UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

MÔNICA LETICIA TURIBIO MARTINS

“IDENTIFICAÇÃO DE GENES CANDIDATOS PARA ACÚMULO DE AÇÚCAR EM CULTIVARES DE CANA-DE-AÇÚCAR (*Saccharum* spp.) A PARTIR DE UMA REGIÃO GENÔMICA SINTÊNICA COM SORGO (*Sorghum bicolor*) E DE UMA REDE DE COEXPRESSÃO.”

“IDENTIFICATION OF CANDIDATE GENES FOR SUGAR ACCUMULATION TRAIT IN SUGARCANE (*Saccharum* spp.) CULTIVARS FROM A SYNTENIC GENOMIC REGION WITH SORGHUM (*Sorghum* bicolor) AND A CO-EXPRESSION NETWORK.”

CAMPINAS

(2024)

MÔNICA LETICIA TURIBIO MARTINS

"IDENTIFICAÇÃO DE GENES CANDIDATOS PARA ACÚMULO DE AÇÚCAR EM CULTIVARES DE CANA-DE-AÇÚCAR (*Saccharum* spp.) A PARTIR DE UMA REGIÃO GENÔMICA SINTÊNICA COM SORGO (*Sorghum bicolor*) E DE UMA REDE DE COEXPRESSÃO."

"IDENTIFICATION OF CANDIDATE GENES FOR SUGAR ACCUMULATION TRAIT IN SUGARCANE (*Saccharum* spp.) CULTIVARS FROM A SYNTENIC GENOMIC REGION WITH SORGHUM (*Sorghum* bicolor) AND A CO-EXPRESSION NETWORK."

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Doutora em Genética e Biologia Molecular, na área de Genética Vegetal e Melhoramento.

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Breeding.

Orientadora: Anete Pereira de Souza

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA MÔNICA LETICIA TURIBIO MARTINS E ORIENTADA PELA PROFESSORA ANETE PREREIRA DE SOUZA.

CAMPINAS

(2024)

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

M366i

Martins, Mônica Letícia Turíbio, 1976-
Identificação de genes candidatos para acúmulo de açúcar em cultivares de cana-de-açúcar (*Saccharum* spp.) a partir de uma região genômica sintênica com sorgo (*Sorghum bicolor*) e de uma rede de coexpressão / Mônica Leticia Turíbio Martins. – Campinas, SP : [s.n.], 2024.

Orientador: Anete Pereira de Souza.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Biologia.

1. Cana-de-açúcar. 2. Genômica. 3. Sacarose. 4. Expressão gênica. 5. Poliploide. I. Souza, Anete Pereira de, 1962-. II. Universidade Estadual de Campinas (UNICAMP). Instituto de Biologia. III. Título.

Informações Complementares

Título em outro idioma: Identification of candidate genes for sugar accumulation trait in sugarcane (*Saccharum* spp.) cultivars from a syntenic genomic region with sorghum (*Sorghum bicolor*) and a co-expression network
Palavras-chave em inglês:
Sugarcane
Genomics
Sucrose
Gene expression
Polyploidy
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutora em Genética e Biologia Molecular
Banca examinadora:
Anete Pereira de Souza [Orientador]
Luciana Rossini Pinto
Rosangela Maria Simeão Resende
Aline Crucello
Bianca Baccili Zanotto Vigna
Data de defesa: 18-06-2024
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0009-0002-0966-8042
- Currículo Lattes do autor: http://lattes.cnpq.br/8658215534453018

Campinas, 18 de junho de 2024.

COMISSÃO EXAMINADORA

Profa. Dra. Anete Pereira de Souza

Profa. Dra. Luciana Rossini Pinto

Dra. Rosângela Maria Simeão Resende

Dra. Aline Crucello

Profa. Dra. Bianca Baccili Zanotto Vigna

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Genética e Biologia Molecular da Unidade Instituto de Biologia.

Dedicatória

João Ramiro, Rosa Branca e Ana Melissa.

Foi por mim. Mas principalmente foi e é por vocês.

Dedico a vocês, minha prole.

Meu sucesso reprodutivo. Meu F1. Minha descendência, minha progênie.

Meus grandes amores.

Meus padawans.

Meu 42.

AGRADECIMENTOS

Agradeço primeiro a Exu, que nunca me deixou faltar o chão, mesmo quando parecia que eu não teria onde pisar. “Nlundi nzo ka u natanga muntu mosi ko” provérbio bantu que diz: “O teto da casa nunca é levantado por uma pessoa só.” E assim tem sido.

Por isso, agradeço a toda a minha família, sobretudo à minha mãe e ao meu irmão Roberto: o apoio, presença e confiança de vocês em mim tem sido mais que fundamental, ouso dizer que sem o suporte de vocês eu não teria conseguido. Ao meu pai, por ser o primeiro biólogo que admirei. Aos meus filhos, os três do meu ventre e meu filho do coração, meu sobrinho Robertinho. Cada passo que dou para ser um ser humano melhor é para vocês.

À Professora Anete, minha orientadora e mentora. Pela oportunidade em aprender a fazer pesquisa, por me ensinar ciência e me inspirar como pessoa. Por estar presente não apenas quando tudo correu bem, por me impulsionar também nos momentos difíceis, por acreditar em mim. Sou profundamente grata.

À Dra. Carla, por me coorientar, pelas reuniões semanais, pela presença, aprendizados, pela amizade construída e momentos de descontração. Agradeço demais Carlinha.

Ao Dr. Danilo, pelas primeiras lições sobre cana-de-açúcar e genética de poliplóides, pela presença constante ensinando, sugerindo e discutindo. E também por ser um grande amigo. O melhor.

Ao querido Ricardo, pela amizade, carinho, confiança e companheirismo. Por estar comigo durante toda a jornada, inclusive e principalmente nos momentos desafiadores. Não é todo mundo que fica. Agradeço por embarcar comigo nas ideias loucas sobre cana-de-açúcar se tornando um colaborador importante neste projeto.

Aos amigos queridos do LAGM que estão e que já foram: Prianda, Ramir, Thamiris, Yohans, Guilherme, Paulo, Isabela, Geovani, Chico, Michele e Aline; pelas conversas sobre experimentos, aprendizados, apoio, reflexões sobre a vida e tudo o mais regadas a café com bolo.

Aos colegas Luís Paulo, Melina, Alexandre e Benício por toda a contribuição, pela paciência e companheirismo. Vocês me ensinaram muito.

A todos colegas e funcionários do LAGM, foram alguns anos juntos.

Às amigas que seguraram minhas mãos e não me deixaram esquecer que eu era capaz. E riram comigo quando eu precisava rir, e dançaram comigo quando eu precisava dançar. Choraram comigo quando eu precisava chorar. As irmãs que eu escolhi. Obrigada Sabrina, Juliane e Suzana por existirem e me escolherem também. Vocês sabem que fazem parte dessa aventura.

À minha família de terreiro. Dividir uma fé nos aproxima e fortalece.

O presente trabalho foi realizado com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 - por meio do Programa Biocomputacional (Biologia Computacional) - (88887.177148/2018-00 e 88887.337878/2019-00) e de Excelência Acadêmica - PROEX 88882.329492/2019-01 (Migrated - SACPAIS e 88887.661406/2022-00) e da Fundação de Amparo à Pesquisa do Estado de São Paulo -FAPESP (2008/52197-4 bioen).

Agradeço.

RESUMO

A cana-de-açúcar (*Saccharum* spp.) é uma planta pertencente à família Poaceae, apta a climas tropicais e subtropicais, cujo centro de origem é possivelmente o sudeste asiático. As plantas cultivadas são derivadas de hibridações interespecíficas que resultaram em híbridos com genomas complexos, grandes (10 Gb), altamente poliploides e aneuplóides. Cada variedade tem características únicas devido a este perfil diferenciado. A cana-de-açúcar é responsável por 24% dos produtos e subprodutos alimentícios consumido mundialmente, ocupando uma área cultivada de aproximadamente 22 milhões de hectares. O sorgo (*Sorghum bicolor*) e a cana-de-açúcar conservam forte sintenia genômica devido a sua proximidade filogenética e o sorgo ainda possui capacidade de acumular açúcar, característica desejada pelos programas de melhoramento genético e produtores de cana-de-açúcar. Este estudo partiu de um *Quantitative Trait Locus* (QTL) para Brix previamente mapeado no genoma de sorgo, e usando a sintenia entre os dois organismos, a região ortóloga foi acessada e recuperada em duas variedades de cana-de-açúcar, a saber SP80-3280 e IACSP93-3046. A região recuperada foi analisada quanto a sua estrutura e as duas variedades foram comparadas entre si, com a sequência monoplóide disponível da variedade R570 e com uma das espécies ancestrais dos híbridos, *Saccharum spontaneum*. Para encontrar genes candidatos a acúmulo de açúcar na região-alvo foram combinadas as seguintes abordagens: análise comparativa da região genômica, levantamento dos genes encontrados e intersecção das informações genéticas e possíveis papéis biológicos das proteínas resultantes de cada gene, análise da expressão desses genes com transcriptoma dos internódios jovens e maduros, análise de genes diferencialmente expressos (DEGs) nesses tecidos e a construção de uma rede de coexpressão gênica para detectar possíveis interações genéticas dos genes da região-alvo. Tais abordagens foram escolhidas para avaliar a região dado o caráter poligênico e altamente complexo do acúmulo de açúcar e juntas formaram uma estratégia integrativa que evidenciou três genes candidatos contidos apenas nessa região: Prolil Oligopeptidase (POP), ABA 8'hidroxilase (ABA 8' OH) e Fator de Resposta ao Etileno 109 (ERF109).

Palavras-chave:

Cana-de-açúcar, Genômica, Sacarose, Expressão Gênica, Poliploidia

ABSTRACT

Sugarcane (*Saccharum* spp.) is a plant belonging to Poaceae family, suited to tropical and subtropical climates, with its center of origin possibly in Southeast Asia. Cultivated plants are derived from interspecific hybridizations that have resulted in hybrids with complex, large genomes (10 Gb), highly polyploid, and aneuploid. Each variety has unique characteristics due to this differentiated profile. Sugarcane accounts for 24% of globally consumed food and food by-products, occupying a cultivated area of approximately 22 million hectares. Sorghum (*Sorghum bicolor*) and sugarcane maintain strong genomic synteny due to their phylogenetic proximity, and sorghum still has the ability to accumulate sugar, a characteristic strongly desired by sugarcane genetic breeding programs and producers. This study started from a previously mapped Quantitative Trait Locus (QTL) for Brix in the sorghum genome, and using synteny between the two organisms, the orthologous region was accessed and retrieved in two sugarcane varieties, namely SP80-3280 and IACSP93-3046. The recovered region was analyzed for its structure, and the two varieties were compared with each other, with the available monoploid sequence of the R570 variety, and with one of the ancestral species of hybrids, Saccharum spontaneum. To find candidate genes for sugar accumulation in the target region, the following approachs were combined: comparative analysis of the genomic region, survey of the genes found, intersection of genetic information and possible biological roles of the resulting proteins from each gene, analysis of the expression of these genes with transcriptome of young and mature internodes, analysis of differentially expressed genes (DEG) in these tissues, and construction of a gene coexpression network to detect possible genetic interactions of genes in the target region. These approachs were chosen to evaluate the region given the polygenic and highly complex nature of sugar accumulation and together made one integrative strategy. This strategy identified three candidate genes contained only in this region: Prolil Oligopeptidase (POP), ABA 8'hydroxylase (ABA 8' OH), and Ethylene Response Factor 109 (ERF109).

Keywords:

Sugarcane, Genomics, Sucrose, Gene expression., Polyploid

Sumário

Revisão Bibliográfica

1- Aspectos Gerais e Cultivares Modernas

A cana-de-açúcar (*Saccharum* spp.) é uma planta semi-perene, monocotiledônea, alógama, herbácea, apta a regiões tropicais e subtropicais, cujo centro de origem é possivelmente o sudeste asiático; notadamente a Nova Guiné, Índia e China (revisado por Amalraj e Balasundaram, 2006). Pertence à família Poaceae, subfamília Panicoideae, tribo Andropogoneae, subtribo Saccharinae e ao gênero *Saccharum* L. (Thorne et al., 1992). Reconhecidamente seis espécies pertencem ao gênero *Saccharum* L.: *Saccharum officinarum* L. (2n=8x=80), *Saccharum spontaneum* L. (2n=40-128), *Saccharum robustum* J. (2n=60-205), *Saccharum barberi* J. (2n=81-124), *Saccharum sinense* R. (2n=111-120) e *Saccharum edule* H. (2n=60-80). Essas espécies formam um agrupamento botânico notável que compartilha características genéticas marcantes como a alta poliploidia (Daniels e Roach, 1987). O gênero *Saccharum*, juntamente com os gêneros *Erianthus*, *Miscanthus*, *Narenga* e *Sclerostachya* fazem parte do grupo informal conhecido como "Complexo *Saccharum*".

As variedades cultivadas são plantas híbridas que possuem como espécies parentais majoritariamente *S. officinarum* e *S. spontaneum* (Bremer et al., 1961; D'Hont et al., 1996; Garsmeur et al., 2018 e Healey et al., 2024). *S. officinarum* é uma espécie rica em açúcar, contudo suscetível a doenças e estresses abióticos. Ao cruzar com *S. spontaneum* esperava-se a introgressão de suas características favoráveis de alta capacidade de perfilhamento e rebrota, maior resistência estresses bióticos e abióticos, mas sem perder a capacidade em acumular açúcar de *S. officinarum*, o que foi alcançado com sucessivos cruzamentos com acessos de *S. officinarum* (Bremer, 1961, D'Hont et al., 1996; Garsmeur et al., 2018; Babu et al., 2022). Este processo ficou conhecido como nobilização, ou seja, nada mais que sucessivos cruzamentos do híbrido interespecífico obtido com acessos de *S. officinarum* (conhecida como "cana nobre") direcionados para a restauração do teor de sacarose e consequentemente seleção dos híbridos combinando características favoráveis de *S. spontaneum*, mas com o teor de sacarose das canas nobres e doces (Loh et al., 1951; Bhat et al., 1985 e Babu et al., 2022). Embora haja indícios que a domesticação da cana-de-açúcar (*S. officinarum*) tenha sido iniciada há milhares de anos (Moore et al., 2013), a nobilização foi iniciada nas décadas finais do século XIX, com o início do melhoramento genético da cana, dando origem às cultivares modernas da cana-de-açúcar

(Bremer et al., 1961; D'Hont et al., 1996; Garsmeur et al., 2018). E foi com esse objetivo que agricultores holandeses, em Java, desenvolveram cultivares resistentes à doença conhecida como ‘sereh’, mas também com alta produtividade em açúcar (Bremer et al., 1961).

2- Importância econômica

Segundo o Departamento de Agricultura dos Estados Unidos, 24% da produção de produtos para alimentos no mundo, incluindo os de origem animal, provém da cana-de-açúcar, e o Brasil é o maior produtor mundial de cana-de-açúcar, seguido pela Índia (USDA, 2021 - consulta em fev/2024). A extensão total mundial da área cultivada da cana-de-açúcar é de aproximadamente 22 milhões de hectares, cuja produtividade média é de 70,9 toneladas por hectare, resultando em uma produção anual estimada em torno de 1.559,8 milhões de toneladas (FAO, consulta em fev/2024 - dados de 2021). De todo o açúcar consumido mundialmente, 80% é de origem da cana-de-açúcar, sendo que Brasil e Índia são responsáveis por 61% (FAO, 2023; Luo et al., 2023).

No Brasil, é projetado que a produção da cana-de-açúcar deverá crescer na safra 2023/2024 em 4,4% em relação ao período anterior. Em 2022/2023 foram aproximadamente 638 milhões de toneladas produzidas. O Sudeste é a maior região produtora e a previsão é que na safra de 2023/2024 a produção alcance até 405 milhões de toneladas (Conab, 2023 - consulta em fev/2024). Segundo a Conab (2023), a produção de açúcar no Brasil atingirá 38,77 milhões de toneladas na safra de 2024, a segunda maior da série histórica. A produção de etanol também deve aumentar em 5,9%, totalizando 33,17 bilhões de litros. No entanto, o etanol da cana-de-açúcar terá aumento de 0,6%, alcançando 27,53 bilhões de litros, mantendo o domínio do mercado. São números expressivos que mostram que o mercado de cana-de-açúcar tende a expandir, sobretudo considerando seus maiores produtos derivados, o açúcar e o bioetanol.

Um dos principais produtos oriundos da cana-de-açúcar é o açúcar e 40% do açúcar originado da cana-de-açúcar consumido no mundo provém do Brasil. (Voora et al., 2023). Embora o mercado internacional de açúcar indique um possível déficit na produção do açúcar oriundo da cana-de-açúcar para a temporada 2023/2024, com previsão de queda na produção global para 175,5 milhões de toneladas, uma redução de 2% em relação à temporada anterior, tal perspectiva se deve às expectativas de diminuição na produção de países-chave, como Tailândia e Índia, principalmente devido a condições climáticas com previsão de fortes secas

relacionadas ao fenômeno *El Niño*. No entanto, o consumo global tem tendência a crescer, a previsão é que haja um aumento de 0,8% nesta temporada (2023/24) (FAO, 2023; consulta em fev/2024). Em países com forte avanço na indústria açucareira, esse progresso se deve em boa parte à implementação de variedades bem adaptadas juntamente com tecnologias eficientes de produção (Babu et al., 2022). Na Índia, por exemplo, sabe-se que há necessidade de aumento na produtividade a fim de atender a demanda projetada para 2050 (Babu et al., 2022).

Os impactos causados pela utilização de fontes de energia não renováveis como o petróleo, carvão e gás natural são considerados um sério problema ambiental (Tulchinsky et al., 2023). A humanidade tem buscado por fontes alternativas e renováveis que atendam a demanda mundial por fontes de energia. "Garantir o acesso a energia disponível, confiável, sustentável e moderna" é um dos Objetivos de Desenvolvimento Sustentável (SDG - Sustainable Development Goals) propostos pela Organização das Nações Unidas (ONU) para a construção de um mundo mais sustentável até 2030 (United Nations - UN, 2015). O etanol advindo da cana-de-açúcar é produzido através da fermentação dos açúcares de seus colmos (Waclawovsky et al., 2010; Ayodele et al., 2020) e embora nos últimos anos o etanol de segunda geração ganhou espaço e estudos a respeito (Manmai et al., 2020; Ayodele et al., 2020), o volume de produção ainda é baixo para o mercado brasileiro (Lorenzi et al., 2019). Além disso, o etanol proveniente da sacarose dos colmos da cana-de-açúcar continua com grande importância para abastecer as necessidades do mercado internacional. (Manmai et al., 2020; Ayodele et al., 2020) e faz necessario tecnologias aplicadas e cana-de-açúcar com qualidade que otimizem este processo (Bertrand et al., 2016; Ayodele et al., 2020).

Neste cenário, o bioetanol emerge como produto com alta relevância, e a cultura da cana-de-açúcar se destaca pelo seu forte potencial para abastecer boa parte desse mercado que se abre. No Brasil, a produção de biocombustíveis é uma necessidade do mercado, mas também está ligada à política de governo e às demandas sociais e ambientais, como diminuição da emissão de poluentes, o uso de combustíveis renováveis, a produção agrícola e a busca pela independência energética. A busca de fonte de energia renovável e contínua, que minimize problemas ambientais e atenda a necessidade da população é tendência mundial, mas pensando em termos de Brasil, é importante considerar também que a garantia de abastecimento energético é uma questão de soberania nacional e atendimento das necessidades básicas da população (revisado por Karp et al., 2021).

Para que todos os países produtores de cana-de-açúcar tenham seus interesses atendidos no mercado mundial de açúcar e bioetanol, é essencial compreender melhor a cultura como um todo, incluindo sua genômica complexa e seus genes de interesse agronômico. Isso permitirá a busca e implementação de cultivares cada vez mais aprimoradas.

3- A Cultura

A cultura da cana-de-açúcar é muito bem-sucedida no Brasil e uma justificativa para tal é que ela prospera em regiões de clima quente. A cana-de-açúcar precisa de estações de crescimento prolongadas, clima quente e luminosidade intensa. Quanto à umidade, precisa ser adequada, pois a cana-de-açúcar não apresenta um acúmulo significativo de açúcar em condições de seca intensa ou solo encharcado, com disponibilidade ótima de água em torno de 15% (FAO, 2023, consulta em fev. 2024). Para brotação e emergência das mudas é preferível que a temperatura se situe entre 32°C e 38°C; para a fase de crescimento, temperaturas diárias médias entre 22° e 30°C são ideais, sendo que uma estação longa de crescimento é essencial para altos rendimentos, variando entre 9 a 24 meses (FAO, 2021, consulta em fev/2024). A cana-de-açúcar prefere solos com mais de um metro de profundidade e que sejam bem aerados (FAO, 2023 - consulta em fev/2024), o que garante o máximo de absorção de nutrientes essenciais do solo pelas raízes. Há necessidade de pelo menos 10% de volume de ar no solo (Toppa et al., 2010). É moderadamente sensível à salinidade, de fato o estresse salino pode inclusive afetar a acumulação de açúcar. Costuma-se realizar 2 a 4 cortes subsequentes de rebrota, após o corte da emergência inicial, mas isso depende da variedade e local de plantio (FAO, 2023 - consulta em fev/2024).

A deficiência hídrica tem um efeito adverso sobre o rendimento da cultura em suas fases iniciais, da emergência das mudas até a elongação do caule, podendo atrasar o desenvolvimento das plantas (FAO, 2021, consulta em fev/2024). Por outro lado, solos com excesso de umidade ou mal drenados não são favoráveis para a cana-de-açúcar, prejudicando seu desenvolvimento desde o perfilhamento (Toppa et al., 2010). Dessa forma, é necessária baixa umidade, mas não severa, para haver formação e acúmulo de açúcar (FAO, consulta em fev/2024). É relatado que o estresse abiótico, desde que não severo, favoreça a acumulação de sacarose nos colmos, sobretudo na fase de maturação (Van der Merwe et al., 2014).

4- Aspectos gerais em fisiologia

O conhecimento sobre a fisiologia da cana-de-açúcar tem sido ampliado e se interconecta com a genômica. Para que haja sacarose acumulada nos colmos, muitos caminhos precisam ser percorridos (revisado por Meena et al., 2022). A respiração faz parte do processo de uma planta na produção de biomassa, ao utilizar carbono armazenado como substrato, fornecer carbono e equivalentes em energia para alimentar fenômenos importantes, como o transporte através do floema e a acumulação de produtos vacuolares nos internódios. Já foi observado que a diminuição do fluxo glicolítico leva a um maior acúmulo de sacarose nos colmos da cana-de-açúcar, pois a respiração concorre com a acumulação de sacarose (Van der Merwe et al., 2014). Portanto, eventos que retiram carbono da respiração supostamente poderiam ser manipulados para aumentar o acúmulo de sacarose. O processo de translocação da sacarose pelo floema, por exemplo, parece ser dependente da respiração na cana-de-açúcar. A atividade da enzima UDP-glicose desidrogenase (UDGDH) inicia uma cascata fisio-química que influencia tanto na respiração, quanto na biossíntese de componentes da parede celular, e também no transporte de açúcares pelo apoplasto do floema e divisão no armazenamento de carboidratos em órgãos sumidouro, sugerindo que há relações entre todos esses eventos biológicos (Van der Merwe et al., 2014).

A distribuição de açúcares na cana-de-açúcar requer uma dinâmica fonte-dreno muito peculiar. A folha atua como fonte de fotoassimilados e no caule estarão os tecidos de armazenamento de sacarose, este é o dreno ou sumidouro. Também faz parte do sumidouro outros tecidos como raízes, brotos e folhas jovens, frutos e sementes em desenvolvimento (Watt et al., 2014 e Singh et al., 2021). A relação fonte-sumidouro é altamente complexa, englobando muitos genes, sequências reguladoras e sinalizadores envolvidos na biossíntese, transporte e acúmulo da sacarose (Singh et al., 2021).

Outro ponto importante é a alta capacidade da cana-de-açúcar em aproveitar água, inclusive por ser uma planta que realiza a via fotossintética C4. Em outras palavras, mesmo em casos de estresse hídrico, esta planta consegue fazer a fotossíntese de maneira eficiente, mobilizando o carbono atmosférico mesmo em condições ambientais desafiadoras (Chandra et al., 2011 e Yasmeen et al., 2020). A captura inicial de CO_2 ocorre nas células do mesofilo e o carbono será fixado pela ação de enzimas específicas, entre elas a fosfoenolpiruvato (PEP), em ácido oxalacético (OAA), um composto de quatro carbonos, de onde vem o nome da via. O

OAA é convertido em ácido málico ou aspártico, e esses ácidos é que serão transportados para as células do feixe vascular. Nos cloroplastos destas células, estes ácidos serão decarboxilados liberando CO_2, e este será enfim fixado pelo ciclo de Calvin nas células do feixe vascular. O piruvato resultante do ciclo de Calvin é transportado de volta para as células do mesofilo para então ser regenerado em PEP (Figura 1). O processo de separação espacial entre a fixação inicial de CO_2 e o ciclo de Calvin ajuda a minimizar a fotorrespiração, aumentando a eficiência da fixação do carbono em condições como calor e baixa concentração de CO_2 (Sage et al., 2014 e Yasmeen et al., 2020). A taxa fotossintética da cana-de-açúcar se relaciona com o acúmulo de sacarose em seus colmos, de maneira que quanto maior a eficiência fotossintética da planta, mais sacarose acumulada nos internódios haverá (Sage et al., 2014). A cana-de-açúcar possui alta eficiência no uso de radiação, conseguindo converter até 2% da energia solar incidente em biomassa (Sage et al., 2014 e Watt et al., 2014).

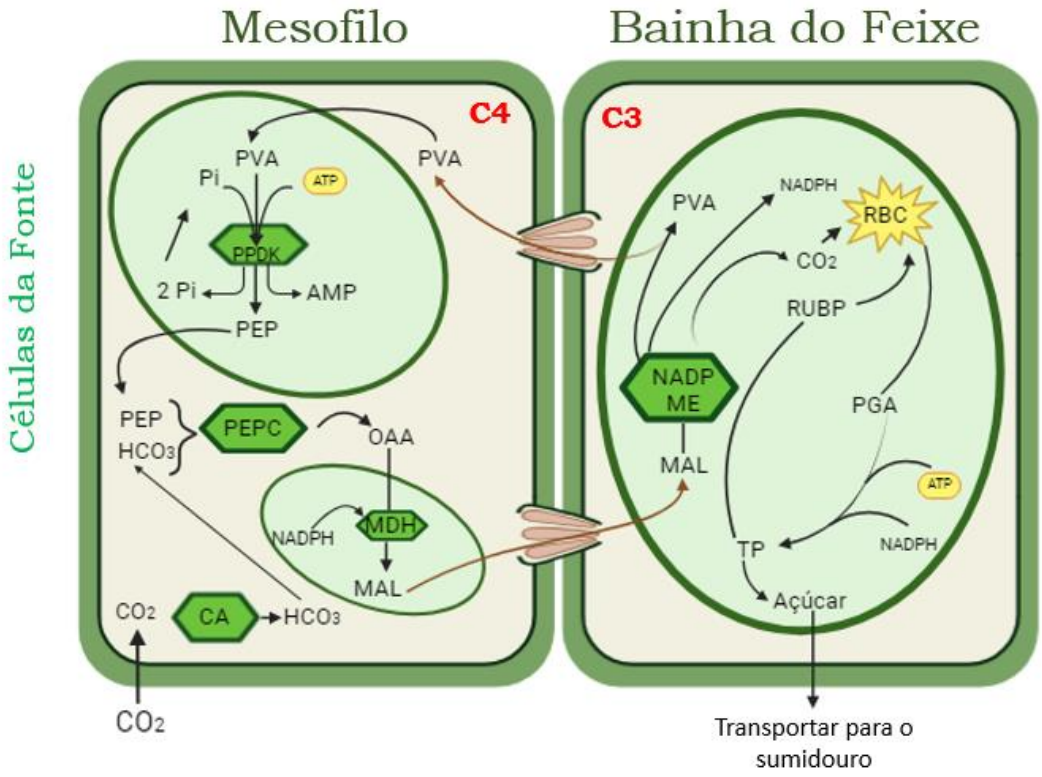


Figura 1 - Esquema simplificado da via fotossintética na cana-de-açúcar. A representação oval em verde claro é de cloroplastos dentro de células vegetais das folhas. Nas células do mesofilo: conversão de CO2 em bicarbonato (HCO3). Fixação do HCO3 pela PEPC resultando em OAA. Este será convertido em MAL pela enzima MDH. O malato será transportado através dos plasmodesmas para células da bainha do feixe. Nas células da bainha do feixe: MAL é convertido em PVA pela enzima ME. Desta reação resultará PVA, NADH e CO2. PVA será transportado e reciclado pela PPDK no mesofilo enquanto ainda nas células da bainha ocorrerá a fixação do CO2 no Ciclo de Calvin iniciando pela ação da RBC, o PGA resultado

desta reação será reduzido para TP que será usado na síntese de açúcares como a sacarose. que serão transportados para os órgãos sumidouros.

***Pi**: fósforo inorgânico*
***PVA**: piruvato*
***AMP**: adenosina monofosfato*
***PPDK:** fosfoenol-piruvato desidrogenase quinase*
***PEP**: fosfoenolpiruvato*
***HCO3**: bicarbonato*
***CA**: anidrase carbônica*
***PEPC**: fosfoenolpiruvato carboxilase*
***OAA:** oxalacetato*
***MDH**: malato desidrogenase*
***MAL**: malato*
***ME:** enzima málica*
***RBC:** rubisco*
***PGA:** PEP carboxilase*
***TP:** triose fosfato*

A quantidade de sacarose estocada nos colmos da cana-de-açúcar pode ser desencadeada pela capacidade de produção da fonte e pelo aumento da utilização de fotoassimilados pelo sumidouro (Singh et al., 2021). Estudos sugerem que o controle da dinâmica fonte-dreno é direcionado sobretudo pela força do dreno (McCormick et al., 2006; Chandra et al. 2015 e Singh et al., 2021). Em tal dinâmica, quando há alta concentração de açúcares dentro das células do floema de órgãos-fonte, é criado um gradiente de potencial hídrico que atrai água para dentro das células do floema, aumentando a pressão de turgor. Para diminuir a pressão de turgor, esses açúcares são descarregados em órgãos-dreno ou sumidouros juntamente com água, o que resulta na diminuição do turgor do floema. A diferença do gradiente do turgor entre fonte e dreno é o que impulsiona o fluxo de açúcares das fontes para os sumidouros através dos condutos interconectados de floema (Watt et al., 2014). A distribuição dos açúcares requer o uso de transportadores de açúcar para administrar a carga e descarga da sacarose no tecido do floema. Os transportadores influenciam a capacidade de sobrevivência da cana-de-açúcar ao impacto de estresses bióticos e abióticos ao interagirem com enzimas e hormônios que contribuem com o balanço osmótico dos órgãos-fonte e órgãos-dreno (Misra et al., 2023).

O controle da transferência de metabólitos pode ser regulado por alterações na demanda dos sumidouros, resultando em alterações globais na expressão gênica e na atividade fotossintética das folhas de cana-de-açúcar. Existem muitos componentes de sinalização implicados na detecção de açúcares, alguns conhecidos como a trealose 6-fosfato que se integram com o metabolismo (Watt et al., 2014). Alguns destes componentes de sinalização podem eventualmente limitar a foto-assimilação e a acumulação de sacarose em colmos de cana-de-açúcar (Sage et al., 2014 e Watt et al., 2014). A obtenção de alto teor de sacarose na

cana-de-açúcar envolve processos fisiológicos que incluem a produção da sacarose na fonte, o transporte para o dreno e a acumulação nos órgãos-dreno. Neste caminho há interação com hormônios, enzimas, transportadores de açúcar e outras moléculas (Misra et al., 2023).

