University of São Paulo "Luiz de Queiroz"
College of Agriculture

De novo assembly and analysis of energy cane transcripts for sugarcane smut disease studies

Gabriela Romêro Campos

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2024

Gabriela Romêro Campos
Agronomic Engineering

De novo assembly and analysis of energy cane transcripts for sugarcane smut disease studies

Advisor:
Dr. CLAUDIA BARROS MONTEIRO-VITORELLO

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2024

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Campos, Gabriela Romêro

De novo assembly and analysis of energy cane transcripts for sugarcane smut disease studies / Gabriela Romêro Campos. - - Piracicaba, 2024.

63 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. *Saccharum* spp. 2. *Sporisorium scitamineum* 3. RNAseq 4. Transcritos diferencialmente expressos 5. Ortologia I título.

With all my love,

To Thamires and Claudio, who illuminate my path with love and support every day.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to everybody and all institutions that contributed to the success of this work.

I am thankful to the "Luiz de Queiroz" College of Agriculture at the University of São Paulo for the opportunity to participate in the Plant Genetics and Breeding program as a master's student and for their academic support.

I am grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil(CAPES) for the financial support, and I extend my appreciation to BioVertis/GranBio and the Agronomic Institute of Campinas for providing the necessary plant material for this study.

I am immensely grateful to my advisor, Dr. Claudia Barros Monteiro Vitorello, who welcomed and supported me throughout my master's journey. I am extremely thankful for her trust and teachings. It was an honor to learn from a professional who is completely passionate about her work and who, with enthusiasm and dedication, cares to pass on this passion.

I also want to thank all my colleagues from the Genomics Group: Enrico Diniz, Thaís Dal'Sasso, Renato Bombardelli, Pedro Vilanova, Ana Julia Franceschi, Jonathan Macedo, Joyce Ferreti, Jéssica Mendes, Gustavo Crestana, João Vitor Ferro, Gustavo Husein, Gustavo Lima, Marcella Ferreira, Tiarla Souto, Maurício de Souza, and Elaine Vidotto. I deeply appreciate all the support they offered during my master's, taught me a lot, and whenever I needed help, there was someone willing to assist. Thank you for making the work environment pleasant and fun and, above all, for helping me grow professionally and personally.

I thank my husband, Claudio Dias da Silva Jr., for always being by my side and supporting me every day, giving me strength and courage to face each new challenge. I also express my gratitude to my sister, Thamires Romêro Campos, and my friends, Laura Panzarin Nerastri, Maria Fernanda Verissimo de Oliveira, and Nathan Barreto. Thank you for making life lighter and being there whenever I need you. Your love gives me the strength to continue.

Finally, I thank my parents, Marilene Romêro Neves Campos and Alexandre Costa Campos, for everything they have done for me that has led me to where I am today. Without you, none of this would be possible.

“A vida em seu nível mais simples e mais fundamental demonstra um nível impressionante de complexidade.”

- Unknow

SUMMARY

RESUMO

Montagem de novo e análise de transcritos de cana-energia para estudos de doenças do carvão da cana-de-açúcar

A cana energia, ao contrário da cana-de-açúcar convencional, foi selecionada para ter um maior teor de fibra do que sacarose. Nas últimas décadas, o cultivo de cana-energia aumentou devido ao grande potencial da cultura para a produção de energia renovável, possibilitando a redução na dependência de combustíveis fósseis e nos impactos ambientais causados por esses. Até o momento, nenhum genoma ou transcriptoma desta cultura foi relatado, principalmente no que diz respeito à interação com o fungo *Sporisorium scitamineum* ou outros patógenos. O *S. scetamineum* é um dos principais patógenos que atacam a cana energia, causando impactos significativos na produtividade. Ha informações disponíveis na interação entre o patógeno e a cana-de-açúcar, mas muito é necessário para entender a interação do patógeno com a cana energia. A fim de lançar alguma luz nesta interação, foi realizado duas montagens *de novo* do transcriptoma de genótipos Vertix 1 suscetíveis e Vertix 2 resistentes de cana-energia com dados de RNA-seq obtidos de gemas de cana-energia 48 horas após a inoculação (48hpi). Análises comparativas foram conduzidas utilizando os transcriptomas *de novo* de ambos os genótipos, cultivares modernas de cana-de-açúcar e espécies relacionadas. Além disso, foi realizada uma análise de expressão diferencial. Identificamos que a cana energia apresenta semelhanças com cultivares modernas de cana-de-açúcar e com a espécie *S. officinarum* e as respostas de defesa são semelhantes às observadas nas interações entre cana-de-açúcar e *S. scitamineum*. No entanto, ela ainda possui características próprias. Esses resultados trazem informações importantes sobre a interação cana-energia x *S. scitamineum*.

Palavras-chave: *Sporisorium scitamineum*, RNAseq, Interação planta-patógeno, Complexo *Saccharum*, Melhoramento genético de cana energia

ABSTRACT

De novo assembly and analysis of energy cane transcripts for sugarcane smut disease studies

Energy sugarcane has emerged as a significant resource of biomass for several countries, offering promising potential for the production of biofuels and other industrial products. To date, no genomes or transcriptomes of this crop have been reported, especially regarding interaction with the fungus *Sporisorium scitamineum* or other pathogens. The fungus *S. scetamineum* is one of the main pathogens that attack energy cane, causing significant impacts on yield. A good amount of information is available in the interaction between the pathogen and sugarcane, but much is needed to understand the interaction of the pathogen with energy cane. In order to shed some light in this interaction, two de novo assembly of the transcriptome of susceptible Vertix 1, and resistant Vertix 2 genotypes of energy cane was assembled with RNA-seq data obtained from energy cane buds 48 hours post inoculation (48hpi). Comparative analyses were conducted using both de novo transcriptomes, modern sugarcane cultivars, and related species. Additionally, a differential expression analysis was performed. We identified that despite energy cane presenting similarities with modern sugarcane cultivars and the species *S. officinarum* and as the defense responses are similar to those observed in interactions between sugarcane and *S. scitamineum*, it still has its own characteristics. These results bring important insights on the energy cane - *S. scitamineum* interaction.

Keywords: *Sporisorium scitamineum*, RNAseq, Plant-pathogen interaction, *Saccharum* complex, Energy cane breeding

1. LITERATURE REVIEW

1.1. Energy cane

Energy cane has been developed as a promising bioenergy crop due to its high fiber content and potential for bioethanol production (Duval et al., 2013). The initial emergence of this crop dates back to the second half of the 1970s in Puerto Rico (Matsuoka et al., 2014), and since then, there has been an increase in the annual production of energy cane. One of the key events that has shaped the current scenario was the 1970 petroleum embargo led by the member countries of the Organization of the Petroleum Exporting Countries (OPEC), which led to the urgent need for alternative energy sources (Matsuoka et al., 2014). Different countries adopted distinct strategies in response to this situation. For instance, the United States opted for ethanol production from corn crops (Mumm et al., 2014), while Brazil chose sugarcane as the primary crop for ethanol production (Coombs, 1984; Goldemberg, 2008; Xavier, 2007). Through the National Alcohol Program (ProAlcool), implemented by the Brazilian government in 1975, the sugarcane industry was able to develop, becoming one of the most efficient systems in converting photosynthates into energy in the present day (Cursi et al., 2022; Matsuoka et al., 2014).

The ProÁlcool program was initiated to produce anhydrous ethanol from sugarcane (*Saccharum officinarum*) for blending with gasoline. Following the second oil crisis in 1979, the program was expanded with the goal of producing ethanol as a substitute fuel for gasoline (Michellon et al., 2008). In this phase, ethanol was produced through the fermentation of sugarcane sugar. However, the industry faced challenges due to high operational costs, limited productivity, and competition between energy generation and food production (Santos et al., 2016). The use of cellulose as a raw material has opened up exciting opportunities for ethanol production. To meet industry demands for second-generation ethanol, a variety of technologies for biomass processing have been developed. These include genetically modified microorganisms that can efficiently convert C5 and C6 sugars into ethanol (Demeke et al., 2013), more streamlined processes for cellulosic ethanol production, and the creation of new hybrids known as energy cane that have higher biomass yields (Carvalho-Netto et al., 2014; Cursi et al., 2022; Silveira et al., 2016; Santos et al., 2016; Matsuoka, 2017).

The first attempts to create these new hybrids were carried out in Louisiana and Puerto Rico with the aim of increasing biomass production, regardless of sucrose content (Santos et al., 2016). Energy cane cultivars are obtained by crossing modern cultivars of *S. officinarum* with genotypes of *S. spontaneum* (Tew & Cobill, 2008). Compared to traditional sugarcane cultivars, energy cane has unique characteristics such as narrower leaf blades, thinner stems, increased

tillering, and high productivity with rapid and vigorous initial development (Abreu et al., 2020; Matsuoka et al., 2014; Surendra et al., 2018). Moreover, energy cane is possessing increased resistance to biotic and abiotic stresses, allowing cultivation in regions not traditionally reserved for food production (Carvalho-Netto et al., 2014). Therefore, energy cane is an environmentally friendly crop as it requires fewer agrochemicals and fertilizers, captures carbon efficiently, and generates renewable energy (Carvalho-Netto et al., 2014; Barbosa et al., 2020).

Breeding programs aimed at developing energy cane hybrids have four main objectives. Firstly, increasing the potential for biomass production enhances the capacity for converting atmospheric carbon into organic carbon. Secondly, enhancing energy density makes energy cane more efficient than food plants. Thirdly, enhancing resistance to pathogens and biotic stresses reduces the crop's dependence on agrochemicals. Lastly, adapting the root system improves carbon fixation efficiency in the soil and controls erosion (Cursi et al., 2022). There are variations in the strategies defined by breeding programs in the development of different energy cane cultivars. However, most institutions establish their strategies based on the desired relative proportion of sugar and fiber (Barbosa et al., 2020). Fiber and sugar content are the primary factors used to classify genotypes into two types (Tew & Cobill, 2008).

Type I cultivars are varieties selected to maximize sugar and fiber content. Comparatively, energy cane Type I presents higher fiber content than the traditional sugarcane, while low variation in sugar content (Tew & Cobill, 2008). Therefore, Type I cultivars can be used for both sugar and ethanol production. In addition, due to their higher fiber content, they contribute more to the production of second-generation ethanol (E2G) and electricity (Barbosa et al., 2020). Type II cultivars are selected to maximize fiber content only (Tew & Cobill, 2008). The demand for this type of material primarily comes from industries generating energy through biomass (Barbosa et al., 2020). Table 1 provides a comparison between two energy cane cultivars produced in Brazil by the company GranBio, Vertix 1 (Type I) and Vertix 2 (Type II), and traditional sugarcane.

