UNIVERSIDADE FEDERAL DE VIÇOSA

Constitutive transcriptional differences in sugarcane genotypes reveal tolerance mechanisms against *Diatrea saccharalis*.

Uilian Stefanello de Mello
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2024

UILIAN STEFANELLO DE MELLO

Constitutive transcriptional differences in sugarcane genotypes reveal tolerance mechanisms against *Diatrea saccharalis*.

Thesis submitted to the Plant Production Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Marcio H. Pereira Barbosa

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
M527c
2024

Mello, Uilian Stefanello de, 1993-
Constitutive transcriptional differences in sugarcane genotypes reveal tolerance mechanisms against *Diatrea saccharalis* / Uilian Stefanello de Mello. – Viçosa, MG, 2024.
1 tese eletrônica (68 f.): il. (algumas color.).

Texto em inglês.
Orientador: Márcio Henrique Pereira Barbosa.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2024.
Referências bibliográficas: f. 55-64.
DOI: https://doi.org/10.47328/ufvbbt.2025.463
Modo de acesso: World Wide Web.

1. Cana-de-açúcar - Melhoramento genético. 2. RNA-seq. 3. Cana-de-açúcar - Doenças e pragas. 4. Cana-de-açúcar - Resistência a doenças e pragas. 5. *Diatraea saccharalis*. 6. Fridelina. I. Barbosa, Márcio Henrique Pereira, 1966-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Fitotecnia. III. Título.

CDD 22. ed. 633.613

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

UILIAN STEFANELLO DE MELLO

Constitutive transcriptional differences in sugarcane genotypes reveal tolerance mechanisms against *Diatrea saccharalis*.

Thesis submitted to the Plant Production Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: July 26, 2024.

Assent:

Uilian Stefanello de Mello
Author

Marcio Henrique Pereira Barbosa
Adviser

ACKNOWLEDGMENTS

This work has been sponsored by the following Brazilian research agencies: Coordination for the Improvement of Higher Education Personnel (CAPES; Financing code 001), Minas Gerais State Foundation for Research Aid (FAPEMIG) and National Council of Scientific and Technological Development (CNPq).

I would like to thank the Universidade Federal de Viçosa (UFV), especially to the Departamento de Fitotecnia for giving me this opportunity, for the teaching excellence and infrastructure granted.

My gratitude goes to the Sugarcane Breeding Program of the Federal University of Viçosa (PMGCA/UFV-RIDESA) and the Laboratory of Biotechnology and Plant Breeding (LBMV) for all the necessary support.

A very special thanks to my adviser Dr. Márcio Henrique Pereira Barbosa, for all the guidance, knowledge and assistance granted. I algo acknowledge the members of my D.Sc. committee for their time and collaboration. I am thankful to my friends and labmates for all the collaborative discussions, knowledge exchanged and help carrying out the research experiments.

ABSTRACT

MELLO, Uilian Stefanello de, D.Sc., Universidade Federal de Viçosa, July, 2024. **Constitutive transcriptional differences in sugarcane genotypes reveal tolerance mechanisms against *Diatrea saccharalis*.**. Adviser: Marcio Henrique Pereira Barbosa.

Diatrea saccharalis constitutes a threat to sugarcane productivity, and obtaining borer tolerant cultivars is the best method of control. Although there are studies about the relationship between the interaction of D. saccharalis with sugarcane, little is known about the molecular and genomic basis of defense mechanisms that confer pest tolerance to sugarcane cultivars. In this sense, the objective of this work aimed at analyzing the leaf constitutional transcriptional profile of six contrasting sugarcane genotypes to borer attack: RB867515, RB047212, IM76228, RB057249, RB047016 and RB047248. A total of 4,366 Differentially Expressed Transcripts (DETs) were identified between tolerant and susceptible genotypes. The tolerant sugarcane genotypes demonstrated a broader up-regulation of the jasmonic acid (JA) biosynthetic pathway and transcription factor-related transcripts. Interestingly, transcripts putatively responsible for the synthesis of the triterpenoid Friedelin were highly expressed in the tolerant genotypes. Our results reveal borer tolerance molecular mechanisms in sugarcane and indicate that genes coding for friedelin synthase are potential targets for sugarcane breeding programs.

Keywords: Sugarcane; RNA-seq; Transcriptome responses; Borer tolerance; Fiedelin

RESUMO

MELLO, Uilian Stefanello de, D.Sc., Universidade Federal de Viçosa, julho de 2024. **Diferenças transcricionais constitutivas em genótipos de cana-de-açúcar revelam mecanismos de tolerância contra *Diatrea saccharalis*.**. Orientador: Marcio Henrique Pereira Barbosa.

Diatrea saccharalis constitui uma ameaça à produtividade da cana-de-açúcar e a obtenção de cultivares tolerantes à broca é o melhor método de controle. Embora existam estudos sobre a relação entre a interação da D. saccharalis com a cana-de-açúcar, pouco se sabe sobre as bases moleculares e genômicas dos mecanismos de defesa que conferem tolerância contra pragas em cultivares de cana-de-açúcar. Nesse sentido, o objetivo desse trabalho foi analisar o perfil transcricional constitucional foliar de seis genótipos de cana-de-açúcar contrastantes ao ataque da broca: RB867515, RB047212, IM76228, RB057249, RB047016 e RB047248. Um total de 4.366 transcritos diferencialmente expressos (do inglês, DETs) foram identificados entre genótipos tolerantes e suscetíveis. Os genótipos de cana-de-açúcar tolerantes demonstraram uma regulação positiva mais ampla da via biossintética do ácido jasmônico (JA) e de transcritos relacionados a fatores de transcrição. Curiosamente, os transcritos supostamente responsáveis pela síntese do triterpenóide Fridelina foram altamente expressos nos genótipos tolerantes. Nossos resultados revelam mecanismos moleculares de tolerância à broca da cana e indicam que os genes que codificam a fridelina sintase são alvos potenciais para programas de melhoramento genético da cana-de-açúcar.

Palavras-chave: Cana-de-açúcar; RNA-seq; Respostas transcricionais; Tolerância à broca; Fridelina

SUMMARY

1 INTRODUCTION

Sugarcane (*Saccharum* sp.), a perennially cultivated crop, clonally propagated and an out-crossing species, derived from South-East Asia and New Guinea where it was first domesticated (Lebot, 1999), plays a crucial role in global agriculture, economy, and sustainability by providing raw material for sugar and energy production, which makes it a vital component in the increasing global interest in renewable biomass crops and a low carbon economy. The worldwide sugarcane production area corresponded to more than twenty-six million hectares distributed in approximately 112 countries and an estimated production of 1,9 billion tons (FAOSTAT, 2022).

Brazil accounts for most of the world's sugarcane production with a planting area of nearly 8.35 million hectares with a production outcome of 677,6 million tons, with 27,9 billion liters of ethanol and 46,8 million tons of sugar produced, according to the 2023/2024 national harvest data (CONAB, 2023).

The *Saccharum* genus includes cultivated and wild species from which the modern varieties of sugarcane were bred through hybridization processes. The modern sugarcane cultivars are typically interspecific hybrids derived from backcrosses between the cultivated *S. officinarum* (2n=10x=80) and wild relative *S. spontaneum* (2n=8x=40-128) species. The genomes of these hybrids are composed basically of 80- 85% from chromosomes derived from *S. officinarum*, 10-15% from *S. spontaneum*, and 5 - 10% from recombinant chromosomes between both species (D'Hont et al., 1996). Sugarcane hybrids, consequently, possess polyploid and aneuploid genomes, which hinders the understanding of the molecular and genetic traits of this crop (Hotta et al., 2010).

A threat to sugarcane productivity is the presence of pests that cause economic losses. The sugarcane borer (*Diatraea saccharalis* Fabr. (Lepidoptera: Crambidae)) is the main pest in sugarcane Brazilian fields. Early-instar larvae feed on the leaf parenchyma and sheaths while older larvae bore into the stalks. As a direct impact, this pest causes yield loss and plant lodging because of holes and galleries formed in sugarcane culms and, indirectly, it favors the spread of the opportunistic fungi (*Fusarium moniliforme* (Went) and *Colletotrichum falcatum* (Sheldon)) responsible for the sugarcane red rot disease, which causes sucrose inversion (chemical breakdown into glucose and fructose) and overall decreased sugarcane quality.