O transporte de água, açúcares, aminoácidos e íons minerais está intimamente ligado dentro da planta, sendo conduzido pelo xilema distribuindo água e minerais principalmente e floema transportando os açúcares. No xilema, a água é impulsionada pela tensão negativa gerada pela transpiração ativa nas folhas. Enquanto isso, o transporte pelo floema cria um gradiente osmótico e hidrostático essencial para que nutrientes e assimilados possam ser transportados nas plantas. Xilema e floema são condutores nas plantas através de tensão negativa e pressão de turgor, respectivamente (Grantz, 2014). O transporte da sacarose ocorre de duas maneiras, pelo movimento simplástico, via plasmodesmos e movimento apoplástico, que requer transportadores específicos. Nos colmos mais maduros da planta em tecidos de armazenamento, simplasto e apoplasto apresentam concentração insignificante de sacarose, pois a grande parte dela já terá sido transportada do citoplasma para o vacúolo, em um processo de transporte ativo (Misra et al., 2023).

5- Complexidade Genômica

O decorrer do desenvolvimento das cultivares modernas levou à formação de híbridos com genomas de alta complexidade. Tal genoma possui características notáveis: tamanho de aproximadamente 10 Gb, altamente poliploide e aneuploide. Estima-se que pelo menos metade do genoma seja composto de elementos repetitivos, principalmente em elementos transponíveis (Cuadrado, 2004; Piperidis et al., 2010; Garcia et al., 2013; Thirugnanasambandam et al., 2018; Healey et al., 2024). Cada híbrido possui uma mistura de cromossomos oriundos tanto de *S. officinarum* (70-80% do número total de todos os cromossomos) e *S. spontaneum* (10-20% também de todos os cromossomos), além de cromossomos recombinantes entre estas duas espécies (5-10% de todos os cromossomos) (Figura 2). Apesar de os números de cromossomos em cada conjunto cromossômico poder variar de um genótipo para outro, as progênies são férteis (D'Hont et al., 1996; Cuadrado, 2004; Piperidis et al., 2010; Garsmeur et al., 2018; Healey et al., 2024).

Figura 2 (adaptada de Garsmeur et al., 2018) - Representação esquemática do cariótipo da cultivar moderna R570 mostrando em amarelo os cromossomos oriundos de S. officinarum estimados em 70-80%, em laranja os cromossomos oriundos de S. spontaneum estimados em 10-20%, e os recombinantes estimados em 5-10%, demonstrados pelos cromossomos em amarelo e laranja.

Tal complexidade genômica tem dificultado muito as pesquisas a respeito da cana-de-açúcar, ainda assim os avanços foram feitos e são bem estabelecidos. Já se conhece o tamanho estimado do genoma (Garsmeur et al., 2018; Wang et al., 2022 e Healey et al., 2024; Bao et al., 2024), o caminho evolutivo que levou a poliploidização de uma das espécies ancestrais dos híbridos, *S. spontaneum* (Zhang et al., 2018), a composição dos centrômeros (Wang et al., 2022), bibliotecas de BAC de variedades híbridas (Sforça et al., 2019), e muitas outras características biológicas essenciais (revisado por Wang et al., 2022).

Devido às características altamente desafiadoras do genoma da cana-de-açúcar, poderia se considerar o uso de genomas de espécies evolutivamente próximas como referência. Genomas de outras gramíneas, como o arroz (*Oryza sativa* L.) (International Rice Genome Sequencing Project e Sasaki, 2005), milho (*Zea mays* L.) (Schnable et al., 2009), trigo (*Triticum aestivum* L.) (IWGSC et al., 2018), miscanto (*Miscanthus sinensis* A.) (Kim et al., 2014;

Tsuruta et al., 2017) e sorgo (*Sorghum bicolor L.*) (McCormick et al., 2018) podem ser considerados para estudar o genoma desafiador da cana-de-açúcar. Além disso, estão disponíveis o genoma com quatro alelos definidos da *S. spontaneum* (Zhang et al., 2018) e a sequência monoplóide de referência da variedade R570 (Garsmeur et al., 2018). Recentemente, um genoma da variedade R570 foi apresentado com definição de alelos (Healey et al., 2024) e um genoma a nível cromossômico da variedade ZZ1 (Bao et al., 2024).

Ao retomar os genomas disponíveis de gramíneas que podem ser boas referências para cana-de-açúcar, pode-se concluir qual é a melhor referência. Começando pelo arroz, atualmente a maior parte desta planta cultivada é diploide (Chen et al., 2021) e possui um genoma relativamente pequeno, estimado em aproximadamente 389 milhões de pares de bases (Mb) e bem mapeado. No entanto, essa espécie divergiu de um ancestral comum do sorgo e do milho há cerca de 50 milhões de anos (International Rice Genome Sequencing Project e Sasaki, 2005, Meyer et al., 2013; Gallaher et al., 2022). O trigo é hexaplóide, com genoma estimado em aproximadamente 17 Gb, composto por três genomas intimamente relacionados e mantidos de certa forma, independentes, graças a eventos de hibridação que ocorreram naturalmente (International Wheat Genome Sequencing Consortium). Além de possuir um genoma desafiador por si só, esta espécie está ainda mais distante filogeneticamente da cana-de-açúcar que o arroz (Meyer et al., 2013; Gallaher et al., 2022). O milho é diplóide e tem o genoma pequeno, com cerca de 2.365 Mb (Haberer et al., 2005) e é mais próximo da cana-de-açúcar que o arroz em termos evolutivos. Todas essas características poderiam tornar o milho uma referência interessante para estudar a genômica da cana-de-açúcar; no entanto, o milho está em outro clado da tribo Andropogonae (Meyer et al., 2013; Gallaher et al., 2022), e por isso esta planta pode não ser a mais adequada a ser usada como modelo para a cana-de-açúcar.

Os genomas das espécies do gênero *Miscanthus* têm sido amplamente estudados. Dentre as espécies do gênero há aquelas com genomas diplóide, triplóide e tetraplóide, e que hibridizam com muita facilidade (Nishiwaki et al., 2011; Moon et al., 2013; Mitros et al., 2020; Zhang et al., 2021). Evidências sugerem que sorgo (*S. bicolor* L.), cana-de-açúcar e miscanto (*Miscanthus sinensis* A.) tem um ancestral em comum (Sheng et al., 2017; Zhang et al., 2018), sendo o sorgo filogeneticamente mais próximo da cana-de-açúcar (Sheng et al., 2017; Zhang et al., 2018; Gallaher et al, 2022) do que o miscanto. O genoma do sorgo é diplóide (2n=20, x=10), com aproximadamente 730 Mb, está sequenciado, montado e anotado (Paterson *et al.* 2009), disponível no Phytozome v13. A proximidade filogenética do sorgo e da cana-de-açúcar (Figura 2A e Figura 2B) pode explicar a alta colinearidade e sintenia entre ambos (Garsmeur et al.,

2017; Zhang et al., 2018; Thirugnanasambandam et al, 2018; Babu et al., 2022; Healey et al., 2024). Além disso, o genoma do sorgo possui revisões das anotações e sequenciamento, diferentemente dos genomas de *Saccharum,* que estão em suas primeiras versões. (Goodstein et al., 2012; McCormick et al., 2018; Garsmeur et al., 2018, Healey et al., 2024, Bao et al., 2024). Vale ressaltar que o genoma do miscanto teve sorgo como referência importante para sua montagem e anotação (Mitros et al., 2020). Assim, o genoma do sorgo tem sido usado com sucesso e segue sendo uma excelente referência para o complexo e desafiador genoma da cana-de-açúcar (Thirugnanasambandam et al, 2018; Wang et al., 2021; Babu et al., 2022), resultando em diversos estudos sobre a genômica e genética de *Saccharum* spp. (Mancini et al., 2018; Sforça et al., 2019; Garsmeur et al., 2018; Zhang et al., 2018, Zhang et al., 2022).Vale ressaltar que o recente genoma da R570 constitui uma referência de estudos, ainda que o conhecimento de que os híbridos possuem ploidias variáveis e dependendo do seu *pedigree* podem ter diferenças genético-genômicas muito significativas (Healey et al., 2024). Ainda assim, há evidências de que todos possuem sintenia com sorgo (Thirugnanasambandam et al, 2018; Wang et al., 2021; Babu et al., 2022).

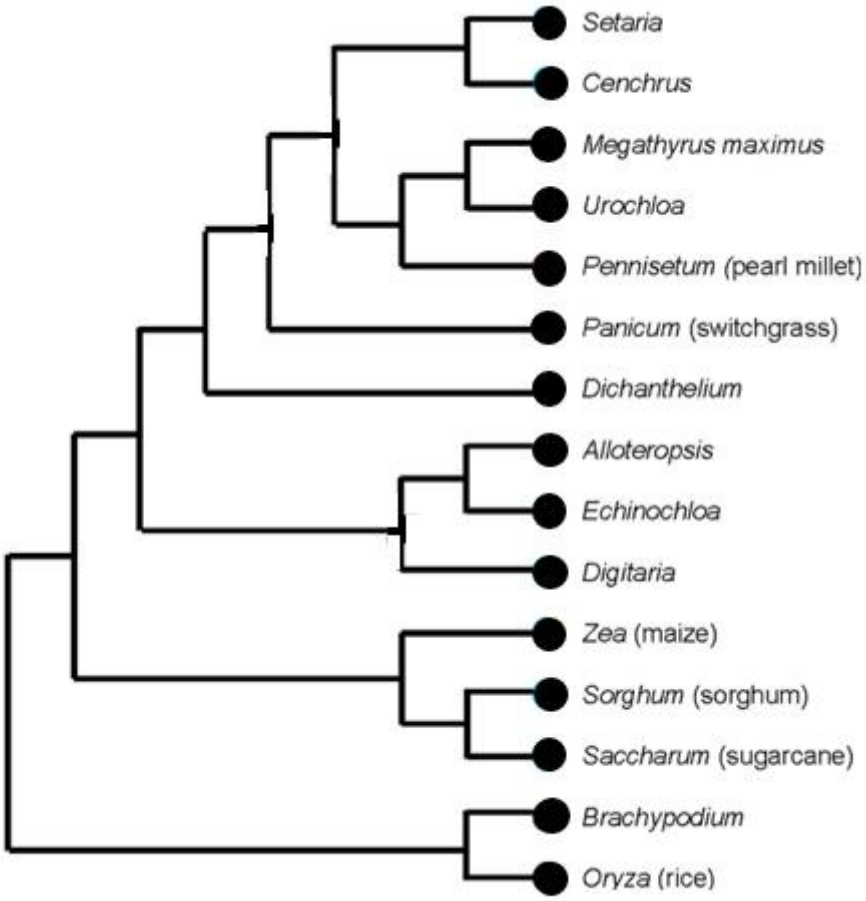


Figura 2.A (adaptado de Rao et al., 2016). Árvore filogenética simplificada das gramíneas demonstrando a proximidade do clado Saccharum e Sorghum, com um ancestral em comum, porém vindo do clado que também originou o milho, enquanto outras gramíneas estão em clados mais distantes como arroz.

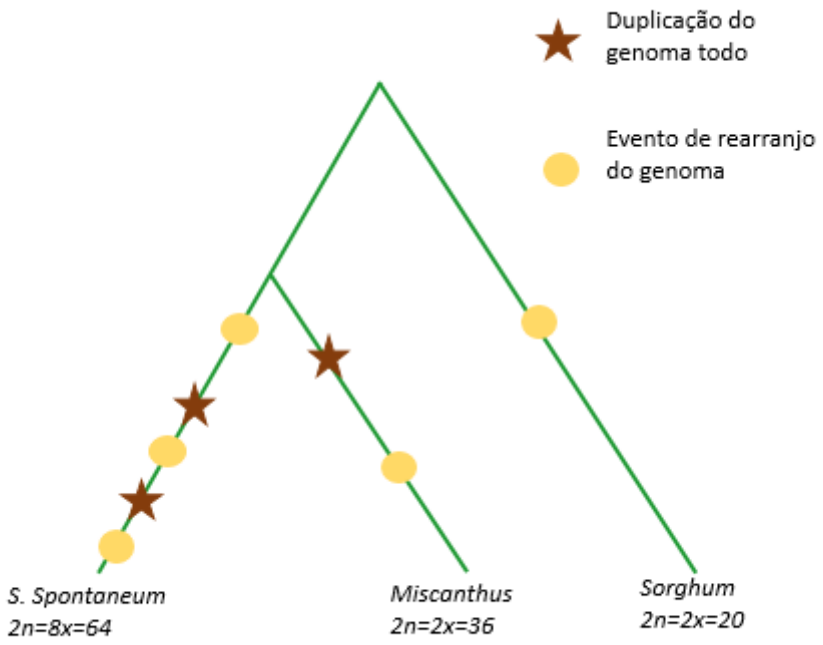


Figura 2.B (adaptado de Zhang et al., 2018). Esquema simplificado da evolução dos gêneros Miscanthus e Sorghum, e Saccharum spontaneum; mostrando as duplicações do genoma todo e eventos de rearranjo do genoma.

6- Quantitative trait loci (QTL)

Marcadores moleculares são sequências específicas de DNA com possibilidade de fornecer informações diretas sobre a diversidade genética (revisado pela Sociedade Brasileira de Genética, 2017). Em plantas, marcadores desempenham um papel fundamental para pesquisa básica e melhoramento vegetal genético sendo empregados na caracterização de germoplasma, no monitoramento da introgressão de alelos de interesse, no isolamento e busca de genes, e como ferramenta auxiliar na preservação da variabilidade genética (revisado por Andersen & Lübberstedt, 2003).

A herança de características fenotípicas que exibem uma variação contínua (como altura, peso, produtividade e acúmulo de açúcar) é controlada por uma interação complexa de múltiplos genes na qual o efeito individual do gene sobre a característica é pequeno. Tal herança é chamada de herança poligênica. As regiões específicas do genoma onde se localizam os genes

que potencialmente influenciam esses traços são denominadas Loci de Características Quantitativas, em inglês *Quantitative Trait Loci* (QTL) (Lander e Botstein, 1989).

Marcadores moleculares são essenciais para localizar QTLs por possuírem variações associadas a diferenças em traços fenotípicos (Lander e Botstein, 1989). A técnica de mapeamento de QTLs envolve a associação de variações fenotípicas com marcadores moleculares, permitindo a identificação de regiões genômicas responsáveis por características quantitativas (Lander e Botstein, 1989) e seus efeitos podem ser aditivos, de dominância e epistáticos, além de serem influenciados pelo ambiente (Alard, 1971). Para avaliar se a distância genética presumida é próxima da distância real entre os genes é utilizada a análise de escore do *Logarithm of Odds* (LOD *score*), que é empregado para determinar a probabilidade da associação entre os genes ser ou não ao acaso. Um *LOD score* alto indica uma maior evidência de ligação genética entre os marcadores analisados, sugerindo uma possível proximidade da distância genética (van Ooijen et al., 1999).

Existem diversos estudos de mapeamento de QTL em cana-de-açúcar; entretanto, o complexo genoma da planta contribui para que seus mapas genéticos disponíveis sejam pouco saturados (Sills et al., 1995; Ming et al., 2001; Aitken et al., 2008; Costa et al., 2016; Zhang et al., 2023). A maioria dos marcadores moleculares usados nestes estudos são co-dominantes. Contudo, devido à poliploidia combinada à aneuploidia da cana-de-açúcar, eles são interpretados como dominantes, sem a possibilidade de distinguir a condição homozigota da condição heterozigota, resultando na perda de muitas informações (Aitken et al., 2022). A complexidade na determinação alélica em um genoma onde cada conjunto cromossômico que geralmente pode abrigar de 8 a 12 cromossomos (Garcia et al., 2013), somada às dificuldades na fenotipagem precisa, à necessidade de populações com tamanho adequado e aos desafios dos modelos de mapeamento para um organismo altamente poliplóide tornam a utilização de QTLs na busca de genes candidatos para uma determinada característica muito desafiadora. Em estudo com a variedade R570, que resultou em um genoma, foi inferido que ao menos metade dos alelos provenientes de *S. officinarum*, de grande interesse por trazerem o aspecto do acúmulo de açúcar, são possivelmente idênticos devido à alta taxa de endogamia da variedade, consequentemente fora do alcance de estudos com mapeamento de QTL (Healey et al., 2024). Nesse cenário, o uso de um organismo com genoma mais acessível se mostra promissor.

7- Bacterial Artificial Chromosome (BAC)

BACs na verdade não são cromossomos, são plasmídeos, moléculas de DNA circulares extracromossomais que se replicam autonomamente em células bacterianas. Receberam este nome por serem capazes de carregar um grande inserto de DNA de um organismo de interesse inserido por transformação no plasmídeo (Shizuya et al., 1992).

BACs permitem a clonagem de fragmentos de DNA de até 300 kb (Shizuya et al. 1992, Zhang et al. 1996). É uma tecnologia muito utilizada em estudos de genômica por apresentar estabilidade do inserto no hospedeiro bacteriano, baixo quimerismo e facilidade de manipulação do DNA BAC (Shizuia et al. 1992, Paterson et al. 2000, Yuksel e Paterson, 2005). A primeira biblioteca de BACs para cana-de-açúcar foi construída em 1999 usando o DNA da variedade R570 e representa em torno de 1,3 vezes o seu genoma total (Tomkins et al. 1999). Figueira et al. (2012) construíram uma biblioteca de BACs de domínio privado para a variedade SP80-3280, representando 0,46 vezes a cobertura do seu genoma. No Brasil foram construídas duas bibliotecas de BACs para as variedades brasileiras SP80-3280 e IAC93-3046 que possuem respectivamente a cobertura de 2,4 e 1,8 vezes o genoma estimado da cana-de-açúcar (Sforça et al. 2019). Estas bibliotecas foram desenvolvidas e estão acondicionadas, disponíveis para o uso em pesquisas, no Laboratório de Análises Genéticas e Moleculares (LAGM) - Unicamp (Universidade Estadual de Campinas).

8- RNA-seq, transcriptômica e análise de expressão

A transcriptômica é anterior às técnicas de NGS (*Next-Generation Sequencing*), mas as técnicas usadas anteriormente como, por exemplo, bibliotecas de EST (*expressed sequence tags*) - coleções de sequências correspondentes a transcritos expressos em um determinado tecido/órgão; e microarranjos - usados para detectar a presença e quantidade de sequências específicas em uma amostra - possuem algumas limitações. Dentre elas, destacam-se o alto custo, ruído elevado nos dados, baixa cobertura e eventualmente exigência de muito conhecimento prévio da genética/genômica da espécie, além do baixo rendimento. Embora estas técnicas continuem sendo usadas, o NGS permite uma análise global de milhares de transcritos simultaneamente, superando muitas das limitações das técnicas anteriores com maior eficiência e velocidade (revisado por Mahood et al., 2020).

O sequenciamento de nova geração (*next generation sequencing* - NGS) revolucionou a genômica, transcriptômica e epigenômica com a possibilidade de sequenciar de maneira mais rápida e eficiente grandes quantidades de material genético (revisado por Kim et al., 2016). O sequenciamento de RNA (RNA-Seq) é uma técnica que gera conjuntos de dados transcriptômicos e permite a identificação e quantificação de transcritos, que podem ser usados para realizar análises de coexpressão e identificar genes diferencialmente expressos (Movahedi et al., 2012). O conhecimento de quais transcritos podem ser encontrados em determinados momento e tecido/órgão de um organismo abre uma janela de conhecimentos que tem sido amplamente utilizada pela ciência (Movahedi et al., 2012; Depuydt & Vandepoele, 2021).

Os dados de expressão obtidos via RNA-Seq podem ser gerados com o objetivo de comparar, em nível transcricional, diferentes condições biológicas, diferentes tecidos/órgãos ou fenótipos contrastantes. O estudo da expressão gênica permite identificar genes envolvidos em respostas aos estresses bióticos e abióticos; revelar mecanismos moleculares envolvidos nessas respostas, compreender melhor os processos de desenvolvimento das plantas como formação de raízes, floração, crescimento e outros; e elucidar vias metabólicas complexas como as que envolvem o acúmulo de sacarose nos colmos da cana-de-açúcar (Rao et al., 2019). Como o acúmulo de açúcar é uma característica poligênica, um grupo de genes candidatos pode ter seu padrão de expressão analisado para buscar entender o papel destes na determinação do fenótipo em diferentes genótipos e tecidos/órgãos (Kvam et al., 2012 e Movahedi et al., 2012).

9- Genes Diferencialmente Expressos, Coexpressão e Primeiros Vizinhos

Para entender a variação fenotípica é fundamental elucidar a sua base molecular. Um dos caminhos é a identificação de genes diferencialmente expressos (DEGs, do inglês *Differentially Expressed Genes*) sob diferentes condições e a identificação desses genes tem sido um objetivo de suma importância ao analisar um conjunto de dados transcriptômicos (Rao et al., 2019 e Shahjaman et al., 2020). Um gene é considerado diferencialmente expresso caso haja alterações ou diferenças nas leituras dos níveis de expressão estatisticamente significativas entre pelo menos duas condições (Anjum et al., 2016). Muitos trabalhos relacionam a identificação de DEGs com a prospecção de genes candidatos para características diversas (Pagariya et al., 2012; Shahjaman et al., 2020 e Mirador Mansuri et al., 2020).

A análise do perfil de expressão gênica tem sido usada com vários objetivos, como análise de vias metabólicas e monitoramento de respostas a estímulos ambientais e estresses; o que sugere que identificar e analisar a coexpressão de genes pode ser útil na identificação de possíveis interações entre genes candidatos entre si e/ou outros genes presentes no transcriptoma (Movahedi et al., 2012). A coexpressão gênica sugere a presença de uma ligação funcional entre os genes, embora não implique necessariamente em uma interação direta (Aoki et al., 2007). Ainda que haja certa abundância de dados públicos em amostras, genes e espécies, há necessidade de projetar estratégias específicas para integrar grandes volumes de dados que possam abordar e responder perguntas biológicas (Rao et al., 2019).

Redes de coexpressão são ferramentas poderosas para interpretar dados transcriptômicos, permitem identificar módulos de genes coexpressos que compartilham ligações funcionais biológicas, além de serem úteis para explorar tais associações a partir de perfis de expressão gênica (Rao et al., 2019), além de serem utilizadas para representar a organização funcional complexa de sistemas biológicos (Serin et al., 2016). Em uma rede de coexpressão, os nós são os genes e as arestas que os ligam indicam a semelhança no perfil de expressão (Langelder & Horvath, 2008). Ao analisar uma rede de coexpressão há a possibilidade de identificar, agrupar e explorar milhares de genes com padrões de expressão semelhantes (Serin et al., 2016).

A análise de redes tem como principal abordagem a detecção de grupos que mostrem alta coordenação transcricional em uma ou várias condições (Rao et al., 2019). A análise de coexpressão segue o princípio da “culpa por associação”. Ao observar um comportamento semelhante no perfil de expressão gênica, assumindo que não pode ser causado por coincidência, pode-se inferir que há ligação potencial de função biológica entre os genes (Serin et a., 2016; Rao et al., 2019). Por isso, pode ser relevante observar os primeiros vizinhos de genes candidatos para uma certa característica ou fenótipo em uma rede de coexpressão (Rao et al., 2019). São chamados de primeiros vizinhos dos genes de uma rede aqueles altamente associados através de padrões de expressão semelhantes (Langelder & Horvath, 2008).

Uma aplicação prática das redes de coexpressão gênica é a descoberta de genes baseada em hipóteses, e esta é das aplicações mais promissoras do uso das redes, pois observar a expressão de genes envolvidos com resposta ao estresse e com transdução de sinal, pode revelar adaptações biológicas e conectar a regulação gênica com respostas ambientais e também com fenótipos específicos (Movahedi et al., 2012). O método a ser usado para avaliar e analisar a

coexpressão será determinado pelas limitações do tamanho da amostra, o número de genes e os valores de expressão nos conjuntos de dados disponíveis (Movahedi et al., 2012 e Rao et al., 2019).

Quando certos genes compartilham uma forte correlação em uma rede de coexpressão com um gene de interesse, estes genes são chamados de primeiros vizinhos ou vizinhos próximos. Estes genes podem fornecer *insights* sobre as funções biológicas que os genes de interesse possam ter e as vias que possam estar envolvidas (Rao et al., 2019). Às análises de rede e expressão pode ser associado o tecido/órgão, as proteínas que esses genes originam e a partir disso a inferência sobre seu papel biológico em uma característica ou fenótipo (Movahedi et al., 2012 e Rao et al., 2019). Desta maneira, podemos inferir que os primeiros vizinhos de um gene de interesse podem contribuir para a expansão da identificação de genes candidatos a um fenótipo de interesse.

Uma abordagem multidisciplinar combinando diferentes estratégias, incluindo análise de expressão comparativa dos genes de interesse, identificação de DEGs, redes de coexpressão para verificar a presença e identificar os primeiros vizinhos e outras técnicas ômicas associadas pode aumentar a confiança na identificação de genes candidatos e na compreensão dos mecanismos subjacentes aos fenótipos estudados.

Objetivos

1- Objetivo geral

- Identificar genes candidatos ao acúmulo de açúcar, integrando análises de genômica estrutural e análises de expressão.

2- Objetivos específicos

- Identificar a região genômica que compõem o QTL de sorgo localizado no cromossomo SBI-02 entre os marcadores moleculares Xtxp56 e Stgnhsbm36 e os genes dentro da região delimitada entre as posições 61.500 Kb a 62.000 Kb.
- Identificar, sequenciar, montar e anotar manualmente os clones BAC das variedades SP80-3280 e IACSP93-3046 utilizando os genes ortólogos ao QTL de sorgo.
- Determinar a organização estrutural dos haplótipos das variedades SP80-3280 e IACSP93-3046.
- Realizar anotação manual da região de interesse em *S. spontaneum.*
- Comparar as regiões genômicas recuperadas das variedades SP80-3280 e IACSP93-3046 entre si, com sorgo, *S. spontaneum* e R570.
- Investigar a expressão dos genes da região nos internódios determinando quais são os Genes Diferencialmente Expressos (DEGs) entre plantas acumuladoras de açúcar e não acumuladoras de açúcar; e identificar genes coexpressos primeiros vizinhos através de uma rede de coexpressão.
- Desenvolver uma estratégia integrada que utilize abordagens de recuperação de regiões genômicas, análises comparativas de genomas e estudos de expressão gênica, juntamente com a investigação da literatura científica, para identificar genes candidatos relacionados ao acúmulo de açúcar em cana-de-açúcar

Capítulo 1 - Identifying Candidate Genes for Sugar Accumulation in Sugarcane: An Integrative Approach

Submitted article to the journal BMC Genomics - Springer Nature group.

Mônica Letícia Turibio Martins[1]; Danilo Augusto Sforça[2]; Luís Paulo dos Santos[1]; Ricardo José Gonzaga Pimenta[1]; Melina Cristina Mancini[1]; Alexandre Hild Aono[1]; Cláudio Benício Cardoso-Silva[1,3]; Sonia Vautrin[4]; Arnaud Bellec[4]; Renato Vicentini dos Santos[1]; Helene Bérgès[4]; Carla Cristina da Silva[1,5]; and Anete Pereira de Souza[1,2*]

[1] Institute of Biology (IB)–State University of Campinas (UNICAMP), Campinas, SP, Brazil
[2] Center for Molecular Biology and Genetic Engineering (CBMEG)–State University of Campinas (UNICAMP), Campinas, SP, Brazil
[3] National Laboratory of Biorenewables—LNBR/CNPEM, Campinas, SP, Brazil
[4] Centre National de Resources Génomiques Végétales (CNRGV/INRA) Toulouse, France
[5] Agronomy Department, Federal University of Viçosa, Viçosa, MG, Brazil

* **Correspondence:**
Anete Pereira de Souza
Departamento de Biologia Vegetal
Universidade Estadual de Campinas
CEP 13083-875
Campinas, São Paulo, Brazil
Tel.: +55-19-3521-1132
E-mail: anete@unicamp.br

Abstract

Background: Elucidating the intricacies of the sugarcane genome is essential for breeding superior cultivars. This economically important crop originates from hybridizations of highly polyploid *Saccharum* species. However, the large size (10 Gb), high degree of polyploidy, and aneuploidy of the sugarcane genome pose significant challenges to complete genome sequencing, assembly, and annotation. One successful strategy for identifying candidate genes linked to agronomic traits, particularly those associated with sugar accumulation, leverages synteny and potential collinearity with related species.

Results: In this study, we explored synteny between sorghum and sugarcane. Genes from a sorghum Brix QTL were used to screen bacterial artificial chromosome (BAC) libraries from two Brazilian sugarcane varieties (IACSP93-3046 and SP80-3280). The entire region was

successfully recovered, confirming synteny and collinearity between the species. Manual annotation identified 51 genes in the hybrid varieties that were subsequently confirmed to be present in *Saccharum* spontaneum. This study employed a multifaceted approach to identify candidate genes for sugar accumulation, including retrieving the genomic region of interest, performing a gene-by-gene analysis, analyzing RNA-seq data for internodes from *Saccharum* officinarum and *S. spontaneum* accessions, constructing a coexpression network to examine the expression patterns of genes within the studied region and their neighbors, and finally identifying differentially expressed genes (DEGs).

Conclusions: This comprehensive approach led to the discovery of three candidate genes potentially involved in sugar accumulation: an ethylene-responsive transcription factor (ERF), an ABA 8'-hydroxylase, and a prolyl oligopeptidase (POP). These findings could be valuable for identifying additional candidate genes for other important agricultural traits and directly targeting candidate genes for further work in molecular breeding.

Keywords: Brix, Polyploidy, Candidate genes, Sugar accumulation, Gene expression

Background

In the 1880s, sugarcane (*Saccharum* spp.) farmers crossed *Saccharum spontaneum* (2n=5x=40 to 16x=128, x=8), which is resistant to biotic and abiotic stress, with Saccharum officinarum (2n=8x=80, x=10), which is considered a noble sugarcane due to its high sugar content. Successive backcrosses with *S. officinarum* have been performed to maintain a high sugar content in hybrids [1] [2] [3] [4] [5]. The crosses between both species have generated modern sugarcane cultivars: plants with large genomes (10 Gb) that are highly polyploid and aneuploid, with at least 50% repetitive regions [6] [7] [8] [9] [5]. The hybrid genome is a mixture of chromosomes originating from *S. officinarum* (70–80% of all chromosomes of the hybrids) and *S. spontaneum* (10–20%) and recombinant chromosomes (5–10%) [2] [6] [7] [3] [5]. The variable ploidy intrinsic to each genotype creates a unique genomic structure with chromosome numbers varying between 40 and 128 [2] [10] [3] [9], which makes the study of the sugarcane genome challenging [9] [4].

Grasses with a reference genome, such as rice [11], maize [12], wheat [13], and miscanthus (*Miscanthus sinensis*) [14] [15], even sugarcane with an allele-defined genome of *S. spontaneum* [16] or a monoploid sequence reference for sugarcane [3] and sorghum [17], are

commonly used as references for studies of the sugarcane genome [9] [18] [4] [19] [3] [16] [20] [5]. Recently, a highly representative genome of the R570 hybrid variety was presented to the scientific community [5]. The *Miscanthus* genome has sorghum as an important reference for its assembly and annotation [21], and sorghum is an ancestor of the *Saccharum* and *Miscanthus* genera. Sorghum has an assembled and annotated diploid genome that is one-tenth of the sugarcane genome size and diverged approximately eight million years ago; nevertheless, its genome has maintained strong synteny and collinearity with the *Saccharum* genus [22] [23] [17] [3] [9] [4]. R570 has a high inbreeding coefficient, with approximately half being identical by descent. Furthermore, in its latest version, sorghum was also used as an important genomic reference [5].

