. *Frontiers in Plant Science*, v. 9, p. 616, 14 maio 2018.
- LAXA, M. Intron-Mediated Enhancement: A Tool for Heterologous Gene Expression in Plants? *Frontiers in Plant Science*, v. 7, 6 jan. 2017.
- SCALZITTI, N. et al. Spliceator: multi-species splice site prediction using convolutional neural networks. *BMC Bioinformatics*, v. 22, n. 1, 23 nov. 2021.
- ALBARADEI, S. et al. Splice2Deep: An ensemble of deep convolutional neural networks for improved splice site prediction in genomic DNA. *Gene*, v. 763, p. 100035, 1 dez. 2020.

FINANCIAMENTO E AGRADECIMENTOS



FAPESP 2022/10264-4
FAPESP 2023/02452-8
FAPESP RCGI 20/15230-5
CNPq 311558/2021-6
Prof. Dr. Marcelo Menossi (UNICAMP)
Prof. Dr. Dirk Walther (MPIMP)