The presence of sugarcane borer in the fields can have significant impacts on productivity. Stalk yield losses are reported to be positively correlated to borer infestation intensity (White et al., 2008). Estimates indicate that each 1% internode infested with sugarcane borer results in 1.5%, 0.49% and 0.28% loss productivity of stem, sugar and alcohol, respectively (ARRIGONI, 2002). Also, sugar yield losses of 8.83% and 19.80% per 1% bored internode were observed (Rossato et al., 2013). Besides that, it also has been shown that despite the borer infestation intensity, no statistically significant stalk and sucrose yield reduction was observed, which could be associated with sugarcane varieties that possess mechanisms to prevent yield losses (Rossato et al., 2019).

Up until now, selection and development of tolerant varieties through conventional breeding methods, focusing on the phenotypic tolerance characterization of sugarcane, has assisted in the control of this pest (Demetrio; Zonetti; Munhoz, 2008; Dinardo-Miranda et al., 2012a, 2012b; Tomaz et al., 2017). However, despite the importance of sugarcane borer and the improvements in breeding programs related to borer tolerance, little is known about genes and molecular mechanisms that confer constitutive tolerance for sugarcane against insect damage mainly at early stages after oviposition and feeding on leaves.

A drawback for the molecular and genetic advances in sugarcane was the absence of a reference genome. Its highly polyploid, aneuploid, heterozygous, and interspecific genome have been a major challenge for producing a reference genome sequence (Garsmeur et al., 2018; Thirugnanasambandam; Hoang; Henry, 2018). For years, the fully sequenced and annotated genome of the diploid species *Sorghum bicolor* had been used for comparative analysis (De Setta et al., 2014; Garsmeur et al., 2018; Le Cunff et al., 2008; Wang et al., 2010a). Only recently, the polyploid reference genome of the modern R570 sugarcane cultivar became available. It contains a representation of unique DNA sequences across the approximately 12 chromosome copies of the polyploid genome and encompasses 194,593 protein-coding genes (Healey et al., 2024).

Over the years of coevolution with insects, plants have developed a complex and dynamic defense mechanism against herbivory. The anti-herbivory mechanisms consist of direct and indirect lines of defense which are constitutively present or induced and produced upon damage or feeding (Mitchell et al., 2016; War et al., 2012). A strategy to discover resistance traits is the exploitation of constitutive defenses since

biosynthesis of respective compounds does not rely on insect attack. Smith and Clement (2012) define host plant resistance as the sum of genetically inherited traits that lead to a plant cultivar being less damaged by a pest arthropod compared to a susceptible one lacking these traits. This resistance is mediated by constitutively produced and pest-induced plant proteins synthesized by resistance gene products.

The analysis of sugarcane transcriptome to identify differentially expressed transcripts (DETs) related to insect defense-related genes is a powerful approach to unravel tolerance mechanisms. Among available methods, RNA sequencing (*RNA-seq*) is a tool which allows the analysis of transcriptional profiles of any species of interest that present contrasting characteristics by quantifying the expression of genes activated or inactivated under certain conditions (Li et al., 2012).

Overall reports on characterizing sugarcane transcriptome have focused on unravelling tolerance mechanisms against abiotic and non-pest-related biotic stresses (Agisha et al., 2022; Belesini et al., 2017; Ma et al., 2023; Pereira-Santana et al., 2017; Santa Brigida et al., 2016; Vignesh et al., 2021; Vital et al., 2017). To the best of our knowledge, the first transcriptome study conducted aiming at discovering sugarcane tolerance mechanisms against *D. saccharalis* was conducted by Mello et al. (2020). According to the authors, transcriptome analysis of contrasting sugarcane varieties (RB867515 and SP803280, characterized by low and high survival of young larvae feeding on leaves, respectively) demonstrated that hormone and triterpenoid biosynthetic pathways are positively influenced and significantly regulated upon *D. saccharalis* infestation.

Furthermore, a study on the leaf wax composition of sugarcane genotypes revealed an association between triterpenes and tolerance to the sugarcane borer. Notably, Friedelin, a pentacyclic terpene, was consistently present in most tolerant genotypes, even in the absence of borer infestation, while being absent in susceptible ones (Wartha, 2018; Wartha et al., 2022). This suggests that friedelin's constitutive biosynthesis may vary between genotypes and contribute to sugarcane tolerance.

Here we present a pioneer report on characterizing the changes in leaf constitutional transcriptome profiles of six *Saccharum* genotypes (RB867515, RB047212, RB057249, RB047016, RB047248, and IM76228). Under the presumption that those varieties have contrasting phenotypes to borer infestation, we focused on characterizing the constitutive transcriptional changes of tolerant plants relative to

susceptible ones and on identifying differences related to defense mechanisms possibly involved in the sugarcane tolerance to borer.

2 MATERIAL AND METHODS

2.1 Sugarcane variety selection and cultivation

The genetic material used in this study was composed of a panel of genotypes belonging to the germplasm bank of the Sugarcane Breeding Program at the Federal University of Viçosa (PMGCA/UFV) along with the Inter-University Network for the Development of Sugarcane Industry (RIDESA). Selected genotypes included unreleased clones from the 04 (RB047212, RB047016 and RB047248) and 05 (RB057249) series of the advanced stage of experimental evaluation in the PMGCA/UFV, the RB867515 cultivar and the IM76228, an interspecific hybrid of *Saccharum robustum*.

The experimental clones were extensively field-based rated at four locations and three agricultural seasons, exhibiting different levels of resistance to natural infestations of sugarcane borer. The resistance classification was based on the infestation index, i.e., the percentage of bored internodes divided by the total internodes (Tomaz et al., 2019, 2020). Field results allowed discrimination of tolerant and susceptible genotypes across several tested locations and agricultural seasons, demonstrating persistence and repeatability of tolerance against sugarcane borer. In general terms, the genotypes IM7622, RB867515 and RB047212 are tolerant and the genotypes RB057249, RB047016 and RB047248 are susceptible to *D. saccharalis* attack (see Supplementary Table1). Moreover, RB867515 and IM76228, *S. robustum* hybrid, are characterized by high mortality of early-stage larvae feeding on leaves, and IM76228 presents minor stalk and leaf damage (Pimentel et al., 2017; Tomaz et al., 2017, 2020).

Experimental sugarcane materials were obtained from single-node stem cuttings containing one lateral bud and planted in 15 L pots filled with a mixture soil/manure (2:1 v/v) and were grown in greenhouse benches with controlled environment (average temperature 28.5 °C and natural light conditions) with soil humidity kept at soil water capacity.

The experiment was conducted with a complete randomized design containing six genotypes (RB867515, RB047212, IM76228, RB057249, RB047016 and RB047248). Three biological replicates of each genotype were used (Figure 1). After two months from planting, vegetal material consisting of the leaves +1, following the Kuijper system (DILLEWIJN, 1952), were harvested for RNA sequencing.