Therefore, choosing a sorghum quantitative trait locus (QTL) for a trait of interest and recovering its orthologous region in sugarcane can be an efficient strategy to retrieve a potentially targeted sugarcane genomic region [18]. Recent studies with the R570 variety confirmed that a significant portion of the alleles originated from *S. officinarum*, and thus the sugar-accumulating origin is identical and thus largely inaccessible to QTL mapping efforts [5]. However, the genomic sequence of a cultivar does not fully reflect the genetic information about the species [24], where a cultivar may not be representative of the entire genomic content of the species [3] [9] [5]. The development of new commercial sugarcane varieties with relatively high yields is the main goal of most breeding programs. Plant breeders seek genotypes that can tolerate biotic and abiotic stress but also have increased sugar accumulation to increase yield [25] [5]. The exploration of candidate genes can be aided by the use of other omic technologies, such as RNA sequencing (RNA-seq), which provides a differential assessment not only between specific tissues but also between notably different varieties and related species [26]. The transcriptome allows for the identification of differentially expressed genes (DEGs) and can be leveraged to construct coexpression networks, aiming to identify coexpressed genes and expand the evidence leading to candidate genes.

Many genes involved in the synthesis and transport of sucrose have been identified in sugarcane [27] [28] [29] [30] [31]. Although sucrose is synthesized in the cytosol of mesophyll cells in most plants, sugarcane requires the involvement of two cell types: the bundle sheath and mesophyll cells. Sucrose synthesis occurs predominantly in the mesophyll, utilizing glucose phosphates, which are then translocated through the conducting strands of the sheath to the vascular compartments of internodal tissues, where they finally accumulate [31]. Moreover, during the plant maturation phase, the sucrose concentration in culms increases,

whereas the proportions of glucose and fructose decrease [32]. However, many processes related to sugar accumulation in sugarcane internodes are not fully understood, and possible pathways and related genes have yet to be identified.

In this context, a sorghum QTL for Brix [33] in sorghum was chosen as a target, and its orthologous region was recovered from two Brazilian cultivars (SP80-3280 and IACSP93-3046) and compared with R570 [3] and the *S. spontaneum* genome [16]. These genomic regions were compared to understand the level of genomic structural variation and genetic differences between sugarcane and sorghum. The genes found in this region were used to search for candidate genes for sugar accumulation through gene annotation evaluation, differential expression analysis of sugarcane stem transcriptomes [34] and a coexpression network. The combination of such strategies provides a more comprehensive and robust perspective in the search for and identification of the most promising candidate genes related to sugar accumulation in sugarcane.

Methods

Sorghum region of interest

A partial QTL that was mapped in sorghum for Brix was selected based on phenotypic variation and concurrently had a high LOD score; this QTL is genetically located on chromosome SBI-02 between the EST-SSR markers Xtxp56 and Stgnhsbm36 [33]. Shiringani et al. [33] mapped two major QTLs for Brix, one on chromosome SBI-02 and another on chromosome SBI-06, both of which significantly influence soluble solids in sorghum. However, both explain the phenotypic variation very well (QTL SBI-02 explains 21.9% of the phenotypic variation, and QTL SBI-06 explains 22.3%) and have high LOD scores (≥ 10); both also exhibit negative additive effects, but the QTL on chromosome SBI-02 reduces the number of units per allele dosage, thus making it preferable. Additionally, the ease of transferring information from the genetic map to the physical map was considered, facilitated by the closer proximity of known molecular markers within the QTL. The QTL on SBI-02 was located at position 82 and flanked by markers at position 80. In contrast, the QTL on chromosome SBI-06 was at position 16, with a known marker at position 13. The marker sequences were used to define the physical chromosomal position using v3.1 of the *Sorghum bicolor* genome [23] available in the Phytozome 13.0 database (https://phytozome-next.jgi.doe.gov/) [35]. The QTL

has a phenotypic variation of 21.9% explained by genotype (R2) and a logarithm of odds (LOD) value of 10.08 [33]. It spans from 61,568 kb to 61,952 kb on sorghum chromosome SBI-02, with an approximate total length of 385 kb. The target region was defined between 61,500 kb and 62,000 kb to ensure the recovery of the genes flanking the region.

Plant material

Two Brazilian sugarcane cultivars were analyzed in the present study. SP80-3280 is known for its high production of sucrose and good tillering ability. It is resistant to smut, mosaic, and rust and is tolerant to scald [36]. The SP80-3280 variety has been widely used in studies to understand sugarcane genomics and genetics. This variety has a collection of sugarcane expressed sequence tags (SUCESTs) [37], transcriptomes [38] [39] [40], mapped QTLs [41] [42], a draft genome [43], and bacterial artificial chromosome (BAC) libraries [44] [45]. The use of sorghum synteny and collinearity has also been the focus of an approach for restoring genomic regions of agronomic interest [18]. The economic importance of the IACSP93-3046 cultivar is due to its high sucrose content, good tillering, resistance to rust and suitability for mechanized harvesting [46]. This cultivar also has a transcriptome [38] and a BAC library [45].

Recovering the sorghum ortholog region in sugarcane

Primer design: The coding sequences (CDSs) of genes within the target sorghum genomic region were retrieved, along with five genes upstream and two genes downstream of the delimited region, totaling 58 genes. The BLASTn algorithm [47] was used to align the CDSs against sugarcane leaf transcripts [38] with a cutoff of E < 1e-10. Sorghum gene sequences that did not have similar transcripts in the sugarcane leaf transcriptome were compared to those in the SUCEST database [37] and the NCBI database [47]. For primer pair development, only sorghum gene sequences that aligned with sugarcane leaf transcripts or with the SUCEST [37] or NCBI [47] databases were used. Additional criteria were established as follows: the genes needed to have putative exons of 200 base pairs or larger to increase the likelihood that the primer pairs would be designed within the exon, given that the intron sizes were unknown. Furthermore, we opted not to design primer pairs for genes that were duplicated in sorghum, assuming that if a gene was duplicated in sorghum, it was likely to be duplicated in sugarcane

as well, which could increase the possibility of amplifying a different region with the duplicated gene in sugarcane. The sequences of the primer pairs can be found in Supplementary Table 1.

Identification of BAC clones, sequencing and assembly: BAC libraries from the varieties SP80-3280 and IACSP93-3046 [45] were used to recover the sequences of interest in the equivalent target region of sorghum in sugarcane varieties. Positive clone selection and preparation of BAC DNA for sequencing and pooling were performed using the steps described previously [19]. Sequencing was performed on the PacBio® Sequel platform (Pacific Biosciences) at the Arizona Genomics Institute (AGI, Tucson, AZ, USA). The vector and Escherichia coli genomic sequences were removed with the BBtools package (https://sourceforge.net/projects/bbmap/). Assembly was performed with the Canu v2.1 program [48] with the default parameters, except for corOutCoverage = 200. The refinement of the final contig consensus sequence was performed by aligning the raw reads to the assembled contigs with the pbalign program, and error correction was performed with the Arrow program. Both programs are present in the SMRTLink v7.0 package (Pacific Biosciences).

Annotation of contig sequences: The annotation of BACs for repetitive elements was performed using the LTR FINDER retrotransposon predictor [49] and the giriREPBASE database [50]. Gene annotation was performed with the NCBI [47] and Phytozome v12.0 [35] databases. The Artemis program of the Sanger Institute [51] was used to visualize genes and repetitive elements. The sorghum CDSs and the manually annotated sugarcane variety CDSs were used to perform similarity searches with BLASTn tools [47] against the following databases: NCBI [52], UniProt [53] and Pfam (Protein Families) [54]. Genes were considered similar if they exhibited a sequence identity of 80% or greater. Contigs that did not have genes, had only one gene or were smaller than 25 kb in size were discarded.

Manual curation of orthologous regions in S. spontaneum: A manual homology curation of the orthologous regions in *S. spontaneum* was performed. The CDSs of each QTL sorghum gene were aligned to the four alleles of the Sspon02 chromosomal set using BLASTn tools [47]. This approach allowed for enhanced accuracy of the automated annotation performed previously [16], including the identification of pseudogenes, thereby providing a more precise definition of the genomic architecture in this specific region in *S. spontaneum*.

The information obtained was used for a detailed literature review of each gene and their derived proteins (Supplementary Table 2), as well as their functions, the biological pathways in which they were presumably implicated, and their potential role in sugar accumulation in plants, particularly grasses and sugarcane.

Comparative genomic analyses: Comparative analyses were performed between the genes present in the target region in both varieties. In addition, the genes of the orthologous region in the *S. bicolor*, *S. spontaneum* [16] and the sugarcane hybrid variety R570 [3] genomes (Phytozome 12) [35] were also used for a comparative analysis. The analyses were performed to determine the synteny, collinearity and genomic structure of the region.

The homologous region in *S. spontaneum* was located using the BLASTn tool [47] by searching against the four homologous chromosome sequences of the *S. spontaneum* homologous chromosome 02 group (Sspon02), and are thus called Sspon2A, Sspon2B, Sspon2C and Sspon2D [16]. Genes were manually curated only in the sorghum orthologous region using the Artemis program from the Sanger Institute [51] for visualization, utilizing the NCBI [52] and Phytozome 12.0 [35] databases for assistance.

Differential gene expression analysis

The expression of genes within the QTL region was analyzed in internode tissues using sugarcane RNA-Seq data generated by Aono et al. [34]. Briefly, the IACSP93-3046 and SP80-3280 varieties, as well as the parental species *S. officinarum* (Badila de Java) and *S. spontaneum* (Krakatau), were grown in a field in Ribeirão Preto, SP, Brazil, with three replicates. The top (3rd) and bottom (8th) internodes were sampled in 2016, when the plants were six months old. RNA was extracted from these tissues and used to construct RNA-seq libraries, which were sequenced on the HiSeq 2500 platform (Illumina, San Diego, CA, United States). The RNA-Seq reads were subjected to quality control via FastQC software [55] and trimmed with Trimmomatic v.0.39 [56]. Gene expression was quantified with Salmon v.1.1.0 software [57] using the longest isoforms of *S. spontaneum* CDSs as a reference and automatic annotations provided by Zhang et al. [16]. A heatmap depicting the expression of all genes within the QTL that was automatically annotated by Zhang et al. [16] was generated using the pheatmap R package [58] in R software [59].

Differentially expressed genes (DEGs) were identified using the edgeR package version 3.38.4 [60]. The raw count data obtained from Salmon software first underwent normalization using the counts per million (CPM) method. Genes with a CPM value ≥ 1 in all samples of at least one biological condition were retained. Counts were subsequently normalized using the trimmed mean of M-values (TMM) method to identify DEGs. Statistical comparisons were

conducted between *S. spontaneum* samples and all other samples. DEGs were determined using a false discovery rate (FDR) threshold of $p \leq 0.05$ and a log2 fold change (FC) cutoff of ≥ 1.

Gene coexpression network analyses

A gene coexpression network was constructed with R software employing the highest reciprocal rank (HRR) methodology to further investigate the biological processes associated with the genes within the QTL [61]. The raw count data were normalized using the transcripts per million (TPM) method, and genes with a TPM > 0 in all samples of at least one biological condition were retained. Pairwise Pearson's correlation coefficients (R) were calculated for pairs of filtered genes. A minimum absolute correlation coefficient threshold of 0.8, which is frequently employed in gene coexpression network studies [62] [63], was used to consider two genes as connected and to ensure robust associations. The generated network was visualized via Cytoscape version 3.10.0 [64].

Results

QTL gene identification, BAC clone selection, sequencing, assembly and annotation

In the QTL for Brix in sorghum, 58 genes were identified, 21 of which were aligned with sugarcane leaf transcripts and presented exons with sizes equal to or greater than 200 bp. Primer pairs were developed for these 21 genes, and one pair failed to produce amplicons. The 20 primer pairs developed were used to screen clones of interest in the BAC libraries of the SP80-3280 and IACSP93-3046 varieties. Among the remaining 38 genes, 14 were found to be duplicated in the sorghum genome, and 24 did not meet the other selection criteria. For each gene, a number was assigned, except for two tandemly duplicated genes, which were assigned a single number (25), as shown in Supplementary Table 2, for a total of 57 genes.

In the screening of the BAC library of IACSP93-3046, 37 clones were positive for at least two genes, and 30 clones were sequenced. Among these contigs, 28 were assembled and manually annotated, representing 26 BACs. In the screening of the BAC library of SP80-3280, 56 clones were positive for at least two genes, and 31 clones were sequenced. Of these clones, genes from the region were present in 16 assembled contigs and were manually annotated,

representing 16 BACs. The size of all the contigs varied between 3,960 bp (pool 25) and 192,924 bp (pool 17), and the total length of the contigs was 5,850,46 bp (Supplementary Table 3). From the 71 contigs that were generated, 43 carried the genes from the target region. Each contig was related to a BAC, and some BACs were represented by two contigs (Supplementary Table 4).

The sorghum orthologous region was recovered from the variety IACSP93-3046 (Figure 1), which has 50 annotated genes. Seven sorghum genes were not found in the recovered sequence (Figure 1). A gap was present between genes 35 and 36. In one of the haplotypes, one annotated gene did not belong to this region in sorghum, although it is located in another region of sorghum chromosome SBI-02. In the SP80-3280 variety (Figure 2), the region was recovered almost in its entirety, with 44 annotated genes. Of the seven genes that were not found in the IACSP93-3046 contig sequences, six were not found in this variety, and one was annotated as a pseudogene. Two gaps were observed, one between genes 24 and 26 and the other between genes 43 and 47.

Forty-five pseudogenes were identified among the homologous genes in the variety IACSP93-3046 and nine pseudogenes in the variety SP80-3280. In the variety IACSP93-3046, genes with insertions of transposons in intronic regions (6–13.3%), insertions/deletions of one or more nucleotides (36–80%) and partial gene sequences (3–6.7%) were considered pseudogenes. Among nine homologous genes considered pseudogenes in SP80-3280, four (44.5%) presented an insertion/deletion of one or more nucleotides, three (33.5%) presented a transposon insertion in intronic regions, and in two (22.2%) of these genes, the pseudogene was a fragment of the gene.

Main differences in sorghum–sugarcane synteny and collinearity in the target region

Chromosome Sspon2A (Supplementary Figure 1): The orthologous region on chromosome Sspon2A is 794,054 bp long and is the closest in size to chromosome SbI-02 of sorghum. It is located between bases 35,019,101 and 35,813,155. Among the 57 genes present in sorghum, 50 orthologs were found in Sspon2A. The seven missing orthologous genes (03, 14, 24, 40, 46, 55 and 56) were not detected throughout the chromosome and not only in the delimited region; they were not detected in the IACSP93-3046, SP80-3280 and R570 varieties. In the region delimited in Sspon2A, the gene Sspon.02G0013290 was found, and it is orthologous to a sorghum gene from chromosome Sb10 (Sobic.010G093001). Three additional

genes were not detected in the IACSP93-3046 and R570 varieties, 31 and 43; these genes were detected on chromosome Sspon2A but as pseudogenes. Gene 51 was also detected as a pseudogene in the SP80-3280 variety but was detected in the IACSP93-3046 and R570 varieties. Eight inversions were observed, two of which were common to the varieties IACSP93-3046 and SP80-3280, and they involved two to eight genes. Duplications, some in tandem, were also observed. Therefore, synteny occurs, as almost all the genes are present, but many breaks in collinearity are observed.

Chromosome Sspon2B (Supplementary Figure 2)—On this chromosome, 1,134,230 bp are present between the first and last genes of the studied region, more than double the region in sorghum, which is located between bases 32,312,667 and 33,446,897. In this chromosome, many collinearity breaks, with inversions and insertions within a cluster of 12 genes, were observed. In this orthologous region, rearrangements and reorganizations were observed, but most of the genes were present, guaranteeing synteny. Of the 57 genes present in sorghum, six were absent from the entire chromosome: 07, 26, 27 and 28. Genes 14 and 31 were also missing and were not found in the IACSP93-3046, SP80-3280 or R570 varieties. An insertion containing a cluster of 16 genes was observed in the genomic region of this homologous chromosome, with eight of these genes being orthologous to genes located on sorghum chromosome SBI-02 between positions 62,116 kb and 62,216 kb, which is close to the target region investigate din this study.

Chromosome Sspon2C (Supplementary Figure 3): The chromosome Sspon2C region is the region that most resembles sorghum chromosome SBI-02. Considering synteny, although it is located between the first gene and the last gene, it is almost twice the size of the region, reaching 969,275 bp, and is located between bases 37,413,201 and 38,382,476. As in the other alleles, Sspon2C also has collinearity breaks with inversions and insertions, and gene sequences from the region are displaced and inserted in other stretches. Among the 57 sorghum genes in the region, two (genes 14 and 31) are absent on this chromosome, and these genes are also absent in the hybrid varieties IACSP93-3046, SP80-3280 and R570. Although the region is quite large, compared with sorghum, no insertions with genes similar to those of other chromosomes in *S. spontaneum* are observed.

Chromosome Sspon2D (Supplementary Figure 4): This chromosome also maintains synteny with sorghum. Among the 57 genes in the sorghum region, 51 remained in Sspon2D. Among the six missing genes, four were not detected in the IACSP93-3046, SP80-3280 or R570 varieties. The region was divided into two subregions. The first subregion is between bases

28,425,371 and 29,097,118 (671,747 bp), and the second subregion is between bases 49,915,069 and 50,120,906 (205,837 bp). These two subregions are separated by approximately 21 Mb in length.

IACSP93-3046 (Figure 1)—This region was recovered with 29 annotated BACs. The synteny between sorghum and sugarcane in this specific region was confirmed, but some collinearity breaks were detected. Two inversions were observed, one between genes 11 and 13 and the other between genes 34 and 35. These inversions were observed in all haplotypes where these genes could be present. Gene 25, which is duplicated in sorghum, appeared in a single copy in the annotated haplotypes; however, gene 26 was duplicated in one of the three haplotypes observed. A sequence of three genes (53, 54 and 55) was duplicated exactly in this sequence, resulting in a collinearity break; however, this finding appeared in only one of the seven haplotypes that could have these genes. In one annotated BAC, an insertion of gene 57 between genes 52 and 53 was observed. Another interesting insertion was found between genes 16 and 17; it was a gene similar to Sobic.002G135950 from sorghum, chromosome SBI-02 at position 20,517,004-20,519,005, and Sobic.002G195033 from sorghum and located at position SBI-02 58,315,718-58,317,644.

SP80-3280 variety (Figure 2)—Of the 31 sequenced clones, 16 were recognized as part of the target region using BLASTn. The orthologous region was partially recovered using BACs belonging to this region. Among the recovered genes, synteny and collinearity may be presumed. Although breaks in collinearity are observed with the variety IACSP93-3046, some of those observed are similar in both varieties, such as an inversion between genes 11 and 13 and another between genes 34 and 35. Of the 15 sequenced and annotated BACs, 11 were positive for the Sobic.002G223900 gene (gene 08), eight of which were also positive for one of the last 20 genes in the target region (genes 38 to 57). Some genes were not observed in the annotations of the varieties IACSP93-3046 and R570; this situation also occurred with the variety SP80-3280, except for one gene (51) that was observed as a possible pseudogene. Two pronounced gaps were detected: the absence of BACs containing genes 24 to 26 and the absence of BACs containing genes 43 to 47. The last gene flanking the region, 57, was also not recovered.

R570 (Supplementary Figure 5)—Upon comparing the findings of Brazilian varieties with those of R570, some commonalities were observed. The inversion between genes 11 and 13 is present in all three hybrid cultivars, indicating that this observation is characteristic of the *Saccharum* genus, as it is also observed in *S. spontaneum*. Tandem duplications, such as those

of gene 25, were noted, mirroring observations in sorghum. Interestingly, IACSP93-3046 lacks this duplication, and due to a gap in the sequencing of this region, this duplication could not be detected in SP80-3280. In R570, genes 48 and 49 are duplicated in tandem, a feature not observed in sorghum. Similar findings were not observed in *S. spontaneum* or in the varieties SP80-3280 and IACSP93-3046.

Expression analysis and search for candidate genes related to sugar accumulation: Investigation of selected genes

A summary of genes 01 to 57, their orthologs in *S. spontaneum* and *S. bicolor*, as well as their proteins, is provided in Supplementary Table 2. Among those genes, a total of 51 genes were observed in the SP80-3280 and IACSP93-3046 varieties. For each of these genes, an investigative approach was adopted utilizing relevant published studies to elucidate their biological roles, molecular interactions, and possible functions in physiological processes. Based on this investigation, 10 candidate genes for sugar accumulation were selected: 02, 06, 09, 10, 15, 19, 20, 23, 25, and 43 (Supplementary Table 2). These genes are possibly involved directly, indirectly, or in fundamental upstream steps involved in some phase of the process of sugar accumulation, which begins with carbon fixation from the atmosphere (photosynthesis), sucrose biosynthesis and transport to the stems, and subsequent accumulation (Supplementary Table 2)

DEG analyses

The expression of the genes within the QTL was evaluated using RNA-Seq data from internodes 3 (younger) and 8 (more mature) of IACSP93-3046 and SP80-3280, as well as data from accessions of the two species considered the main ancestors of modern cultivars, namely, *S. spontaneum* and *S. officinarum*. Seven of the 51 sorghum genes being analyzed had no orthologs represented in the *S. spontaneum* genome, which was used for the gene quantification procedures; therefore, they are not represented in the expression data.

A heatmap depicting the expression of the remaining 44 genes normalized by TPM is shown in Figure 3. This approach allowed us to observe internode gene expression patterns in two commercial sugarcane varieties, IACSP93-3046 and SP80-3280, and the parental species *S. officinarum* and *S. spontaneum*. For ten genes (10, 11, 12, 13, 15, 22, 25, 34, 45, 46, and 56), no expression was detected in any biological replicate, or minimal expression was detected in

up to three biological replicates. These genes may play crucial roles in other plant organs, such as leaves or roots, or could also be relevant in other stages of plant maturation. However, due to a lack of evidence of expression in the organ/tissue and maturation stages being analyzed, these genes were not considered candidates involved in sugar production.

After filtering, 22,859 of the 35,471 genes present in the *S. spontaneum* CDSs were stably expressed under at least one biological condition and were thus retained for DEG analyses. By comparing varieties with high (*S. officinarum*, IACSP93-3046, and SP80-3280) and low (*S. spontaneum*) sugar contents, 6,264 DEGs were identified (Supplementary Table 4). Seven genes within the QTL region were DEGs; their log2(FC) values, FDR-corrected p values and annotations are available in Table 1.

Gene coexpression network analyses

Based on the expression data, an HRR coexpression network (Supplementary Figure 5) was constructed to explore new evidence that could contribute to the search for candidate genes. During the filtering procedures, 7,565 genes were excluded, and the remaining 27,906 genes were used as inputs to construct the network. The final network had 6,809 connected nodes (genes) and an average of 17 neighbors per node. Among these genes, 3,397 genes were identified as DEGs, and six genes were identified within the QTL. A first neighbor search was employed to identify genes related to potential sugar accumulation candidates and to assess whether these genes could support their roles in this process. The first neighbors of the genes within the QTL represented in the network can be seen in Table 2. Three of these genes—01, 23, and 26—were also identified as DEGs (Table 1).

Gene 01 (prolyl oligopeptidase—POP) exhibited relatively low expression in the stems of *S. spontaneum* and relatively high expression in samples from sugar-accumulating plants. Gene 23 (abscisic acid 8'-hydroxylase 3—ABA8’OH) has virtually no expression in the internodes of the sugar-accumulating plants sampled and is expressed at low levels in *S. spontaneum*. Gene 26 (ethylene responsive factor 109—ERF109) also has almost no expression in sugarcane plants, whereas it is expressed in *S. spontaneum*; however, in this case, significantly higher expression is detected in the more mature internodes of *S. spontaneum* (I8) than in the less mature internodes (I3). This evidence led to the selection of genes 01, 23, and 26 as the primary candidate genes in the QTL for sugar accumulation.

Discussion

Main differences in genomic architecture

Synteny and collinearity have been used to compare and recover genomic regions of interest in sugarcane using sorghum [22] [23] [17] [5] [18] [3] [9] [15] [45] [34] [65] [5] and *Miscanthus* [66] [21] [67] genomes as references, revealing high gene retention [18] [3] [45] [68]. The comparison of the same region between sugarcane varieties and their ancestral species can provide insights into the genomic complexity of sugarcane. The region evaluated in this work showed substantial differences among the genotypes studied, such as gene duplications, loss of gene exons, pseudogenization, gene inversions, gene deletions and insertions.

For example, gene 25 (similar to alpha-amylase—AMY—Supplementary Table 2) is duplicated in tandem in sorghum and in R570, but in four IACSP93-3046 haplotypes, it is present in a single copy (a gap exists in the region SP80-3280). The sequences of genes 53, 54, and 55 (Supplementary Table 2) are duplicated in tandem in BAC Shy141H03 of IACSP93-3046 (Figure 1), but they are not duplicated in sorghum and are not present in the recovered SP80-3280 haplotypes or in any of the alleles of the *S. spontaneum* genome. In addition to gene duplications, gene inversions were detected in the orthologous region between sorghum and all the *Saccharum* accessions evaluated (Figure 2), which suggests that both inversions occurred after sorghum–sugarcane divergence. Overall, when inversions do not significantly disrupt the gene balance of an organism, the direct consequences tend to be minimal. Documented cases exist where inversion results in pseudogenization or even deletion of one of the genes [69] [70].

In BAC Shy411A07 of the IACSP93-3046 variety, gene 11 (Supplementary Table 2) was absent, yet the remaining genes (12 and 13—Supplementary Table 2) indicated that an inversion occurred. Fragments of Harbinger-type (HARB) repetitive elements were found near pseudogenes 12 and 13 (Supplementary Table 2). HARB transposons are classified as class II transposable elements (TEs) that carry out the cleavage and transfer of single DNA strands mediated by transposases [69] [70]. The presence of these HARB transposons suggests a possible relationship between these elements and these inversions, which were present in all the examined varieties, especially with the probable pseudogenization of genes 12 and 13 (Supplementary Table 2). The process of cleavage followed by fusion may have led to the deletion of bases, resulting in the truncation of genes and, consequently, the loss of their functions. In the SP80-3280 variety, gene 13 (Supplementary Table 2) contains a single exon spanning 2,682 base pairs. On the other hand, gene 12 (Supplementary Table 2) maintains two

introns, even in its pseudogenized state, and in this case, it is situated between two TEs, similar to the HARB type (BAC Shy260G24). In BAC Shy492F12, gene 12 (Supplementary Table 2) is also close to a HARB-type TE flanking the last exon. In this case, gene 12 (Supplementary Table 2) presents characteristics indicative of a functional gene. The gene had different CDS base pair compositions among the haplotypes but was always between 1347 bp and 1488 bp. Additionally, gene 11 (Supplementary Table 2) retained a single intron, with a length ranging between 1,200 and 1,209 bp. However, neither variation was detected in sorghum, suggesting that it might be a unique characteristic of the *Saccharum* genus. Gene 11 on chromosome Sspon2B of *S. spontaneum* (Supplementary Table 2) contains a single exon.

In a haplotype of the variety IACSP93-3046, represented by BAC Shy112C03, a gene whose ortholog in sorghum is not found in the QTL studied was detected. Notably, this gene is similar to the two sorghum genes Sobic.002G135950 and Sobic.002G195033. These sorghum genes have 91% sequence identity, and both have a zinc finger domain. The probable orthologous gene in the IACSP93-3046 variety is inserted in a retrotransposon similar to Copia22-ZM_I/LTR. Interestingly, this gene has all the characteristic features of being a functional gene, even when it is inserted into a TE. One possible explanation for this insertion is that the gene was cotransported with the retrotransposon. As Class II TEs, they can replicate a copy of themselves, which is subsequently inserted into different genomic regions. As such, this haplotype is highly likely to be a copy of the gene. The presence of a TE within an expressed gene (CENP-C) in sugarcane has been previously described [45], demonstrating that the proximity or overlap of TEs and genes does not hinder the function of the gene, at least in sugarcane. This gene, which is specific to the IACSP93-3046 haplotype (BAC Shy112C03), has a zinc finger domain; in plants, proteins featuring this domain are transcription factors (TFs) related to the control of cell division in totipotent tissues (petunias), histone–DNA binding (wheat), leaf budding (Chinese cabbage), soil salinity tolerance (Arabidopsis) and carbon metabolism (potato) [71].

This gene has also been detected in the orthologous region of the S. spontaneum chromosome Sspon2B but as a gene fragment. In IACSP93-3046, the gene is located between genes 16 and 17 (Supplementary Table 2) in a reverse orientation, and in Sspon2B, it is located between genes 39 and 41 (Supplementary Table 2), with gene 40 (Supplementary Table 2) being inserted into another fragment of the orthologous region in a strand orientation. This gene could also have been transposed with a TE, as possibly occurred with the hybrid. Importantly, *S. spontaneum* is a wild species that has been evolving under the pressure of natural selection,

without the same level of human interference that fully domesticated plants undergo, as is the case with modern sugarcane cultivars—commercial hybrids. Despite such variability, we can observe the presence of potential genomic structure characteristics of *S. spontaneum* in commercial hybrids, such as inversions 11–13 (Supplementary Table 2) and 34–35 (Supplementary Table 2), which are present in at least three of the four alleles and are also present in all recovered haplotypes of the SP80-3280 and IACSP94-3046 varieties, where these inversions can be observed, as well as in the R570 variety [3].

The orthologous regions in the hybrid varieties appear to be more similar to those in sorghum than to those in *S. spontaneum*. Some fundamental characteristics are shared, such as synteny. However, differences such as inversions, duplications, insertions of orthologous genes from the same genomic region and even from sequences that are similar to genes from sorghum chromosomes other than SBI-02, possible pseudogenization and translocation were detected. However, this finding is not surprising considering that the chromosomes originating from *S. spontaneum* found in hybrids constitute only 10% to 20% of the chromosomes of modern hybrids [2] [6] [7] [3]. Furthermore, wild species such as *S. spontaneum* have been subjected to natural selection pressure, resulting in a high level of expected heterozygosity for wild plants. While wild species have evolved naturally, commercial varieties have been selected and improved over the past 120 years to meet human needs [72], which has led to substantial genomic differences between them.

Investigation of genes involved in sugar accumulation

Among the 51 studied genes, 17 were annotated as genes associated with tolerance or response to stress (genes 01, 04, 11, 12, 13, 20, 22, 23, 24, 26, 32, 33, 34, 36, 37, 41, and 57—Supplementary Table 2). The period when sugarcane accumulates sucrose in its stems coincides with the dry and high luminosity period in a significant portion of the crops. During this period, leaves gradually fall, while sugar accumulates in culms [73]. Sucrose accumulation occurs in response to stressful conditions [74]. Therefore, a connection may exist between sucrose accumulation in sugarcane and genes related to abiotic stress.

Considering the characteristics of the crop, five genes related to the abscisic acid (ABA) response were also observed (genes 20, 23, 32, 33 and 37—Supplementary Table 2). In addition to being a crucial hormone for the photosynthetic process by regulating stomatal closure and opening, ABA is highly responsive to stress, particularly water stress [75]. Leaf water potential

and stomatal conductance are crucial factors for sugarcane to be able to produce carbohydrates that are converted into sucrose, transported to stalks, and subsequently accumulate [76] [77].

In addition, five genes (genes 17, 18, 45, 46, and 47—Supplementary Table 2) belonging to the lipolytic enzyme GDXG family were detected. This enzyme family is characterized by two consensus sequences containing a histidine residue and a serine residue as putative active site residues [78]. In some genes encoding these enzymes, the alpha/beta hydrolase (ABH) domain, which is known to play a role in catalyzing the cleavage of carbon double bonds and decarboxylation, may be present [79]. Additionally, six genes (8, 49, 50, 52, 53, and 54—Supplementary Table 2) containing the ABH domain but belonging to the carboxylesterase (CXE) family were identified, five of which are sequential. The specificities of genes within the same family characterized by shared domains may vary significantly, necessitating further exploration of the biological roles of each gene. Notably, a group of genes sharing closely related domains and families on the same chromosome, even sequentially, may suggest a potential origin through duplication events that diverged during evolution into distinct genes while maintaining some similarity [80].

Candidate genes for sugar accumulation

The ERF109 (gene 26—Supplementary Table 2), ABA 8' OH (gene 23—Supplementary Table 2), and POP (gene 01—Supplementary Table 2) genes are candidates for sucrose accumulation in sugarcane, considering that sugarcane needs soil with low humidity, approximately 15%, for greater sugar accumulation [81].