Tabela 1. Table 1: Main differences between Type I, Type II energy cane and sugar cane. (Adapted from Cursi et al. (2022)).

	Sugarcane	Type I energy cane	Type II energy cane
Productivity (x)	x	> 1.5 x	> 2.0 x
Sugars (kg/t)	150	> 100	< 100
Fiber (%)	15	18 to 22	> 25

Number of cuts	4 to 5	8 to 10	> 10
Resistance to pest and diseases	+	++	+++
Industrial use	Sugar and ethanol	Sugar, ethanol and energy	Ethanol 1G, E2G, biochemicals, energy and biomethane

A third type of energy cane was proposed due to the identification of intrinsic variations in sugarcane and energy cane genotypes' biomass characteristics (Santchurn et al., 2014). Categorizing energy cane into three types improves the understanding and utilization of genetic variability present in genetic materials, allowing new strategies for breeding and more effective selection (Santchurn et al., 2014). The classification is based on the proposal by Tew and Cobil in 2008, however, Type I is further divided into two categories: genotypes with a higher fiber concentration without compromising sugar content and genotypes with a significant reduction in sugar content and a fiber content of no more than 22% (Santchurn et al., 2014) (Figure 1).

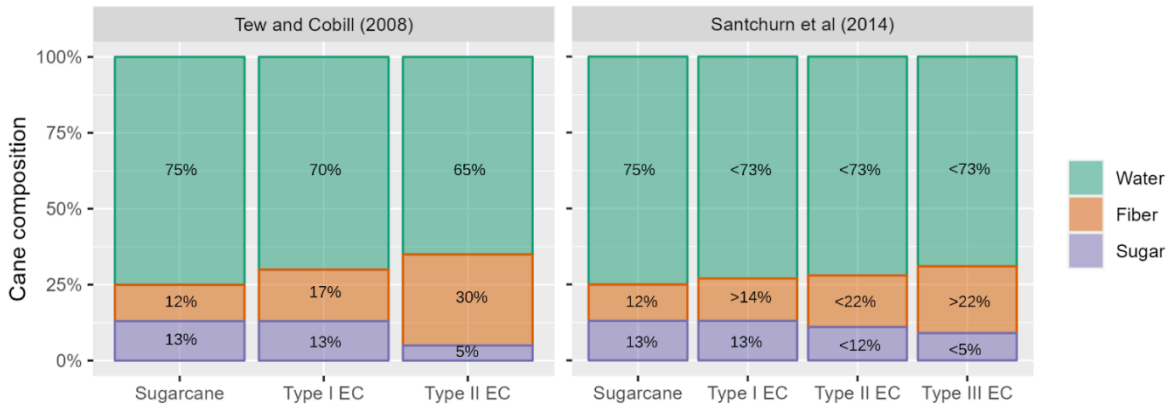

Figure 1. Composition and categories of energy cane according to Tew and Cobil (2008) and Santchurn et al. (2013). (Adapted from Tew e Cobil (2008) and Santchurn et al. (2013)).

There are different types of energy cane that play distinct roles in the sugarcane industry (Matsuoka et al., 2014; Cursi, 2022). The industry requires exploration of significant genetic diversity to meet the demands of various sectors and to develop different cultivars. This is especially important for the sugar and fiber composition of species within the Saccharum and Miscanthus complexes (Barbosa et al., 2020). Major sugarcane-producing countries use accessions from the *Saccharum* genus and modern cultivars to produce new hybrids. In Brazil, key germplasms are maintained by the Sugarcane Technology Center (CTC), BioVertis/GranBio, Agronomic Institute of Campinas (IAC), and the Inter-university Network for the Development

of the Sugarcane Industry (RIDESA) (Cursi et al., 2022). Other significant centers for research and crop improvement include The Sugarcane Breeding Institute (SBI) in India, the United States Department of Agriculture (USDA) in the United States, Sugar Research Australia (SRA) in Australia, the South African Sugarcane Research Institute (SASRI) in South Africa, the Mauritius Sugar Industry Research Institute (MSIRI) in Mauritius, and the West Indies Central Sugarcane Breeding Station (WICSCBS) in Barbados (Barbosa et al., 2020).

Specific production data for energy cane are scarce, as it is a relatively new crop. Generally, the production of energy cane is reported as a part of sugarcane production, which was the most produced crop worldwide until 2022, with Brazil leading as the largest producer globally, followed by India, China, and Thailand (FAOSTAT, 2024). The production capacity of energy cane varies significantly depending on the environmental conditions of each country. However, many countries that cultivate sugarcane also produce E2G and have specific breeding programs for energy cane (Figure 2). This indicates that energy cane has great potential as a sustainable alternative to power the global energy matrix.

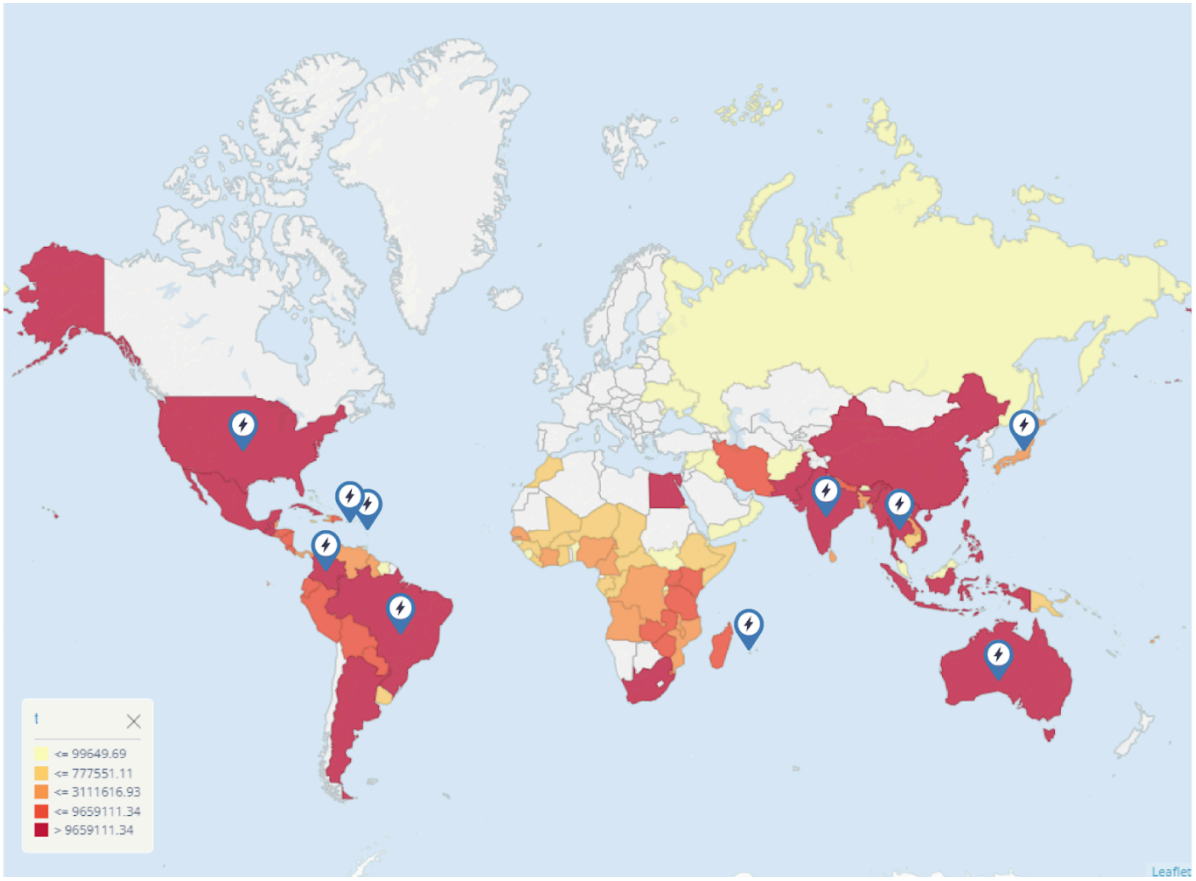

Figure 2. World sugarcane production in 2022 in tons. Each red dot on the map indicates that the country has a sugarcane breeding program and countries with the biofuel symbol produce ethanol. (Adapted from Barbosa et al. (2020), Diniz et al. (2019), FAOSTAT (2024)).

During the 2022/23 harvest in Brazil, approximately 54.1% of the total sugarcane and energy cane produced were used to ethanol production. Notably, 57% of this production consisted of E2G (UNICA, 2024). Additionally, bioelectricity production from sugarcane is the fourth most important source in the Brazil's electrical matrix. It generates 72% (18.4 thousand GWh) of all biomass electricity (UNICA, 2024). If the biomass available in sugarcane fields were fully used, bioelectricity could potentially reach 151 thousand GWh, representing over 30% of the National Integrated System (SIN)'s energy consumption (UNICA, 2024). This highlights the increasing trend of using biomass for bioenergy production. Energy cane has the potential to further increase bioenergy production and meet the national demand for it.

The potential of energy cane biomass to generate biofuels and bioelectricity can be increased by managing the factors that limit its growth. Diseases affecting sugarcane and energy cane are among the main factors responsible for productivity losses, with sugarcane smut, caused by the biotrophic fungus *Sporisorium scitamineum*, being particularly detrimental, capable of causing losses of up to 100% when highly susceptible varieties are cultivated (Tokeshi & Rago, 2016). Therefore, it is essential to investigate the interaction between smut disease and energy cane to develop new strategies that mitigate these losses and promote more sustainable and productive cultivation.

1.2. Sugarcane smut

The sugarcane smut disease is caused by fungi belonging to the Basidiomycota phylum, which can attack various species, including both cultivated and wild (Singh et al., 2004). *Sporisorium scitamineum* (Syd.) (Syn: *Ustilago scitaminea*) is the agent responsible for smut disease in sugarcane and energy cane, representing one of the main threats to the crop (Bhuiyan et al., 2021; Monteiro Vitorello et al., 2018; Rajput et al., 2021). Anecdotal evidence suggests that the fungus originated in India, with species *S. barberi* and *S. spontaneum* acting as hosts (Bhuiyan et al., 2021). Southeast Asia was predicted as the center of origin for sugarcane smut through study of genetic diversity of different isolates (Cortes et al., 2024) since showed greater genetic diversity compared to isolates collected from 15 other countries, particularly in African and American populations (Braithwaite et al., 2024; Que et al., 2014a; Raboin et al., 2007; Shen et al., 2016; Xu et al., 2004).