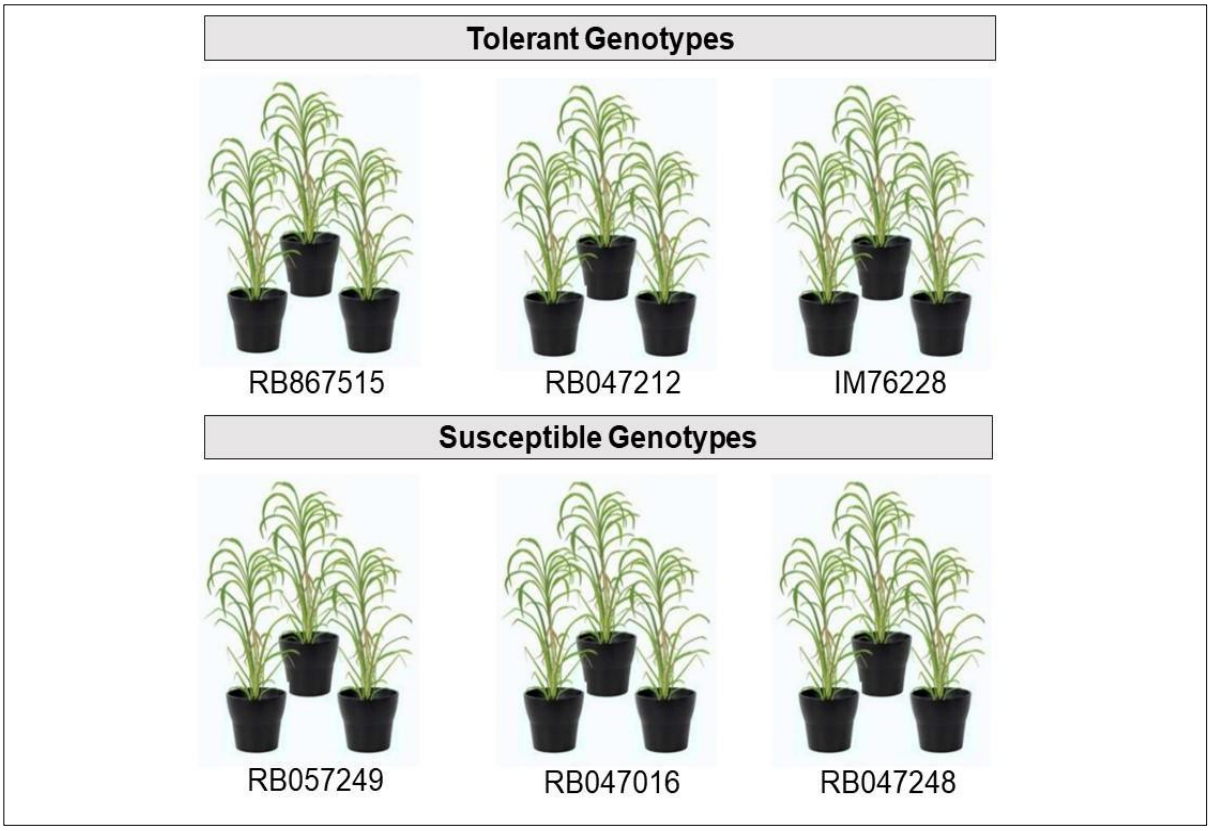


Figure 1 – Schematic representation of the experiment design, depicting the sugarcane genotypes with their respective three biological replicates.

2.2 RNA extraction, RNA-seq library construction and sequencing

One leaf sample of each biological replicate was ground to fine powder under liquid nitrogen using pistils and crucibles. RNA was extracted using the TRIzol™ Reagent (*Thermo Fisher Scientific*) following protocol as described by the manufacturer. The quantification of total extracted RNA was accessed using a Qubit® 2.0 Fluorometer kit (*Life technologies*). Total RNA was sent out to Genoma (USP – University of São Paulo) for constructing mRNA-seq libraries and sequencing using paired-end mode (2x150bp) in the *Illumina* equipment *NovaSeq 6000* (>80M reads/sample). RNA-seq library was constructed using the ZYMO-SEQ Ribofree Total RNA Library kit.

2.3 Reads quality assessment, trimming, ribosomal RNA removal and genomic features

The FASTQ libraries underwent filtering steps to obtain clean reads, which included: (1) raw reads quality assessment using *FASTQC* version 0.12.1 (https://github.com/sandrews/FastQC), (2) trimming for adapters and removal of low-quality reads using *Trimmomatic* version 0.39 (Bolger; Lohse; Usadel, 2014) with the following parameters: *ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:12 TRALING:12 SLIDINGWINDOW:5:20 MINLEN:75*; and (3) removal of ribosomal RNA using *SortMeRNA* version 4.3.6 (Kopylova; Noé; Touzet, 2012) using the following databases: *smr_v4.3_default_db.fasta*, *smr_v4.3_sensitive_db.fasta, smr_v4.3_fast_db.fasta*, *smr_v4.3_sensitive_db_rfam_seeds.fasta*, *SILVA_138.1_SSURef_NR99_tax_silva.fasta*. The quality results were aggregated across all samples into one single report with *MultiQC* (Ewels et al., 2016).

To characterize genomic origin of reads (exonic, intronic, intergenic and exon/intron junction regions), clean reads from each sample were mapped against the annotated R570 v2.1 genome (*https://phytozome-next.jgi.doe.gov/info/SofficinarumxspontaneumR570_v2_1*) using *STAR* software version 2.7.11b (Dobin et al., 2013). Alignment data relative to the features of the mapped reads and their genomic properties was processed and visualized using *Qualimap* tool version 2.3 (García-Alcalde et al., 2012) aggregated with *MultiQC*. Exonic features, according to *Qualimap*, included 5'UTR, protein coding and 3'UTR regions.

2.4 Transcriptome assembly

A transcriptome for each genotype biological replicate was *de novo* assembled using *rnaSpades* program version 3.13.1 (Bushmanova et al., 2019) using respective clean and filtered reads. For this step, 18 transcriptomes were assembled using default parameters (k-mer sizes equal to 49 and 73) and chosen based on the assembler built-in hard filtration procedure, which removes low coverage and relatively small length transcripts (for more details on rnaSPAdes filtration parameters, see Bushmanova (2019), Table S13). Transcripts with less than 300bp were removed using *Seqkit* tool from each transcriptome (Shen et al., 2016). After assembly step, the 18 outputs with transcripts with lengths of at least 300bp were merged using *Trans-AbySS* (Robertson

et al., 2010) version 2.01, which, during the merging process, incorporates a built-in filtration procedure to remove redundant contigs.

To obtain a reference transcriptome for downstream analysis, a clustering procedure was performed on the merged assemblies based on the similarity of sequences: a pre-processing step involving sequence complexity reduction, which creates a unified reference containing canonical sequences representatives of the transcriptome. For that purpose, *MMseqs2* version 13.45111 (https://github.com/soedinglab/MMseqs2) (Steinegger; Söding, 2017) was used to cluster transcript sequences sharing 90% of identity with the following arguments: easy-cluster and COV-MODE 1. Assembly integrity was assessed by N50 statistics using *Seqkit* tool and transcriptome assembly and annotation completeness against the *Viridiplantae* database was evaluated using BUSCO (Benchmarking Universal Single-Copy Orthologs) version 5.2.2 (Simão et al., 2015).

2.5 Differential expression analysis

Differential expression analysis was performed using all samples (18) designed in a model with two groups contrasting in tolerance to *D. saccharalis*. Total reads of each biological replicate were mapped to the reference transcriptome using *Kallisto* version 0.50.1 (Bray et al., 2016) and *Salmon* version 1.10.2 (Patro et al., 2017) programs and the abundance of mapped reads were analyzed using the *DESeq2* package version 1.6.3 (Love; Huber; Anders, 2014) implemented in R version 3.5.1 (R Development CoreTeam (2024), 2024).

Abundance data was imported to *R* software using the *tximport* package (Soneson; Love; Robinson, 2015) and data were filtered for transcripts with expression levels of at least one count per million (CPM) in at least three biological replicates of the same genotype. Analysis was performed comparing transcript abundance from the tolerant group relative to the susceptible one. Heatmap and PCA (Principal Component Analysis) clustering analysis were done using rlog transformation of normalized counts. An adjusted p-value ≤ 0.01 and a $\log_2$FoldChange ≥ 2 was used to select significant differentially expressed transcripts commonly screened by *Kallisto* and *Salmon* mapping results. In addition, Pearson´s Correlation Coefficient analysis was performed to compare the rlog-transformed data from read abundances obtained by

Kallisto and *Salmon* software. This analysis included pairwise comparisons relative to all biological replicates of each genotype.

2.6 Characterization and functional annotation

Reference transcriptome sequences were aligned with DIAMOND (Buchfink; Xie; Huson, 2014), version 0.9.14.115, on NCBI RefSeq protein database in conjunction with protein sequences from GenBank database to retrieve corresponding best hit annotations for each transcript. The database used encompassed 8,917,582 protein sequences from 12,680 plant species.

Conserved protein domains were assigned to the proteins encoded by DETs using the online platform TRAPID: Rapid Analysis of Transcriptome Data (Bucchini et al., 2021), which uses the PLAZA 2.5 database (*https://bioinformatics.psb.ugent.be/plaza/)* to assign functional annotations based on sequence similarity and the *InterPro* database to functionally annotate sequences to protein families, domains, and functional sites (Hunter et al., 2009). The search parameters included the monocot database and an e-value of 1.0e-5.

Gene ontology (GO) terms were assigned to DETs using the online platform PlantRegMap (Plant Transcriptional Regulatory Map) (Jin et al., 2017) which adopt the topGo package and Fisher's exact tests to find significantly over-represented GO terms (p-value≤0.01). GO enrichment analysis was done using the *Sorghum bicolor* database as reference. We also used KOBAS (KEGG Orthology Based Annotation System) version 3.0 to identify enriched metabolic pathways among DETs (Xie et al., 2011). Enriched KEGG pathways were determined by using Fisher's exact tests followed by Benjamini-Hochberg FDR correction method (FDR≤0.01). The workflow of the bioinformatic pipelines conducted in this study is shown in Figure 2.