Gene 26 (Supplementary Table 2) is similar to ERF109, an ethylene-responsive TF. ERF109 expression is related to anthocyanin accumulation in apples, as ERF109 directly binds to the promoters of anthocyanin synthesis genes [82]. Jasmonic acid (JA) accumulation in plant wounds also activates the expression of ERF109. ERF109 induces the biosynthesis of the auxin protein ASA1, which aids in the process of secondary root formation mediated by JA-dependent ERF109 signaling [83]. In sugarcane, more than 16,000 genes have been identified as potential targets regulated by ERF109, indicating that ERF109 has a broad influence on gene expression. Functions are diverse and include metabolic activities such as Rubisco activity, triggering the biosynthesis of hormones such as cytokinins, and gibberellin-mediated responses [84]. ERF109 was not expressed in the internodes of any of the analyzed sugar-accumulating plants or in the younger internodes of *S. spontaneum*, but significant expression was detected in its older

internodes. In transgenic lemon, the overexpression of ERF109 causes global reprogramming of plant expression [85]. ERF109 acts as a stress-responsive TF; however, the proposed role of this gene in sugar accumulation, given its promiscuity, requires further investigation.

Proline-dependent genes

Genes 01 and 23 (Supplementary Table 2) are related to proline accumulation. Proline is an amino acid with a unique configuration, restricting its free rotation at the α-carbon because the nitrogen and α-carbon are combined in a pyrrolidine ring. This structure contributes to the rigidity of proteins containing proline residues and requires specific enzymes, including POP, for cleavage [86]. Enzymes that specifically cleave proline are known to be involved in proline accumulation in the cytosol of plant cells. This accumulation is essential for the adaptive response of plants to adverse situations [87]. Plants accumulate proline to maintain cellular homeostasis, aid in water absorption, and better adapt to abiotic stresses such as drought, salinity, and heavy metals. These adverse conditions lead to the excess production of reactive oxygen species (ROS) and consequences such as lipid peroxidation, increased osmolyte levels, and activation of antioxidant systems [88] [89] [90] [91] [92] [93]. As such, proline accumulation enhances adaptive responses in plants. Plants may increase proline biosynthesis in response to the above conditions or reuse presynthesized proline from proteins and peptides that are not essential [94].

The exogenous application of proline in maize has been reported to increase sugar, oil, moisture, and protein levels in seeds under drought conditions [95] [91]. In sugarcane, the efficiency of photosynthesis and stomatal conductance is specifically related to sucrose accumulation in stalks [96]. In mature plants, the relationship between leaves as a sugar source and other organs, including internodes, is critical for the regulation of photosynthesis rates and sucrose accumulation in stalks [74]. Proline interacts with other metabolites, including soluble sugars. The phenomenon of accumulated proline interacting with sucrose, for example, to adjust the osmotic balance during salt stress, has been reported [91] [92] [93]. However, whether this interaction could be driven by other physiological changes in different plants in response to different stimuli remains unclear. Although the connection between stress and plants has been clarified, much still needs to be elucidated. In sugarcane, the accumulation of sucrose and starch in leaves coincides with a reduction in photosynthetic rates, which occurs during low water availability [73].

Gene 23 (Supplementary Table 2), which shares similarities with ABA 8'hydroxylase enzymes from the cytochrome P450 family, converts ABA into 8'-hydroxy ABA and then into phaseic acid [98], regulating ABA metabolism and influencing plant responses to environmental stress and development, including germination, root growth, and fruit maturation [99]. Inhibition of this enzyme affects the balance of processes involving ABA [99], such as stomatal closure in response to water, salt, and thermal stresses. Studies in grapes have shown that inhibiting ABA 8'-hydroxylase results in reduced leaf water potential and stomatal conductance [100], accompanied by proline accumulation in leaves and the growth of adventitious roots [100].

In sugar-accumulating plants, ABA 8'-hydroxylase is virtually not expressed, suggesting a potential role for ABA regulation in sugar storage tissues (Figure 3— genes 13240-23). The dry climate during the sugar accumulation period in sugarcane, as observed during sample collection, indicates potential moderate water stress. Under water stress conditions, the sugarcane genotypes with the most efficient sugar accumulation tend to maintain greater stomatal conductance [101]. The lack of expression of the ABA catabolic gene suggested that the need for stomatal conductance regulation in these plants may be linked to maintaining open stomata. Additionally, in grapes, inhibition of the ABA 8'-hydroxylase gene improves tolerance to dehydration and promotes adventitious root formation, indicating an effective strategy for coping with water stress. Gene 23 was not expressed in the internodes (gene 13240–23—Figure 3), suggesting that the plant may use this strategy to improve its tolerance to a potential water deficit.

Gene 01 (Supplementary Table 2) shares similarities with POP, which belongs to the serine protease family (clan SC, family S9) that includes various peptidases. POP is a cytoplasmic enzyme that hydrolyzes peptide bonds at the C-terminal side of proline residues [102] [103]. The enzyme's three-dimensional structure allows for the postproline cleavage of peptides containing up to 30 amino acid residues [102] [103]. Post- or preproline-cleaving enzymes can belong to different peptidase families, including aminopeptidases, endopeptidases, or oligopeptidases (PAP/PEP/POP). The most common domain in family S9 is a substrate-limiting β-propeller domain preventing unwanted digestion, while the α/β hydrolase domain catalyzes the reaction at the carboxy terminus of proline residues [103]. POP is a ubiquitous protein with a well-established structure and mechanism of action. However, its biological role in plants has not been fully elucidated. Increased POP expression in plants is known to be associated with tolerance to various types of abiotic stresses [102] [104] [105]. In

flax, this phenomenon seems to be related to a fundamental mechanism for embryo growth in seeds [102]. In coffee, POP overexpression is linked to a significant increase in the number of branches in transgenic plants [104]. Other peptidases that hydrolyze with proline specificity are related to plant development, such as pollen development [90], flowering, increased ABA activity, protection of photosynthetic activity during salt stress, elimination of reactive oxygen species, and overall osmotic potential adjustment [91].

Gene 01 is a DEG (Table 1) that is highly expressed in the internodes of the sampled sugar-accumulating plants (SP80-3280 and IACSP93-3046 varieties, and *S. officinarum*) and expressed at significantly lower levels in *S. spontaneum*, a sugar-nonaccumulating sugarcane species known for its resistance to various stress types. Although POP is related to stress resistance and tolerance, more evidence suggests that this gene is involved in this process. Proline-cleaving enzymes are expressed when a plant needs to accumulate proline, a phenomenon that usually occurs when the plant requires osmotic regulation due to stresses such as water and salinity stress. Proline can also bind to soluble sugars such as sucrose when the homeostasis of plant cells must be regulated. Gene 01 has three first neighbors, one of which possesses the ENTH domain—a lipid-binding region crucial for clathrin-coated vesicle formation, endocytosis at the trans-Golgi network (TGN), and vacuolar transport. This gene could play a role in the transport of proline and sucrose, given its correlation with POP.

Conclusions

This study examined the genomic architecture of a specific region in two sugarcane cultivars, validating sorghum as a key reference genome for identifying candidate genes for quantitative traits. Using a method that integrated comparative genomic analysis, RNA-seq data, and a literature review, three candidate genes involved in sugar accumulation were identified: ERF-109, ABA 8’OH, and POP. ERF-109, a versatile transcription factor, is known to be involved in metabolic processes and stress resistance. This study revealed that it potentially influences sugar accumulation. ABA 8’OH and POP are linked to proline accumulation, which is associated with stress responses and sucrose binding, indicating a possible association between sucrose and proline accumulation. This integrative method is recommended for identifying new candidate genes for sugar accumulation and other traits in sugarcane, with the aim of developing cultivars with relatively high sugar contents. We

encourage more research on these three genes, believing that we can learn more about the intricate process of storing and accumulating sugar in sugarcane.

List of abbreviations

ABA 8'OH—Abscisic acid 8'-hydroxylase

AMY—Alpha amylase

BAC—Bacterial artificial chromosome

BLAST—Basic Local Alignment Search Tool

CDS—Coding sequence

CPM—Counts per million

DEG—Differentially expressed gene

DGE—Differential gene expression

ERF-109—Ethylene responsive factor 109

FDR—false discovery rate

HARB—Harbinger-type transposon

HRR—Highest reciprocal rank

LOD—Logarithm of the odds value

PAP—Prolyl aminopeptidase

PEP—Prolyl endopeptidase

POP—Prolyl oligopeptidase

QTL—Quantitative trait locus

RNA-seq—RNA sequencing

TE—Transposable element

TMM—Trimmed mean of M values

TPM—Transcripts per million

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Supplementary Table 5 for this article has been deposited in Figshare and is available at https://figshare.com/articles/dataset/Supplementary_Table_5_for_Martins_et_al_2024_-_Identifying_Candidate_Genes_for_Sugar_Accumulation_in_Sugarcane_an_integrative_approach/26236361

The raw BAC sequencing data used in this article have been submitted to the SRA/NCBI under BioProject PRJNA1028979 and BioProject PRJNA1108670.

The RNA sequencing data used in the present work are available from the SRA under BioProject ID PRJNA681593.

Funding

This work was supported by grants from the São Paulo Research Foundation (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program). Mônica Leticia Turibio Martins received an MSc. fellowship from CAPES (grant 88887.177148/2018-00) and Ph.D. fellowships from CAPES (grants 88887.337878/2019-00, 88882.329492/2019-01 and 88887.661406/2022-00). Luís Paulo dos Santos received an MSc. fellowship from FAPESP, Ricardo José Gonzaga Pimenta and Alexandre Hild Aono received Ph.D. fellowships from FAPESP (grants 2017/18225-0, 2019/21682-9 and 2019/03232-6). Anete Pereira de Souza received a research fellowship from CNPq (grant 312777/2018-3).

Authors' contributions

MLTLM: conceptualization, benchwork, data curation, methodology, formal analysis, visualization, investigation, writing—original draft and writing—review and editing. DAS: conceptualization, methodology, visualization and writing—review and editing. LPS: benchwork, data curation, formal analysis, investigation. RJPG: formal analysis, bioinformatics methodology, and writing—review and editing. MCM: benchwork, visualization and methodology. AHA: formal analysis and bioinformatics methodology. CBCS: bioinformatics methodology and formal analysis. SV: conceptualization, methodology, writing—review. AB: conceptualization, methodology, writing—review. RVS: bioinformatics methodology. HB: conceptualization, methodology, writing—review. CCS: conceptualization, methodology, visualization and writing—review and editing. APS: conceptualization, visualization, writing—review and editing, funding acquisition, project administration, resources, supervision.

Acknowledgments

The authors acknowledge the Improvement of Higher Education Personnel (CAPES) via the Genetic and Molecular Biology Post-Graduate Program of Biology Institute from Unicamp

(Universidade Estadual de Campinas) and the São Paulo Research Foundation (FAPESP) for financial support.

References

1. Bremer G. Problems in breeding and cytology of sugar cane. Euphytica. 1961;10:59-78.
2. D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Molec Gen Genet. 1996;250:405–413. doi:10.1007/BF02174028.
3. Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun. 2018;9:2638. doi:10.1038/s41467-018-05051-5.
4. Babu KSD, Janakiraman V, Palaniswamy H, Kasirajan L, Gomathi R, Ramkumar TR. A short review on sugarcane: its domestication, molecular manipulations, and future perspectives. Genet Resour Crop Evol. 2022;69:2623-2643.
5. Healey AL, Garsmeur O, Lovell JT, Shengquiang S, Sreedasyam A, Jenkins J, et al. The complex polyploid genome architecture of sugarcane. Nature. 2024. doi:10.1038/s41586-024-07231-4.
6. Cuadrado A. Genome remodelling in three modern *S. officinarumxS. spontaneum* sugarcane cultivars. J Exp Bot. 2004;55:847–854. doi:10.1093/jxb/erh093.
7. Piperidis G, Piperidis N, D'Hont A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Genet Genomics. 2010;284:65–73. doi:10.1007/s00438-010-0546-3.
8. Garcia AAF, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira MLC, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep. 2013;3:3399. doi:10.1038/srep03399.
9. Thirugnanasambandam PP, Hoang NV, Henry RJ. The Challenge of Analyzing the Sugarcane Genome. Front Plant Sci. 2018;9:616. doi:10.3389/fpls.2018.00616.
10. Zhang, J., Nagai, C., Yu, Q., Pan, Y.-B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three *Saccharum* species. Euphytica 185, 511–519. doi:10.1007/s10681-012-0664-6.

11. International Rice Genome Sequencing Project, Sasaki T. The map-based sequence of the rice genome. Nature. 2005;436:793–800. doi:10.1038/nature03895.
12. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science. 2009;326:1112–1115. doi:10.1126/science.1178534.
13. The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, Stein N, Feuillet C, Keller B, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361:eaar7191. doi:10.1126/science.aar7191.
14. Kim C, Wang X, Lee TH, Jakob K, Lee GJ, Paterson AH. Comparative Analysis of *Miscanthus* and *Saccharum* Reveals a Shared Whole-Genome Duplication but Different Evolutionary Fates. Plant Cell. 2014;26:2420–2429. doi:10.1105/tpc.114.125583.
15. Tsuruta S, Ebina M, Kobayashi M, Takahashi W. Complete Chloroplast Genomes of Erianthus arundinaceus and *Miscanthus sinensis*: Comparative Genomics and Evolution of the Saccharum Complex. PLoS ONE. 2017;12:e0169992. doi:10.1371/journal.pone.0169992.
16. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum* spontaneum L. Nat Genet. 2018;50:1565–1573. doi:10.1038/s41588-018-0237-2.
17. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A, Morishige DT, Grimwood J, Schmutz J, Mullet JE. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J. 2017 Nov 21.
18. Wang J, Roe B, Macmil S, Yu Q, Murray JE, Tang H, et al. Linearity between autopolyploid sugarcane and diploid sorghum genomes. 2010;17.
19. Mancini MC, Cardoso-Silva CB, Sforça DA, Pereira de Souza A. "Targeted Sequencing by Gene Synteny," a New Strategy for Polyploid Species: Sequencing and Physical Structure of a Complex Sugarcane Region. Front Plant Sci. 2018;9:397. doi:10.3389/fpls.2018.00397.
20. Wang T, Fang J, Zhang J. Advances in Sugarcane Genomics and Genetics. Sugar Tech. 2022;24:354–368. https://doi.org/10.1007/s12355-021-01065-4.

21. Mitros T, Session AM, James BT, et al. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. Nat Commun. 2020;11:5442. doi:10.1038/s41467-020-18923-6.
22. Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH. Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics. 1998 Dec;150(4):1663-82. doi: 10.1093/genetics/150.4.1663. PMID: 9832541; PMCID: PMC1460436.
23. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The *Sorghum bicolor* genome and the diversification of grasses. Nature. 2009;457:551–556. doi:10.1038/nature07723.
24. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CKK, et al. The pangenome of hexaploid bread wheat. Plant J. 2017;90:1007–1013. doi:10.1111/tpj.13515.
25. Cursi DE, et al. History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. Sugar Tech. 2021:1–22.
26. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019;20:631–656. doi:10.1038/s41576-019-0150-2.
27. Zhu YJ, Albert HH, Moore PH. Differential expression of soluble acid invertase genes in the shoots of high-sucrose and low-sucrose species of *Saccharum* and their hybrids. Aust J Plant Physiol. 2000;27:193–199.
28. Carson DL, Botha FC. Genes expressed in sugarcane maturing internodal tissue. Plant Cell Rep. 2002;20:1075–1081.
29. Grivet L, Arruda P. Sugarcane genomics: Depicting the complex genome of an important tropical crop. Curr Opin Plant Biol. 2002;5:122–127.
30. Casu RE, Grof CPL, Rae AL, McIntyre CL, Dimmock CM, Manners JM. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol. 2003;52:371–386.
31. Vasantha S, Kumar RA, Tayade AS, et al. Physiology of Sucrose Productivity and Implications of Ripeners in Sugarcane. Sugar Tech. 2022;24:715–731. doi:10.1007/s12355-021-01062-7.

32. Chandra A, Jain R, Solomon S. Complexities of invertases controlling sucrose accumulation and retention in sugarcane. Curr Sci. 2012;102(6):857–866. Available from: http://www.jstor.org/stable/24084500.
33. Shiringani AL, Frisch M, Friedt W. Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. Theor Appl Genet. 2010;121:323–336. doi:10.1007/s00122-010-1312-y.
34. Aono AH, Pimenta RJG, Garcia ALB, Correr FH, Hosaka GK, Carrasco MM, Cardoso-Silva CB, Mancini MC, Sforça DA, Dos Santos LB, Nagai JS, Pinto LR, Landell MGA, Carneiro MS, Balsalobre TW, Quiles MG, Pereira WA, Margarido GRA, de Souza AP. The Wild Sugarcane and Sorghum Kinomes: Insights Into Expansion, Diversification, and Expression Patterns. Front Plant Sci. 2021 Jul 7;12:668623. doi: 10.3389/fpls.2021.668623. PMID: 34305969; PMCID: PMC8294386.
35. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178–D1186. doi:10.1093/nar/gkr944.
36. Marin FR. Variedades - EMBRAPA. 2022. Available from: https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana/pre-producao/caracteristicas/variedades. Accessed March 2024.
37. Vettore AL. Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. Genome Research. 2003;13:2725–2735. doi:10.1101/gr.1532103.
38. Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, et al. De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. PLoS ONE. 2014;9:e88462. https://doi.org/10.1371/journal.pone.0088462.
39. Nishiyama MY, Ferreira SS, Tang P-Z, Becker S, Pörtner-Taliana A, Souza GM. Full-Length Enriched cDNA Libraries and ORFeome Analysis of Sugarcane Hybrid and Ancestor Genotypes. PLoS ONE. 2014;9:e107351. doi:10.1371/journal.pone.0107351.
40. Mattiello L, Riaño-Pachón DM, Martins MCM, da Cruz LP, Bassi D, Marchiori PER, et al. Physiological and transcriptional analyses of developmental stages along sugarcane leaf. BMC Plant Biol. 2015;15:300. doi:10.1186/s12870-015-0694-z
41. Aitken KS, Jackson PA, McIntyre CL. Quantitative trait loci identified for sugar related traits in a sugarcane (Saccharum spp.) cultivar × *Saccharum* officinarum population. Theor Appl Genet. 2006;112:1306–1317. doi:10.1007/s00122-006-0233-2.

42. Costa EA, Anoni CO, Mancini MC, Santos FRC, Marconi TG, Gazaffi R, et al. QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny. Euphytica. 2016;211:1–16. doi:10.1007/s10681-016-1746-7.
43. Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280 [version 2; peer review: 2 approved]. F1000 Research. 2017;6:861.
44. Figueira TR, Okura V, Rodrigues da Silva F, Jose da Silva M, Kudrna D, Ammiraju JS, et al. A BAC library of the SP80-3280 sugarcane variety (Saccharum sp.) and its inferred microsynteny with the sorghum genome. BMC Res Notes. 2012;5:185. doi:10.1186/1756-0500-5-185.
45. Sforça DA, Vautrin S, Cardoso-Silva CB, Mancini MC, Romero-da Cruz MV, Pereira GS, et al. Gene Duplication in the Sugarcane Genome: A Case Study of Allele Interactions and Evolutionary Patterns in Two Genic Regions. Front Plant Sci. 2019;10:553. doi:10.3389/fpls.2019.00553.
46. Mancini MC, Leite DC, Perecin D, Bidóia MAP, Xavier MA, Landell MGA, et al. Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. Sugar Tech. 2012;14:119–125. doi:10.1007/s12355-012-0141-5.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215(3):403-410. doi:10.1016/s0022-2836(05)80360-2.
48. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–736. doi:10.1101/gr.215087.116.
49. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–W268. doi:10.1093/nar/gkm286.
50. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 2006;7:474. doi:10.1186/1471-2105-7-474.
51. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–945. doi:10.1093/bioinformatics/16.10.944.
52. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the National Center for Biotechnology

Information. Nucleic Acids Res. 2022 Jan 7;50(D1):D20-D26. doi:10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

53. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan 6;51(D1):D523-D531. doi:10.1093/nar/gkac1052.
54. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D412-D419. doi:10.1093/nar/gkaa913.
55. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: www.bioinformatics.babraham.ac.uk/projects/fastqc.
56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
57. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-419. doi:10.1038/nmeth.4197
58. Kolde R. Pheatmap: pretty heatmaps. R package version. 2012;1(2):726.
59. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2011.
60. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140.
61. Mutwil M, Usadel B, Schütte M, Loraine A, Ebenhöh O, Persson S. Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. Plant Physiology. 2009;152:29-43. doi:10.1104/pp.109.145318.
62. de Oliveira FA, Vigna BB, da Silva CC, Fávero AP, de Matta FP, Azevedo AL, de Souza AP. Coexpression and transcriptome analyses identify active apomixis-related genes in Paspalum notatum leaves. BMC Genomics. 2020;21:1-15.
63. Rosolen RR, Aono AH, Almeida DA, Ferreira Filho JA, Horta MAC, De Souza AP. Network analysis reveals different cellulose degradation strategies across Trichoderma harzianum strains associated with XYR1 and CRE1. Front Genet. 2022;13:807243.
64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

65. Federico ML, Diniz AL, Souza GM, Snowdon R, Erazzú L. Translational genomics from sorghum (*Sorghum bicolor*) to sugarcane (*Saccharum* spp.) for bioenergy breeding. Plant Breeding. 2022;141(3):389-398.
66. Phytozome 13 - Miscanthus sinensis v7.1 DOE-JGI [Internet]. Available from: http://phytozome.jgi.doe.gov/.
67. Zhang G, Ge C, Xu P, Wang S, Cheng S, Han Y, Wang Y, Zhuang Y, Hou X, Yu T, Xu X, Deng S, Li Q, Yang Y, Yin X, Wang W, Liu W, Zheng C, Sun X, Wang Z, Ming R, Dong S, Ma J, Zhang X, Chen C. The reference genome of Miscanthus floridulus illuminates the evolution of Saccharinae. Nat Plants. 2021 May;7(5):608-618. doi:10.1038/s41477-021-00908-y. Epub 2021 May 6. Erratum in: Nat Plants. 2021 Jul;7(7):991. Erratum in: Nat Plants. 2021 Jul;7(7):990. PMID: 33958777; PMCID: PMC8238680.
68. Miao J, Feng Q, Li Y, et al. Chromosome-scale assembly and analysis of biomass crop Miscanthus lutarioriparius genome. Nat Commun. 2021;12:2458. doi:10.1038/s41467-021-22738-4.
69. Zhao D, Fergunson AA, Jiang N. What makes up plant genomes: The vanishing line between transposable element and genes. Biochim Biophys Acta. 2016;1859:366–380.
70. Redd PS, Payero L, Gilbert DM, Page CA, King R, McAssey EV, Bodie D, Diaz S, Hancock N. Transposase expression, element abundance, element size, and DNA repair determine the mobility and heritability of PIF/Pong/Harbinger transposable elements. Front Cell Dev Biol. 2023. doi:10.3389/fcell.2023.1184046.
71. Takatsuji H. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. Plant Mol Biol. 1999.
72. Singh A. Benefits of crop diversification in Fiji's sugarcane farming. Asia & the Pacific Policy Studies. 2020;7(1):65-80. doi:10.1002/app5.291.
73. Garcia FHS, Mendonça AMC, Rodrigues M, Matias FI, Filho MPS, Santos HRB, Taffner J, Barbosa JPRAD. Water deficit tolerance in sugarcane is dependent on the accumulation of sugar in the leaf. Ann Appl Biol. 2020;176(1). doi:10.1111/aab.12559.
74. De Souza AP, Grandis A, Arenque-Musa BC, Buckeridge MS. Diurnal variation in gas exchange and nonstructural carbohydrates throughout sugarcane development. Funct Plant Biol. 2018 Jul;45(8):865-876. doi:10.1071/FP17268. PMID: 32291068.
75. Chen K, Li GJ, Bressan RA, Song CP, Zhu JK, Zhao Y. Abscisic acid dynamics, signaling, and functions in plants. J Integr Plant Biol. 2020 Jan;62(1):25-54. doi:10.1111/jipb.12899. PMID: 31850654.

76. Smit MA, Singels A. The response of sugarcane canopy development to water stress. Field Crops Res. 2006;98(2–3):91-97. doi:10.1016/j.fcr.2005.12.009.
77. Aluko OO, Li C, Wang Q, Liu H. Sucrose utilization for improved crop yields: A review article. Int J Mol Sci. 2021;22(9):4704. doi:10.3390/ijms22094704.
78. van der Vlugt-Bergmans CJ, van der Werf MJ. Genetic and biochemical characterization of a novel monoterpene epsilon-lactone hydrolase from Rhodococcus erythropolis DCL14. Appl Environ Microbiol. 2001 Feb;67(2):733-741. doi:10.1128/AEM.67.2.733-741.2001. PMID: 11157238; PMCID: PMC92642
79. Mindrebo JT, Nartey CM, Seto Y, Burkart MD, Noel JP. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Curr Opin Struct Biol. 2016 Dec;41:233-246
80. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18(6):292-298. doi:10.1016/S0169-5347(03)00033-8.
81. FAO. The State of Food Security and Nutrition in the World 2023. Food and Agriculture Organization of the United Nations. 2023. Disponível em: https://www.fao.org/3/cc8589en/cc8589en_sugar.pdf.
82. Ma H, Yang T, Li Y, Zhang J, Wu T, Song T, Yao Y, Tian J. The long noncoding RNA MdLNC499 bridges MdWRKY1 and MdERF109 function to regulate early-stage light-induced anthocyanin accumulation in apple fruit. Plant Cell. 2021 Oct 11;33(10):3309-3330. doi:10.1093/plcell/koab188. PMID: 34270784; PMCID: PMC8505877.
83. Guarneri N, Willig JJ, Sterken MG, Zhou W, Hasan MS, Sharon L, Grundler FMW, Willemsen V, Goverse A, Smant G, Lozano-Torres JL. Root architecture plasticity in response to endoparasitic cyst nematodes is mediated by damage signaling. New Phytol. 2023 Feb;237(3):807-822. doi:10.1111/nph.18570. Epub 2022 Dec 1. PMID: 36285401; PMCID: PMC10108316.
84. Yu G, Sun B, Zhu Z, Mehareb EM, Teng A, Han J, Zhang H, Liu J, Liu X, Raza G, Zhang B, Zhang Y, Wang K. Genome-wide DNase I-hypersensitive site assay reveals distinct genomic distributions and functional features of open chromatin in autopolyploid sugarcane. Plant J. 2024 Jan;117(2):573-589. doi:10.1111/tpj.16513. Epub 2023 Oct 27. PMID: 37897092.
85. Wang M, Dai W, Du J, Ming R, Dahro B, Liu JH. ERF109 of trifoliate orange (Poncirus trifoliata (L.) Raf.) contributes to cold tolerance by directly regulating expression of Prx1 involved in antioxidative process. Plant Biotechnol J. 2019;17:1316–1332.

86. Dong Z, Yang S, Zhang Z, Tang C, Kan Y, Yao L. Prolyl aminopeptidases: Reclassification, properties, production and industrial applications. Process Biochem. 2022;118:121-132. doi:10.1016/j.procbio.2022.04.025.
87. Ghifari AS, Teixeira PF, Kmiec B, Singh N, Glaser E, Murcha MW. The dual-targeted prolyl aminopeptidase PAP1 is involved in proline accumulation in response to stress and during pollen development. J Exp Bot. 2022 Jan 5;73(1):78-93. doi:10.1093/jxb/erab397. PMID: 34460901.
88. Rejeb KB, Abdelly C, Savouré A. How reactive oxygen species and proline face stress together. Plant Physiol Biochem. 2014;80:278-284. doi:10.1016/j.plaphy.2014.04.007.
89. Vasantha S, Kumar RA, Tayade AS, et al. Physiology of Sucrose Productivity and Implications of Ripeners in Sugarcane. Sugar Tech. 2022;24:715–731. doi:10.1007/s12355-021-01062-7.
90. Kazemi-Shahandashti K, Maali-Amiri R. Global insights of protein responses to cold stress in plants: Signaling, defence, and degradation. J Plant Physiol. 2018;226:123-135. doi:10.1016/j.jplph.2018.03.022.
91. Sharma P, Sharma P, Arora P, Verma V, Khanna K, Saini P, et al. Role and Regulation of ROS and Antioxidants as Signaling Molecules in Response to Abiotic Stresses. In: Iqbal M, Khan R, Sudhakar Reddy P, Ferrante A, Khan NA, editors. Plant Signaling Molecules. Woodhead Publishing; 2019. p. 141-156. doi:10.1016/B978-0-12-816451-8.00008-3.
92. Kavi Kishor PB, Suravajhala P, Rathnagiri P, Sreenivasulu N. Intriguing Role of Proline in Redox Potential Conferring High Temperature Stress Tolerance. Front Plant Sci. 2022 Jun 10;13:867531. doi: 10.3389/fpls.2022.867531. PMID: 35795343; PMCID: PMC9252438
93. Ghosh UK, Islam MN, Siddiqui MN, Cao X, Khan MAR. Proline, a multifaceted signalling molecule in plant responses to abiotic stress: understanding the physiological mechanisms. Plant Biol (Stuttg). 2022 Mar;24(2):227-239. doi: 10.1111/plb.13363. Epub 2021 Nov 18. PMID: 34796604.
94. Ghifari AS, Teixeira PF, Kmiec B, Singh N, Glaser E, Murcha MW. The dual-targeted prolyl aminopeptidase PAP1 is involved in proline accumulation in response to stress and during pollen development. J Exp Bot. 2022 Jan 5;73(1):78-93. doi:10.1093/jxb/erab397. PMID: 34460901.
95. Ali Q, Anwar F, Ashraf M, Saari N, Perveen R. Ameliorating effects of exogenously applied proline on seed composition, seed oil quality and oil antioxidant activity of

maize (*Zea mays* L.) under drought stress. Int J Mol Sci. 2013;14:818-835. doi:10.3390/ijms14010818.

96. Singels A, Jackson P, Inman-Bamber G. Sugarcane. In: Sadras VO, Calderini DF, editors. Crop Physiology Case Histories for Major Crops. Academic Press; 2021. p. 674-713. doi:10.1016/B978-0-12-819194-1.00021-9
97. Gurrieri L, Merico M, Trost P, Forlani G, Sparla F. Impact of drought on soluble sugars and free proline content in selected Arabidopsis mutants. Biology. 2020;9:367. DOI: 10.3390/biology9110367.
98. Krochko JE, Abrams GD, Loewen MK, Abrams SR, Cutler AJ. (+)-Abscisic acid 8'-hydroxylase is a cytochrome P450 monooxygenase. Plant Physiol. 1998 Nov;118(3):849-860. doi:10.1104/pp.118.3.849. PMID: 9808729; PMCID: PMC34795.
99. Wang X, Jin B, Yan W, Wang J, Xu J, Cai C, Qi X, Xu Q, Yang X, Xu X, Chen X. Cucumber abscisic acid 8′-hydroxylase Csyf2 regulates yellow flesh by modulating carotenoid biosynthesis. Plant Physiol. 2023;193(2):1001–1015. doi:10.1093/plphys/kiad383.
100. Tomiyama H, Sato M, Opio P, et al. Inhibition of Abscisic Acid 8′-Hydroxylase Affects Dehydration Tolerance and Root Formation in Cuttings of Grapes (Vitis labrusca L. × Vitis vinifera L. cv. Kyoho) Under Drought Stress Conditions. J Plant Growth Regul. 2020;39:1577–1586. doi:10.1007/s00344-020-10171-8.
101. Sajid M, Amjid M, Munir H, Ahmad M, Zulfiqar U, Ali MF, Farah MA, Ahmed MAA, Artyszak A. Comparative Analysis of Growth and Physiological Responses of Sugarcane Elite Genotypes to Water Stress and Sandy Loam Soils. Plants. 2023; 12(15):2759. https://doi.org/10.3390/plants12152759
102. Gutierrez L, Castelain M, Verdeil JL, Conejero G, Van Wuytswinkel O. A possible role of prolyl oligopeptidase during Linum usitatissimum (flax) seed development. Plant Biol (Stuttg). 2008 May;10(3):398-402. doi: 10.1111/j.1438-8677.2008.00038.x. PMID: 18426487.
103. Baharin A, Ting TY, Goh HH. Post-Proline Cleaving Enzymes (PPCEs): Classification, Structure, Molecular Properties, and Applications. Plants (Basel). 2022;11(10):1330. https://doi.org/10.3390/plants11101330.
104. Singh R, Irikura B, Nagai C, et al. Characterization of Prolyl Oligopeptidase Genes Differentially Expressed Between Two Cultivars of Coffea arabica L.. Tropical Plant Biol. 2011;4:203–216. https://doi.org/10.1007/s12042-011-9082-5.