The first report of the disease was in 1877 in Natal, South Africa (Lee-Lovick, 1978). Since then, it has spread to other producing countries in Central, East, and West Africa, Indonesia, Central and South America, Australia (Croft et al., 2008; Thokoane & Rutherford, 2001) and, recently it has been reported in Papua New Guinea (Tom et al., 2017). In Brazil, the first report was in São Paulo in 1946 (Rago et al., 2009)and genetic evidence suggests that the

dispersion was mediated by humans within Brazilian territory and among neighboring countries (Benevenuto et al., 2016). Currently, Fiji is the only sugarcane-producing country that has not been affected by this fungus (Bhuiyan et al., 2021; Monteiro Vitorello et al., 2018; Rajput et al., 2021).

The damages caused by sugarcane smut disease vary depending on the pathogen races, environmental conditions, cultivar genotype, as well as the interaction between these factors (Rajput et al., 2021). When plants get infected, their contribution to crop yield is minimal, which can result in total losses at harvest (Cortes et al., 2024; Tokeshi & Rago, 2016). The infection affects plant physiology, including changes in photosynthetic rate and significant reductions in sucrose content, brix, purity, and other indicators of juice quality. Plant morphology is negatively affected by the fungi, resulting in reduced height, stem diameter, number of internodes, and increased fibrous content of the stems (Bhuiyan et al., 2021; Mansoor et al., 2016; Marchelo-d'Ragga, 2016; Rajput et al., 2021). Productivity losses are correlated with the number of whips per area (Rajput et al., 2021). The whips are structures formed due to modifications in the apical meristem of the stem and can be easily identified in the field (Ferreira & Comstock, 1989). They vary in length from a few centimeters to over 1.5 meters (Monteiro Vitorello et al., 2018; Tokeshi & Rago, 2016), and contain a central core of parenchyma and vascular elements, surrounded by a cylinder of teliospores (Ferreira & Comstock, 1989; Marques et al., 2018).

Teliospores are resistance spores that result from karyogamy and hyphae fragmentation (Benevenuto et al., 2016). They are responsible for primary infection when they are spread through soil or planting infected stubble, and secondary infection when they are carried by air and infect healthy crops (Rajput et al., 2021). Teliospore germination under favorable environmental conditions, leading to germination in the probasidium and meiosis, resulting in the production of haploid sporidia. These sporidia grow as yeast-like haploid cells, and host infection occurs when two haploid sporidia of different mating types undergo plasmogamy, giving rise to a dikaryotic infective hypha (Monteiro Vitorello et al., 2018; Taniguti et al., 2015; Tokeshi & Rago, 2016). The infective hypha is capable of infecting the host through the formation of an appressorium, then colonizing it inter- and intracellularly systemically until reaching the apical meristem, eventually undergoing karyogamy, initiating sporogenesis, and inducing the formation of a new whip (Monteiro Vitorello et al., 2018; Taniguti et al., 2015; Tokeshi & Rago, 2016).

During the interaction, the pathogen faces several challenges imposed by the plant, known as defense mechanisms. Such mechanisms can be classified as external and internal or a combination of both (McNeil et al., 2018). External resistance mechanisms are related to physical and chemical barriers, such as the number of scales and quantity of trichomes in the bud region

(Longatto et al., 2015; Monteiro Vitorello et al., 2018) in addition to the presence of flavonoid compounds and phenylpropanoids present in these same regions (Fontaniella et al., 2002; Millanes et al., 2005) that act to inhibit and/or delay the germination of teliospores (Monteiro Vitorello et al., 2018). Internal resistance mechanisms are those expressed when the plant recognizes the presence of the fungus in internal tissues and triggers various defense responses. Among them is an increase in lignification (Santiago et al., 2012), the production of glycoproteins, phytoalexins and polyamines and hormonal changes (McNeil et al., 2018; Monteiro Vitorello et al., 2018). However, the presence and composition of such mechanisms vary among different sugarcane genotypes, resulting in different levels of plant resistance to the pathogen. It has already been observed that smut-resistant sugarcane genotypes perceive the presence of the fungus earlier compared to susceptible genotypes, which leads to extensive transcriptional reprogramming, mainly a strong oxidative burst and modulations in the antioxidant system (Peters et al., 2017).

Different methods can be used to manage sugarcane smut disease, including physically treating the setts with hot water, thinning out diseased plants, or applying fungicides. However, these strategies are not very efficiency (Sundar et al., 2012). The most efffective way to minimize the disease's impact on the crop in by using resistant varieties (Cortes et al., 2024; Rajput et al., 2021). Sugarcane and energy cane varieties exhibit different levels of resistance, and the disease's progression is influenced by the variety's resistance level (Bhuiyan et al., 2021). The resistance level is determined by the number of whips emitted by the fungus in an artificially inoculated population, with the scale ranging from no whips emitted to 12.5% of plants with whips (resistant varieties) to 25.6 to 100% of plants with whips (susceptible varieties) (Lemma et al., 2015).

Several studies have been conducted to determine factors that make a sugarcane variety resistant or another susceptible to the smut. A complex network of defense responses to pathogen attack is being discovered. It has been identified that resistant cultivars exhibit higher levels of ethylene, salicylic acid, and jasmonic acid compared to susceptible cultivars (Chen et al., 2023). Moreover, a reduction in lipid peroxidation has been observed, along with a decrease in catalase activity, an increase in glutathione S-transferase activity (Peters et al., 2017), and a significant increase in hydrogen peroxide accumulation, coinciding with the germination of the fungus's teliospores and the formation of appressoria, essential structures for pathogen infection (Peters et al., 2017). Transcriptome analysis of the *S. scitamineum*-sugarcane interaction has led to the identification of differentially expressed genes associated with resistance in infected plants (Agisha et al., 2022; Huang et al., 2018; McNeil et al., 2018; Que et al., 2014). This information,

along with other data on the genetic architecture of energy cane, can help in the development of more resistant genotypes against smut.

1.3. Genomics and Transcriptomics of Energy Sugarcane

The genome of sugarcane is considered to be the most intricate among cultivated species (Piperidis & D'Hont, 2020). This is because of its complicated evolutionary history. The term "*Saccharum* complex" was coined to describe the closely related genera that contribute to the genetic modern cultivars (Mukherjee, 1957). This complex includes the genera *Erianthus, Miscanthus, Narenga, Saccharum*, and *Sclerostachya* (Tew & Cobill, 2008). Among these, *Saccharum* and *Miscanthus* are more closely related to each other than to the other genera (Hodkinson et al., 2002). There are six species included in the *Saccharum* genus: *S. officinarum* (x = 10, 2n = 80); *S. robustum* (x = 10, 2n = 60, 80); *S. edule* (2n = 60–80); *S. barberi* (2n = 111–120); *S. sinense* (2n = 81–124); and *S. spontaneum* (x = 8, 2n = 40–128) (Tew & Cobill, 2008). However, recently study proposed that there are three founder genomes in the *Saccharum* genus (A, B, and C), with A and B unevenly distributed in *S. officinarum* and its wild ancestor *S. robustum*, while the third, C, is observed in wild *S. spontaneum* (Pompidor et al., 2021). On the other hand, *S. barberi* and *S. sinense* exhibit a lesser relationship with the main *Saccharum* germplasm pool, suggesting possible introgression from other genera (Tew & Cobill, 2008) (Figure 3).

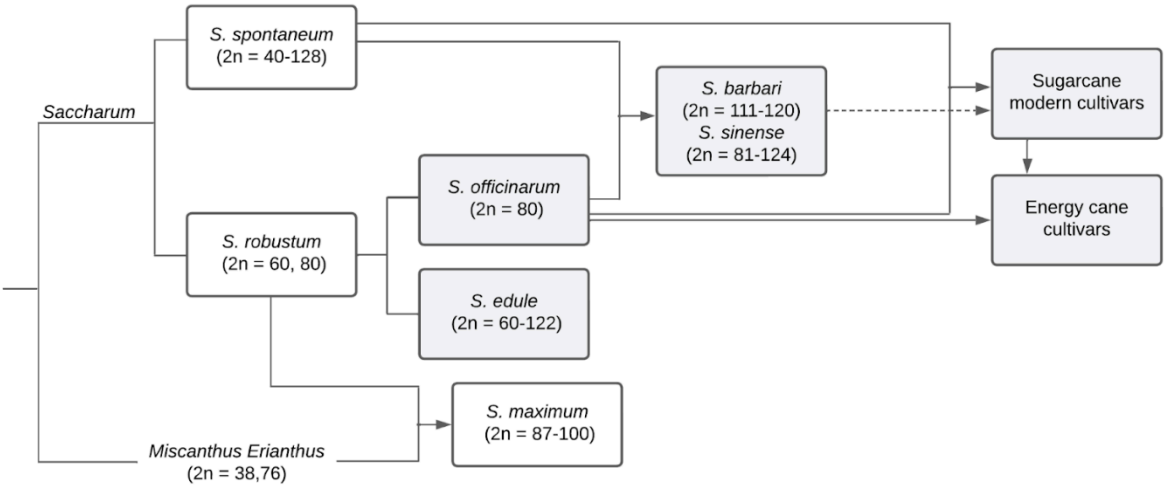

Figure 3. The evolutionary lineage of Saccharum hybrids is depicted as follows: Wild species are represented by white boxes, while cultivated species are depicted in gray. Arrows indicate hybridization events, with dashed lines indicating minor contributions to modern sugarcane cultivars. (Adapted of Setta et al. (2012), D'Hont et al. (2008), Thirugnanasambandam et al. (2018).

Modern sugarcane cultivars are interspecific hybrids resulting from crosses between accessions of *S. officinarum* and *S. spontaneum* (D'Hont et al., 2008). These hybrids have a mixture of genetic contributions from each of these parental species, along with undergoing multiple

events of aneuploidy and recombination between their genomes. In terms of polyploidy, they typically possess between 10 to 13 sets of their 10 basic chromosomes and generally consist of 70-80% of the genetic material from *S. officinarum*, 10-20% from *S. spontaneum*, and approximately 10% with chromosomes recombined between these two ancestors (Piperidis & D'Hont, 2020). Energy cane cultivars originate from interspecific hybridization between modern sugarcane cultivars and accessions of closely related wild species, with the most commonly used being *S. spontaneum* (Silva, 2017). *S. spontaneum* possesses greater genetic diversity within the *Saccharum* complex, is highly polymorphic (Silva, 2017; Tew & Cobill, 2008), and is utilized as a source of adaptability to different environments and resistance to biotic stresses (Silva, 2017).

The genome of modern sugarcane cultivars is complex due to crosses and hybridization. This complexity is further increased by high ploidy and frequent recombination, which can cause molecular and epigenetic changes that affect gene expression and phenotype (Madlung & Wendel, 2013). As a result of this complexity, sugarcane genetics has been slower to advance compared to other economically important crops, and breeding still primarily depends on conventional methods (Piperidis & D'Hont, 2020). While some progress has been made, such as the production of genetic maps based on single-dose markers (Aitken et al., 2014), the lack of high-quality multiple-dose markers makes it difficult to precisely identify homologs and assign them to homology groups. Although sugarcane genetic maps have become denser and cover larger fractions of the genome (Barreto et al., 2019; You et al., 2019).