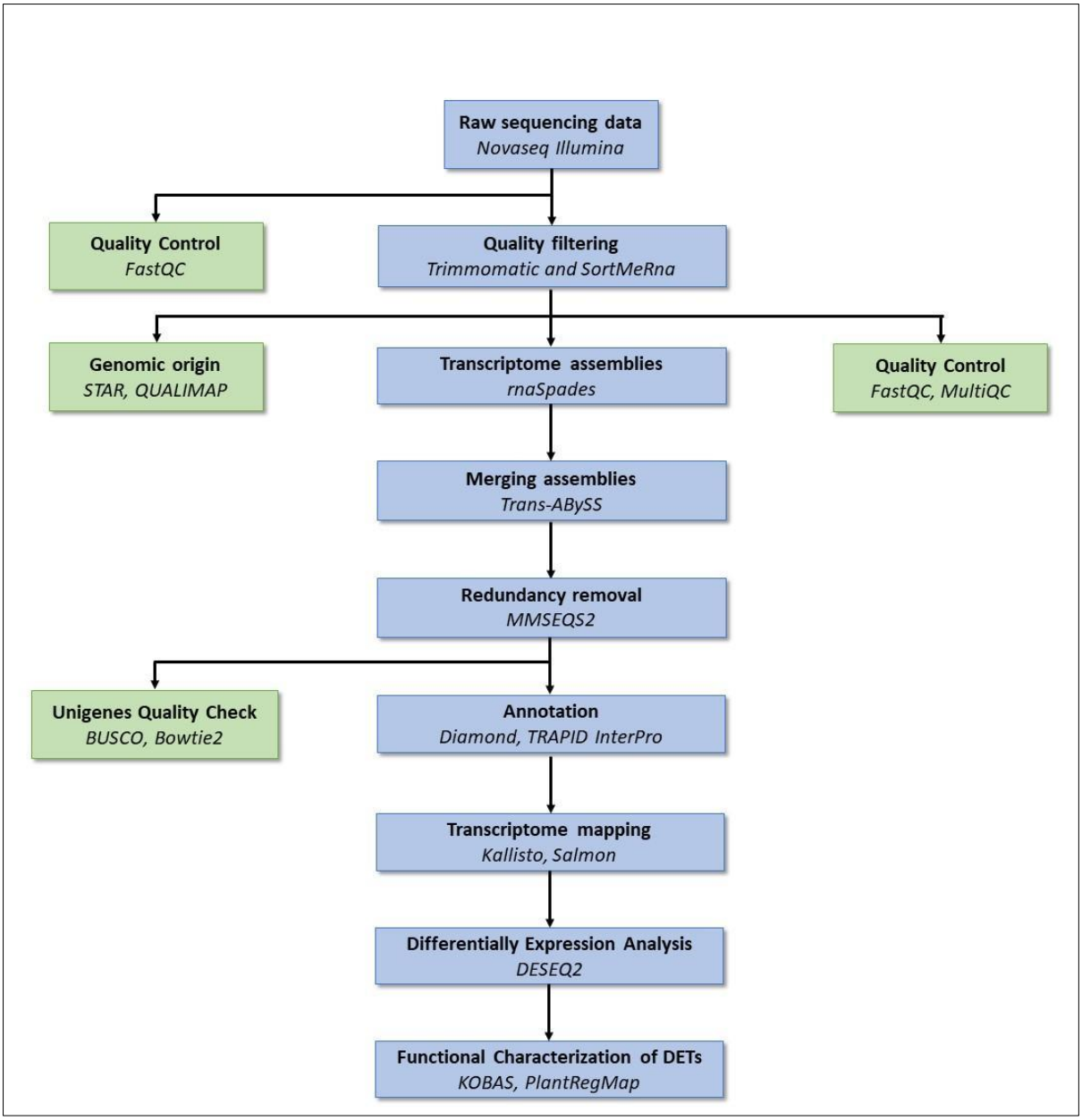


Figure 2 – Pipeline of procedure analysis. Raw reads were obtained from RNA-seq libraries of tolerant and susceptible sugarcane plants to *D. saccharalis*. First, a quality filter was applied to all raw reads. Next, clean reads were mapped to transcriptome reference. Last, differential expression analysis was conducted and DETs were functionally annotated using PlantRegMap and KOBAS.

3 RESULTS AND DISCUSSION

3.1 RNA sequencing, *de novo* assembly and characterization of sequencing libraries and transcriptome

From the sequencing procedure, thirty-six paired-end libraries were generated (two per biological replicate), totaling 4.01 billion reads, with individual library sizes ranging from 47 to 71 million reads. Raw reads underwent quality trimming and filtering to remove low-quality regions (Q<20), ambiguous nucleotides, adapters, and ribosomal RNA. This process removed approximately 15.06% of low-quality reads and 42.2% of ribosomal RNA reads from all RNA libraries (Table 1).

High-quality reads from individual biological replicates were *de novo* assembled to generate 18 transcriptomes with transcripts with lengths of at least 300bp. Spades generated assemblies with transcript number ranging from 112,841 to 183,841 (based on hard-filtering mode), totaling 2,543,819 transcripts (Table 2). Assemblies were merged using Trans-AbySS into one single fasta file containing 2,296,214 transcripts (9.73% reduction after contig redundancy removal). The merged assembly was diminished in approximately 70.09% after sequence redundancy and complexity reduction using MMseqs2 based on 90% search similarity, rendering a reference transcriptome containing 686,859 of representative sequences with a transcript N50 length of 1,539bp. Trimmed and filtered libraries had approximately 96.4% of mapped reads to the reference transcriptome. The number of raw, filtered, and mapped reads are listed in Table 1.

Table 1 – Summary of sequencing, trimming, filtering, and mapping data in each genotype, representing the total of three paired-end libraries per biological replicate.

Samples	Raw reads[a]	Trimming and filtering				% bowtie2 mapping[d]
		Trimmomatic[b]	%	SortMeRna[c]	%	
RB867515	615,052,691	502,183,080	81.65	305,251,426	60.78	96.93
RB047212	671,160,408	566,115,194	84.35	344,579,236	60.87	95.68
IM76228	642,131,026	546,242,338	85.07	293,299,126	53.69	96.02
RB047016	764,849,040	682,087,342	89.18	284,279,814	41.68	94.90
RB047248	621,705,134	537,099,676	86.39	371,047,206	69.08	97.62
RB057249	695,411,818	572,500,252	82.33	370,310,340	64.68	97.27
Total	4,010,310,117	3,406,227,882	84.94	1,968,767,148	57.80	96.4

Notes: [a] total number of reads after *Illumina* sequencing; [b] total number of clean reads obtained after quality filter using Trimmomatic; [c] total number of clean reads after ribosomal RNA removal; [d] number of clean reads mapped to reference transcriptome.

Table 2 – Number of transcripts per assembly for each genotype´s biological replicate.

Samples	Biological replicate			Total
	1	2	3	
RB867515	132,847	135,907	133,768	
RB047212	148,064	139,482	150,559	
IM76228	119,491	153,046	151,266	
RB047016	174,006	183,841	135,171	
RB047248	121,681	142,640	146,025	
RB057249	144,749	118,435	112,841	
Total				2,543,819

Comparing the 686,859 transcripts against the *Viridiplantae* database, we found 98.35% complete sequences (52 single-copy and 366 duplicated BUSCOS) and 1.65% fragmented sequences (7 BUSCOS). The functional annotation analysis with DIAMOND (blastx option) assigned protein names to a total of 301,729 transcripts (43.93%) and 318,246 transcripts (46.33%) had at least one protein domain/motif InterPro annotation after submission to the TRAPID platform. In total, 279,569 (40.7%) transcripts shared a NCBI RefSeq/Genbank protein hit and an InterPro protein domain/motif annotation. Exactly 346,453 transcripts (50.44%) did not match any protein sequence/domain. Despite the representative number of protein sequences and species in the database used, there may still be novel or rare isoforms, species-specific transcripts, or splicing variants that are not present in existing databases.

The non-annotated transcripts might also be intronic transcripts, transcribed from non-coding regions within or outside genes. Intronic reads have been associated with transcriptional activity, related to nascent transcription in combination with co-transcriptional splicing events (Gaidatzis et al., 2015). These intronic transcribed sequences appear to harbor a yet unexplored reservoir of novel, functional RNAs (Laurent et al., 2012). Concerning RNA-seq approaches, a small number of long non-coding RNAs (lncRNAs), small RNAs and intronic reads have been demonstrated to constitute a large fraction of the reads in the rRNA depletion RNA sequencing data (Zhao et al., 2018). Also, artificial transcripts (artifacts) such as chimeric and misassembled transcripts, erroneous insertions because of low-complexity and repetitive regions, and multiple similar genes being incorrectly reconstructed as fused transcripts might arise in the assembly process (Venturini et al., 2018).