105. Tan CM, Chen RJ, Zhang JH, Gao XL, Li LH, Wang PR, Deng XJ, Xu ZJ. OsPOP5, a prolyl oligopeptidase family gene from rice confers abiotic stress tolerance in Escherichia coli. Int J Mol Sci. 2013 Oct 10;14(10):20204-19. doi: 10.3390/ijms141020204. PMID: 24152437; PMCID: PMC3821611.

106. Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, et al. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. New Phytol. 2011;189:629–642. doi:10.1111/j.1469-8137.2010.03497.x.

107. DiMario RJ, Giuliani R, Ubierna N, Slack AD, Cousins AB, Studer AJ. Lack of leaf carbonic anhydrase activity eliminates the C4 carbon-concentrating mechanism requiring direct diffusion of CO2 into bundle sheath cells. Plant Cell Environ. 2022 May;45(5):1382-1397. doi:10.1111/pce.14291.

108. Raffaele S, Perraki A, Mongrand S. The Remorin C-terminal Anchor was shaped by convergent evolution among membrane binding domains. Plant Signaling & Behavior. 2013;8(3). doi:10.4161/psb.23207

109. Gui J, Zheng S, Shen J, Li L. Grain setting defect1 (GSD1) function in rice depends on S-acylation and interacts with actin 1 (OsACT1) at its C-terminal. Front Plant Sci. 2015 Oct 2;6:804.

110. Ikeda M, Mitsuda N, Ohme-Takagi M. Arabidopsis HsfB1 and HsfB2b Act as Repressors of the Expression of Heat-Inducible Hsfs But Positively Regulate the Acquired Thermotolerance. Plant Physiol. November 2011;157:1243–1254. doi:10.1104/pp.111.183659.

111. Fragkostefanakis S, Simm S, El-Shershaby A, Hu Y, Bublak D, Mesihovic A, Darm K, Mishra SK, Tschiersch B, Theres K, Scharf C, Schleiff E, Scharf KD. The repressor and co-activator HsfB1 regulates the major heat stress transcription factors in tomato. Plant Cell Environ. 2019;42:874–890. doi:10.1111/pce.13467

112. Chen H, Liu X, Li S, Yuan L, Mu H, Wang Y, Li Y, Duan W, Fan P, Liang Z, Wang L. The class B heat shock factor HSFB1 regulates heat tolerance in grapevine. Hortic Res. 2023;10:uad001. doi:10.1093/hr/uhad001.

113. Ng LM, Melcher K, Teh BT, Xu HE. Abscisic acid perception and signaling: structural mechanisms and applications. Acta Pharmacol Sin. 2014 May;35(5):567-84. doi:10.1038/aps.2014.5. PMID: 24786231; PMCID: PMC4813750.

114. Wang M, Dai W, Du J, Ming R, Dahro B, Liu JH. ERF109 of trifoliate orange (Poncirus trifoliata (L.) Raf.) contributes to cold tolerance by directly regulating

expression of Prx1 involved in antioxidative process. Plant Biotechnol J. 2019;17:1316–1332.

115. Bahieldin A, Atef A, Edris S, Gadalla NO, Ramadan AM, Hassan SM, Al Attas SG, Al-Kordy MA, Al-Hajar AS, Sabir JSM, Nasr ME, Osman GH, El-Domyati FM. Multifunctional activities of ERF109 as affected by salt stress in Arabidopsis. Sci Rep. 2018;8:6403. doi:10.1038/s41598-018-24452-6.

116. Zhao M, Ma L, Song N, Cheng J, Zhao Z, Wu J. The regulation of Alternaria alternata resistance by LRR-RK4 through ERF109, defensin19 and phytoalexin scopoletin in Nicotiana attenuata. Plant Sci. 2022;323:111414. doi:10.1016/j.plantsci.2022.111414.

117. Phytozome. Phytozome Homepage [Internet]. Berkeley (CA): Joint Genome Institute; c2022 [cited 2024 May 13]. Available from: https://phytozome-next.jgi.doe.gov/

118. Teresinski HJ, Hau B, Symonds K, Kilburn R, Munro KA, Doner NM, Mullen R, Li VH, Snedden WA. Arabidopsis calmodulin-like proteins CML13 and CML14 interact with proteins that have IQ domains. Plant Cell Environ. 2023;46:2470–2491. DOI: 10.1111/pce.14616. Received: 15 September 2022 | Accepted: 9 May 2023.

119. EMBL. EMBL Homepage [Internet]. Heidelberg (DE): European Molecular Biology Laboratory; c2022 [cited 2024 May 13]. Available from: https://www.embl.org/

120. Feng Y, Hiwatashi T, Minamino N, Ebine K, Ueda T. Membrane trafficking functions of the ANTH/ENTH/VHS domain-containing proteins in plants. FEBS Lett. 2022;596:2256–2268.

121. Jiang H, Ma QJ, Zhong MS, Gao HN, Li YY, Hao YJ. The apple palmitoyltransferase MdPAT16 influences sugar content and salt tolerance via an MdCBL1–MdCIPK13–MdSUT2.2 pathway. The Plant Journal. 2021;106:689–705.

122. Bozzi, A. T., & Gaudet, R. (2021). Molecular Mechanism of Nramp-Family Transition Metal Transport. Journal of Molecular Biology, 433(16), 166991. https://doi.org/10.1016/j.jmb.2021.166991

123. Zhou XN, Zheng YY, Cai Z, Wang XY, Liu Y, Yu A, Chen XL, Liu J, Zhang Y, Wang AX. Identification and functional analysis of tomato TPR gene family. Int J Mol Sci. 2021;22:758. DOI: 10.3390/ijms22020758

Figure 1

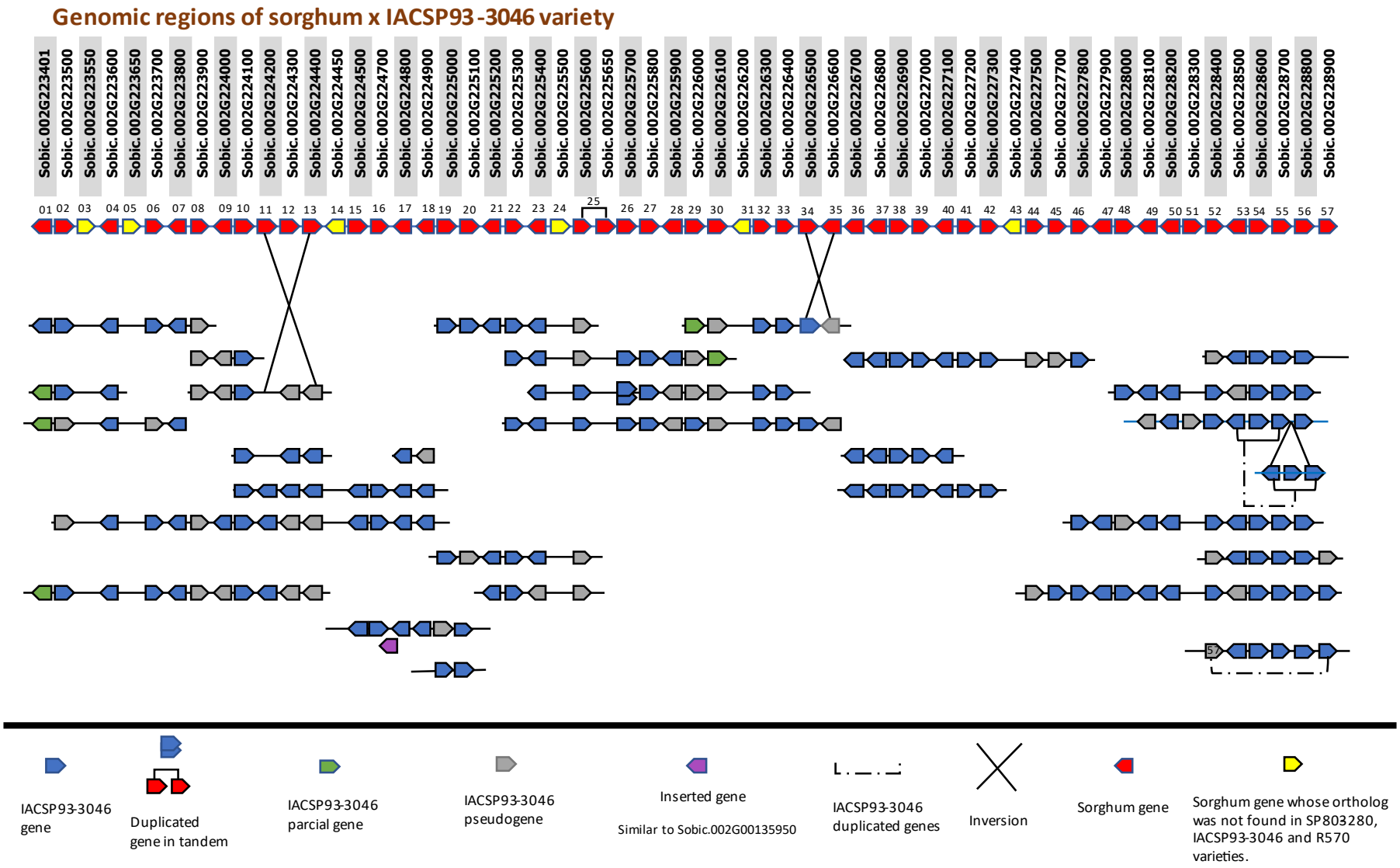

Genomic regions of sorghum x IACSP93-3046 variety
Sobic.002G223401
Sobic.002G223500
Sobic.002G223550
Sobic.002G223600
Sobic.002G223650
Sobic.002G223700
Sobic.002G223800
Sobic.002G223900
Sobic.002G224000
Sobic.002G224100
Sobic.002G224200
Sobic.002G224300
Sobic.002G224400
Sobic.002G224450
Sobic.002G224500
Sobic.002G224700
Sobic.002G224800
Sobic.002G224900
Sobic.002G225000
Sobic.002G225100
Sobic.002G225200
Sobic.002G225300
Sobic.002G225400
Sobic.002G225500
Sobic.002G225600
Sobic.002G225650
Sobic.002G225700
Sobic.002G225800
Sobic.002G225900
Sobic.002G226000
Sobic.002G226100
Sobic.002G226200
Sobic.002G226300
Sobic.002G226400
Sobic.002G226500
Sobic.002G226600
Sobic.002G226700
Sobic.002G226800
Sobic.002G226900
Sobic.002G227000
Sobic.002G227100
Sobic.002G227200
Sobic.002G227300
Sobic.002G227400
Sobic.002G227500
Sobic.002G227700
Sobic.002G227800
Sobic.002G227900
Sobic.002G228000
Sobic.002G228100
Sobic.002G228200
Sobic.002G228300
Sobic.002G228400
Sobic.002G228500
Sobic.002G228600
Sobic.002G228700
Sobic.002G228800
Sobic.002G228900
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
IACSP93-3046 gene
Duplicated gene in tandem
IACSP93-3046 parcial gene
IACSP93-3046 pseudogene
Inserted gene
Similar to Sobic.002G00135950
IACSP93-3046 duplicated genes
Inversion
Sorghum gene
Sorghum gene whose ortholog was not found in SP803280, IACSP93-3046 and R570 varieties.

Figure 1. *Genomic regions of the sorghum × IACSP93-3046 variety.* Genomic region of sorghum × IACSP93-3046. Each pentagon represents a gene and shows the orientation of the gene in the genome, where right-facing arrows indicate the sense direction and left-facing arrows indicate the antisense direction. The solid lines with pentagons in red and yellow represent the QTL genes of sorghum, and the yellow genes lacking orthologs in the SP80-3280, IACSP93-3046, and R570 varieties. For each gene, a number has been designated, and the name assigned to it in Phytozome v.13 is indicated above each gene. The genomic region of the IACSP93-3046 variety is shown below. Each solid line below the genomic region of sorghum represents a haplotype, and each haplotype may be formed by one or two BACs (Supplementary Table 4). In the IACSP93-3046 genomic region, 50 genes between genes and pseudogenes were recovered. The gene represented in purple, just below the line of one of the haplotypes, is an exclusive finding within this BAC, indicating that it was observed between the two genes immediately above, namely, genes 16 and 17. This gene is not one of the 57 QTL genes of sorghum; however, it is similar to the sorghum gene Sobic.002G00135950 (Phytozome v.13). A pseudogene orthologous to this gene was observed on chromosome Sspon2B (Supplementary Figure 2). The inversions are indicated by crossed solid lines and occurred between genes 11 and 13 and between genes 34 and 35; the haplotypes below also have the same inversion represented above. Gene 57 is duplicated, and the duplicate is pseudogenized out of the QTL order and is indicated by a gray pentagon with the number 57. The haplotypes were used to reconstruct the genomic region of the IACSP93-3046 variety.

Figure 2

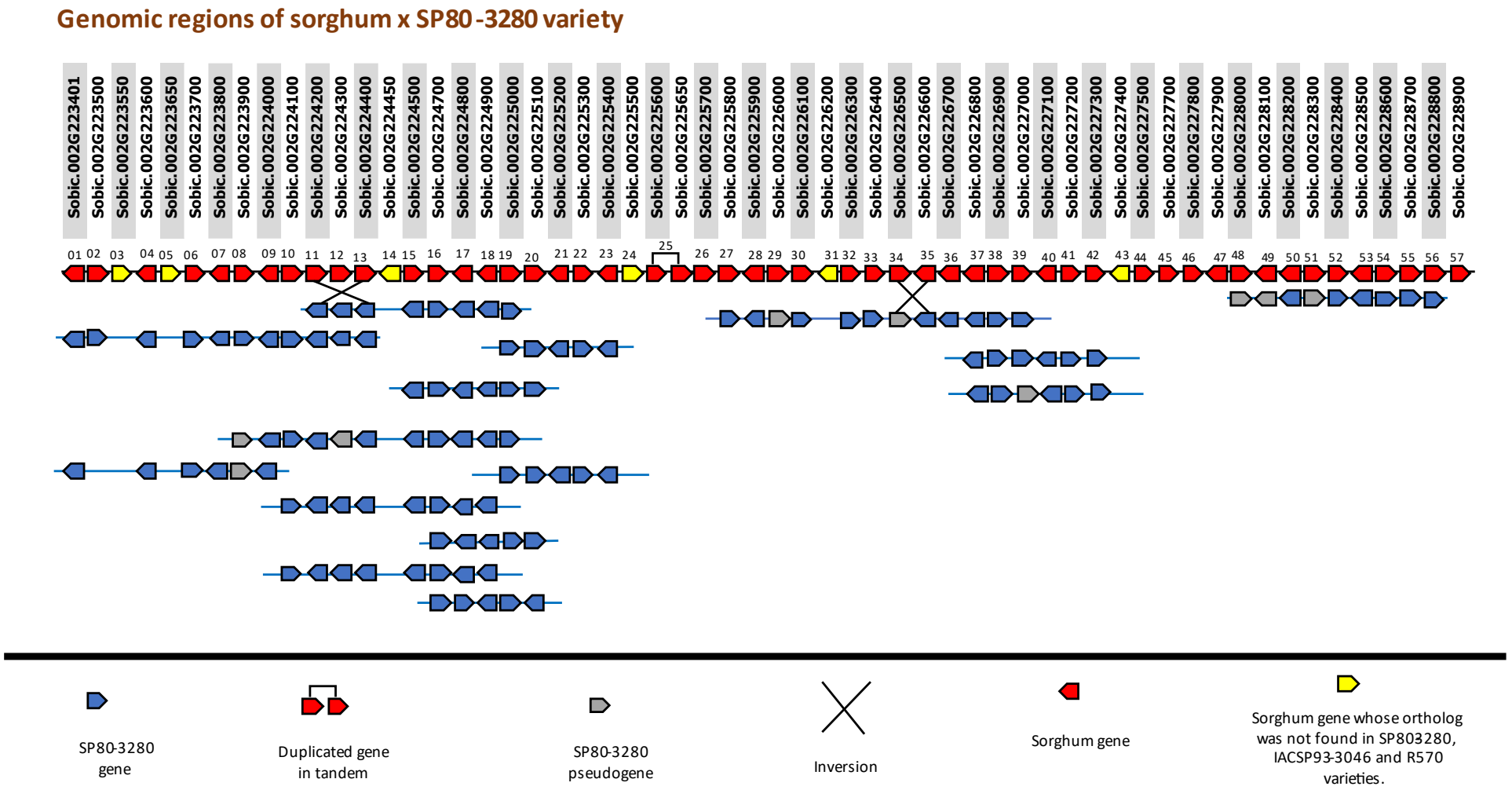

Genomic regions of sorghum x SP80-3280 variety
Sobic.002G223401
Sobic.002G223500
Sobic.002G223550
Sobic.002G223600
Sobic.002G223650
Sobic.002G223700
Sobic.002G223800
Sobic.002G223900
Sobic.002G224000
Sobic.002G224100
Sobic.002G224200
Sobic.002G224300
Sobic.002G224400
Sobic.002G224450
Sobic.002G224500
Sobic.002G224700
Sobic.002G224800
Sobic.002G224900
Sobic.002G225000
Sobic.002G225100
Sobic.002G225200
Sobic.002G225300
Sobic.002G225400
Sobic.002G225500
Sobic.002G225600
Sobic.002G225650
Sobic.002G225700
Sobic.002G225800
Sobic.002G225900
Sobic.002G226000
Sobic.002G226100
Sobic.002G226200
Sobic.002G226300
Sobic.002G226400
Sobic.002G226500
Sobic.002G226600
Sobic.002G226700
Sobic.002G226800
Sobic.002G226900
Sobic.002G227000
Sobic.002G227100
Sobic.002G227200
Sobic.002G227300
Sobic.002G227400
Sobic.002G227500
Sobic.002G227700
Sobic.002G227800
Sobic.002G227900
Sobic.002G228000
Sobic.002G228100
Sobic.002G228200
Sobic.002G228300
Sobic.002G228400
Sobic.002G228500
Sobic.002G228600
Sobic.002G228700
Sobic.002G228800
Sobic.002G228900
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
SP80-3280 gene
Duplicated gene in tandem
SP80-3280 pseudogene
Inversion
Sorghum gene
Sorghum gene whose ortholog was not found in SP803280, IACSP93-3046 and R570 varieties.

Figure 2. *Genomic regions of the sorghum × SP80-3280 variety.* Each pentagon represents a gene and shows the orientation of the gene in the genome, where right-facing arrows indicate the sense direction and left-facing arrows indicate the antisense direction. The solid lines with pentagons in red and yellow represent the QTL genes of sorghum, and the yellow genes lacking orthologs in the SP80-3280, IACSP93-3046, and R570 varieties. For each gene, a number has been designated (Supplementary Table 2), and the name assigned to it in Phytozome v.13 is indicated above each gene. A genomic region of the SP80–3280 variety is presented below. Each solid line represents a haplotype, and the blue and gray pentagons represent the genes and pseudogenes of the SP80-3280 variety simultaneously. Each haplotype is formed by one or two BACs (Supplementary Table 4). A total of 44 genes were recovered between genes and pseudogenes. Inversions are represented by solid crossed lines and occur between genes 11 and 13 and between genes 34 and 35. All haplotypes below the first inversion representation have similar inversions. The haplotypes reconstructed the genomic region of the SP80–3280 variety.

Figure 3

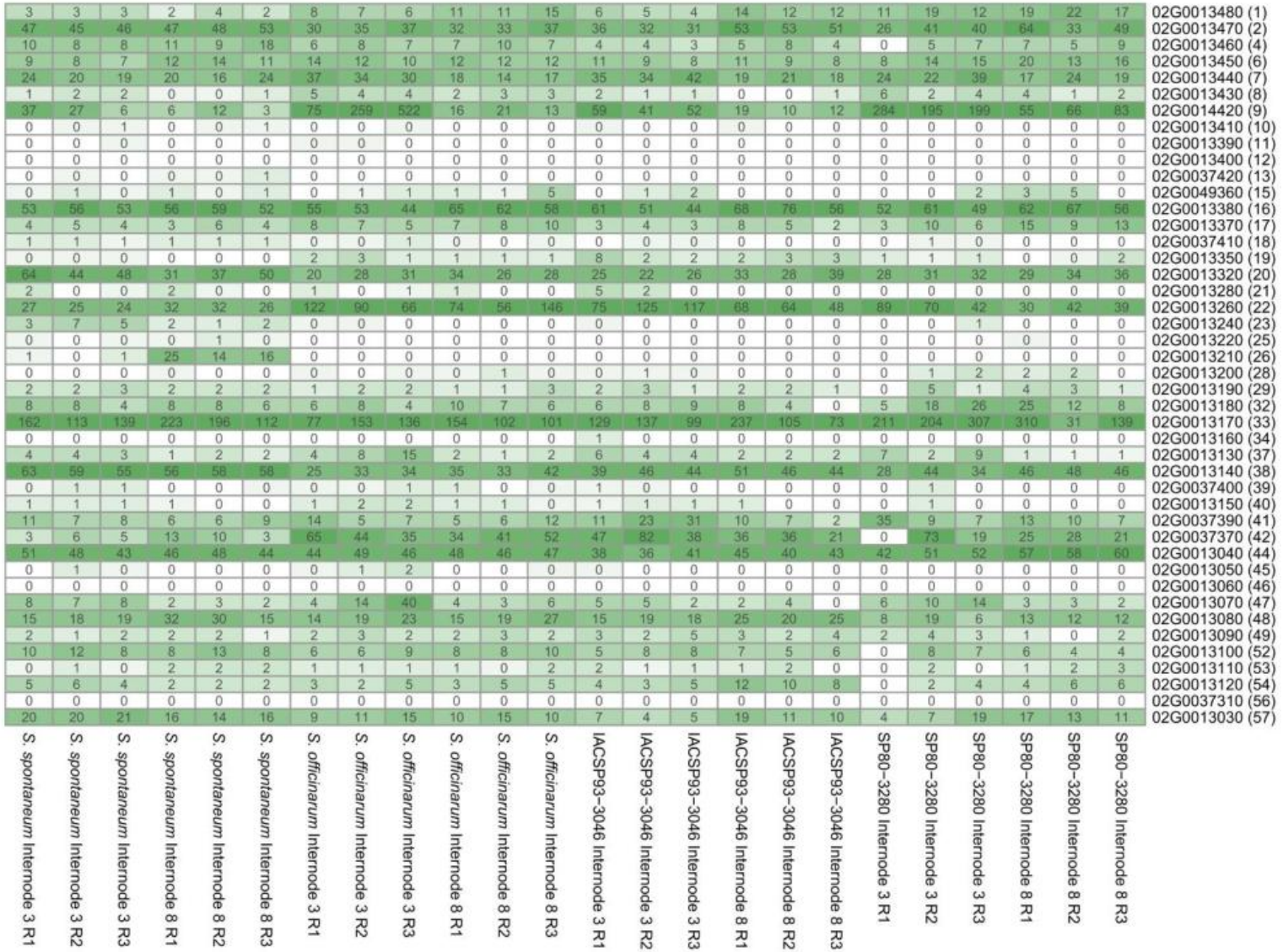


Figure 3. *Heatmap representing the level of gene expression in the internodes of SP80-3280, IACSP93-3046, S. officinarum and S. spontaneum.* A heatmap depicting gene expression levels across tissues and internodes 3 (top) and 8 (bottom) of sugarcane plants (varieties SP80-3280 and IACSP93-3046*; S. officinarum* and *S. spontaneum*) in triplicate. The darker the shade of green is, the higher the expression level.

Table 1. Differentially expressed genes (DEGs) between non-accumulating sugar plants (*S. spontaneum*) and accumulating sugar plants (SP80-3280 and IACSP93-3046 varieties, and *S. officinarum*) located within the QTL under analysis.

Gene number	log2(FC)	P value	*S. spontaneum* gene/protein and description
01	2.049208548	1.5015E-06	**Sspon.02G0013480—**Similar to prolyl oligopeptidase (POP), an enzyme that cleaves oligopeptides up to 30 amino acid residues postproline [102].
09	2.826212214	0.003517002	**Sspon.02G0014420—**The conversion of CO_2 and water to bicarbonate and the release of a proton is catalyzed by carbonic anhydrase (CA), the first enzymatic step of photosynthesis in C4 plants. This reaction takes place in the mesophyll cells. Bicarbonate will initiate the first carboxylation of C4 [106].
19	3.472221834	0.006550842	**Sspon.02G0013350—**They are membrane proteins, and part of the so-called lipid rafts (raft proteins—protein groups of membrane proteins that resemble a boat on the lipid group) [107]. Remorines may be associated with the regulation and translocation of photoassimilates. A specific type, GSD1, belonging to group 6 of remorines, has the function of regulating the conductance of photoassimilates through plasmodesmata in rice [108].
22	1.306514099	0.00043504	**Sspon.02G0013260—**B-type HSF are classically transcriptional repressor proteins [110]. In grape HSFB1 has expression induced by heat stress. In rice, HSFB1 expression was related to cold. The gene has increased transcription when the plant is undergoing abiotic stresses, however, they act mainly as transcriptional repressors [111] [112].
23	-5.246110673	0.009820214	**Sspon.02G0013240—**ABA 8'-hydroxylase (ABA 8’OH) is an enzyme involved in the catabolism of the hormone ABA (abscisic acid). ABA is inactive when it has a hydroxyl group (OH) at its 8' position, while its removal makes it active. ABA is involved in the closure of stomata, which are crucial structures in photosynthetic processes as they regulate gas exchange between the environment and the plant [113].
26	-17.55155671	5.38665E-05	**Sspon.02G0013210—**Studies suggest that ERF109 is a positive regulator of cold tolerance [114]. Other classic abiotic stresses also regulate ERF109 [115]. The gene, in joint action, also regulates defense against some pathogens [116].
42	2.410548623	0.005106978	**Sspon.02G0037370—**Uncharacterized protein—predicted [117]

Table 2. Sugarcane genes and their first neighbors identified by coexpression network analysis. The first neighbor genes are defined by their names in *S. spontaneum*, and their descriptions are based on the annotations of the EMBL database.

Gene Number	Protein	First neighbors/description

	(Phytozome v.13) [45]	
01	**POP -**Similar to prolyl oligopeptidase family	**Sspon.01G0011190—**The gene shares similarity with the IQ motif, known for its interaction with calmodulins in plant cells. This interaction influences the function of the target protein, particularly in cytoskeletal processes and cellular development, with a primary role in regulating signaling alongside CaM, CML, and CAMTA proteins [118].
		Sspon.01G0011270—No translating CDS [119]. The sequence has 73.7% identity with a protein containing the RRM domain from Miscanthus lutarioparius [117].
		Sspon.01G0011300—This gene is equipped with a ENTH domain responsible for lipid binding, pivotal in the creation of clathrin-coated vesicles, trans-Golgi network (TGN) endocytosis, and vacuolar transport, contributing to the plant's immune response. Proteins within the ANTH/ENTH/VHS family display functional redundancy, likely shaped by natural selection, necessitating mutations in at least two genes to significantly affect plant function [120].
07	Similar to a membrane protein, possibly structural. (GO:0016020)	**Sspon.01G0010650—**similar to MYB46, a transcription factor, activates genes for cellulose, hemicellulose, and lignin synthesis, interacting with other transcription factors. In apples, MYB46 overexpression enhances salt tolerance, stress response, and promotes secondary cell wall biosynthesis, including lignin deposition by binding directly to relevant gene promoters. Activated during plant stress, MYB46 regulates genes involved in both biotic and abiotic stress responses [75].
18	Protein belonging to the **GDXG** family. Similar to alpha/beta hydrolase (**ABH**).	**Sspon.01G0019580—**No translating CDS, the intronic sequence of this gene is 100% identical to the Sobic.001G217300 gene in sorghum. However, in sorghum, this gene consists of three exons, whereas in *S. spontaneum*, it contains seven exons, with the additional four exons being the first four in the sequence [117].
		Sspon.01G0019780—Segments of the gene exhibit similarities to various segments on sorghum chromosome 1, encompassing both small gene fragments and intergenic regions [117].
23	**ABA8' OH**- Similar to Abscisic acid 8'-hydroxylase	**Sspon.01G0005990—**This gene shares 94.1% similarity with *Miscanthus lutarioriparius*' s-acetyltransferase, identified as a palmitoyltransferase (PAT16). Acetyltransferases, influencing protein modification in plants, transfer acetyl groups from Acetyl-CoA. Palmitoyltransferases specifically add a 16-carbon palmitate to proteins, crucial for membrane protein function by anchoring to the cell membrane [121].
26	**ERF109-** Similar to Ethylene Responsive factor 109	**Sspon.01G0005860—**NRAMP genes in plants are key players in selectively absorbing and transporting essential transition metals during heavy metal stress. Regulated by phytohormones, these genes maintain metal balance. A research with potatoes unveils molecular insights for potential development of low-metal-accumulating plant varieties. In essence, NRAMP genes are vital for plant resilience to heavy metal stress [122].
37	**PP2C**- Similar to protein phosphatase 2C	**Sspon.01G0005530—**In S. spontaneum, there are two isoforms of this gene, one encoding NPG1 (NO pollen germination 1) and a nontranslating CDS. The latter is annotated with two additional exons compared to its sorghum counterpart. This isoform features the tetratricopeptide repeat (TPR) domain. The TPR domain is explored in a study with tomatoes, influencing cell regulation, gene expression, and stress responses. The same study with tomatoes suggests the TPR gene has a potential link to energy metabolism and acts as a mediator in disease resistance [123].