The genomes of various sugarcane cultivars have been studied to understand their structure and evolution. The assembly of monoploid sequences based on the R570 cultivar (Garsmeur et al., 2018), the construction of 32 pseudo-chromosomes from the haploid genome of the AP85-441 cultivar of *Saccharum spontaneum* (Zhang et al., 2018), along with the draft genome sequence of the SP80-3280 cultivar (Riaño-Pachón & Mattiello, 2017), partial sequences of the same cultivar (Souza et al., 2019), and the genome of the Khon Kaen 3 cultivar of sugarcane (Shearman et al., 2022), have greatly contributed to unraveling the genome structure and its evolutionary aspects. However, understanding the origins of multiple copies of a gene and its different alleles still presents significant challenges due to the high level of polyploidy and aneuploidy, along with structural variations in chromosomes. Especially when specific allelic expression occurs, which has been observed in sugarcane, mainly in genes related to the defense response and cell wall biosynthesis (Correr et al., 2022; Margarido et al., 2022).

Transcriptomic analysis is a highly valuable tool in studying sugarcane and energy cane characteristics. It provides insights into gene expression, isoform diversity, and elucidating the genetic mechanisms underlying specific genotype characteristics (Hoang et al., 2017).

Comparative transcriptome analysis allows for the identification of conserved and divergent gene expression patterns, evolutionary relationships, functional adaptations, and responses to environmental stimuli. Research on sugarcane transcript studies began in South Africa, analyzing 7,242 expressed sequence tags (ESTs) (Carson & Botha, 2002). The largest EST collection was developed by the Brazilian project SUCEST and consists of approximately 238,000 ESTs . Additionally, three more sets of ESTs were generated by researchers in Australia (Bower et al., 2005; Casu et al., 2003) and the United States (Ma et al., 2004). Currently, there are over 280,000 ESTs described and incorporated into the Sugarcane Gene Index (Cardoso-Silva et al., 2014).

Several studies involving transcriptome analysis have been developed to explore the genome and understand the molecular mechanisms underlying different phenotypic characteristics, such as sucrose accumulation, resistance to biotic and abiotic stresses, and vegetative development, among others (Huang et al., 2018; Taniguti et al., 2015; Thirugnanasambandam et al., 2017; Yan et al., 2021). Sugarcane smut disease is a significant biotic stressor, and the first publicly available transcriptome sequencing study of sugarcane using next-generation sequencing (NGS) used six sugarcane genotypes for *de novo* transcriptome analysis and assembly, identifying putative genes related to disease and sucrose accumulation (Cardoso-Silva et al., 2014). Other transcriptome analyses of the sugarcane-*S. scitamineum* interaction led to the identification of similarly differentially expressed genes induced by *S. scitamineum* in resistant and susceptible genotypes (Huang et al., 2018; Que et al., 2014; Schaker et al., 2016), as well as the identification of different resistance mechanisms (McNeil et al., 2018). Various defense signaling pathways activated based on pathogen virulence, hormonal signaling, and defense-related metabolite synthesis involved in defense mechanisms have also been identified in this pathosystem through transcriptomics (Agisha et al., 2022). However, there is still much unknown information in the sugarcane and energy cane genomes, highlighting the need for ongoing exploration.

1.4. *De novo* Assembly of transcriptomes from polyploid organism

RNA-Seq is a widely used methodology for identifying differentially expressed genes and discovering SNPs (Single Nucleotide Polymorphisms) in both model and non-model organisms (Chopra et al., 2014; Duan et al., 2012). When working with organisms that have known reference genomes, the mapping approach is used. This involves aligning reads to annotated references, followed by transcript assembly, SNP identification, and quantification of transcription expression levels based on this mapping information (Chopra et al., 2014). When there is no well-defined reference genome, various strategies can be employed, such as using

references from related species, EST assembly or *de novo* assembly of RNA-Seq data (Chopra et al., 2014). However, using related species as a reference may result in a loss of species-specific information, while EST assembly requires extensive information on ESTs or a genomic database (Chopra et al., 2014).

RNA-Seq proves to be a valuable option for *de novo* transcriptome assembly and it helps in understanding molecular and functional mechanisms (Hölzer & Marz, 2019; Madritsch et al., 2021). However, *de novo* transcriptome assembly has challenges due to the presence of various alleles, closely related paralogs, and homologs, leading to considerable diversity of isoforms (Chopra et al., 2014; Góngora-Castillo & Buell, 2013; Madritsch et al., 2021).This challenge is even more significant in complex polyploid plants due to gene duplications, dosage imbalance, and allele-specific gene expression and the presence of multiple homoeologs in allopolyploids, which adds a level of complexity (Madritsch et al., 2021; Voshall & Moriyama, 2020), as is the case with sugarcane and energy cane. These characteristics may lead to increased rates of fused or redundant transcripts (Madritsch et al., 2021)

To improve the accuracy of *de novo* assembly in transcriptome with polyploid characteristics, it is essential to utilize high-quality materials such as non-fragmented RNA and long sequencing reads (Gutierrez-Gonzalez & Garvin, 2017). It is also important to ensure satisfactory coverage and use an appropriate analysis pipeline to avoid confusion between homoeologous nucleotide differences and sequencing errors (Gutierrez-Gonzalez & Garvin, 2017). Polyploidy has significant impacts on transcript assembly and quantification, especially in separate expression levels between homoeologs or subgenomes (Voshall & Moriyama, 2020). Indeed, allopolyploids may exhibit unequal expression among duplicated genes, known as homoeolog expression bias, which can result in differences in proteome composition and gene expression, even in phenotypically similar allopolyploids (Voshall & Moriyama, 2020; Yoo et al., 2012). Although mapping to a reference genome can correct expression levels between genes in most cases sequence variations among homologs can still lead to inaccurate expression estimates (Voshall & Moriyama, 2020).

In response to these challenges, several strategies and tools have been developed. One alternative to traditional alignment for RNA-seq quantification analyses is the use of pseudo-alignment methods such as Salmon (Patro et al., 2017) and Kallisto (Bray et al., 2016). By employing an expectation-maximization approach (Li et al., 2010) to assign reads to transcripts in situations of ambiguity, these methods can resolve mappings between various homoeologs. This is possible when a sufficient number of reads are exclusively mapped to only one homoeolog, allowing for the estimation of overall expression levels (Voshall & Moriyama, 2020).

Furthermore, pseudo-alignment is more computationally efficient as it does not require exact alignment to genomic references, resulting in faster processing and reduced resource usage (Bray et al., 2016; Patro et al., 2017). By overlapping information between reads and transcripts, pseudo-alignment enables more precise assignment of reads, contributing to more reliable estimates of gene expression levels in polyploid organisms.

Most software developed for assembly based only on RNA-seq data, such as Trinity (Grabherr et al., 2011), employs de Bruijn graphs (Compeau et al., 2011) for assembly. This method involves breaking reads into shorter overlapping sequences of a given length k, called kmers. This approach helps in efficient traversal of overlaps to reconstruct the original sequence (Voshall & Moriyama, 2020). It is also essential to balance the length of kmers with their limitations. Shorter kmers are more likely to cover the transcription sequences that need to be assembled completely. Still, they also increase the risk of incorrect assemblies when the same k-mer is present in reads from multiple transcripts (Voshall & Moriyama, 2020) . These limitations must be balanced with the length of kmers. Shorter kmers are more likely to completely cover the transcription sequences that need to be assembled, but they also have a higher likelihood of resulting in incorrect assemblies when the same k-mer is present in reads from multiple transcripts (Voshall & Moriyama, 2020) . These issues become more complicated for polyploid transcriptome assembly, as different yet highly similar sequences between alleles, as well as homoeologs, lead to more ambiguity in the graph. This greatly enlarges the graph size and increases the probability of traversal errors (Voshall & Moriyama, 2020).

There are software tools available freely or commercially that have been successful in complex organisms (Chopra et al., 2014). These include SOAPdenovo-Trans (Xie et al., 2014), rnaSPAdes (Bushmanova et al., 2019) and Trinity (Grabherr et al., 2011). Trinity stands out as one of the most popular software tools for generating high-quality *de novo* transcriptomes with low base error rates and the ability to capture multiple isoforms, which are crucial for maintaining acceptable levels of accuracy when characterizing genes (Chopra et al., 2014). Despite these strategies, assembling allopolyploid transcriptomes *de novo* remains a challenging task. However, the *de novo* assembly of the energy cane transcriptome still offers significant advantages for scientific studies. This is primarily attributed to the lack of a genomic sequence that fully represents a complete hybrid genome of these plants. Energy cane is a crop that faces these challenges as they are allopolyploid hybrids, and up to this moment, genomes or transcriptomes have not been reported, especially regarding interaction with the fungus *S. scitamineum* or other pathogens. In this scenario, we aim to perform de novo assembly of two transcriptomes, using genotypes contrasting in resistance to the pathogen. What made it possible to explore the

differences in response between the transcriptomes of both genotypes and, through comparative analyses with modern sugarcane cultivars and species related to energy cane, identify whether there is specific information regarding energy cane.

1.5. Objective

1.5.1. General objective

This work aims to construct a reference transcriptome of contrasting energy cane varieties for infection with *S. scitamineum*. We aim to compare the molecular events involved in the response of resistant and susceptible plants and to compare them to those described for sugarcane. We anticipate that these signatures will differ from those previously described for sugarcane, particularly considering the discrepancy in initial growth rates between sugarcane and energy cane. By comparing these transcriptomic profiles, we aim to illuminate the path to understanding fundamental molecular mechanisms underlying resistance and susceptibility to *S. scitamineum* in energy cane varieties.

1.5.2. Specific objectives

1) Process the sequencing data obtained from experiments involving plants infected and non-infected with *S. scitamineum* regarding parameters such as quality, sequencing depth, completeness, and others for subsequent analyses.
2) Perform assemblies using custom programs for transcriptomic data analysis and select those that best fit the analysis of the obtained sequences.
3) Compare contrasting varieties regarding their response to infection among themselves and other selected genotypes of closely related species with available sequences.