Mapping RNA-seq reads to the R570 genome yielded an average alignment rate of approximately 90% (Table 3). According to Qualimap results, a substantial number of reads also aligned to intronic and intergenic regions, roughly similar to

exonic regions regarding percentages (Figures 3 and 4). It is worth mentioning that high intronic and intergenic read mapping might be partially explained by isoform variants of genes, small or long non-coding and translated regions still not annotated in the R570 reference genome.

Additionally, RNA-seq experiments can generate reads derived not only from mature RNA transcripts but also from pre-mRNA. Ribosomal RNA depletion methods remove cytoplasmic and mitochondrial rRNA, while preserving poly(A) + mRNA, non-coding RNAs, and protein-coding mRNAs that are not polyadenylated, which include fractions of incompletely transcribed mRNAs (partially spliced or unspliced but not polyadenylated) as well as undegraded intronic transcripts (Kapranov et al., 2010; Morlan; Qu; Sinicropi, 2012).

Table 3 – STAR results following the alignment of tolerant and susceptible RNA-seq libraries to the R570 genome.

	RB867515			RB047212			IM76228			Average
Aligned pairs (M)	47,9	50,6	40,9	47,7	60,5	52,3	36,9	40,3	36,2	
Alignment (%)	92,52	93,01	90,14	93,86	93,67	93,90	85,63	77,34	71,80	87.98
	RB047016			**RB047248**			**RB057249**			
Aligned pairs (M)	48,2	33,4	38,3	50,5	65,2	58,4	62,4	64,03	48,1	
Alignment (%)	83,73	80,85	90,98	94,29	94,36	94,35	95,04	95,05	93,99	93.99
Average										**89.69**

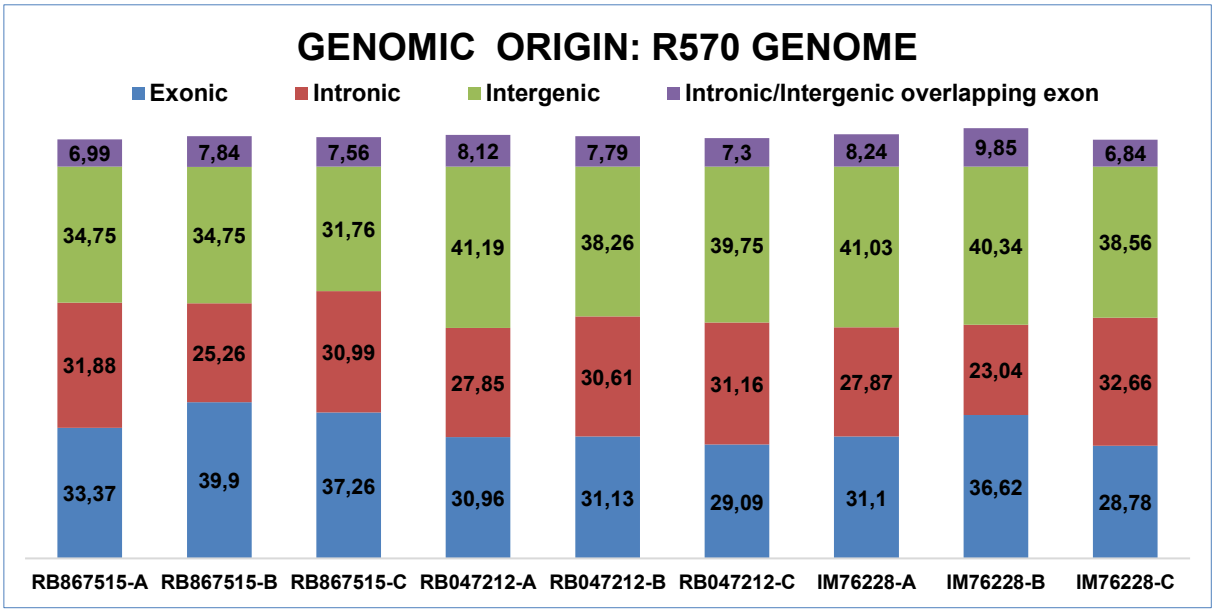


Figure 3 – Genomic origin of RNA-seq reads from tolerant genotypes, showing their percentage distribution based on alignment to the R570 polyploid genome. Letters A, B and C represent biological replicates.

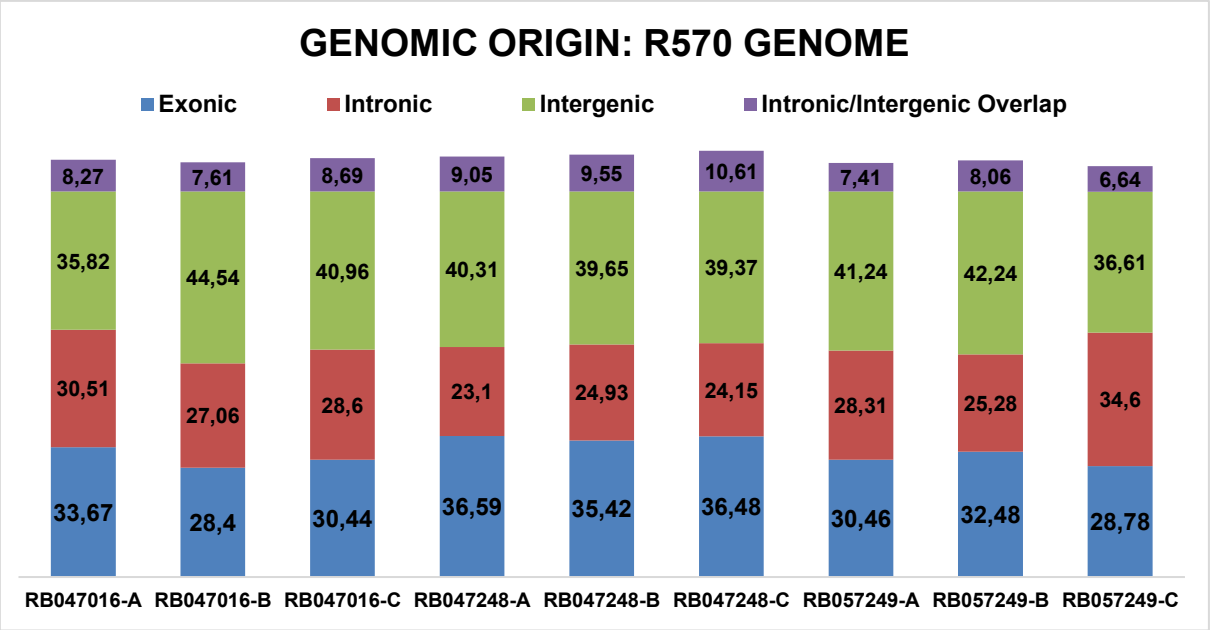


Figure 4 – Genomic origin of RNA-seq reads from susceptible genotypes, showing their percentage distribution based on alignment to the R570 polyploid genome. Letters A, B and C represent biological replicates.

A reciprocal blastn analysis, using coding sequences (CDS) from the polyploid R570 genomes (299,731 sequences) and transcripts from the reference transcriptome generated in this study, revealed that 50.65% (347,957) of transcripts had at least one blast hit in R570 CDS database. Conversely, 97.37% (291,851) of R570 coding sequences matched at least one transcript in the reference transcriptome. This suggests the presence of protein-coding genes, splicing variants, isoforms, and other forms of transcripts not yet annotated in the sugarcane genome. Reciprocal blastn analysis results are shown in Table 4.

Table 4 – Results of the reciprocal blastn analysis performed between the reference transcriptome generated in this study and the R570 coding sequences (CDS).

Database	Total sequences	Blastn hits	%
Reference transcriptome	686,859	347,957	50.65
R570	299,731	291,851	97.37

3.2 Clustering analysis and differential gene expression

Following the removal of transcripts expressed at low levels (CPM≤1), the resulting transcript expression matrices comprised 101,606 transcripts quantified by Kallisto and 107,995 transcripts quantified by Salmon. These matrices were subsequently subjected to exploratory data analysis of the sequenced libraries. Initial

evaluation of the RNA-seq data was conducted using Principal Component Analysis (PCA), which provides insights into the association and similarity among samples concerning gene expression. The data underwent regularized log (rlog) transformation to stabilize variance across expression levels. The plots illustrate the projection of samples onto the two-dimensional space defined by the first and second principal components of the covariance matrices.