Supplementary Material

Supplementary Figures

Supplementary Figure 1. Genomic regions of sorghum x *Saccharum spontaneum* chromosome 2A, with a size of 794.054 bp

Supplementary Figure 2. Genomic region of sorghum × *Saccharum spontaneum* chromosome 2B, with a size of 1.134.230 bp

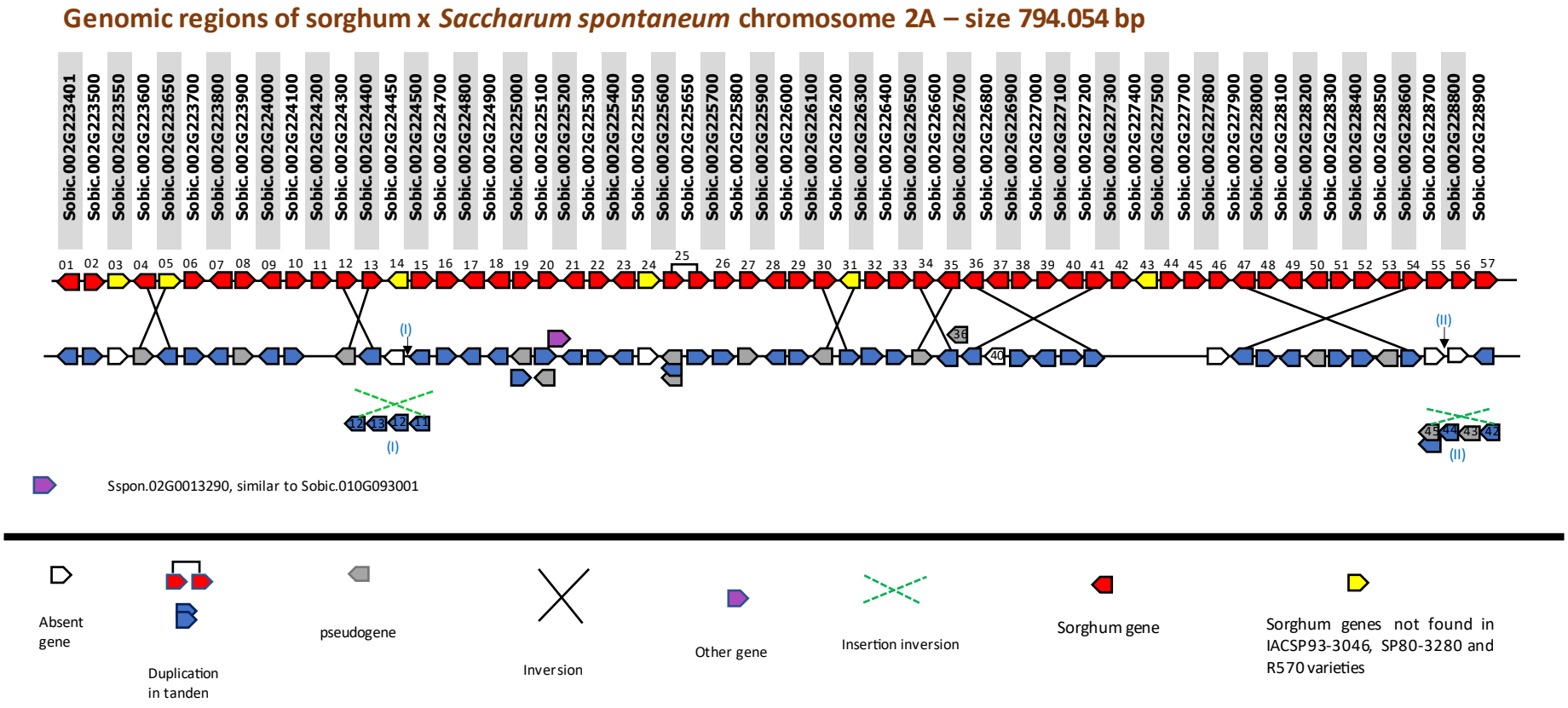
Supplementary Figure 3. Genomic region of sorghum × *Saccharum spontaneum* chromosome 2C, with a size of 969.275 bp

Supplementary Figure 4. Genomic region of sorghum x *Saccharum spontaneum* chromosome 2D – size 205.837 bp - (21 Mb) - 671.747 bp

Supplementary Figure 5. Sorghum × R570 variety genomic regions

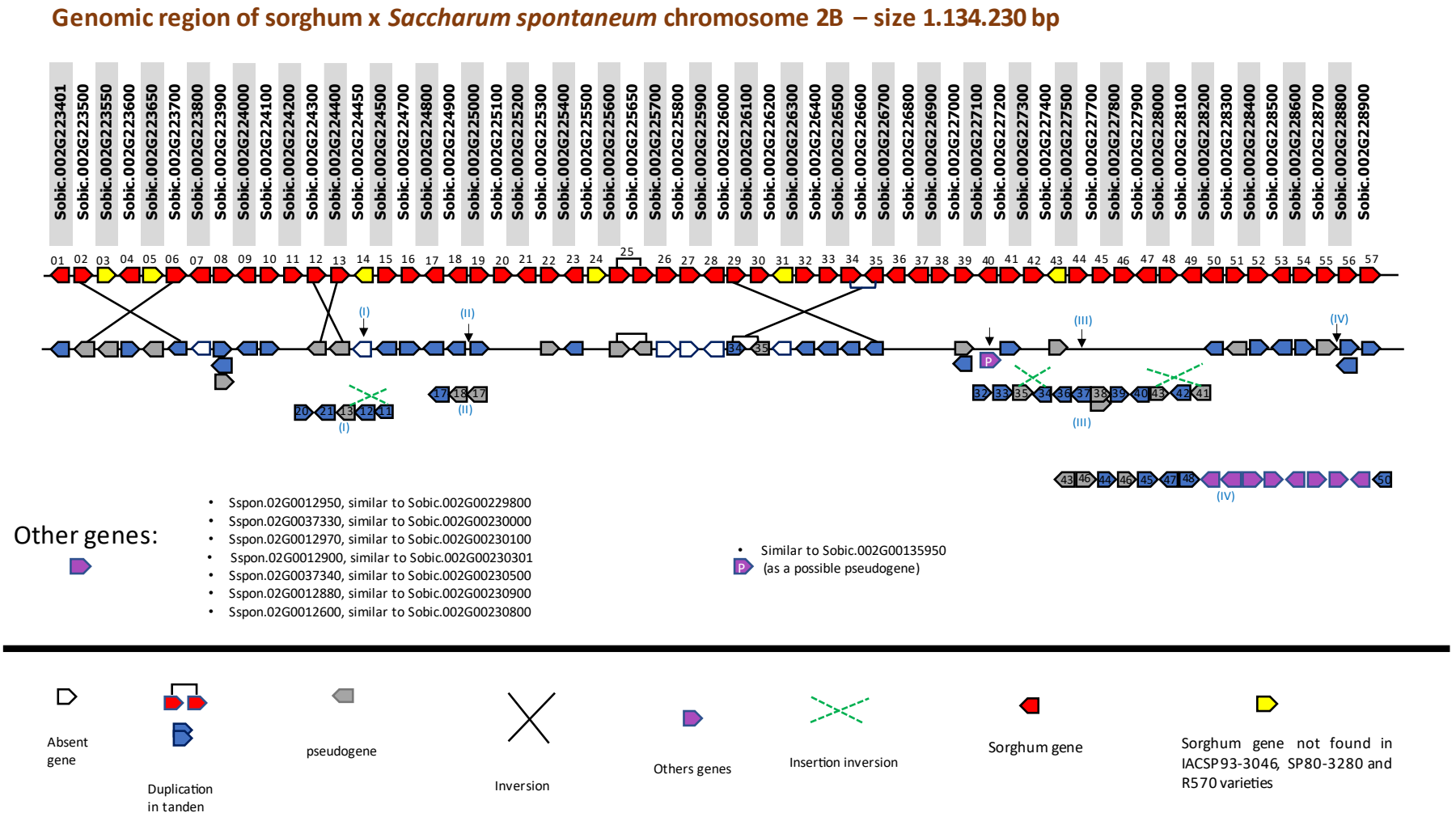
Supplementary Figure 6 Representation of the highest reciprocal rank (HRR)

Supplementary Figure 1



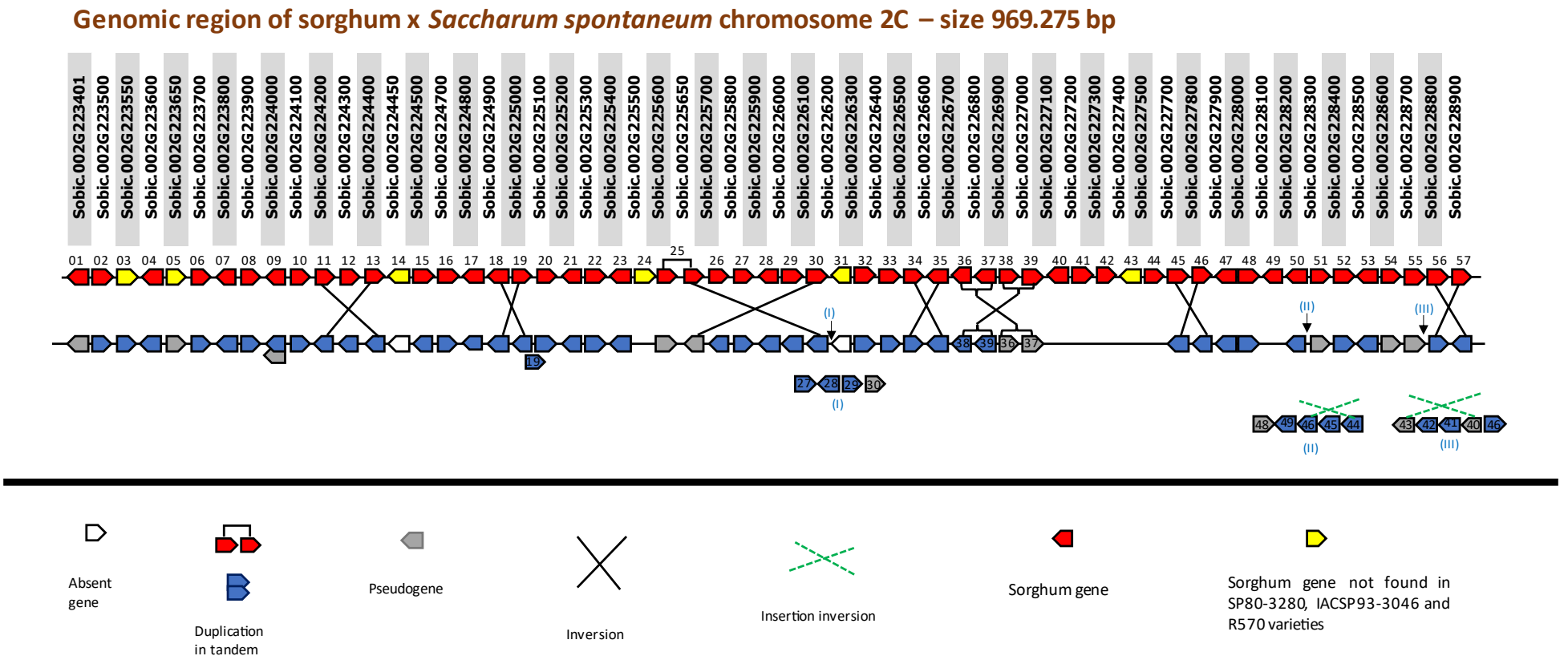
Supplementary Figure 1. Genomic regions of sorghum x *Saccharum spontaneum* chromosome 2A, with a size of 794.054 bp. *Genomic region of sorghum:* The solid line with numbered pentagons in red and yellow represents the genomic region of sorghum and the QTL genes. The gene names of sorghum (Phytozome v.13) are presented above. Each pentagon represents a gene and its orientation in the sorghum genome. The right-facing arrow indicates the sense orientation, and the left-facing arrow indicates the antisense direction. The yellow pentagons represent the sorghum genes not found in the SP80-3280, IACSP93-3046, and R570 varieties. The genes were numbered from 1 to 57 (Supplementary Table 2). Genomic region of chromosome Sspon2A: The orthologous genomic region in *S. spontaneum* for chromosome Sspon2A is presented below the representation of the sorghum genomic region. Genes are shown in blue pentagons on a solid line, and pseudogenes are shown in gray. Each gene follows the sorghum numbering from 1 to 57 listed above; when they are not in the same order as the sorghum genomic region, they can be numbered to indicate to which gene they are orthologous. The genes absent from the Sspon2A chromosome are represented by a white pentagon. The insertion represented by the purple pentagon between genes 20 and 21 does not have an ortholog in the sorghum QTL but has an ortholog observed on sorghum chromosome SBI-10 (Phytozome v13). In *S. spontaneum*, it was named Sspon.02G0013290, which is specific to the genomic region of Sspon2A. The insertions of gene clusters are represented below the main line representing the genomic region of Sspon2A. The clusters were formed by genes orthologous to those in the sorghum QTL. Each inserted cluster received a Roman numeral, and on the main line, the equivalent Roman numeral above a black arrow indicates where the cluster was inserted. Inversions within the insertions are indicated by crossed green dotted lines. Synteny can be visually observed, although many breaks in collinearity are observed.

Supplementary Figure 2



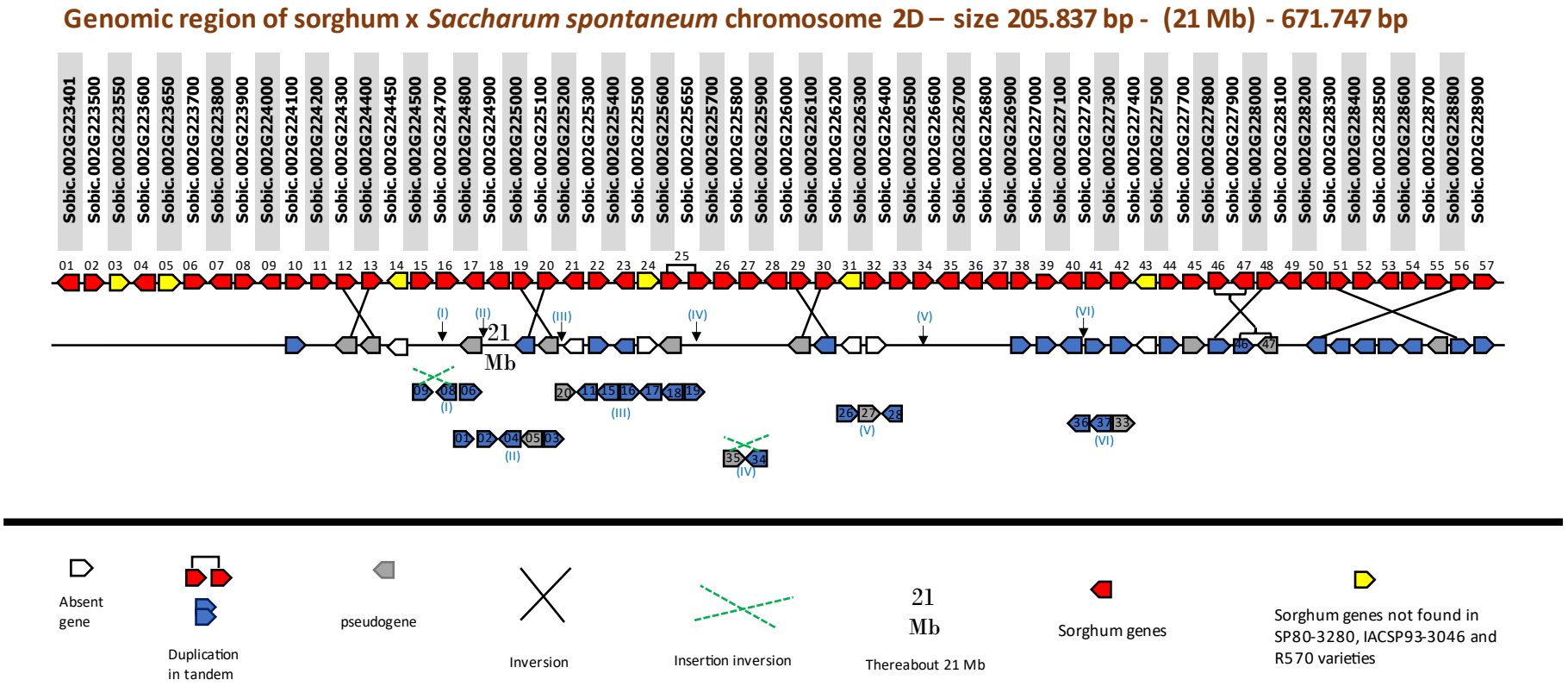
Supplementary Figure 2. Genomic region of sorghum × *Saccharum spontaneum* chromosome 2B, with a size of 1.134.230 bp. *Genomic region of sorghum:* The solid line with numbered pentagons in red and yellow represents the genomic region of sorghum and the QTL genes. The gene names of sorghum (Phytozome v.13) are presented above. Each pentagon represents a gene and its orientation in the sorghum genome. The right-facing arrow indicates the sense orientation, and the left-facing arrow indicates the antisense direction. The yellow pentagons represent the sorghum genes that were not found in the SP80-3280, IACSP93-3046 and R570 varieties. The genes were numbered from 1 to 57 (Supplementary Table 2). *Genomic region of chromosome Sspon2B:* Below the representation of the sorghum genomic region, the orthologous genomic region in *S. spontaneum* for chromosome Sspon2B is shown. The main solid line represents the region whose orthologous genes to sorghum are shown in blue pentagons and pseudogenes are shown in gray, following the numbering proposed for sorghum. The genes absent from the Sspon2B chromosome are represented by a white pentagon. Insertions of gene clusters are represented below the main line, and each cluster is named with a Roman numeral. The location of the insertion in the genomic region is indicated by a black arrow, and above it, the Roman numeral indicates which cluster was located there. Cluster IV contains 16 genes, including pseudogenes, eight of which are "other genes" shown in purple. Although they are orthologous to genes on the sorghum chromosome SBI-02 (Phytozome v.13), orthologs in the genomic region of the sorghum QTL were not identified. The absence of similar findings in the genomic regions of other homologous chromosomes of *S. spontaneum*, as well as in the haplotypes of SP80-3280, IACSP93-3046, and R570, suggests the possibility of a specific duplication on this particular chromosome. Inversions are indicated by crossed solid lines, and when the inversions are within an insertion, they are indicated by green dotted lines. Synteny can be observed, despite the breaks in collinearity.

Supplementary Figure 3



Supplementary Figure 3. Genomic region of sorghum × *Saccharum spontaneum* chromosome 2C, with a size of 969.275 bp. *Genomic region of sorghum:* The solid line with numbered pentagons in red and yellow represents the genomic region of sorghum and the QTL genes. The gene names of sorghum (Phytozome v.13) are presented above. Each pentagon represents a gene and its orientation in the sorghum genome. The right-facing arrow indicates the sense orientation, and the left-facing arrow indicates the antisense direction. The yellow pentagons represent the sorghum genes that were not found in the SP80-3280, IACSP93-3046 and R570 varieties. The genes were numbered from 1 to 57 (Supplementary Table 2). *Genomic region of chromosome Sspon2C*: The genomic region of chromosome Sspon2C of *S. spontaneum* is shown on a continuous line below the representation of the sorghum genomic region. The orthologous genes are represented as blue pentagons, and the pseudogenes are represented in gray. The orientation of the genes in the genome was indicated by the direction of the pentagon, as was the case for the sorghum genomic region. The genes also follow the numbering of their sorghum orthologs, from 1 to 57 (Supplementary Table 2). The genes absent from the Sspon2C chromosome are represented by white pentagons. An insertion of a duplicate of gene 19 was identified, as represented by a blue pentagon numbered 19 between genes 18 and 20. In this case, gene 18 was part of an inversion between genes 18 and 19. In addition, three more insertions in clusters were identified. The clusters are indicated below the main line by linear groupings of blue pentagons (genes) and gray pentagons (pseudogenes) and are named with Roman numerals. The location of each cluster is indicated by a black arrow on the main line, with the corresponding Roman numeral above it. Synteny was observed with less pronounced breaks in collinearity compared with other homologous Sspon2 chromosomes in *S. spontaneum*.

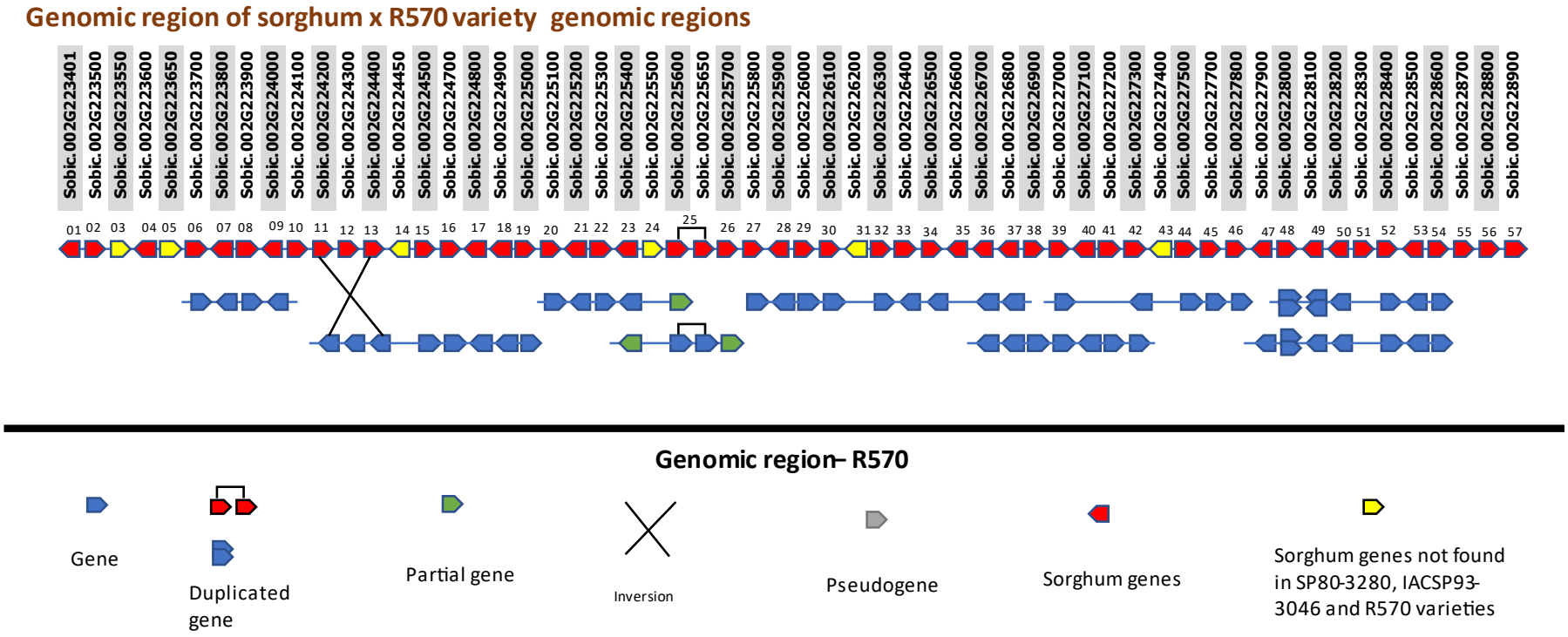
Supplementary Figure 4



Supplementary Figure 4. Genomic region of sorghum x *Saccharum spontaneum* chromosome 2D – size 205.837 bp - (21 Mb) - 671.747 bp. *Genomic region of sorghum:* The solid line with numbered pentagons in red and yellow represents the genomic region of sorghum and the QTL genes. The gene names of sorghum (Phytozome v.13) are presented above. Each pentagon represents a gene and its orientation in the sorghum genome. The right-facing arrow indicates the sense orientation, and the left-facing arrow indicates the antisense direction. The yellow pentagons

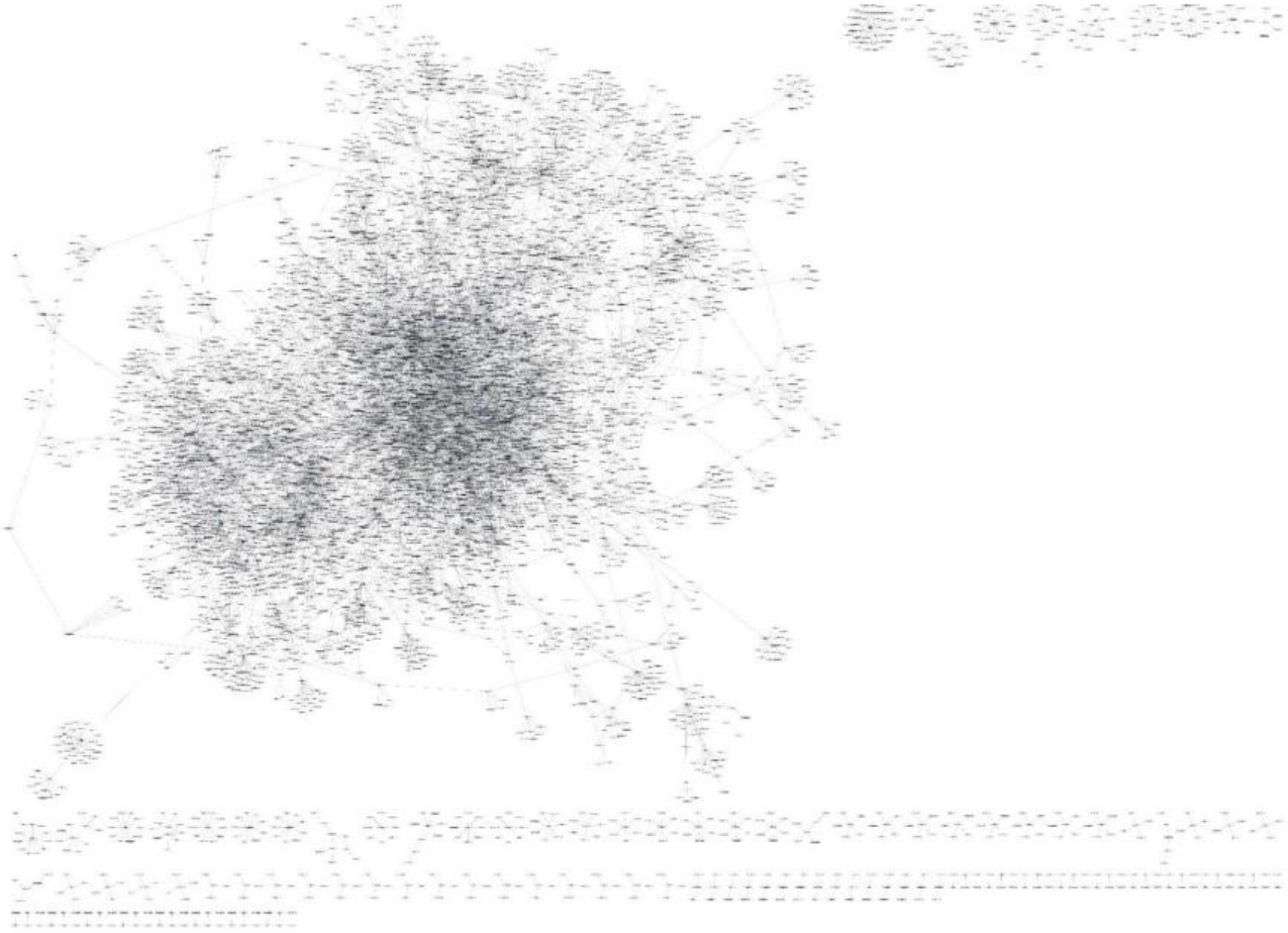
represent the sorghum genes that were not found in the SP80-3280, IACSP93-3046 and R570 varieties. The genes were numbered from 1 to 57 (Supplementary Table 2). *Genomic region of chromosome Sspon2D:* The orthologous genomic region of chromosome Sspon2D of *S. spontaneum* is shown below the representation of the sorghum genomic region. Genes are shown as blue pentagons, and pseudogenes are shown in gray on a solid line. The genes follow the numerical order of their sorghum orthologs, from 1 to 57 (Supplementary Table 2). When they did not follow the order of the sorghum genomic region, they were numbered according to their sorghum ortholog. Genes absent in chromosome Sspon2D are represented by white pentagons. The clusters inserted into the genomic region are presented below the main line. Each cluster was named with a Roman numeral, and its location is indicated by a black arrow on the main line, with the Roman numeral denoting which cluster was inserted at that location. Inversions are indicated by crossed lines. Inversions within insertions are shown by green crossed dashed lines. The orthologous genomic region on chromosome Sspon2D was not linear and was separated by 21 million base pairs. The location of this large insertion is indicated by the text "21 Mb" on the main line. Synteny was observed; however, more pronounced breaks in collinearity were noted than in the chromosomal regions of other homologous chromosomes in *S. spontaneum*.

Supplementary Figure 5



Supplementary Figure 5. *Sorghum × R570 variety genomic regions*: The solid line with numbered pentagons in red and yellow represents the genomic region of sorghum and the QTL genes. The gene names of sorghum (Phytozome v.13) are presented above. Each pentagon represents a gene and its orientation in the sorghum genome. The right-facing arrow indicates the sense orientation, and the left-facing arrow indicates the antisense direction. The yellow pentagons represent the sorghum genes that were not found in the SP80-3280, IACSP93-3046 and R570 varieties. The genes were numbered from 1 to 57 (Supplementary Table 2). *Genomic region of R570:* Below the representation of the sorghum genomic region, the orthologous genomic region in R570 [3] is shown. The genes are represented as blue pentagons and follow the order presented in the sorghum genomic region. In this figure, possible pseudogenes are not indicated. Genes that were not fully recovered are represented by green pentagons, and an inversion is indicated by crossed solid lines.

Supplementary Figure 6



Supplementary Figure 6. Representation of the highest reciprocal rank (HRR) gene coexpression network built with sugarcane transcriptome data. Nodes, representing genes, are sized according to degree.

Supplementary Tables

Supplementary Table 1. Characterization of the primer pairs used for BAC selection with sequence origins in sorghum and similarity in sugarcane transcripts.

Primer pairs	Gene number	Sequences	Efs* (bp)	Sorghum gene	Sugarcane transcript [45] [48]
223401	01	F: GCAGCACGAAACAAATCTGG R:TTGCTGGATGGGGATATAGTGC	79	Sobic.002G223401.1	comp85324_c0_seq10 [45]
223700	06	F: GCCGTCTGGACAACAACTTTC R: CGAAGGATCAGAACCGTTTGC	149	Sobic.002G223700.1	comp82173_c0_seq1 [45]
223900	08	F: AGCGAGGTGAAGCATCTAGTG R: TGGCTGTCGAAGAAGTCAAC	101	Sobic.002G223900.1	SCRUFL1114A11.g [45]
224100	10	F:ATCGTGGAGATGACCTACAAGC R:AACGAGACCAAGTTGTCTGC	149	Sobic.002G224100.1	SCCCRT2002C03.g [45]
Sh 129	11	F: ATGGTGGTGTCGATGGAGAG R:AGCTGGTACGTCTTGGTCAG	71	Sobic.002G224200.1	SCEZSD1080A05.g [57]
Sh 130	13	F: TGGGCTTGGTTCTGATCCTC R: ATGTTAGAGGCGGTGAAGGG	94	Sobic.002G224400.1	SCCCCL4013C01.g [57]
Sh 131	15	F: ACCCATGCAGGTCACAACAG R: TGTACTGCTCTGCTCGGTAC	97	Sobic.002G224500.1	SCBGRT1053C03.g [57]
Sh 121	16	F: TGCCTACTGCCACTCATCAC R: TGCATTGAGGCCTGGAGAAG	144	Sobic.002G224700.1	comp191263_c0_seq1 [48]
Sh 121-2	17	F:CGGACGATCCAAGGATTAACCC R: GAACAGCCTCTCACCCACTAG	74	Sobic.002G224800.1	comp188036_c0_seq1 [48]
Sh 122	19	F: GTCCACCAAGTCTTCCTGGAG R: TCAGTCCCAATTCCAGTGACG	105	Sobic.002G225000.1	comp206011_c0_seq1 [48]
Sh 123	21	F: TCGACCATCTCGTCCACTTC R: GGTGGGTCGACAACTACCTC	85	Sobic.002G225200.1	comp174420_c0_seq1 [48]
Sh 124	23	F: TTGACGATCCAGGTGAGCAC R: AGGTCGCCGACAACATCATC	76	Sobic.002G225400.1	comp196460_c1_seq1 [48]
Sh 125	24	F: GTCTCCTGCAATTTCACCTTCG R: CCAAGACAAGATTGCGCAAC	86	Sobic.002G225500.1	comp163627_c1_seq1 [48]
Sh 126	25	F: ACGACCATCATGAGCTCTGC R: TTCTTCACGGGCCTCAACG	149	Sobic.002G226000.1	comp197793_c1_seq1 [48]
Sh 127	30	F: AAGCAGGGTAAGAGGAGGAG R: CACATCATCCACCAGGTAGC	72	Sobic.002G226100.1	comp194226_c0_seq1 [48]
Sh 128	37	F: TTCCACCACATCCGGTTCAG R:TATAGCCCCGAGTCCTCCTAAG	124	Sobic.002G226800.1	comp167884_c1_seq1 [48]

Primer pairs	**Gene number**	**Sequences**	**Efs* (bp)**	**Sorghum gene**	**Sugarcane transcript** [45] [48]
Sh 132	39	F: CTCTTGCTGCTGCTTTTCGG R:TACAGCACGCTTCCAAGTGG	137	Sobic.002G227000.2	SCACAD1035G02.g [57]
Sh 133	44 **	F: GCACTGTGCAATGGTTACCC R:GCCAGAGCCAGTTTTTGTGG	109	Sobic.002G227500.1	SCBFSD1036B01.b [57]
Sh 134	48	F: ACACTCCGTACTTCCGCATC R: CGGTGTTGCTGTCTATGACG	129	Sobic.002G228000.1	SCJFHR1C04B11.g [57]
Sh 135	52	F: CAACATCGCGCACAACATGG R: AAGTAAGGGTCCAGGAGCAG	105	Sobic.002G228400.1	SCVPRZ2035E02.g [57]
Sh 136	55	F: ATTCCGATCGCTGTCAGGAC R: TCATGCAGCCTGGTGATTCC	105	Sobic.002G228700.1	SCQSST1040G01.g [57]

*Expected fragment size

**The primer pair did not amplify genomic DNA from the SP80-3280 and IACSP93SP-3046 varieties and was therefore not used.