REFERENCES

Abreu, L. G. F., Grassi, M. C. B., de Carvalho, L. M., da Silva, J. J. B., Oliveira, J. V. C., Bressiani, J. A., & Pereira, G. A. G. (2020). Energy cane vs sugarcane: Watching the race in plant development. *Industrial Crops and Products*, *156*, 112868. https://doi.org/10.1016/j.indcrop.2020.112868

Agisha, V. N., Ashwin, N. M. R., Vinodhini, R. T., Nalayeni, K., Ramesh Sundar, A., Malathi, P., & Viswanathan, R. (2022). Transcriptome analysis of sugarcane reveals differential switching of major defense signaling pathways in response to Sporisorium scitamineum isolates with varying virulent attributes. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.969826

Aitken, K. S., McNeil, M. D., Hermann, S., Bundock, P. C., Kilian, A., Heller-Uszynska, K., Henry, R. J., & Li, J. (2014). A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. *BMC Genomics*, *15*(1), 1–12. https://doi.org/10.1186/1471-2164-15-152/TABLES/6

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*.

Apezzato-da-Gloria, B. A., Albernas, M. C. C., & Amorim, L. (1995). Structural characteristics of buds of sugarcane cultivars with different levels for resistance to smut/Strukturelle Charakteristika der Knospen von Zuckerrohrsorten mit unterschiedlichen Resistenzgraden gegen Zuckerrohrbrand. *Zeitschrift Für Pflanzenkrankheiten Und Pflanzenschutz/Journal of Plant Diseases and Protection*, 502–508.

Barbosa, G. V. S., Dos Santos, J. M., Diniz, C. A., Cursi, D. E., & Hoffmann, H. P. (2020). Energy cane breeding. In *Sugarcane Biorefinery, Technology and Perspectives* (pp. 103–116). Elsevier. https://doi.org/10.1016/B978-0-12-814236-3.00006-8

Barreto, F. Z., Bachega Feijó Rosa, J. R., Almeida Balsalobre, T. W., Pastina, M. M., Silva, R. R., Hoffmann, H. P., de Souza, A. P., Franco Garcia, A. A., & Carneiro, M. S. (2019). A genome-wide association study identified loci for yield component traits in sugarcane (Saccharum spp.). *PLOS ONE*, *14*(7), e0219843. https://doi.org/10.1371/JOURNAL.PONE.0219843

Benevenuto, J., Longatto, D. P., Reis, G. V., Mielnichuk, N., Palhares, A. C., Carvalho, G., Saito, S., Quecine, M. C., Sanguino, A., Vieira, M. L. C., Camargo, L. E. A., Creste, S., &

Monteiro-Vitorello, C. B. (2016). Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus. *FEMS Microbiology Letters*, *363*(24). https://doi.org/10.1093/FEMSLE/FNW277

Bhuiyan, S. A., Magarey, R. C., McNeil, M. D., & Aitken, K. S. (2021). Sugarcane Smut, Caused by Sporisorium scitamineum, a Major Disease of Sugarcane: A Contemporary Review. *Phytopathology*, *111*(11), 1905–1917. https://doi.org/10.1094/PHYTO-05-21-0221-RVW/ASSET/IMAGES/LARGE/PHYTO-05-21-0221-RVWF3.JPEG

Bower, N. I., Casu, R. E., Maclean, D. J., Reverter, A., Chapman, S. C., & Manners, J. M. (2005). Transcriptional response of sugarcane roots to methyl jasmonate. *Plant Science*, *168*(3), 761–772. https://doi.org/10.1016/J.PLANTSCI.2004.10.006

Braithwaite, K. S., Bakkeren, G., Croft, B. J., & Brumbley, S. M. (2024). Genetic variation in a worldwide collection of the sugarcane smut fungus Ustilago scitaminea. In *Proceedings-Australian Society Sugar Cane Technologist* (Vol. 26, pp. 48–48). PK Editorial Services.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology 2016 34:5*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), 1–13. https://doi.org/10.1093/GIGASCIENCE/GIZ100

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Costa Canesin, L. E., Pinto, L. R., Carneiro, M. S., Garcia, A. A. F., De Souza, A. P., & Vicentini, R. (2014). De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. *PLOS ONE*, *9*(2), e88462. https://doi.org/10.1371/JOURNAL.PONE.0088462

Carson, D., & Botha, F. (2002). Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Reports 2002 20:11*, *20*(11), 1075–1081. https://doi.org/10.1007/S00299-002-0444-1

Carvalho-Netto, O. V., Bressiani, J. A., Soriano, H. L., Fiori, C. S., Santos, J. M., Barbosa, G. V., Xavier, M. A., Landell, M. G., & Pereira, G. A. (2014). The potential of the energy cane as the main biomass crop for the cellulosic industry. *Chemical and Biological Technologies in Agriculture*, *1*(1), 1–8. https://doi.org/10.1186/S40538-014-0020-2/FIGURES/1

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:1023957214644

Chao, C. P., & Hoy, J. W. (1988). Estimation of heritability of sugarcane smut resistance in Louisiana. *Phytopathology*, *77*(12), 1723–1724.

Chen, S., Chen, Z., Lin, X., Zhou, X., Yang, S., & Tan, H. (2023). Why different sugarcane cultivars show different resistant abilities to smut?: Comparisons of endophytic microbial compositions and metabolic functions in stems of sugarcane cultivars with different abilities to resist smut. *BMC Plant Biology*, *23*(1), 1–16. https://doi.org/10.1186/S12870-023-04446-X/FIGURES/7

Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., & Burow, M. D. (2014). Comparisons of De Novo Transcriptome Assemblers in Diploid and Polyploid Species Using Peanut (Arachis spp.) RNA-Seq Data. *PLOS ONE*, *9*(12), e115055. https://doi.org/10.1371/JOURNAL.PONE.0115055

Compeau, P. E. C., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology 2011 29:11*, *29*(11), 987–991. https://doi.org/10.1038/nbt.2023

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Coombs, J. (1984). Sugar-cane as an Energy Crop. *Biotechnology and Genetic Engineering Reviews*, *1*(1), 311–346. https://doi.org/10.1080/02648725.1984.10647789

Correr, F. H., Furtado, A., Franco Garcia, A. A., Henry, R. J., & Rodrigues Alves Margarido, G. (2022). Allele expression biases in mixed-ploid sugarcane accessions. *Scientific Reports 2022 12:1*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-12725-0

Cortes, J. D., Gutierrez, A. F., Hoy, J. W., Hale, A. L., & Baisakh, N. (2024). Genetic mapping of quantitative trait loci controlling smut resistance in Louisiana sugarcane using a bi-parental mapping population. *Plant Gene*, *37*, 100445. https://doi.org/10.1016/J.PLGENE.2023.100445

Coudert, E., Gehant, S., de Castro, E., Pozzato, M., Baratin, D., Neto, T., Sigrist, C. J. A., Redaschi, N., Bridge, A., Bridge, A. J., Aimo, L., Argoud-Puy, G., Auchincloss, A. H., Axelsen, K. B., Bansal, P., Baratin, D., Neto, T. M. B., Blatter, M.-C., Bolleman, J. T., … Wang, Y.

(2023). Annotation of biologically relevant ligands in UniProtKB using ChEBI. *Bioinformatics*, *39*(1). https://doi.org/10.1093/bioinformatics/btac793

Croft, B. J., Berding, N. I. L. S., Cox, M. C., Bhuiyan, S., & Bruce, R. C. (2008). Breeding smut-resistant sugarcane varieties in Australia: progress and future directions. *Proceedings of the Australian Society of Sugar Cane Technologists*, *30*, 125–134.

Cursi, D. E., Hoffmann, H. P., Barbosa, G. V. S., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A. R., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., & Carneiro, M. S. (2022). History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. *Sugar Tech*, *24*(1), 112–133. https://doi.org/10.1007/S12355-021-00951-1/TABLES/6

Demeke, M. M., Dietz, H., Li, Y., Foulquié-Moreno, M. R., Mutturi, S., Deprez, S., Den Abt, T., Bonini, B. M., Liden, G., Dumortier, F., Verplaetse, A., Boles, E., & Thevelein, J. M. (2013). Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnology for Biofuels*, *6*(1), 1–24. https://doi.org/10.1186/1754-6834-6-89/TABLES/4

D. 'Hont, A., Mendes Souza, G., Menossi, M., Vincentz, M., Van-Sluys, M.-A., Glaszmann, J. C., & Ulian, E. (2008). Sugarcane: A Major Source of Sweetness, Alcohol, and Bio-energy. *Genomics of Tropical Crop Plants*, 483–513. https://doi.org/10.1007/978-0-387-71219-2_21

Diniz, A. L., Ferreira, S. S., ten-Caten, F., Margarido, G. R. A., dos Santos, J. M., Barbosa, G. V. de S., Carneiro, M. S., & Souza, G. M. (2019). Genomic resources for energy cane breeding in the post genomics era. *Computational and Structural Biotechnology Journal*, *17*, 1404–1414. https://doi.org/10.1016/J.CSBJ.2019.10.006

Duan, J., Xia, C., Zhao, G., Jia, J., & Kong, X. (2012). Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *BMC Genomics*, *13*(1), 1–12. https://doi.org/10.1186/1471-2164-13-392/TABLES/8

Duval, B. D., Anderson-Teixeira, K. J., Davis, S. C., Keogh, C., Long, S. P., Parton, W. J., & DeLucia, E. H. (2013). Predicting Greenhouse Gas Emissions and Soil Carbon from Changing Pasture to an Energy Crop. *PLOS ONE*, *8*(8), e72019. https://doi.org/10.1371/JOURNAL.PONE.0072019

Emms, D. M., Kelly, & Affiliations. (2018). STAG: Species Tree Inference from All Genes. *BioRxiv*, 267914. https://doi.org/10.1101/267914

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Emms, D. M., & Kelly, S. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. *Molecular Biology and Evolution*, *34*(12), 3267–3278. https://doi.org/10.1093/molbev/msx259

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/BIOINFORMATICS/BTW354

Fan, J., Yang, J., Wang, Y., Li, G., Li, Y., Huang, F., & Wang, W. (2016). Current understanding on *Villosiclava virens* , a unique flower‐infecting fungus causing rice false smut disease. *Molecular Plant Pathology*, *17*(9), 1321–1330. https://doi.org/10.1111/mpp.12362

FAOSTAT. (2024). *Crops and livestock products: Sugarcane*. https://www.fao.org/faostat/en/#data/QCL/visualize

Ferreira, S. A., & Comstock, J. C. (1989). Smut. *Diseases of Sugarcane*, 211–229. https://doi.org/10.1016/B978-0-444-42797-7.50018-1

Fontaniella, B., Márquez, A., Rodríguez, C. W., Piñón, D., Solas, M. T., Vicente, C., & Legaz, M. E. (2002). A role for sugarcane glycoproteins in the resistance of sugarcane to Ustilago scitaminea. *Plant Physiology and Biochemistry*, *40*(10), 881–889. https://doi.org/10.1016/S0981-9428(02)01443-2

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M. A., … D'Hont, A. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications 2018 9:1*, *9*(1), 1–10. https://doi.org/10.1038/s41467-018-05051-5