Clustering results, including PCA and heatmap, from the Kallisto-filtered transcript expression matrix are shown in Figure 5. Comparable findings were observed with Salmon data (Supplementary Figure 1). The first two principal components, PC1 and PC2, accounted for 45% and 19% of the total variance, respectively, cumulatively explaining 64% of the overall variability in the dataset. A clear separation of genotypes was observed along PC1, which primarily distinguished between tolerant (represented by red dots) and susceptible (blue dots) groups, suggesting that borer tolerance is associated with significant alterations in gene expression profiles.

In general, tolerant genotypes (IM76228, RB047212, and RB867515) clustered within the right and upper left quadrants of the PCA space. Notably, the IM76228 group formed a distinct cluster at the far left along PC1, indicating unique transcriptional signature relative to other tolerant genotypes. Susceptible genotypes (RB047016, RB047248, and RB057249) were distributed across separate regions, with RB047016 clustering in the lower left and RB047248 in the upper right portions of the plot, suggesting transcriptomic heterogeneity within this group as well. Importantly, clustering patterns did not strictly correspond to tolerance classification, as the PCA was conducted using global transcriptomic profiles from individual libraries, which reflect overall variation in gene expression independent of predefined phenotypic categories.

Biological replicates of each genotype consistently clustered together in the PCA space, demonstrating high reproducibility of transcriptomic profiles within genotypes. The observed grouping pattern likely reflects underlying genetic divergence among the genotypes, particularly evident in the distinct separation of IM76228, a representative of *Saccharum robustum*, which is a wild relative of cultivated sugarcane.

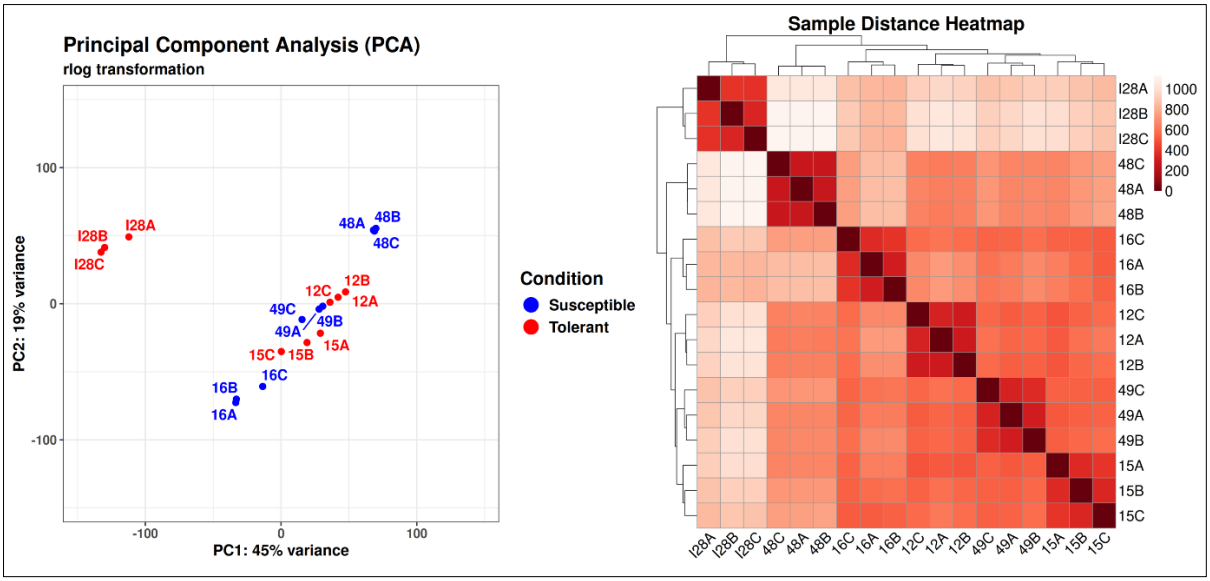


Figure 5 - rlog-transformation PCA plot and heatmap clustering generated using kallisto-filtered transcript expression matrix. Numbers 12, 15, I28, 16, 48, and 49 represent the genotypes RB867515, RB047212, IM76228, RB057249, RB047016, and RB047248. Letters A, B and C denote biological replicates.

Pairwise comparison of Pearson's correlation Coefficient was applied to the rlog-transformed read counts generated from Kallisto and Salmon filtered data (CPM≥1). The correlation analysis revealed R values between 0.87 and 0.95 among biological replicates for both input datasets (Figure 6). Different library sizes (total number of mapped reads per sample) were observed among biological samples after quantifying abundances of transcripts from RNA-seq data with Kallisto and Salmon (Table 5). While both mapping software exhibited very similar alignment pattern, Salmon consistently aligned more reads.

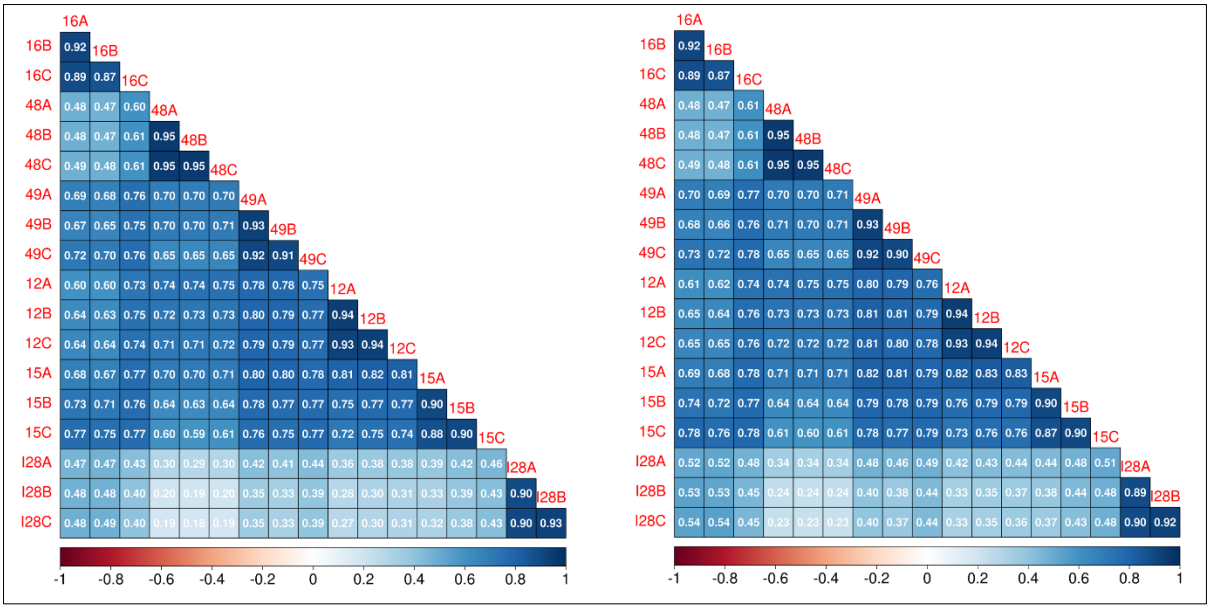


Figure 6 – Pairwise Pearson's correlation Coefficient results for rlog-transformed read counts from Kallisto (left) and Salmon (right) data. R values varied from 0.87 to 0.95 for both inputs.

.

Table 5 – Reads mapped per sample (library sizes) by Kallisto and Salmon.

Sample	Kallisto Mapped reads	Kallisto Mapping %	Salmon Mapped reads	Salmon Mapping %
12A	43,129,005	84.2	48,057,820	93.7
12B	56,449,757	86.9	62,293,598	95.8
12C	48,151,409	85.9	52,800,110	94.1
15A	44,820,593	85.9	49,475,175	94.8
15B	48,407,135	88.3	52,422,073	95.6
15C	39,692,831	87.0	44,158,979	96.7
I28A	37,781,014	87.1	41,584,902	95.8
I28B	45,426,378	86.6	49,705,277	94.7
I28C	43,594,174	85.8	47,923,964	94.2
16A	49,914,759	86.1	55,045,547	94.8
16B	34,628,198	83.0	38,328,650	91.8
16C	35,716,273	84.2	39,975,047	94.2
48A	48,331,377	89.7	52,314,893	97.0
48B	60,967,780	87.7	66,850,057	96.2
48C	54,330,333	87.4	60,160,753	96.8
49A	58,198,338	88.1	63,535,299	96.1
49B	60,719,016	89.8	65,713,432	97.1
49C	45,122,703	87.7	49,199,209	95.6

Log_2 fold changes (log_2FCs) were calculated between tolerant and susceptible sample libraries to determine differential expression of transcripts (DETs). DETs were screened using a false discovery rate (FDR)≤0.01 and a log_2FC≥2. Utilizing transcript abundance matrices from Kallisto and Salmon, 7,149 (4,426 up-regulated and 2,723 down-regulated) and 5,702 (3,394 up-regulated and 2,308 down-regulated) were identified, respectively. A total of 4,366 DETs (2,662 up-regulated and 1,704 down-regulated) were commonly screened and were used for downstream analysis. Venn

Diagrams were built to demonstrate the overlap and uniqueness of these DETs (Figure 7).