Supplementary Table 2. Summary of orthologous genes in sorghum and their proteins. Each gene was assigned a number according to the order in which it appeared in the sorghum QTL. Some genes are absent in the column listing the names of the genes in *S. spontaneum*; however, their absence does not indicate that the genes are absent in the genomic region but rather that they were not detected by automated annotation [16]. In fact, only gene 14 was absent from chromosome Sspon.2 in *S. spontaneum.* The other genes (3, 5, 24, 27, 30, 31, 35, 36, 43, 51 and 55) were visualized and annotated manually (Supplementary Figures 1, 2, 3 and 4).

Gene number	***S. bicolor* gene**	***S. spontaneum* gene**	**Protein (Phytozome 13)**
01	Sobic.002G223401	Sspon.02G0013480	Similar to a protein of the prolyl oligopeptidase family **(POP)**
02	Sobic.002G223500	Sspon.02G0013470	Predicted **AST**-like aspartate aminotransferase protein
03	Sobic.002G223550	-	Predicted, uncharacterized protein
04	Sobic.002G223600	Sspon.02G0013460	Predicted phosphatidylinositol-4-phosphate 5-kinase-related protein **(PIP5Ks)**
05	Sobic.002G223650	-	Hypothetical, uncharacterized protein
06	Sobic.002G223700	Sspon.02G0013450	Protein inferred by homology as a rhomboid-like protein, **RBL10**
07	Sobic.002G223800	Sspon.02G0013440	Hypothetical uncharacterized protein
08	Sobic.002G223900	Sspon.02G0013430	Hypothetical protein similar to F-box-like proteins (**FBPs**)
09	Sobic.002G224000	Sspon.02G0014420	Similar to alpha-carbonic anhydrase domain-containing protein (**CA**)
10	Sobic.002G224100	Sspon.02G0013410	Similar to phosphate transporter 3, **PHT3**
11	Sobic.002G224200	Sspon.02G0013390	Similar to the heat shock factor B 4 **(HSFB4)** protein

Gene number	***S. bicolor* gene**	***S. spontaneum* gene**	**Protein (Phytozome 13)**
12	Sobic.002G224300	Sspon.02G0013400	Predicted protein similar to the **NAM** protein, an NAC transcription factor
13	Sobic.002G224400	Sspon.02G0037420	Receptor-like serine/threonine-protein kinase, lectin type G
14	Sobic.002G224450	-	Uncharacterized predicted protein
15	Sobic.002G224500	Sspon.02G0049360	Similar to a BHLH domain-containing protein (basic helix-loop-helix), **HEC**
16	Sobic.002G224700	Sspon.02G0013380	Inference by homology, **NF-Kappa B activating protein**
17	Sobic.002G224800	Sspon.02G0013370	Protein belonging to the **GDXG** family, a family of lipolytic enzymes that is similar to alpha/beta hydrolase (with folded domain): alpha/beta hydrolase fold domain-containing (ABH)
18	Sobic.002G224900	Sspon.02G0037410	Protein belonging to the **GDXG** family, a family of lipolytic enzymes, and is similar to alpha/beta hydrolase (with folded domain): alpha/beta hydrolase fold domain-containing (ABH)
19	Sobic.002G225000	Sspon.02G0013350	Protein inferred by homology that is similar to a C-terminal remorine protein: remorine C-terminal like (**REM**)
20	Sobic.002G225100	Sspon.02G0013320	Similar to abscisic acid-insensitive 5-like protein 4, basic-leucine zipper domain **(ABF1)**

Gene number	***S. bicolor* gene**	***S. spontaneum* gene**	**Protein (Phytozome 13)**
21	Sobic.002G225200	Sspon.02G0013280	Similar to putative uncharacterized protein B1342C04.33
22	Sobic.002G225300	Sspon.02G0013260	Similar to the heat shock transcription factor B1 **(HSB1)** protein
23	Sobic.002G225400	Sspon.02G0013240	Similar to the abscisic acid 8'-hydroxylase **(AA8' OH)** protein
24	Sobic.002G225500	-	Similar to a protein containing a zinc finger domain
25	Sobic.002G225600/650	Sspon.02G0013220	Similar to alpha-amylase **(AMY)**
26	Sobic.002G225700	Sspon.02G0013210	Similar to ethylene responsive factor 109 **(ERF109)**
27	Sobic.002G225800	-	**E3 ubiquitin** ligase involved in syntaxin degradation
28	Sobic.002G225900	Sspon.02G0013200	Similar to fucosyltransferase (**FUT**) GO:0008417
29	Sobic.002G226000	Sspon.02G0013190	Malectin-like domain-containing protein **(MLD)**
30	Sobic.002G226100	-	Protein disulfide isomerase **(SCO2)**
31	Sobic.002G226200	-	Uncharacterized protein
32	Sobic.002G226300	Sspon.02G0013180	Calmodulin-like protein **(CML)**
33	Sobic.002G226400	Sspon.02G0013170	Calmodulin-like protein **(CML)**
34	Sobic.002G226500	Sspon.02G0013160	Putative uncharacterized protein similar to a basic leucine zipper (**BZIP**) domain-containing protein
35	Sobic.002G226600	-	Uncharacterized protein

Gene number	***S. bicolor* gene**	***S. spontaneum* gene**	**Protein (Phytozome 13)**
36	Sobic.002G226700	-	Similar to BRI1 kinase inhibitor 1 **(BKI1)**
37	Sobic.002G226800	Sspon.02G0013130	**PP2C**—protein phosphatase 2C//subfamily not named
38	Sobic.002G226900	Sspon.02G0013140	Similar to enoyl-CoA reductase (**ECR**)
39	Sobic.002G227000	Sspon.02G37380/ 37400	Similar to enoyl-CoA reductase (**ECR**)
40	Sobic.002G227100	Sspon.02G0013150	Cyclin-dependent kinase inhibitor (**CDK inhibitor**)
41	Sobic.002G227200	Sspon.02G0037390	Similar to DNAJ-proteins or heat shock protein 40 (**HSP40**)
42	Sobic.002G227300	Sspon.02G0037370	Uncharacterized protein—predicted
43	Sobic.002G227400	-	Similar to pentatricopeptide repeat **(PPR)**
44	Sobic.002G227500	Sspon.02G0013040	Similar to the putative protein transport protein SEC23
45	Sobic.002G227700	Sspon.02G0013050	Protein belonging to the **GDXG** family, a family of lipolytic enzymes, and is similar to alpha/beta hydrolase (with a folded domain): alpha/beta hydrolase fold domain-containing (ABH)
46	Sobic.002G227800	Sspon.02G0013060	Protein belonging to the **GDXG** family, a family of lipolytic enzymes, and is similar to alpha/beta hydrolase (with a folded domain): alpha/beta hydrolase fold domain-containing (ABH)

Gene number	*S. bicolor* gene	*S. spontaneum* gene	Protein (Phytozome 13)
47	Sobic.002G227900	Sspon.02G0013070	Similar to alpha/beta hydrolase fold domain-containing (ABH) and gibberellin receptor (**GID1L2)**
48	Sobic.002G228000	Sspon.02G0013080	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)
49	Sobic.002G228100	Sspon.02G0013090	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)**.**
50	Sobic.002G228200	Sspon.02G0013090	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)**.**
51	Sobic.002G228300	-	Uncharacterized and predicted protein
52	Sobic.002G228400	Sspon.02G0013100	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)
53	Sobic.002G228500	Sspon.02G0013110	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)
54	Sobic.002G228600	Sspon.02G0013120	Similar to alpha/beta hydrolase fold domain-containing (ABH) and carboxylesterase (**CXE**)
55	Sobic.002G228700	-	Similar to phospholipase A/patatin-related (**pPLA**)
56	Sobic.002G228800	Sspon.02G0037310	Uncharacterized protein—predicted
57	Sobic.002G228900	Sspon.02G0013030	RNA binding protein, **RBP47**

Supplementary Table 3. Summary of PacBio® Sequel sequencing and sequence assembly results.

	Pool Name	BAC	Contigs	Longest contig size (bp)	Shortest contig size (bp)	Total bases (bp)
SP80-3280	2	2	2	138,727	102,149	240,876
	3	2	2	132,039	123,810	255,849
	5	2	1	141,128	141,128	141,128
	6	2	1	126,235	126,235	126,235
	7	2	1	106,318	106,318	106,318
	8	2	1	142,948	142,948	142,948
	9	2	2	180,638	121,873	302,511
	13	2	1	132,828	132,828	132,828
	14	2	2	102,504	87,864	190,368
	15	2	1	84,208	84,208	84,208
	16	2	1	158,108	158,108	158,108
Subtotal	-	22	15	-	-	1,881,377
IACSP93-3046	1	1	1	107,880	107,880	107,880
	2	2	1	126,094	126,094	126,094
	3	1	1	135,702	135,702	135,702
	16	1	1	115,793	115,793	115,793
	17	3	5	192924	8538	391,863
	18	3	4	113852	10714	252,862
	19	3	9	127023	4750	378,918
	20	3	6	183570	4414	482,101
	21	3	8	133902	4038	325,041
	22	3	3	153020	4057	251,633
	23	3	5	181700	4576	506,830
	24	3	5	152382	14478	522,427
	25	3	7	124722	3960	371,939
Subtotal	-	32	56	-	-	3,969,083
Total	-	**63**	**71**	-	-	**5,850,460**

Supplementary Table 4. Results of PacBio® Sequel sequencing and sequence assembly.

Pool Name	Contigs	Contig length	Coverage	BAC clone	Variety
1	1	107880	61.22	Shy112C03	IACSP93-3046
2	4	126094	42.98	Shy123P01	
	7	138727	22.07	Shy488A19	SP80-3280
	17	102149	30.52	Shy361L04	
3	3	135702	52.33	Shy187N11	IACSP93-3046
	10	132039	64.22	Shy255L20	SP80-3280
	15	123810	140.51	Shy289A21	
5	6	141128	17.41	Shy253P08	
6	08	126235	59.89	Shy486B15	
7	17	106318	129.15	Shy041F06	
8	8	142948	28.61	Shy260G24	
				Shy223J17	
9	1	180638	67.16	Shy378L10	
	5	121873	59.36	Shy378L03	
13	7	132828	58.63	Shy368O04	
14	21	102504	47.01	Shy492F12	
	29	87864	164.54	Shy021C23	
15	15	84208	276.17	Shy486F01	
16	1	158108	109.61	Shy504G20	
	10	115793	53.93	Shy141H03	IACSP93-3046
17	3	192924	110.5	Shy273L13	
	29	102135	126.15	Shy130N20	
	74	45538	48.08	Shy012B04	
18	5	113852	199.27	Shy416D13	
	13	106999	99.82	Shy006P16	
19	1	127023	115	Shy120H04	
	48	54704	31.71	Shy411A07	
	114	36851	25.38		

	2804	126947	134.62	Shy178E18	
20	3	183570	85.32	Shy192F11	
	9	170179	204.58	Shy031N20	
	26	72739	212.27	Shy265N09	
	47	46762	220.47		
21	1	680338	42.32	Shy282B05	
	228	21867	9.18	Shy238L17	
	254	8060	11.17		
22	4	153020	375.13	Shy333K17	
	23	94556	372.38	Shy404J16	
23	2	151700	83.2	Shy188C18	
	3357	159080	103.9	Shy191K12	
24	1	152382	124.36	Shy320P14	
	10	134990	72.59	Shy187N11	
25	17	110471	135.26	Shy406H18	
	19	124722	101.93	Shy397C11	
	24	102547	88.29	Shy116E06	

Supplementary Table 5. Differentially expressed genes (DEGs) identified between *Saccharum spontaneum* and *Saccharum officinarum*, IACSP93-3046 and SP80-3280. The $\log_2$(fold change) ($\log_2$(FC)) values and false discovery rate (FDR)-corrected *p* values are provided for each gene.

Supplementary Table 5 for this article has been deposited in Figshare and is available at https://figshare.com/articles/dataset/Supplementary_Table_5_for_Martins_et_al_2024_-_Identifying_Candidate_Genes_for_Sugar_Accumulation_in_Sugarcane_an_integrative_approach/26236361

Resumo dos resultados

- A partir de um QTL já mapeado para Brix em sorgo, uma região ortóloga foi recuperada em cana-de-açúcar. Foram identificados 58 genes em sorgo, um deles duplicado *in tandem*.

- Foram identificados, sequenciados, montados e anotados manualmente 16 clones BACs positivos para pelo menos dois genes da região de interesse da variedade SP80-3280 e 35 clones BACs positivos para pelo menos dois genes da região de interesse da variedade IACSP93-3046. Assim, foi possível recuperar uma região contínua em ambas variedades somando 51 genes. Na variedade IACSP93-3046 foi observado um *gap* e na variedade SP80-3280 foram observados dois *gaps*.

- A organização estrutural dos haplótipos das variedades SP80-3280 e IACSP93-3046 confirmou a sintenia e colinearidade da região de interesse das variedades brasileiras com a região ortóloga em sorgo.

- A anotação manual da região de interesse em *S. spontaneum* demonstrou que há sintenia da espécie com sorgo e com as variedades brasileiras. No entanto há muitas quebras de colinearidade, observadas através das inversões, inserções e translocações. Mesmo entre os cromossomos homólogos há diversas quebras de colinearidade, provavelmente por se tratar de uma espécie silvestre, não sofrendo o processo de domesticação que ocorreu com sorgo e com as cultivares brasileiras.

- Através de análises genômicas comparativas conduzidas entre as variedades SP80-3280, IACSP93-3046 e R570; *S. spontaneum* e *S. bicolor* foram observadas sintenia e colinearidade entre todas, embora visivelmente menos com *S. spontaneum*. As cultivares modernas possuem mais sintenia e colinearidade com sorgo do que com a espécie que faz parte dos seus cruzamentos ancestrais, *S. spontaneum*, oque é justificado com o fato de que as cultivares modernas possuem 10 a 20% de cromossomos oriundos de *S. spontaneum* em seu genoma. As variedades brasileiras também se assemelham mais entre si do que com a variedade francesa R570. Isso pode ser justificado pela origem em comum das variedades brasileiras, vindas do germoplasma brasileiro, e também da alta taxa de endogamia observada na variedade francesa.

- O padrão de expressão de 44 genes ortólogos em cana-de-açúcar dentre os 51 genes recuperados em cana-de-açúcar do QTL de sorgo pode ser observado através de um *heatmap* elaborado utilizando dados de RNA-seq dos internódios 3 (mais jovens) e 8 (mais maduros) das variedades brasileiras (SP80-3280 e IACSP93-3046) e das espécies *S. spontaneum* e *S. officinarum*.

- Foram identificados 6.264 DEGs ao comparar as plantas com alto teor de açúcar acumulado nos colmos com plantas de baixo teor de açúcar acumulado nos colmos. Dentro da região-alvo foram obtidos sete DEGs: genes 01 (Prolil Oligopeptidase – POP), 09 (Anidrase Carbônica – CA), 19 (Remorina – REM), 22 (Fator de Transcrição de choque térmico B1 – HSB1), 23 (Ácido Absísico 8’ hidroxilase – ABA 8’OH), 26 (Fator Responsivo ao Etileno 109 – ERF-109) e 42 (uma proteína não caracterizada).

- Foram identificados seis genes com primeiros vizinhos através de uma rede coexpressão HRR dentro da região-alvo. A rede final tinha 6.809 nós conectados e uma média de 17 vizinhos por nó: 01 (Prolil Oligopeptidase – POP), 07 Mieloblastose 46 – MYB-46), 18 (família GDXG), 23 (Ácido Abscísico 8’ hidroxilase – ABA 8’OH), 26 (Fator Responsivo ao Etileno 109 – ERF-109) e 37 (Fosfatase 2C – PP2C).

- Foram elucidados os papéis biológicos, interações moleculares e funções possíveis em processos fisiológicos através de uma abordagem investigativa para cada um dos 51 genes recuperados nas variedades brasileiras.

- Três genes candidatos para a característica acúmulo de açúcar foram identificados na região-alvo seguindo os seguintes critérios: (1) ter alguma relação fisiológica descrita em literatura de referência relacionada ao acúmulo de açúcar, resposta e/ou tolerância a estresses abióticos ou ambas; (2) ser DEG entre canas-de-açúcar acumuladoras de açúcar vs. não-acumuladoras de açúcar; (3) ter pelo menos um primeiro vizinho na rede de coexpressão HRR. São eles: 01 (Prolil Oligopeptidase – POP), 23 (Ácido Abscísico 8’hidroxilase – ABA 8’OH) e 26 (Fator Responsivo ao Etileno 109 – ERF-109).

Conclusões

- Este estudo examinou a arquitetura genômica de uma região específica em dois cultivares de cana-de-açúcar, validando o sorgo como uma referência genômica chave para identificar genes candidatos para características quantitativas e outros estudos envolvendo a genômica da cana-de-açúcar.
- Utilizando uma estratégia que integrou análise genômica comparativa, dados de RNA-seq e revisão de literatura, foram identificados três genes candidatos para a acumulação de açúcar: ERF-109, ABA 8'OH e POP.
    - ERF-109, um fator de transcrição versátil, é conhecido por estar envolvido em processos metabólicos e resistência ao estresse. Este estudo mostrou que ele potencialmente influencia a acumulação de açúcar.
    - ABA 8'OH e POP estão ligados à acumulação de prolina, que está associada a respostas ao estresse e ligação à sacarose, demonstrando uma possível associação entre acumulação de sacarose e acumulação de prolina livre na planta.
- Esta estratégia integrativa é recomendada para identificar novos genes candidatos para acumulação de açúcar e outras características quantitativas na cana-de-açúcar, visando desenvolver cultivares com maior teor de açúcar

Perspectivas

Aumentar o teor de sacarose nos colmos da cana-de-açúcar mantendo ou mesmo aumentando a resistência aos diferentes estresses é o objetivo de melhoristas e produtores de cana-de-açúcar. Para isso, apostamos no caminho de encontrar genes que tenham um papel fundamental no acúmulo de açúcar, já que acumular açúcar é uma característica altamente desejável, mas também muito complexa e poligênica. Após encontrar três genes candidatos relacionados a fenótipos de caráter quantitativo através de uma estratégia que integra diferentes abordagens, a grande perspectiva é usar esses genes para desenvolver cultivares de cana-de-açúcar com altíssimo teor de sacarose e que sejam competitivas no mercado açucareiro. Para isso, o aprofundamento sobre o papel destes genes é necessário.

Aprofundar as pesquisas a respeito de ERF-109, ABA 8'OH e POP não apenas sob o aspecto genético-genômico, mas associando-os com a fisiologia da cana-de-açúcar, além de analisar a expressão desses genes nos tecidos-dreno associando-os com a fenologia da planta pode incluir informações preciosas que levem a novas cultivares. Sugere-se a inclusão de novas abordagens como transformação, nocauteando e superexpressando genes, para investigar o papel destes no acúmulo de açúcar. Os genes relacionados ao acúmulo de prolina parecem ser os mais promissores, pois não apresentam o caráter promíscuo de ERF-109.

Utilizar a estratégia integrativa apresentada neste estudo em outros tecidos e outras variedades pode trazer um novo olhar sobre estes (ERF-109, ABA 8'OH e POP) e outros genes da região -alvo. Adicionalmente, buscar novos genes candidatos ao acúmulo de açúcar e outras características quantitativas de interesse agronômico pode ser uma boa aposta para pesquisas futuras, utilizando essa mesma estratégia, posto que esta se mostrou bastante eficiente em encontrar genes candidatos promissores.

Ao confirmar o papel desses genes no acúmulo de açúcar, o desenvolvimento de marcadores e/ou teste de qPCR que possam identificar variedades mais ou menos produtivas pode ser desenvolvido, impactando diretamente no melhoramento da cana-de-açúcar.

Referências Bibliográficas Gerais

Aitken, K.S., Hermann, S., Karno, K. *et al.* Genetic control of yield related stalk traits in sugarcane. *Theor Appl Genet* **117**, 1191–1203 (2008).

Aitken, K. S., Jackson, P. A., & McIntyre, C. L. (2006). Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum spp.*) cultivar × *Saccharum officinarum* population. *Theoretical and Applied Genetics*, *112*(7), 1306–1317. https://doi.org/10.1007/s00122-006-0233-2

Aitken, K. S. (2022). History and development of molecular markers for Singh sugarcane breeding. *Sugar Tech*, 24(1), 341-353.

Ali, Q., Anwar, F., Ashraf, M., Saari, N., & Perveen, R. (2013). Ameliorating effects of exogenously applied proline on seed composition, seed oil quality and oil antioxidant activity of maize (*Zea mays* L.) under drought stress. *International Journal of Molecular Sciences*, *14*(1), 818–835. https://doi.org/10.3390/ijms14010818

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2

Aluko, O. O., Li, C., Wang, Q., & Liu, H. (2021). Sucrose utilization for improved crop yields: A review article. *International Journal of Molecular Sciences*, *22*(9), 4704. https://doi.org/10.3390/ijms22094704

Amalraj, V. A., & Balasundaram, N. (2006). On the taxonomy of the members of '*Saccharum* complex'. *Genetic Resources and Crop Evolution*, 53, 35-41.

Andersen, J. R., & Lübberstedt, T. (2003). Functional markers in plants. *Trends in Plant Science, 8*(11), 554-560.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., & Rai, A. (2016). Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound

Distribution Approach. *Journal of Computational Biology*, 23(4), 239-247. https://doi.org/10.1089/cmb.2015.0205

Aoki, K., Ogata, Y., & Shibata, D. (2007). Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. Plant and Cell Physiology, 48(3), 381–390. https://doi.org/10.1093/pcp/pcm013 .

Aono, A. H., Pimenta, R. J. G., Garcia, A. L. B., Correr, F. H., Hosaka, G. K., Carrasco, M. M., Cardoso-Silva, C. B., Mancini, M. C., Sforça, D. A., Dos Santos, L. B., Nagai, J. S., Pinto, L. R., Landell, M. G. A., Carneiro, M. S., Balsalobre, T. W., Quiles, M. G., Pereira, W. A., Margarido, G. R. A., & de Souza, A. P. (2021). The wild sugarcane and sorghum kinomes: Insights into expansion, diversification, and expression patterns. *Frontiers in Plant Science, 12*, 668623. https://doi.org/10.3389/fpls.2021.668623

Ayodele, B. V., Alsaffar, M. A., & Mustapa, S. I. (2020). An overview of integration opportunities for sustainable bioethanol production from first- and second-generation sugar-based feedstocks. Journal of Cleaner Production, 245, 118857. https://doi.org/10.1016/j.jclepro.2019.118857 .

Babu, K. S. D., Janakiraman, V., Palaniswamy, H., Kasirajan, L., Gomathi, R., & Ramkumar, T. R. (2022). A short review on sugarcane: its domestication, molecular manipulations, and future perspectives. *Genetic Resources and Crop Evolution, 69*(8), 2623-2643.

Baharin, A., Ting, T. Y., & Goh, H. H. (2022). Post-Proline Cleaving Enzymes (PPCEs): Classification, structure, molecular properties, and applications. *Plants (Basel), 11*(10), 1330. https://doi.org/10.3390/plants11101330

Bahieldin, A., Atef, A., Edris, S., Gadalla, N. O., Ramadan, A. M., Hassan, S. M., Al Attas, S. G., Al-Kordy, M. A., Al-Hajar, A. S., Sabir, J. S. M., Nasr, M. E., Osman, G. H., & El-Domyati, F. M. (2018). Multifunctional activities of ERF109 as affected by salt stress in Arabidopsis. *Scientific Reports, 8*, 6403. https://doi.org/10.1038/s41598-018-24452-6

Bao, Y., Zhang, Q., Huang, J. *et al.* A chromosomal-scale genome assembly of modern cultivated hybrid sugarcane provides insights into origination and evolution. *Nat Commun* 15, 3041 (2024). https://doi.org/10.1038/s41467-024-47390-6 .

Bertrand, E., Vandenberghe, L. P. S., Soccol, C. R., Sigoillot, J.-C., & Faulds, C. (2016). First generation bioethanol. In Green Fuels Technology (pp. 175–212). Springer. https://doi.org/10.1007/978-3-319-39402-9_5

Bhat, S.R., Gill, S.S. The implications of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica* **34**, 377–384 (1985). https://doi.org/10.1007/BF00022932 .

Bhattramakki, D., Dong, J., Chhabra, A. K., & Hart, G. E. (2000). An integrated SSR and RFLP linkage map of Sorghum bicolor (L.) Moench. *Genome, 43*, 988–1002. https://doi.org/10.1139/g00-063

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bozzi, A. T., & Gaudet, R. (2021). Molecular mechanism of Nramp-family transition metal transport. *Journal of Molecular Biology, 433*(16), 166991. https://doi.org/10.1016/j.jmb.2021.166991

Bremer, G. (1961). Problems in breeding and cytology of sugar cane. *Euphytica, 10*(1), 59-78.

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., et al. (2014). De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS ONE, 9*(3), e88462. https://doi.org/10.1371/journal.pone.0088462

Carson, D. L., & Botha, F. C. (2002). Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Reports, 20*(12), 1075–1081. https://doi.org/10.1007/s00299-002-0474-4

Casu, R. E., Grof, C. P. L., Rae, A. L., McIntyre, C. L., Dimmock, C. M., & Manners, J. M. (2003). Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Molecular Biology, 52*(2), 371–386. https://doi.org/10.1023/A:1022841708985

Chandra, A., Jain, R., & Solomon, S. (2012). Complexities of invertases controlling sucrose accumulation and retention in sugarcane. *Current Science, 102*(6), 857–866. Available from http://www.jstor.org/stable/24084500

Chandra, A.P.K. Verma, M.N. Islam, M.P. Grisham, R. Jain, A. Sharma, K. Roopendra, K. Singh, P. Singh, I. Verma, and S. Solomon. 2015. Expression analysis of genes associated with

sucrose accumulation in sugarcane (*Saccharum* spp. hybrids) varieties differing in content and time of peak sucrose storage. *Plant Biology* 17: 608–617.

Chen, H., Liu, X., Li, S., Yuan, L., Mu, H., Wang, Y., Li, Y., Duan, W., Fan, P., Liang, Z., & Wang, L. (2023). The class B heat shock factor HSFB1 regulates heat tolerance in grapevine. *Horticulture Research*, *10*, uad001. https://doi.org/10.1093/hr/uhad001

Chen, K., Li, G. J., Bressan, R. A., Song, C. P., Zhu, J. K., & Zhao, Y. (2020). Abscisic acid dynamics, signaling, and functions in plants. *Journal of Integrative Plant Biology*, *62*(1), 25–54. https://doi.org/10.1111/jipb.12899

Chen, R., Feng, Z., Zhang, X., Song, Z., & Cai, D. (2021). A New Way of Rice Breeding: Polyploid Rice Breeding. *Plants*, 10(3), 422. https://doi.org/10.3390/plants10030422

CONAB - Companhia Nacional de Abastecimento. (2023, 26 de abril). Produção de cana-de-açúcar deve crescer 4,4% na safra 2023/24, estimada em 637,1 milhões de toneladas [Comunicado de imprensa]. Recuperado de https://www.conab.gov.br/

Costa, E.A., Anoni, C.O., Mancini, M.C. *et al.* QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny. *Euphytica* **211**, 1–16 (2016). https://doi.org/10.1007/s10681-016-1746-7

Cuadrado, A. (2004). Genome remodelling in three modern *S. officinarum* × *S. spontaneum* sugarcane cultivars. *Journal of Experimental Botany*, *55*(397), 847–854. https://doi.org/10.1093/jxb/erh09

Cursi, D. E., et al. (2021). History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. *Sugar Tech*, 23(1), 1–22. https://doi.org/10.1007/s12355-021-01080-5

Daniels, J., & Roach, B. T. (1987). Taxonomy and evolution in sugarcane. In: *Sugarcane improvement through breeding* (pp. 7–84)

de Oliveira, F. A., Vigna, B. B., da Silva, C. C., Fávero, A. P., de Matta, F. P., Azevedo, A. L., & de Souza, A. P. (2020). Coexpression and transcriptome analyses identify active apomixis-related genes in *Paspalum notatum* leaves. *BMC Genomics*, *21*, 1–15. https://doi.org/10.1186/s12864-020-06932-7

de Souza, A. P., Grandis, A., Arenque-Musa, B. C., & Buckeridge, M. S. (2018). Diurnal variation in gas exchange and nonstructural carbohydrates throughout sugarcane development. *Functional Plant Biology, 45*(8), 865–876. https://doi.org/10.1071/FP17268

Depuydt T, Vandepoele K. Multi-omics network-based functional annotation of unknown Arabidopsis genes. *Plant J*. 2021 Nov;108(4):1193-1212. doi: 10.1111/tpj.15507. Epub 2021 Oct 10. PMID: 34562334.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J. C., Rao, S., & Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum spp.*) by molecular cytogenetics. *Molecular and General Genetics, 250*(4), 405–413. https://doi.org/10.1007/BF02174028

DiMario, R. J., Giuliani, R., Ubierna, N., Slack, A. D., Cousins, A. B., & Studer, A. J. (2022). Lack of leaf carbonic anhydrase activity eliminates the C4 carbon-concentrating mechanism requiring direct diffusion of CO2 into bundle sheath cells. *Plant Cell & Environment, 45*(5), 1382–1397. https://doi.org/10.1111/pce.14291

Dong, Z., Yang, S., Zhang, Z., Tang, C., Kan, Y., & Yao, L. (2022). Prolyl aminopeptidases: Reclassification, properties, production and industrial applications. *Process Biochemistry, 118*, 121–132. https://doi.org/10.1016/j.procbio.2022.04.025

EMBL. (2022). *EMBL Homepage*. European Molecular Biology Laboratory. Available from https://www.embl.org/

FAO. (2023). The State of Food Security and Nutrition in the World 2023. Food and Agriculture Organization of the United Nations. https://www.fao.org/3/cc8589en/cc8589en_sugar.pdf

Federico, M. L., Diniz, A. L., Souza, G. M., Snowdon, R., & Erazzú, L. (2022). Translational genomics from sorghum (*Sorghum bicolor*) to sugarcane (*Saccharum spp.*) for bioenergy breeding. *Plant Breeding, 141*(3), 389–398. https://doi.org/10.1111/pbr.12956

Feng, Y., Hiwatashi, T., Minamino, N., Ebine, K., & Ueda, T. (2022). Membrane trafficking functions of the ANTH/ENTH/VHS domain-containing proteins in plants. *FEBS Letters, 596*, 2256–2268. https://doi.org/10.1016/j.febslet.2022.06.014

Figueira, T.R.e.S., Okura, V., Rodrigues da Silva, F. *et al.* A BAC library of the SP80-3280 sugarcane variety (*saccharum* sp.) and its inferred microsynteny with the sorghum genome. *BMC Res Notes* 5, 185 (2012). https://doi.org/10.1186/1756-0500-5-185

Fragkostefanakis, S., Simm, S., El-Shershaby, A., Hu, Y., Bublak, D., Mesihovic, A., Darm, K., Mishra, S. K., Tschiersch, B., Theres, K., Scharf, C., Schleiff, E., & Scharf, K. D. (2019). The repressor and co-activator HsfB1 regulates the major heat stress transcription factors in tomato. *Plant Cell & Environment, 42*, 874–890. https://doi.org/10.1111/pce.13467

Garcia, A., Mollinari, M., Marconi, T. *et al.* SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci Rep* 3, 3399 (2013). https://doi.org/10.1038/srep03399

Garcia, F. H. S., Mendonça, A. M. C., Rodrigues, M., Matias, F. I., Filho, M. P. S., Santos, H. R. B., Taffner, J., & Barbosa, J. P. R. A. D. (2020). Water deficit tolerance in sugarcane is dependent on the accumulation of sugar in the leaf. *Annals of Applied Biology, 176*(1). https://doi.org/10.1111/aab.12559

Garsmeur, O., Charron, C., Bocs, S., Jouffe, V., Samain, S., Couloux, A., et al. (2011). High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. *New Phytologist, 189*, 629–642. https://doi.org/10.1111/j.1469-8137.2010.03497.x

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications, 9*, 2638. https://doi.org/10.1038/s41467-018-05051-5

Ghifari, A. S., Teixeira, P. F., Kmiec, B., Singh, N., Glaser, E., & Murcha, M. W. (2022). The dual-targeted prolyl aminopeptidase PAP1 is involved in proline accumulation in response to stress and during pollen development. *Journal of Experimental Botany, 73*(1), 78–93. https://doi.org/10.1093/jxb/erab397

Ghosh, U. K., Islam, M. N., Siddiqui, M. N., Cao, X., & Khan, M. A. R. (2022). Proline, a multifaceted signalling molecule in plant responses to abiotic stress: Understanding the physiological mechanisms. *Plant Biology, 24*(2), 227–239. https://doi.org/10.1111/plb.13363

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research, 40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Grivet, L., & Arruda, P. (2002). Sugarcane genomics: Depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology*, *5*(2), 122–127. https://doi.org/10.1016/S1369-5266(02)00253-3

Guarneri, N., Willig, J. J., Sterken, M. G., Zhou, W., Hasan, M. S., Sharon, L., Grundler, F. M. W., Willemsen, V., Goverse, A., Smant, G., & Lozano-Torres, J. L. (2023). Root architecture plasticity in response to endoparasitic cyst nematodes is mediated by damage signaling. *New Phytologist*, *237*(3), 807–822. https://doi.org/10.1111/nph.18570

Gui, J., Zheng, S., Shen, J., & Li, L. (2015). Grain setting defect1 (GSD1) function in rice depends on S-acylation and interacts with actin 1 (OsACT1) at its C-terminal. *Frontiers in Plant Science*, *6*, 804. https://doi.org/10.3389/fpls.2015.00804

Gurrieri, L., Merico, M., Trost, P., Forlani, G., & Sparla, F. (2020). Impact of drought on soluble sugars and free proline content in selected Arabidopsis mutants. *Biology*, *9*(11), 367. https://doi.org/10.3390/biology9110367

Gutierrez, L., Castelain, M., Verdeil, J. L., Conejero, G., & Van Wuytswinkel, O. (2008). A possible role of prolyl oligopeptidase during *Linum usitatissimum* (flax) seed development. *Plant Biology*, *10*(3), 398–402. https://doi.org/10.1111/j.1438-8677.2008.00038.x

Haberer, G., Young, S., Bharti, A. K., Gundlach, H., Raymond, C., Fuks, G., Butler, E., Wing, R. A., Rounsley, S., & Birren, B. (2005). Structure and architecture of the maize genome. *Plant Physiology*, *139*(4), 1612–1624. https://doi.org/10.1104/pp.105.068718

Healey, A.L., Garsmeur, O., Lovell, J.T. et al. The complex polyploid genome architecture of sugarcane. Nature (2024). https://doi.org/10.1038/s41586-024-07231-4 Kvam, V. M., Liu, P., & Si, Y. (2012).