Glassop, D., Rae, A. L., & Bonnett, G. D. (2014). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech*, *16*(3), 235–240. https://doi.org/10.1007/s12355-013-0284-z

Goldemberg, J. (2008). The Brazilian biofuels industry. *Biotechnology for Biofuels*, *1*(1), 6. https://doi.org/10.1186/1754-6834-1-6

Góngora-Castillo, E., & Buell, C. R. (2013). Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural Product Reports*, *30*(4), 490–500. https://doi.org/10.1039/C3NP20099J

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology 2011 29:7*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gutierrez-Gonzalez, J. J., & Garvin, D. F. (2017). De Novo Transcriptome Assembly in Polyploid Species. *Methods in Molecular Biology (Clifton, N.J.)*, *1536*, 209–221. https://doi.org/10.1007/978-1-4939-6682-0_15

Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P. P., Botha, F. C., & Henry, R. J. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics*, *18*(1), 1–22. https://doi.org/10.1186/S12864-017-3757-8/FIGURES/9

Hodkinson, T. R., Chase, M. W., & Renvoize, S. A. (2002). Characterization of a Genetic Resource Collection for Miscanthus (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, *89*(5), 627–636. https://doi.org/10.1093/AOB/MCF091

Hölzer, M., & Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, *8*(5), 1–16. https://doi.org/10.1093/GIGASCIENCE/GIZ039

Huang, N., Ling, H., Su, Y., Liu, F., Xu, L., Su, W., Wu, Q., Guo, J., Gao, S., & Que, Y. (2018). Transcriptional analysis identifies major pathways as response components to Sporisorium scitamineum stress in sugarcane. *Gene*, *678*, 207–218. https://doi.org/10.1016/J.GENE.2018.08.043

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/J.JMB.2015.11.006

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/NMETH.3317

Kolde, R., & Kolde, M. R. (2015). Package 'pheatmap.' *R Package*, *1*(7).

Lee-Lovick, G. (1978). Smut of sugarcane-Ustilago scitaminea. *Review of Plant Pathology*, *57*(5), 181–188.

Lemma, A., Hagos, H., Zekarias, Y., & Tekle, A. (2015). Study on the Reaction of Sugarcane Genotypes (CIRAD-2011) to Sugarcane Smut (Sporisorium scitamineum) in the Ethiopian Sugarcane Plantations. *Advances in Crop Science and Technology*, *3*(4), 1–3. https://doi.org/10.4172/2329-8863.1000181

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., & Dewey, C. N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, *26*(4), 493–500. https://doi.org/10.1093/BIOINFORMATICS/BTP692

Longatto, D. P., Carvalho, G., Benevuto, J., Peters, L. P., Schaker, P. D., Taniguti, L. M., & Monteiro-Vitorello, C. B. (2015). Carvão da cana-de-açúcar: avanços na compreensão deste patossistema. *Revisão Anual de Patologia de Plantas*, *23*, 62–89.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Madlung, A., & Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenetic and Genome Research*, *140*(2–4), 270–285. https://doi.org/10.1159/000351430

Madritsch, S., Burg, A., & Sehr, E. M. (2021). Comparing de novo transcriptome assembly tools in di- and autotetraploid non-model plant species. *BMC Bioinformatics*, *22*(1), 1–17. https://doi.org/10.1186/S12859-021-04078-8/TABLES/2

Ma, H. M., Schulze, S., Lee, S., Yang, M., Mirkov, E., Irvine, J., Moore, P., & Paterson, A. (2004). An EST survey of the sugarcane transcriptome. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *108*(5), 851–863. https://doi.org/10.1007/S00122-003-1510-Y

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, *1*(12), e323. https://doi.org/10.1002/CPZ1.323

Mansoor, S., Aslam Khan, M., Aslamkhan, M., Ahmed Khan, N., Rasool Nasir, I., Umer, M., Distributor Pakistan Tobacco Company, B., & Sharif Dist Khan, T. D. (2016). *Effect of Whip Smut Disease on the Quantitative and Qualitative Parameters of Sugarcane Varieties/Lines*. https://doi.org/10.19080/ARTOAJ.2016.02.555588

Marchelo-d'Ragga, P. W. (2016). Estimation of yield loss in sugarcane (Saccharum spp.) due to effects of smut disease in some selected sugarcane genotypes under Sudan conditions. *J. Agric. Res*, *2*, 24–32.

Margarido, G. R. A., Correr, F. H., Furtado, A., Botha, F. C., & Henry, R. J. (2022). Limited allele-specific gene expression in highly polyploid sugarcane. *Genome Research*, *32*(2), 297–308. https://doi.org/10.1101/GR.275904.121

Marques, J. P. R., Hoy, J. W., Appezzato-da-Glória, B., Viveros, A. F. G., Vieira, M. L. C., & Baisakh, N. (2018). Sugarcane cell wall-associated defense responses to infection by Sporisorium scitamineum. *Frontiers in Plant Science*, *9*, 332791. https://doi.org/10.3389/FPLS.2018.00698/BIBTEX

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12.

Matsuoka, S. (2017). Free Fiber Level Drives Resilience and Hybrid Vigor in Energy Cane. *Journal of Scientific Achievements*, *2*(1), 1–35. http://jsciachv.sinaweb.net/article_80746.html

Matsuoka, S., Kennedy, A. J., Santos, E. G. D. dos, Tomazela, A. L., & Rubio, L. C. S. (2014). Energy Cane: Its Concept, Development, Characteristics, and Prospects. *Advances in Botany*, *2014*, 1–13. https://doi.org/10.1155/2014/597275

McNeil, M. D., Bhuiyan, S. A., Berkman, P. J., Croft, B. J., & Aitken, K. S. (2018). Analysis of the resistance mechanisms in sugarcane during Sporisorium scitamineum infection using RNA-seq and microscopy. *PLOS ONE*, *13*(5), e0197840. https://doi.org/10.1371/JOURNAL.PONE.0197840

Michellon, E., Santos, A. A. L., & Rodrigues, J. R. A. (2008). *Breve descrição do Proálcool e perspectivas futuras para o etanol produzido no Brasil*.

Millanes, A. M., Fontaniella, B., Legaz, M. E., & Vicente, C. (2005). Glycoproteins from sugarcane plants regulate cell polarity of Ustilago scitaminea teliospores. *Journal of Plant Physiology*, *162*(3), 253–265. https://doi.org/10.1016/J.JPLPH.2004.05.017

Monteiro Vitorello, C. B., Carvalho Schaker, P. D., Benevenuto, J., Teixeira e Silva, N. de S., & de Almeida, S. S. (2018). *Progress in understanding fungal diseases affecting sugarcane: smut*. 221–243. https://doi.org/10.19103/AS.2017.0035.33

Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, *24*(3), 319–324. https://doi.org/10.1093/bioinformatics/btm585

Mukherjee, S. K. (1957). Origin and Distribution of Saccharum. *Https://Doi.Org/10.1086/335962*, *119*(1), 55–61. https://doi.org/10.1086/335962

Mumm, R. H., Goldsmith, P. D., Rausch, K. D., & Stein, H. H. (2014). Land usage attributed to corn ethanol production in the United States: sensitivity to technological advances in corn

grain yield, ethanol conversion, and co-product utilization. *Biotechnology for Biofuels*, *7*(1), 61. https://doi.org/10.1186/1754-6834-7-61

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 2017 14:4*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Peters, L. P., Carvalho, G., Vilhena, M. B., Creste, S., Azevedo, R. A., & Monteiro-Vitorello, C. B. (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. *Planta*, *245*(4), 749–764. https://doi.org/10.1007/S00425-016-2642-Z/FIGURES/8

Piperidis, N., & D'Hont, A. (2020). Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *The Plant Journal : For Cell and Molecular Biology*, *103*(6), 2039–2051. https://doi.org/10.1111/TPJ.14881

Pompidor, N., Charron, C., Hervouet, C., Bocs, S., Droc, G., Rivallan, R., Manez, A., Mitros, T., Swaminathan, K., Glaszmann, J. C., Garsmeur, O., & D'Hont, A. (2021). Three founding ancestral genomes involved in the origin of sugarcane. *Annals of Botany*, *127*(6), 827–840. https://doi.org/10.1093/AOB/MCAB008

Que, Y., Su, Y., Guo, J., Wu, Q., & Xu, L. (2014a). A Global View of Transcriptome Dynamics during Sporisorium scitamineum Challenge in Sugarcane by RNA-seq. *PLOS ONE*, *9*(8), e106476. https://doi.org/10.1371/JOURNAL.PONE.0106476

Que, Y., Su, Y., Guo, J., Wu, Q., & Xu, L. (2014b). A Global View of Transcriptome Dynamics during Sporisorium scitamineum Challenge in Sugarcane by RNA-seq. *PLOS ONE*, *9*(8), e106476. https://doi.org/10.1371/JOURNAL.PONE.0106476

Raboin, L. M., Selvi, A., Oliveira, K. M., Paulet, F., Calatayud, C., Zapater, M. F., Brottier, P., Luzaran, R., Garsmeur, O., Carlier, J., & D'Hont, A. (2007). Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea. *Fungal Genetics and Biology*, *44*(1), 64–76. https://doi.org/10.1016/J.FGB.2006.07.004

Rago, A. M., Casagrande, M. V., & Massola Júnior, N. S. (2009). Variabilidade patogênica de Ustilago scitaminea no estado de São Paulo. *Summa Phytopathologica*, *35*(2), 93–97. https://doi.org/10.1590/S0100-54052009000200002

Rajput, M. A., Rajput, N. A., Syed, R. N., Lodhi, A. M., & Que, Y. (2021). Sugarcane Smut: Current Knowledge and the Way Forward for Management. *Journal of Fungi 2021, Vol. 7, Page 1095*, *7*(12), 1095. https://doi.org/10.3390/JOF7121095

Riaño-Pachón, D. M., & Mattiello, L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Research*, *6*. https://doi.org/10.12688/F1000RESEARCH.11859.2

Rody, H. V. S., Bombardelli, R. G. H., Creste, S., Camargo, L. E. A., Van Sluys, M. A., & Monteiro-Vitorello, C. B. (2019). Genome survey of resistance gene analogs in sugarcane: Genomic features and differential expression of the innate immune system from a smut-resistant genotype. *BMC Genomics*, *20*(1), 1–17. https://doi.org/10.1186/S12864-019-6207-Y/FIGURES/5

Santchurn, D., Ramdoyal, K., Badaloo, M. G. H., & Labuschagne, M. T. (2014). From sugar industry to cane industry: Evaluation and simultaneous selection of different types of high biomass canes. *Biomass and Bioenergy*, *61*, 82–92. https://doi.org/10.1016/J.BIOMBIOE.2013.11.023