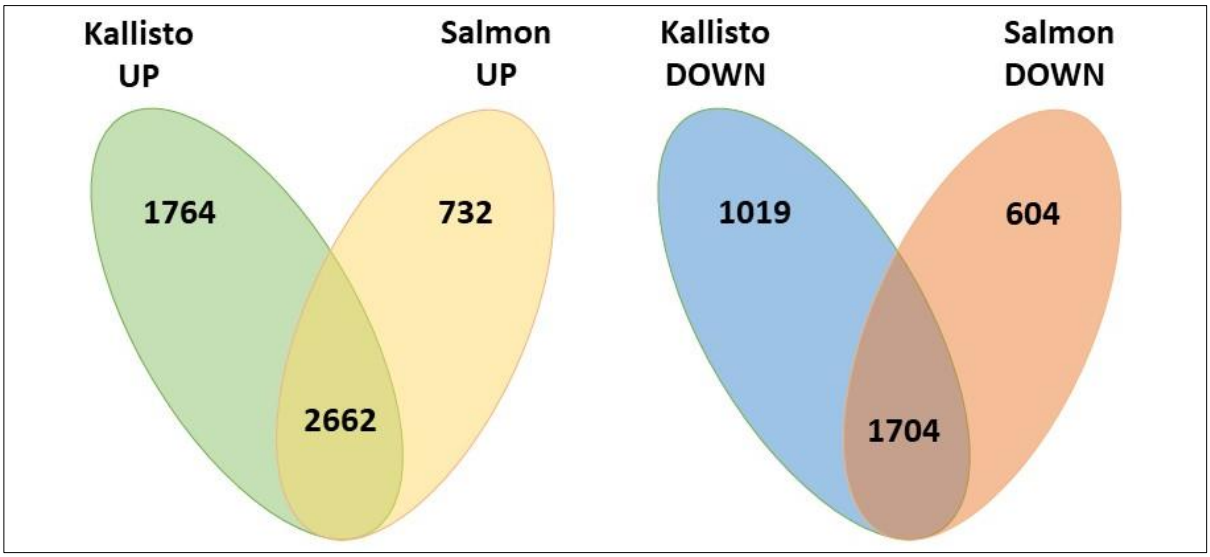


Figure 7 – Number of differentially expressed transcripts (DETs) identified in contrasting sugarcane genotypes against *D. saccharalis*.

The set of differentially expressed transcripts (DETs) was also subjected to Principal Component Analysis (PCA) to explore patterns of gene expression variation among the samples. This analysis revealed that the first two principal components (PC1 and PC2) accounted for 37% and 22% of the total variance, respectively, cumulatively explaining 59% of the overall variability in the dataset (Figure 8). The tolerant genotypes (represented by red dots) clustered closely in the central to upper-right quadrant of the PCA space, indicating a high degree of similarity in their transcriptional profiles. In contrast, susceptible genotypes (blue dots) were more dispersed, largely occupying the bottom-left and top-left quadrants, while still forming a cluster distinct from the tolerant group. The dispersed pattern observed among the susceptible genotypes, despite their selection and classification based on field performance, is expected given the inherent genetic divergence among sugarcane genotypes and the polygenic, multifactorial nature of pest resistance. The pronounced separation along PC1 suggests that this component is primarily responsible for distinguishing between the tolerant and susceptible genotypes. This separation indicates the presence of distinct transcriptional signatures potentially associated with constitutional differential responses that confer pest tolerance.

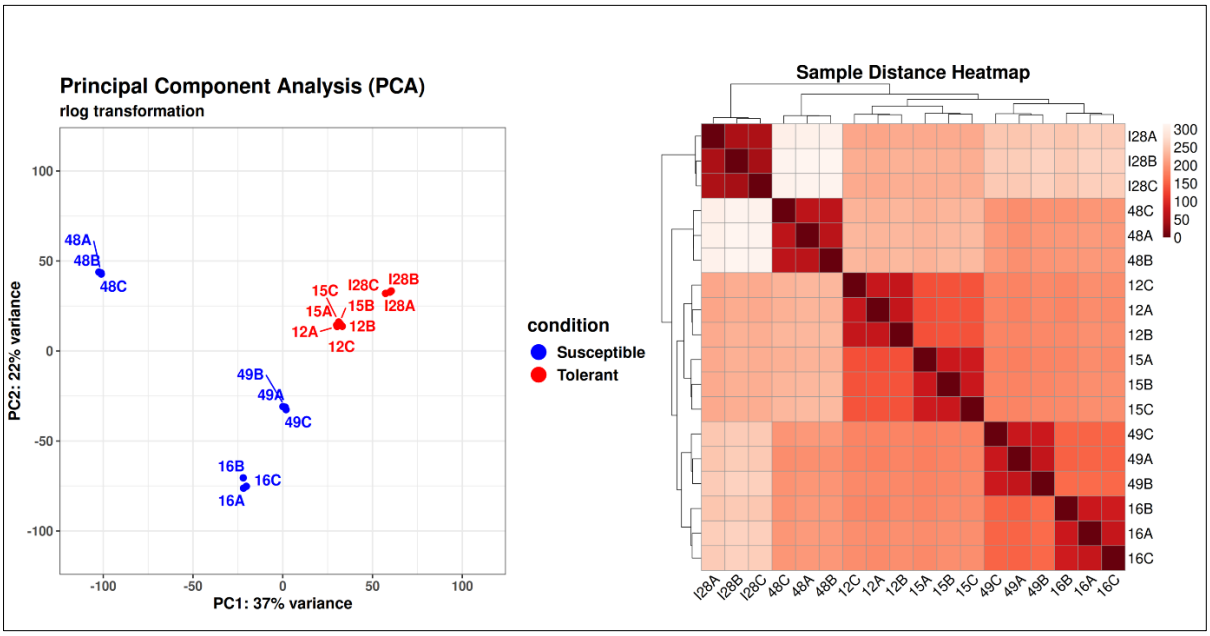


Figura 8 – rlog-transformation PCA plot and heatmap clustering generated using kallisto-filtered transcript expression matrix of DETs. Results were very similar for Salmon alignment data. Numbers 12, 15, I28, 16, 48, and 49 represent the genotypes RB867515, RB047212, IM76228, RB057249, RB047016 and RB047248. Letters A, B and C denote biological replicates.

3.3 Functional annotation of differentially expressed transcripts

Commonly screened DETs were annotated by DIAMOND (blastx option) using the NCBI RefseqPlant protein database in conjunction with protein sequences from the GenBank database. Protein domains/motifs (PDs) were assigned using the online tool TRAPID, which utilizes InterPro, a database encompassing functionally annotated protein families, domains, and functional sites.

As a result of the DIAMOND analysis, 3,571 DETs were annotated (2,174 up-regulated and 1,397 down-regulated, representing 81.67% and 81.98% of total commonly screened DETs, respectively). TRAPID assigned 3,435 DETs to at least one InterPro entry (2,083 up-regulated and 1,352 down-regulated, representing 81.98% and 79.43% of total commonly screened DETs, respectively) (Figure 8). With the adopted parameters, exactly 3,284 DETs were annotated by both databases, while 644 DETs remained unannotated (Figure 9), potentially representing non-coding RNA sequences, transcripts not yet annotated in monocot species, or even assembly artifacts.