Ikeda, M., Mitsuda, N., & Ohme-Takagi, M. (2011). Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiology*, *157*, 1243–1254. https://doi.org/10.1104/pp.111.183659

International Rice Genome Sequencing Project, & Sasaki, T. (2005). The map-based sequence of the rice genome. *Nature*, *436*(7052), 793–800. https://doi.org/10.1038/nature03895

Jiang, H., Ma, Q. J., Zhong, M. S., Gao, H. N., Li, Y. Y., & Hao, Y. J. (2021). The apple palmitoyltransferase MdPAT16 influences sugar content and salt tolerance via an MdCBL1–

MdCIPK13–MdSUT2.2 pathway. *The Plant Journal, 106*, 689–705. https://doi.org/10.1111/tpj.15219

Karp, S. G., Medina, J. D. C., Letti, L. A. J., Woiciechowski, A. L., Carvalho, J. C. de, Schmitt, C. C., Penha, R. de O., Kumlehn, G. S., & Soccol, C. R. (2021). Bioeconomy and biofuels: the case of sugarcane ethanol in Brazil. Biofuels, Bioproducts and Biorefining. Advance online publication. https://doi.org/10.1002/bbb.2195 .

Kavi Kishor, P. B., Suravajhala, P., Rathnagiri, P., & Sreenivasulu, N. (2022). Intriguing role of proline in redox potential conferring high temperature stress tolerance. *Frontiers in Plant Science, 13*, 867531. https://doi.org/10.3389/fpls.2022.867531

Kazemi-Shahandashti, R., & Maali-Amiri, R. (2018). Global insights of protein responses to cold stress in plants: Signaling, defence, and degradation. *Journal of Plant Physiology, 226*, 123–135. https://doi.org/10.1016/j.jplph.2018.03.022

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L. S., & Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science*, 242, 14–22

Kim, C., Wang, X., Lee, T. H., Jakob, K., Lee, G. J., & Paterson, A. H. (2014). Comparative analysis of *Miscanthus* and *Saccharum* reveals a shared whole-genome duplication but different evolutionary fates. *Plant Cell, 26*(6), 2420–2429. https://doi.org/10.1105/tpc.114.125583

Kohany, O., Gentles, A. J., Hankus, L., & Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, *7*, 474. https://doi.org/10.1186/1471-2105-7-474

Kolde, R. (2012). pheatmap: Pretty heatmaps. R package version 1.0.12. Available from https://CRAN.R-project.org/package=pheatmap

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research, 27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Krochko, J. E., Abrams, G. D., Loewen, M. K., Abrams, S. R., & Cutler, A. J. (1998). (+)-Abscisic acid 8'-hydroxylase is a cytochrome P450 monooxygenase. *Plant Physiology, 118*(3), 849–860. https://doi.org/10.1104/pp.118.3.849

Kvam, V. M., Liu, P., & Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany*. Advance online publication. https://doi.org/10.3732/ajb.1100340

Langfelder, P., Horvath, S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1, 54 (2007). https://doi.org/10.1186/1752-0509-1-54

Loh, L., Lo, C. S., & Tseng, P. M. (1951). Notes on sugar cane nobilization methods. In *Proceedings of the Seventh Congress of the International Society of Sugar Cane Technologists, Brisbane, 25th August-15th September, 1950* (pp. 254-265).

Lorenzi, B. R., & Novaes de Andrade, T. H. (2019). O etanol de segunda geração no Brasil: políticas e redes sociotécnicas [Second generation ethanol in Brazil: Policies and socio-technical networks]. Revista Brasileira de Estudos Sociais, 34(100). https://doi.org/10.1590/3410014/2019 .

Luo, T., Liu, X., & Lakshmanan, P. (2023). A combined genomics and phenomics approach is needed to boost breeding in sugarcane. Plant Phenomics, 5, Article 0074. https://doi.org/10.34133/plantphenomics.0074

Ma, H., Yang, T., Li, Y., Zhang, J., Wu, T., Song, T., Yao, Y., & Tian, J. (2021). The long noncoding RNA MdLNC499 bridges MdWRKY1 and MdERF109 function to regulate early-stage light-induced anthocyanin accumulation in apple fruit. *Plant Cell, 33*(10), 3309–3330. https://doi.org/10.1093/plcell/koab188

Mahood EH, Kruse LH, Moghe GD. Machine learning: A powerful tool for gene function prediction in plants. Appl Plant Sci. 2020 Jul 28;8(7):e11376. doi: 10.1002/aps3.11376. PMID: 32765975; PMCID: PMC7394712.

Mancini, M. C., Cardoso-Silva, C. B., Sforça, D. A., & Pereira de Souza, A. (2018). "Targeted sequencing by gene synteny," a new strategy for polyploid species: Sequencing and physical structure of a complex sugarcane region. *Frontiers in Plant Science, 9*, 397. https://doi.org/10.3389/fpls.2018.00397

Mancini MC, Leite DC, Perecin D, Bidóia MAP, Xavier MA, Landell MGA et al. (2012). Characterization of the genetic variability of a sugarcane commercial cross through yield components and quality parameters. *Sugar Tech* 14, 119–125 .

Manmai, N., Unpaprom, Y., Ponnusamy, V. K., Ramaraj, R.. (2020). Bioethanol production from the comparison between optimization of sorghum stalk and sugarcane leaf for sugar production by chemical pretreatment and enzymatic degradation. Fuel, 278, 118262. https://doi.org/10.1016/j.fuel.2020.118262

Marin, F. R. (2022). Variedades - EMBRAPA. Retrieved from https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana/pre-producao/caracteristicas/variedades. Accessed March 2024.

Mattiello, L., Riaño-Pachón, D.M., Martins, M.C.M. *et al.* Physiological and transcriptional analyses of developmental stages along sugarcane leaf. *BMC Plant Biol* 15, 300 (2015). https://doi.org/10.1186/s12870-015-0694-z

McCormick, A.J., D.A. Watt, and M.D. Cramer. 2009. Supply and demand: Sink regulation of sugar accumulation in sugarcane. *Journal of Experimental Botany* 60: 357–364.

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B. D., McKinley, B., Mattison, A., Morishige, D. T., Grimwood, J., Schmutz, J., & Mullet, J. E. (2017). The *Sorghum bicolor* reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant Journal*. Advance online publication. https://doi.org/10.1111/tpj.13909

Meena, M. R., Appunu, C., Kumar, R. A., Manimekalai, R., Vasantha, S., Krishnappa, G., Kumar, R., Pandey, S. K., & Hemaprabha, G. (2022). Recent Advances in Sugarcane Genomics, Physiology, and Phenomics for Superior Agronomic Traits. *Frontiers in Genetics, 13*, Article 854936. doi:10.3389/fgene.2022.854936

Melo, F. A. D. (2015). Sugarcane Agricultural Production, Bioenergy and Ethanol (pp. 407-422). In Sugarcane: Chapter 19 - Remuneration System of Sugarcane.

Miao, J., Feng, Q., Li, Y., et al. (2021). Chromosome-scale assembly and analysis of biomass crop *Miscanthus lutarioriparius* genome. *Nature Communications, 12*, 2458. https://doi.org/10.1038/s41467-021-22738-4

Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D., & Noel, J. P. (2016). Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Current Opinion in Structural Biology*, *41*, 233–246.

Ming, R., Liu, S. C., Lin, Y. R., da Silva, J., Wilson, W., Braga, D., van Deynze, A., Wenslaff, T. F., Wu, K. K., Moore, P. H., Burnquist, W., Sorrells, M. E., Irvine, J. E., & Paterson, A. H. (1998). Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. *Genetics, 150*(4), 1663–1682. https://doi.org/10.1093/genetics/150.4.1663

Ming, R., Liu, S.-C., Moore, P. H., Irvine, J. E., & Paterson, A. H. (2001). QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane. *Genome Research, 11*(11), 2075-2084. https://doi.org/10.1101/gr.198801

Ming, R., Wang, .Y., Draye, .X. *et al.* Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. *Theor Appl Genet* **105**, 332–345 (2002). https://doi.org/10.1007/s00122-001-0861-5

Mirdar Mansuri, R., Shobbar, Z. S., Babaeian Jelodar, N., et al. (2020). Salt tolerance involved candidate genes in rice: An integrative meta-analysis approach. *BMC Plant Biology*, 20(1), 452. https://doi.org/10.1186/s12870-020-02679-8

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mitros, T., Session, A. M., James, B. T., et al. (2020). Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nature Communications, 11*, 5442. https://doi.org/10.1038/s41467-020-18923-6

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant Journal, 90*(6), 1007–1013. https://doi.org/10.1111/tpj.13515

Moon, YH., Cha, YL., Choi, YH. et al. Diversity in ploidy levels and nuclear DNA amounts in Korean Miscanthus species. *Euphytica* 193, 317–326 (2013). https://doi.org/10.1007/s10681-013-0910-6

Moore, P. H., Paterson, A. H., & Tew, T. (2013). Sugarcane: the crop, the plant, and domestication. In *Sugarcane: physiology, biochemistry, and functional biology* (pp. 1-17).

Movahedi, S., Van Bel, M., Heyndrickx, K. S., & Vandepoele, K. (2012). Comparative co-expression analysis in plant biology. *Plant, Cell & Environment*, 35(10), 1787-1798. https://doi.org/10.1111/j.1365-3040.2012.02517.x

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., & Persson, S. (2009). Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiology, 152*(1), 29–43. https://doi.org/10.1104/pp.109.145318

Ng, L. M., Melcher, K., Teh, B. T., & Xu, H. E. (2014). Abscisic acid perception and signaling: Structural mechanisms and applications. *Acta Pharmacologica Sinica, 35*(5), 567-584. https://doi.org/10.1038/aps.2014.5

Nishiyama, M. Y., Ferreira, S. S., Tang, P.-Z., Becker, S., Pörtner-Taliana, A., & Souza, G. M. (2014). Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. *PLoS ONE, 9*(8), e107351. https://doi.org/10.1371/journal.pone.0107351

Nishiwaki, A., Mizuguti, A., Kuwabara, S., Toma, Y., Ishigaki, G., Miyashita, T., Yamada, T., Matuura, H., Yamaguchi, S., Rayburn, A. L., Akashi, R., & Stewart, J. R. (2011). Discovery of natural Miscanthus (Poaceae) triploid plants in sympatric populations of Miscanthus sacchariflorus and Miscanthus sinensis in southern Japan. American Journal of Botany, 98(1), 154-159. https://doi.org/10.3732/ajb.1000258

Pagariya, M. C., Devarumath, R. M., & Kawar, P. G. (2012). Biochemical characterization and identification of differentially expressed candidate genes in salt stressed sugarcane. Plant Science, 184, 1-13. https://doi.org/10.1016/j.plantsci.2011.12.002

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., & Poliakov (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, 457, 551-556.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods, 14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Phytozome. (2022). *Phytozome Homepage*. Joint Genome Institute. Available from https://phytozome-next.jgi.doe.gov/

Phytozome 13 - Miscanthus sinensis v7.1 DOE-JGI [Internet]. Available from http://phytozome.jgi.doe.gov/

Piperidis, G., Piperidis, N., & D'Hont, A. (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics, 284*(1), 65–73. https://doi.org/10.1007/s00438-010-0546-3

United Nations, Department of Economic and Social Affairs. (n.d.). Sustainable Development. Retrieved from https://sdgs.un.org/goals

R Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Raffaele, S., Perraki, A., & Mongrand, S. (2013). The Remorin C-terminal anchor was shaped by convergent evolution among membrane binding domains. *Plant Signaling & Behavior, 8*(3). https://doi.org/10.4161/psb.23207

Rao, X., & Dixon, R. A. (2016). The Differences between NAD-ME and NADP-ME Subtypes of C4 Photosynthesis: More than Decarboxylating Enzymes. *Frontiers in Plant Science, 7*, Article 1525. https://doi.org/10.3389/fpls.2016.01525

Rao, X., & Dixon, R. A. (2019). Co-expression networks for plant biology: Why and how. *Acta Biochimica et Biophysica Sinica*, 51(10), 981–988. https://doi.org/10.1093/abbs/gmz080.

Redd, P. S., Payero, L., Gilbert, D. M., Page, C. A., King, R., McAssey, E. V., Bodie, D., Diaz, S., & Hancock, N. (2023). Transposase expression, element abundance, element size, and DNA repair determine the mobility and heritability of PIF/Pong/Harbinger transposable elements. *Frontiers in Cell and Developmental Biology*. https://doi.org/10.3389/fcell.2023.1184046

Rejeb, K. B., Abdelly, C., & Savouré, A. (2014). How reactive oxygen species and proline face stress together. *Plant Physiology and Biochemistry, 80*, 278–284. https://doi.org/10.1016/j.plaphy.2014.04.007

Riaño-Pachón, D. M., & Mattiello, L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000 Research, 6*, 861. https://doi.org/10.12688/f1000research.12443.2

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rosolen, R. R., Aono, A. H., Almeida, D. A., Ferreira Filho, J. A., Horta, M. A. C., & De Souza, A. P. (2022). Network analysis reveals different cellulose degradation strategies across *Trichoderma harzianum* strains associated with XYR1 and CRE1. *Frontiers in Genetics, 13*, 807243. https://doi.org/10.3389/fgene.2022.807243

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al. (2000). Artemis: Sequence visualization and annotation. *Bioinformatics, 16*(10), 944–945. https://doi.org/10.1093/bioinformatics/16.10.944

Sage, R. F., Peixoto, M. M., & Sage, T. L. (2014). Photosynthesis in sugarcane. In P. H. Moore & F. C. Botha (Eds.), *Sugarcane: Physiology, Biochemistry, and Functional Biology* (1st ed., pp. 122-154). John Wiley & Sons, Inc.

Sajid, M., Amjid, M., Munir, H., Ahmad, M., Zulfiqar, U., Ali, M. F., Farah, M. A., Ahmed, M. A. A., & Artyszak, A. (2023). Comparative analysis of growth and physiological responses of sugarcane elite genotypes to water stress and sandy loam soils. *Plants, 12*(15), 2759. https://doi.org/10.3390/plants12152759

Santos, Luís Paulo dos. (2019) Identificação e caracterização de uma região genômica de cana-de-açúcar utilizando um QTL para Brix em sorgo: Identification and characterization of a genomic region from sugarcane using a QTL for Brix in sorghum. 2019. 1 recurso online (77 p.) Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Biologia, Campinas, SP. Disponível em: https://hdl.handle.net/20.500.12733/1639520. Acesso em: 22 abr. 2024.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C. J., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B. W., Pruitt, K. D., & Sherry, S. T. (2022). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Sharma, P., Sharma, P., Arora, P., Verma, V., Khanna, K., Saini, P., et al. (2019). Role and regulation of ROS and antioxidants as signaling molecules in response to abiotic stresses. In M. Iqbal, R. Khan, P. Sudhakar Reddy, A. Ferrante, & N. A. Khan (Eds.), *Plant Signaling Molecules* (pp. 141–156). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-816451-8.00008-3

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science, 326*(5956), 1112–1115. https://doi.org/10.1126/science.1178534

Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M., & Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science,* 7, Article 444. https://doi.org/10.3389/fpls.2016.00444

Sforça, D. A., Vautrin, S., Cardoso-Silva, C. B., Mancini, M. C., Romero-da Cruz, M. V., Pereira, G. S., Conte, M., Bellec, A., Dahmer, N., Fourment, J., Rodde, N., Van Sluys, M. A., Vicentini, R., Garcia, A. A. F., Forni-Martins, E. R., Carneiro, M. S., Hoffmann, H. P., Pinto, L. R., Landell, M. G. A., Vincentz, M., Berges, H., & de Souza, A. P. (2019). Gene Duplication in the Sugarcane Genome: A Case Study of Allele Interactions and Evolutionary Patterns in Two Genic Regions. *Frontiers in Plant Science,* 10, 553. doi:10.3389/fpls.2019.00553.

Shahjaman, M., Mollah, M. M. H., Rahman, M. R., Islam, S. M. S., & Mollah, M. N. H. (2020). Robust identification of differentially expressed genes from RNA-seq data. *Genomics*, 112(2), 2000-2010. https://doi.org/10.1016/j.ygeno.2019.07.014

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shiringani, A. L., Frisch, M., & Friedt, W. (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. *Theoretical and Applied Genetics, 121*(2), 323–336. https://doi.org/10.1007/s00122-010-1312-y

Shizuya, H., Birren, B., Kim, U., Mancino, V., Slepak, T., Tachiri, Y., & Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the USA*, 89, 8794-8797.

Sills, G.R., Bridges, W., Al-Janabi, S.M. *et al.* Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. ex Grassl). *Mol Breeding* 1, 355–363 (1995). https://doi.org/10.1007/BF01248413

Singels, A., Jackson, P., & Inman-Bamber, G. (2021). Sugarcane. In V. O. Sadras & D. F. Calderini (Eds.), *Crop Physiology Case Histories for Major Crops* (pp. 674–713). Academic Press. https://doi.org/10.1016/B978-0-12-819194-1.00021-9

Singh, A. (2020). Benefits of crop diversification in Fiji's sugarcane farming. *Asia & the Pacific Policy Studies, 7*(1), 65–80. https://doi.org/10.1002/app5.291

Singh, R., Irikura, B., Nagai, C., et al. (2011). Characterization of prolyl oligopeptidase genes differentially expressed between two cultivars of *Coffea arabica* L. *Tropical Plant Biology, 4*, 203–216. https://doi.org/10.1007/s12042-011-9082-5

Smit, M. A., & Singels, A. (2006). The response of sugarcane canopy development to water stress. *Field Crops Research, 98*(2–3), 91–97. https://doi.org/10.1016/j.fcr.2005.12.009

Sociedade Brasileira de Genética. (2017). Marcadores Moleculares na Era Genômica: Metodologias e Aplicações (A. C. Turchetto-Zolet, C. Turchetto, C. M. Zanella, & G. Passaia, Eds.). Ribeirão Preto. ISBN 978-85-89265-26-3

Srinivas, G., Satish, K., Murali Mohan, S., Nagaraja Reddy, R., Madhusudhana, R., Balakrishna, D., Venkatesh Bhat, B., Howarth, C. J., & Seetharama, N. (2008). Development of genic-microsatellite markers for sorghum staygreen QTL using a comparative genomic approach with rice. *Theoretical and Applied Genetics, 117*(2), 283–296. https://doi.org/10.1007/s00122-008-0773-8

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics, 20*(8), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Takatsuji, H. (1999). Zinc-finger proteins: The classical zinc finger emerges in contemporary plant science. *Plant Molecular Biology*.

Tan, C. M., Chen, R. J., Zhang, J. H., Gao, X. L., Li, L. H., Wang, P. R., Deng, X. J., & Xu, Z. J. (2013). OsPOP5, a prolyl oligopeptidase family gene from rice, confers abiotic stress tolerance in *Escherichia coli*. *International Journal of Molecular Sciences, 14*(10), 20204–20219. https://doi.org/10.3390/ijms141020204

Teresinski, H. J., Hau, B., Symonds, K., Kilburn, R., Munro, K. A., Doner, N. M., Mullen, R., Li, V. H., & Snedden, W. A. (2023). Arabidopsis calmodulin-like proteins CML13 and CML14

interact with proteins that have IQ domains. *Plant Cell & Environment, 46*, 2470–2491. https://doi.org/10.1111/pce.14616

The International Wheat Genome Sequencing Consortium, Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(6403), eaar7191. https://doi.org/10.1126/science.aar7191

The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Thirugnanasambandam, P. P., Hoang, N. V., & Henry, R. J. (2018). The Challenge of Analyzing the Sugarcane Genome. *Frontiers in Plant Science, 9*, Article 616. https://doi.org/10.3389/fpls.2018.00616

Thorne, R. F. (1992). An updated phylogenetic classification of the flowering plants. *Aliso: A Journal of Systematic and Floristic Botany*, *13*(2), Article 8. https://scholarship.claremont.edu/aliso/vol13/iss2/8

Tomiyama, H., Sato, M., Opio, P., et al. (2020). Inhibition of Abscisic Acid 8′-Hydroxylase affects dehydration tolerance and root formation in cuttings of grapes (*Vitis labrusca* L. × *Vitis vinifera* L. cv. Kyoho) under drought stress conditions. *Journal of Plant Growth Regulation, 39*, 1577–1586. https://doi.org/10.1007/s00344-020-10171-8

Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing R (1999) A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics* 99, 419-424

Toppa, E. V. B., Jadoski, C. J., Julianetti, A., Hulshof, T., & Ono, E. O. (2010). Physiology development in the vegetative stage of sugarcane. *Applied Research & Agrotechnology*, *3*(2). https://doi.org/10.5777/paet.v3i2.1064

Tsuruta, S., Ebina, M., Kobayashi, M., & Takahashi, W. (2017). Complete chloroplast genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative genomics and evolution of the *Saccharum* complex. *PLoS ONE*, *12*(10), e0169992. https://doi.org/10.1371/journal.pone.0169992

Tulchinsky, T. H., & Cohen, M. J. (2023). Global health. In T. H. Tulchinsky & M. J. Cohen (Eds.), The New Public Health (4th ed., pp. 1097-1158). Elsevier Inc. https://doi.org/10.1016/B978-0-12-822957-6.00002-8

van der Merwe, M. J., & Botha, F. C. (2014). Chapter 7: Respiration as a competitive sink for sucrose accumulation in sugarcane culm: Perspectives and open questions. In P. H. Moore & F. C. Botha (Eds.), *Sugarcane: Physiology, Biochemistry, and Functional Biology*. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/9781118771280.ch7

van der Vlugt-Bergmans, C. J., & van der Werf, M. J. (2001). Genetic and biochemical characterization of a novel monoterpene epsilon-lactone hydrolase from *Rhodococcus erythropolis* DCL14. *Applied and Environmental Microbiology*, *67*(2), 733–741. https://doi.org/10.1128/AEM.67.2.733-741.2001

van Ooijen, J. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83, 613–624 (1999). https://doi.org/10.1038/sj.hdy.6886230

Vasantha, S., Kumar, R. A., Tayade, A. S., et al. (2022). Physiology of sucrose productivity and implications of ripeners in sugarcane. *Sugar Tech*, *24*(5), 715–731. https://doi.org/10.1007/s12355-021-01062-7

Vettore, A. L. (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research*, *13*(12), 2725–2735. https://doi.org/10.1101/gr.1532103

Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H., & Souza, G. M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnology Journal, 8(3), 263-276. https://doi.org/10.1111/j.1467-7652.2009.00491.x .

Wang, J., Roe, B., Macmil, S., Yu, Q., Murray, J. E., Tang, H., et al. (2010). Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics*, *11*, 261. https://doi.org/10.1186/1471-2164-11-261

Wang, M., Dai, W., Du, J., Ming, R., Dahro, B., & Liu, J. H. (2019). ERF109 of trifoliate orange (*Poncirus trifoliata* (L.) Raf.) contributes to cold tolerance by directly regulating expression of Prx1 involved in antioxidative process. *Plant Biotechnology Journal*, *17*(7), 1316–1332. https://doi.org/10.1111/pbi.13119

Wang, T., Fang, J. & Zhang, J. Advances in Sugarcane Genomics and Genetics. *Sugar Tech* **24**, 354–368 (2022). https://doi.org/10.1007/s12355-021-01065-4

Wang, X., Jin, B., Yan, W., Wang, J., Xu, J., Cai, C., Qi, X., Xu, Q., Yang, X., Xu, X., & Chen, X. (2023). Cucumber abscisic acid 8′-hydroxylase Csyf2 regulates yellow flesh by modulating carotenoid biosynthesis. *Plant Physiology*, *193*(2), 1001–1015. https://doi.org/10.1093/plphys/kiad383

Watt, D. A., McCormick, A. J., & Cramer, M. D. (2014). Source and sink physiology. In P. H. Moore & F. C. Botha (Eds.), *Sugarcane: Physiology, Biochemistry, and Functional Biology* (1st ed., pp. 483-520). John Wiley & Sons, Inc.

Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*(W1), W265–W268. https://doi.org/10.1093/nar/gkm286

Yasmeen, S., Khan, M. T., & Khan, I. A. (2020). Revisiting the physical mutagenesis for sugarcane improvement: A stomatal perspective. *Scientific Reports*, 10(1), 16003. https://doi.org/10.1038/s41598-020-73087-z

Yu, G., Sun, B., Zhu, Z., Mehareb, E. M., Teng, A., Han, J., Zhang, H., Liu, J., Liu, X., Raza, G., Zhang, B., Zhang, Y., & Wang, K. (2024). Genome-wide DNase I-hypersensitive site assay reveals distinct genomic distributions and functional features of open chromatin in autopolyploid sugarcane. *Plant Journal*, *117*(2), 573–589. https://doi.org/10.1111/tpj.16513

Yuksel B, Paterson AH (2005). Construction and characterization of a peanut HindII BAC library. *Theoretical and Applied Genetics*. 111, 630-639.

Zhang, B., Huang, Y., Zhang, L., Zhou, Z., Zhou, S., Duan, W., Yang, C., Gao, Y., Li, S., Chen, M., Li, Y., Yang, X., Zhang, G., & Huang, D. (2023). Genome-Wide Association Study Unravels Quantitative Trait Loci and Genes Associated with Yield-Related Traits in Sugarcane. *Journal of Agricultural and Food Chemistry*, *71*(44), 16815–16826. https://doi.org/10.1021/acs.jafc.3c02935

Zhang, G., Ge, C., Xu, P., Wang, S., Cheng, S., Han, Y., Wang, Y., Zhuang, Y., Hou, X., Yu, T., Xu, X., Deng, S., Li, Q., Yang, Y., Yin, X., Wang, W., Liu, W., Zheng, C., Sun, X., Wang, Z., Ming, R., Dong, S., Ma, J., Zhang, X., & Chen, C. (2021). The reference genome of

Miscanthus floridulus illuminates the evolution of Saccharinae. *Nature Plants*, *7*(5), 608–618. https://doi.org/10.1038/s41477-021-00908-y

Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology & Evolution*, *18*(6), 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8

Zhang, J., Nagai, C., Yu, Q., Pan, Y.-B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three *Saccharum* species. *Euphytica*, *185*(3), 511–519. https://doi.org/10.1007/s10681-012-0664-6

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics*, *50*(12), 1565–1573. https://doi.org/10.1038/s41588-018-0237-2

Zhao, D., Ferguson, A. A., & Jiang, N. (2016). What makes up plant genomes: The vanishing line between transposable element and genes. *Biochimica et Biophysica Acta*, *1859*(3), 366–380. https://doi.org/10.1016/j.bbagrm.2015.12.007

Zhao, M., Ma, L., Song, N., Cheng, J., Zhao, Z., & Wu, J. (2022). The regulation of *Alternaria alternata* resistance by LRR-RK4 through ERF109, defensin19, and phytoalexin scopoletin in *Nicotiana attenuata*. *Plant Science*, *323*, 111414. https://doi.org/10.1016/j.plantsci.2022.111414

Zhou, X. N., Zheng, Y. Y., Cai, Z., Wang, X. Y., Liu, Y., Yu, A., Chen, X. L., Liu, J., Zhang, Y., & Wang, A. X. (2021). Identification and functional analysis of tomato TPR gene family. *International Journal of Molecular Sciences*, *22*, 758. https://doi.org/10.3390/ijms22020758

Zhu, Y. J., Albert, H. H., & Moore, P. H. (2000). Differential expression of soluble acid invertase genes in the shoots of high-sucrose and low-sucrose species of *Saccharum* and their hybrids. *Australian Journal of Plant Physiology*, *27*(2), 193–199.

Anexos

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **"IDENTIFICAÇÃO DE GENES CANDIDATOS PARA ACÚMULO DE AÇÚCAR EM CULTIVARES DE CANA-DE AÇÚCAR (Saccharum spp.) A PARTIR DE UMA REGIÃO GENÔMICA SINTÊNICA COM SORGO (Sorghum bicolor) E DE UMA REDE DE COEXPRESSÃO.**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 15 de agosto de 2024

Assinatura : _____
Nome do(a) autor(a): **Mônica Letícia Turíbio Martins**
RG n.° 26463189-4

Assinatura : _____
Nome do(a) orientador(a): **Anete Pereira de Souza**
RG n.° 8680325-6

COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "***"IDENTIFICAÇÃO DE GENES CANDIDATOS PARA ACÚMULO DE AÇÚCAR EM CULTIVARES DE CANA-DE AÇÚCAR (Saccharum spp.) A PARTIR DE UMA REGIÃO GENÔMICA SINTÊNICA COM SORGO (Sorghum bicolor) E DE UMA REDE DE COEXPRESSÃO.***", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: _____
Nome do(a) aluno(a): Mônica Letícia Turíbio Martins

Assinatura: _____
Nome do(a) orientador(a): Anete Pereira de Souza

Data: 15/08/2024