Santiago, R., Alarcón, B., de Armas, R., Vicente, C., & Legaz, M. E. (2012). Changes in cinnamyl alcohol dehydrogenase activities from sugarcane cultivars inoculated with Sporisorium scitamineum sporidia. *Physiologia Plantarum*, *145*(2), 245–259. https://doi.org/10.1111/J.1399-3054.2012.01577.X

Santos, L. V., De Barros Grassi, M. C., Gallardo, J. C. M., Pirolla, R. A. S., Calderón, L. L., De Carvalho-Netto, O. V., Parreiras, L. S., Camargo, E. L. O., Drezza, A. L., Missawa, S. K., Teixeira, G. S., Lunardi, I., Bressiani, J., & Pereira, G. A. G. (2016). Second-Generation Ethanol: The Need is Becoming a Reality. *Https://Home.Liebertpub.Com/Ind*, *12*(1), 40–57. https://doi.org/10.1089/IND.2015.0017

Schaker, P. D. C., Palhares, A. C., Taniguti, L. M., Peters, L. P., Creste, S., Aitken, K. S., Van Sluys, M. A., Kitajima, J. P., Vieira, M. L. C., & Monteiro-Vitorello, C. B. (2016). RNAseq Transcriptional Profiling following Whip Development in Sugarcane Smut Disease. *PLOS ONE*, *11*(9), e0162237. https://doi.org/10.1371/JOURNAL.PONE.0162237

Schmitz, L., McCotter, S., Kretschmer, M., Kronstad, J. W., & Heimel, K. (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. *F1000Research*, *7*, 1812. https://doi.org/10.12688/f1000research.16404.1

Sehgal, D. (2022). *Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane* [Doctoral dissertation]. Universidade de São Paulo.

Setta, N., Metcalfe, C. J., Cruz, G. M. Q., Ochoa, E. A., & Van Sluys, M. A. (2012). Noise or symphony: Comparative evolutionary analysis of sugarcane transposable elements with other grasses. *Topics in Current Genetics*, *24*, 169–192. https://doi.org/10.1007/978-3-642-31842-9_10/FIGURES/00105

Shearman, J. R., Pootakham, W., Sonthirod, C., Naktang, C., Yoocha, T., Sangsrakru, D., Jomchai, N., Tongsima, S., Piriyapongsa, J., Ngamphiw, C., Wanasen, N., Ukoskit, K., Punpee, P., Klomsa-ard, P., Sriroth, K., Zhang, J., Zhang, X., Ming, R., Tragoonrung, S., & Tangphatsornruang, S. (2022). A draft chromosome-scale genome assembly of a commercial sugarcane. *Scientific Reports*, *12*(1). https://doi.org/10.1038/S41598-022-24823-0

Shen, W. kuan, Xu, G. hong, Luo, M. zhu, & Jiang, Z. de. (2016). Genetic diversity of Sporisorium scitamineum in mainland China assessed by SCoT analysis. *Tropical Plant Pathology*, *41*(5), 288–296. https://doi.org/10.1007/S40858-016-0099-Z/METRICS

Silva, J. A. (2017). The Importance of the Wild Cane Saccharum spontaneum for Bioenergy Genetic Breeding. *Sugar Tech*, *19*(3), 229–240. https://doi.org/10.1007/S12355-017-0510-1/METRICS

Silveira, L. C. I., Brasileiro, B. P., Kist, V., Weber, H., Daros, E., Peternelli, L. A., & Barbosa, M. H. P. (2016). Selection in energy cane families. *Crop Breeding and Applied Biotechnology*, *16*(4), 298–306. https://doi.org/10.1590/1984-70332016V16N4A45

Singh, N., Somai, B. M., & Pillay, D. (2004). Smut disease assessment by PCR and microscopy in inoculated tissue cultured sugarcane cultivars. *Plant Science*, *167*(5), 987–994. https://doi.org/10.1016/J.PLANTSCI.2004.05.006

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., Gaiarsa, J. W., Diniz, A. L., Oliveira, M. D. M., Ferreira, S. D. S., Nishiyama, M. Y., Ten-Caten, F., Ragagnin, G. T., Andrade, P. D. M., De Souza, R. F., Nicastro, G. G., Pandya, R., Kim, C., Guo, H., … Heckerman, D. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *GigaScience*, *8*(12), 1–18. https://doi.org/10.1093/GIGASCIENCE/GIZ129

Sundar, A. R., Barnabas, E., Malathi, P., & Viswanathan, R. (2012). A mini review on smut disease of sugarcane caused by Sporisorium scitamineum. *Botany*, *53*, 107–128.

Surendra, K. C., Ogoshi, R., Zaleski, H. M., Hashimoto, A. G., & Khanal, S. K. (2018). High yielding tropical energy crops for bioenergy production: Effects of plant components, harvest years and locations on biomass composition. *Bioresource Technology*, *251*, 218–229. https://doi.org/10.1016/J.BIORTECH.2017.12.044

Taniguti, L. M., Schaker, P. D. C., Benevenuto, J., Peters, L. P., Carvalho, G., Palhares, A., Quecine, M. C., Nunes, F. R. S., Kmit, M. C. P., Wai, A., Hausner, G., Aitken, K. S., Berkman, P. J., Fraser, J. A., Moolhuijzen, P. M., Coutinho, L. L., Creste, S., Vieira, M. L. C., Kitajima, J. P., & Monteiro-Vitorello, C. B. (2015). Complete Genome Sequence of Sporisorium scitamineum and Biotrophic Interaction Transcriptome with Sugarcane. *PLOS ONE*, *10*(6), e0129318. https://doi.org/10.1371/JOURNAL.PONE.0129318

Tao, X., Xia, S., Liu, Q., Li, D., & He, X. (2023). Research progress on microbial control of sugarcane smut. *Ying Yong Sheng Tai Xue Bao = The Journal of Applied Ecology*, *34*(3), 846–852. https://doi.org/10.13287/J.1001-9332.202303.029

Tew, T. L., & Cobill, R. M. (2008). Genetic improvement of sugarcane (Saccharum spp.) as an energy crop. *Genetic Improvement of Bioenergy Crops*, 273–294. https://doi.org/10.1007/978-0-387-70805-8_9/COVER

Thirugnanasambandam, P. P., Hoang, N. V., Furtado, A., Botha, F. C., & Henry, R. J. (2017). Association of variation in the sugarcane transcriptome with sugar content. *BMC Genomics*, *18*(1), 1–22. https://doi.org/10.1186/S12864-017-4302-5/FIGURES/7

Thirugnanasambandam, P. P., Hoang, N. V., & Henry, R. J. (2018). The challenge of analyzing the sugarcane genome. *Frontiers in Plant Science*, *9*, 357194. https://doi.org/10.3389/FPLS.2018.00616/BIBTEX

Thokoane, L. N., & Rutherford, R. S. (2001). cDNA -AFLP differential display of sugarcane (Saccharum spp, hybrids) genes induced by challenge with the fungal pathogen Ustilago scitaminea (sugarcane smut). *South African Sugar Technologists' Association*, *75*, 104–107.

TOKESHI, H., & RAGO, A. (2016). Doenças da cana-de-açúcar. In *Manual de fitopatologia* (Ceres, Vol. 2, pp. 219–233).

Tom, L. E. K. A., Braithwaite, K. A. T. H. Y., & Kuniata, L. S. (2017). Incursion of sugarcane smut in commercial cane crops at Ramu, Papua New Guinea. *39th Conf Aust Soc Sugar Cane Technol ASSCT*, 377–384.

UNICA. (2024, January 20). *União da Indústria de cana-de-açúcar*. https://unicadata.com.br/

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002

van der Linde, K., & Göhre, V. (2021). How Do Smut Fungi Use Plant Signals to Spatiotemporally Orientate on and In Planta? *Journal of Fungi*, *7*(2), 107. https://doi.org/10.3390/jof7020107

Villanueva, R. A. M., & Chen, Z. J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Measurement: Interdisciplinary Research and Perspectives*, *17*(3), 160–167. https://doi.org/10.1080/15366367.2019.1565254

Voshall, A., & Moriyama, E. N. (2020). Next-generation transcriptome assembly and analysis: Impact of ploidy. *Methods*, *176*, 14–24. https://doi.org/10.1016/J.YMETH.2019.06.001

Xavier, M. R. (2007). *The Brazilian Sugarcane Ethanol Experience*.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G. K. S., & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660–1666. https://doi.org/10.1093/BIOINFORMATICS/BTU077

Xu, L., Que, Y., & Chen, R. (2004). Genetic diversity of Ustilago scitaminea in Mainland China. *Sugar Tech*, *6*(4), 267–271. https://doi.org/10.1007/BF02942507/METRICS

Yan, H., Zhou, H., Luo, H., Fan, Y., Zhou, Z., Chen, R., Luo, T., Li, X., Liu, X., Li, Y., Qiu, L., & Wu, J. (2021). Characterization of full-length transcriptome in Saccharum officinarum and

molecular insights into tiller development. *BMC Plant Biology*, *21*(1), 1–12. https://doi.org/10.1186/S12870-021-02989-5/TABLES/4

Yoo, M. J., Szadkowski, E., & Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity 2013 110:2*, *110*(2), 171–180. https://doi.org/10.1038/hdy.2012.94

You, Q., Yang, X., Peng, Z., Islam, M. S., Sood, S., Luo, Z., Comstock, J., Xu, L., & Wang, J. (2019). Development of an Axiom Sugarcane100K SNP array for genetic map construction and QTL identification. *Theoretical and Applied Genetics*, *132*(10), 2829–2845. https://doi.org/10.1007/S00122-019-03391-4/FIGURES/4

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C. M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D. R., Huang, L., Li, Z., … Ming, R. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature Genetics 2018 50:11*, *50*(11), 1565–1573. https://doi.org/10.1038/s41588-018-0237-2

2. *De novo* ASSEMBLY AND ANALYSIS OF ENERGY CANE TRANSCRIPTS FOR SUGARCANE SMUT DISEASE STUDIES

2.1. Abstract

The crescent demand for renewable energy has led to the development of sustainable strategies, such as energy cane (Saccharum spp. hybrids), which emerges as a promising source of biomass for second-generation ethanol (E2G) production. These cultivars result from the crossing between accessions of the *Saccharum* complex and modern sugarcane varieties, aiming to select materials with high biomass production. In addition to E2G production, energy cane has potential for various industrial uses, including sugar production, biodiesel, biopolymers, fibers for thermoplastics, and bioelectricity. However, the interaction between energy cane genotypes and pathogens, such as the fungus *S. scitamineum*, has not been widely explored. In this study, we aimed to define a transcriptome for two energy cane genotypes (Vertix 1 and Vertix 2) contrasting in response to sugarcane smut and created a reference aiming to maintain variability while reducing redundancy through orthology. We performed functional analysis of the transcriptome, identifying differentially expressed and functionally enriched genes. We identified that despite energy cane presenting similarities with modern sugarcane cultivars and the species *S. officinarum,* it still has its own characteristics. Additionally, we observed that the infection and colonization of the fungus caused modifications, and we identified differentially expressed transcripts related to auxin response, reactive oxygen species, lignification, among other defense mechanisms. Our results mainly highlight the complexity of the energy cane transcriptome and provide an initial basis for future investigations into the interactions between this crop and its pathogens, as well as its distinct characteristics compared to other sugarcane varieties.

Keywords: *Saccharum* spp., *Sporisorium scitamineum* , RNAseq, Ortology

REFERENCES

Abreu, L. G. F., Grassi, M. C. B., de Carvalho, L. M., da Silva, J. J. B., Oliveira, J. V. C., Bressiani, J. A., & Pereira, G. A. G. (2020). Energy cane vs sugarcane: Watching the race in plant development. *Industrial Crops and Products*, *156*, 112868. https://doi.org/10.1016/j.indcrop.2020.112868

Agisha, V. N., Ashwin, N. M. R., Vinodhini, R. T., Nalayeni, K., Ramesh Sundar, A., Malathi, P., & Viswanathan, R. (2022). Transcriptome analysis of sugarcane reveals differential switching of major defense signaling pathways in response to Sporisorium scitamineum isolates with varying virulent attributes. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.969826

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*.

Apezzato-da-Gloria, B. A., Albernas, M. C. C., & Amorim, L. (1995). Structural characteristics of buds of sugarcane cultivars with different levels for resistance to smut/Strukturelle Charakteristika der Knospen von Zuckerrohrsorten mit unterschiedlichen Resistenzgraden gegen Zuckerrohrbrand. *Zeitschrift Für Pflanzenkrankheiten Und Pflanzenschutz/Journal of Plant Diseases and Protection*, 502–508.

Bhuiyan, S. A., Magarey, R. C., McNeil, M. D., & Aitken, K. S. (2021). Sugarcane Smut, Caused by Sporisorium scitamineum, a Major Disease of Sugarcane: A Contemporary Review. *Phytopathology*, *111*(11), 1905–1917. https://doi.org/10.1094/PHYTO-05-21-0221-RVW/ASSET/IMAGES/LARGE/PHYTO-05-21-0221-RVWF3.JPEG

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Costa Canesin, L. E., Pinto, L. R., Carneiro, M. S., Garcia, A. A. F., De Souza, A. P., & Vicentini, R. (2014). De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. *PLOS ONE*, *9*(2), e88462. https://doi.org/10.1371/JOURNAL.PONE.0088462

Carvalho-Netto, O. V., Bressiani, J. A., Soriano, H. L., Fiori, C. S., Santos, J. M., Barbosa, G. V., Xavier, M. A., Landell, M. G., & Pereira, G. A. (2014). The potential of the energy cane as the main biomass crop for the cellulosic industry. *Chemical and Biological Technologies in Agriculture*, *1*(1), 1–8. https://doi.org/10.1186/S40538-014-0020-2/FIGURES/1

Chao, C. P., & Hoy, J. W. (1988). Estimation of heritability of sugarcane smut resistance in Louisiana. *Phytopathology*, *77*(12), 1723–1724.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Coudert, E., Gehant, S., de Castro, E., Pozzato, M., Baratin, D., Neto, T., Sigrist, C. J. A., Redaschi, N., Bridge, A., Bridge, A. J., Aimo, L., Argoud-Puy, G., Auchincloss, A. H., Axelsen, K. B., Bansal, P., Baratin, D., Neto, T. M. B., Blatter, M.-C., Bolleman, J. T., … Wang, Y. (2023). Annotation of biologically relevant ligands in UniProtKB using ChEBI. *Bioinformatics*, *39*(1). https://doi.org/10.1093/bioinformatics/btac793

Cursi, D. E., Hoffmann, H. P., Barbosa, G. V. S., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A. R., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., & Carneiro, M. S. (2022). History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. *Sugar Tech*, *24*(1), 112–133. https://doi.org/10.1007/S12355-021-00951-1/TABLES/6

Barbosa, G. V., Dos Santos, J. M., Diniz, C. A., Cursi, D. E., & Hoffmann, H. P. (2020). Energy cane breeding. In *Sugarcane Biorefinery, Technology and Perspectives* (pp. 103–116). Elsevier. https://doi.org/10.1016/B978-0-12-814236-3.00006-8

Emms, D. M., Kelly, & Affiliations. (2018). STAG: Species Tree Inference from All Genes. *BioRxiv*, 267914. https://doi.org/10.1101/267914

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Emms, D. M., & Kelly, S. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. *Molecular Biology and Evolution*, *34*(12), 3267–3278. https://doi.org/10.1093/molbev/msx259

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/BIOINFORMATICS/BTW354

Fan, J., Yang, J., Wang, Y., Li, G., Li, Y., Huang, F., & Wang, W. (2016). Current understanding on *Villosiclava virens* , a unique flower‐infecting fungus causing rice false smut disease. *Molecular Plant Pathology*, *17*(9), 1321–1330. https://doi.org/10.1111/mpp.12362

Ferreira, S. A., & Comstock, J. C. (1989). Smut. *Diseases of Sugarcane*, 211–229. https://doi.org/10.1016/B978-0-444-42797-7.50018-1

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M. A., … D'Hont, A. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications 2018 9:1*, *9*(1), 1–10. https://doi.org/10.1038/s41467-018-05051-5

Glassop, D., Rae, A. L., & Bonnett, G. D. (2014). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech*, *16*(3), 235–240. https://doi.org/10.1007/s12355-013-0284-z

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology 2011 29:7*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/J.JMB.2015.11.006

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/NMETH.3317

Kolde, R., & Kolde, M. R. (2015). Package 'pheatmap.' *R Package*, *1*(7).

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, *1*(12), e323. https://doi.org/10.1002/CPZ1.323

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12.

Matsuoka, S. (2017). Free Fiber Level Drives Resilience and Hybrid Vigor in Energy Cane. *Journal of Scientific Achievements*, *2*(1), 1–35. http://jsciachv.sinaweb.net/article_80746.html

McNeil, M. D., Bhuiyan, S. A., Berkman, P. J., Croft, B. J., & Aitken, K. S. (2018). Analysis of the resistance mechanisms in sugarcane during Sporisorium scitamineum infection using RNA-seq and microscopy. *PLOS ONE*, *13*(5), e0197840. https://doi.org/10.1371/JOURNAL.PONE.0197840

Monteiro Vitorello, C. B., Carvalho Schaker, P. D., Benevenuto, J., Teixeira e Silva, N. de S., & de Almeida, S. S. (2018). *Progress in understanding fungal diseases affecting sugarcane: smut*. 221–243. https://doi.org/10.19103/AS.2017.0035.33

Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, *24*(3), 319–324. https://doi.org/10.1093/bioinformatics/btm585

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 2017 14:4*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Peters, L. P., Carvalho, G., Vilhena, M. B., Creste, S., Azevedo, R. A., & Monteiro-Vitorello, C. B. (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. *Planta*, *245*(4), 749–764. https://doi.org/10.1007/S00425-016-2642-Z/FIGURES/8

Que, Y., Su, Y., Guo, J., Wu, Q., & Xu, L. (2014). A Global View of Transcriptome Dynamics during Sporisorium scitamineum Challenge in Sugarcane by RNA-seq. *PLOS ONE*, *9*(8), e106476. https://doi.org/10.1371/JOURNAL.PONE.0106476

Rajput, M. A., Rajput, N. A., Syed, R. N., Lodhi, A. M., & Que, Y. (2021). Sugarcane Smut: Current Knowledge and the Way Forward for Management. *Journal of Fungi 2021, Vol. 7, Page 1095*, *7*(12), 1095. https://doi.org/10.3390/JOF7121095

Rody, H. V. S., Bombardelli, R. G. H., Creste, S., Camargo, L. E. A., Van Sluys, M. A., & Monteiro-Vitorello, C. B. (2019). Genome survey of resistance gene analogs in sugarcane: Genomic features and differential expression of the innate immune system from a smut-resistant genotype. *BMC Genomics*, *20*(1), 1–17. https://doi.org/10.1186/S12864-019-6207-Y/FIGURES/5

Santchurn, D., Ramdoyal, K., Badaloo, M. G. H., & Labuschagne, M. T. (2014). From sugar industry to cane industry: Evaluation and simultaneous selection of different types of high biomass canes. *Biomass and Bioenergy*, *61*, 82–92. https://doi.org/10.1016/J.BIOMBIOE.2013.11.023

Schaker, P. D. C., Palhares, A. C., Taniguti, L. M., Peters, L. P., Creste, S., Aitken, K. S., Van Sluys, M. A., Kitajima, J. P., Vieira, M. L. C., & Monteiro-Vitorello, C. B. (2016). RNAseq

Transcriptional Profiling following Whip Development in Sugarcane Smut Disease. *PLOS ONE*, *11*(9), e0162237. https://doi.org/10.1371/JOURNAL.PONE.0162237

Schmitz, L., McCotter, S., Kretschmer, M., Kronstad, J. W., & Heimel, K. (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. *F1000Research*, *7*, 1812. https://doi.org/10.12688/f1000research.16404.1

Sehgal, D. (2022). *Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane* [Doctoral dissertation]. Universidade de São Paulo.

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Tao, X., Xia, S., Liu, Q., Li, D., & He, X. (2023). Research progress on microbial control of sugarcane smut. *Ying Yong Sheng Tai Xue Bao = The Journal of Applied Ecology*, *34*(3), 846–852. https://doi.org/10.13287/J.1001-9332.202303.029

TOKESHI, H., & RAGO, A. (2016). Doenças da cana-de-açúcar. In *Manual de fitopatologia* (Ceres, Vol. 2, pp. 219–233).

UNICA. (2024, January 20). *União da Indústria de cana-de-açúcar*. https://unicadata.com.br/

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002

van der Linde, K., & Göhre, V. (2021). How Do Smut Fungi Use Plant Signals to Spatiotemporally Orientate on and In Planta? *Journal of Fungi*, *7*(2), 107. https://doi.org/10.3390/jof7020107

Villanueva, R. A. M., & Chen, Z. J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Measurement: Interdisciplinary Research and Perspectives*, *17*(3), 160–167. https://doi.org/10.1080/15366367.2019.1565254

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C. M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D. R., Huang, L., Li, Z., … Ming, R. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature Genetics 2018 50:11*, *50*(11), 1565–1573. https://doi.org/10.1038/s41588-018-0237-2