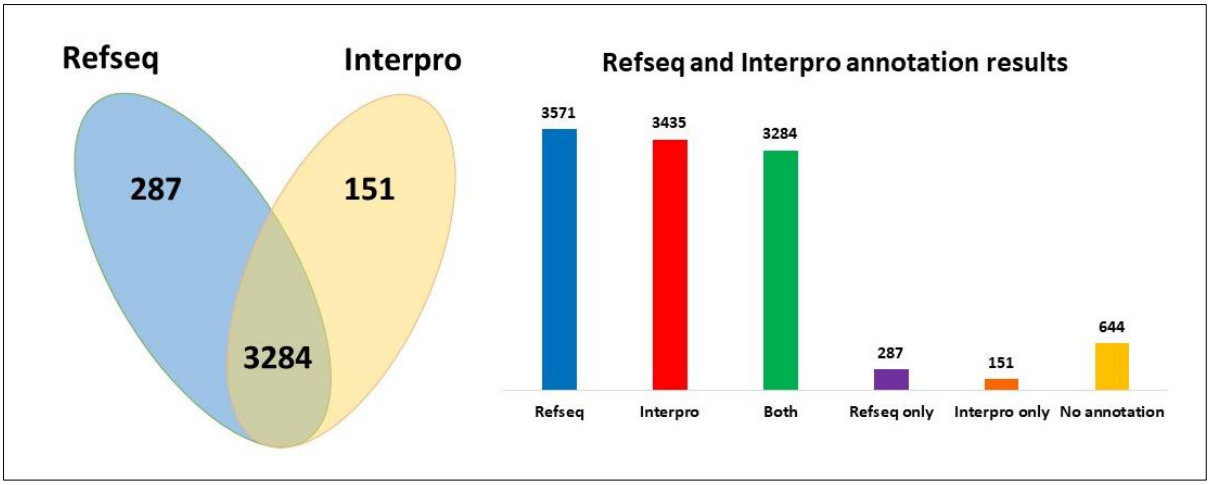


Figure 9 – Refseq and Interpro annotation results. Venn diagram (left) illustrates that 3,284 DETs (out of 4,366) shared both Refseq and Interpro annotations, while the bar plot (right) shows 644 transcripts lacking matches to any protein sequence or domain in the searched databases.

Regarding the similarity search conducted by TRAPID, most of the assigned transcripts had significant matches with sequences from *Saccharum spontaneum,* an ancestral polyploid wild species of modern sugarcane hybrids. Other species exhibiting sequence identity similarity in the search included: *Miscanthum sinensis*, *Sorghum bicolor*, *Oryza sativa* and *Setaria italica*. It is worth noting that TRAPID utilizes the Monocots PLAZA 5.0 database, which does not include R570 reference sequences. A summary of the TRAPID statistical analysis is presented in Table 6.

Table 6 – Summary of TRAPID results.

Description	UP-regulated DETs	DOWN-regulated DETs
Transcript Information		
Total number	2,662	1,704
Average length (bp)	2,184.1	2,359.4
Transcripts with putative frameshift	686	429
Similarity search information (DIAMOND)		
Saccharum spontaneum	1,216	768
Miscanthus sinensis	581	418
Sorghum bicolor	257	166
Oryza sativa ssp. indica	68	35
Setaria italica	55	40
InterPro		
InterPro domains	2,286	1,382
Transcripts with PDs	2,083	1,352

Among the assigned protein domains in up-regulated DETs are transcripts with PDs/motifs involved in physiological and biosynthetic processes of hormones (IPR020834 e IPR001128), defensive compounds and triterpenes (IPR001128, IPR008930 and IPR018333) and transcription factors (IPR001005, IPR011598, IPR003657), as identified by the Interpro database *(https://www.ebi.ac.uk/interpro/entry/InterPro/#table)* (Table 7). Most of these have already been reported to be involved in defense mechanisms against biotic stresses and plant-insect interaction (La Camera et al., 2004; Zebelo; Maffei, 2015). Analysis of these PDs in the down-regulated DETs revealed a lower number of annotated transcripts, except for SANT/Myb and bHLH domains. Notably, down-regulated DETs were not assigned to WRKY and Lipoxygenase PDs (Table 7).

Table 7 – Interpro entries assigned to up- and down-regulated DETs related to hormones, triterpenoids, and transcription factors.

Rank *	Interpro	Description	Transcript Number **
UP-regulated DETs			
14	IPR008930	Terpenoid cyclases	55 (1.26%)
22	IPR001128	Cytochrome P450	37 (0.85%)
28	IPR018333	Squalene cyclase	35 (0.80%)
96	IPR001005	SANT/Myb domain	13 (0.30%)
104	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain	12 (0.27%)
111	IPR003657	WRKY domain	11 (0.25%)
248	IPR020834	Lipoxygenase, conserved site	6 (0.14%)
DOWN-regulated DETs			
17	IPR001005	SANT/Myb domain	27 (0.62%)
46	IPR008930	Terpenoid cyclases	14 (0.32%)
65	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain	12 (0.27%)
134	IPR001128	Cytochrome P450	8 (0.18%)
208	IPR018333	Squalene cyclase	5 (0.11%)
–	IPR003657	WRKY domain	–
–	IPR020834	Lipoxygenase, conserved site	–

* Position among the most frequently assigned Interpro entries.
** Percentage considering the total number of commonly screened DETs (4,366).

3.4 KEGG pathway and GO enrichment of differentially expressed transcripts

The mapping of metabolic pathways available through the Kyoto Encyclopedia of Genes and Genomes (KEGG) provides systematic classifications of gene functions. KEGG enrichment analysis was conducted using the online tool KOBAS 3.0. The enrichment was performed using the group of DETs commonly screened by Kallisto and Salmon software, against the available *Sorghum bicolor* KEGG pathways.

The enrichment analysis of DETs (q-value ≤ 0.01) revealed 16 up-regulated and 16 down-regulated KEGG pathways. In general, the analysis of the contrasting groups (tolerant vs susceptible) showed an up-regulation of pathways associated with biosynthesis of fatty acids, hormones and terpenoids (Figure 10). Conversely, pathways associated with pyruvate, photosynthesis, nitrogen, and carbon metabolism exhibited down-regulation (Figure 11).

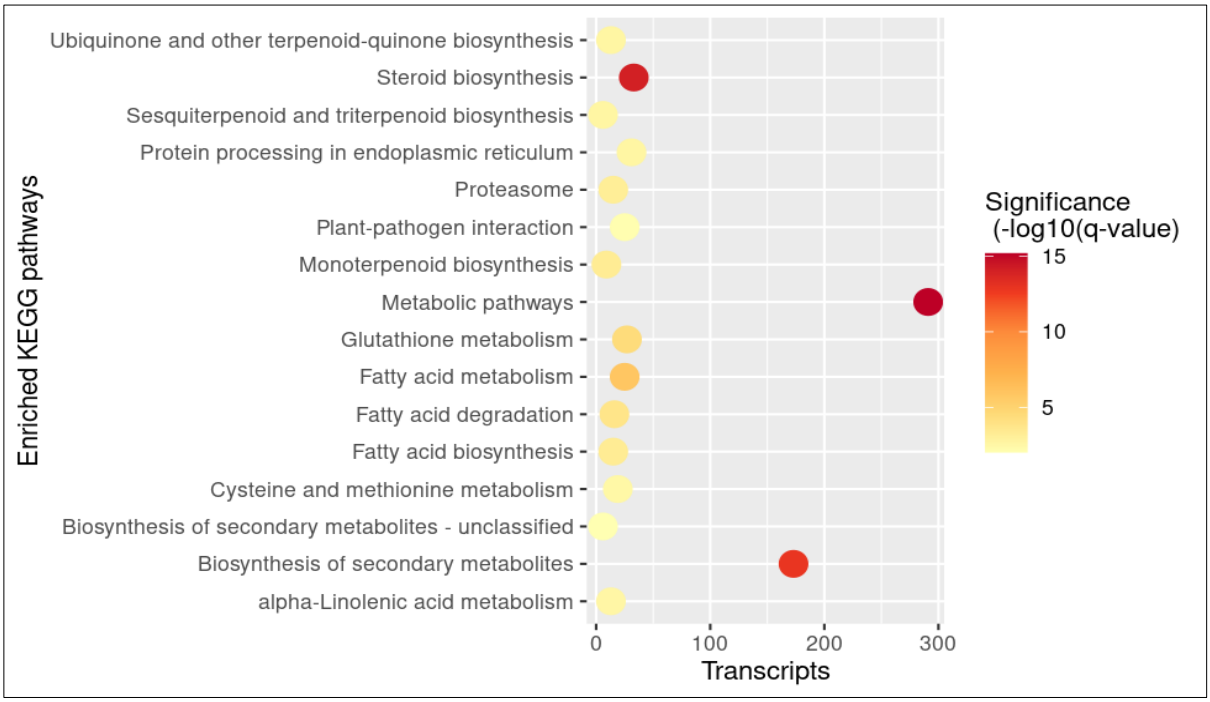


Figure 10 - Dot plot representation of up-regulated KEGG metabolic pathways. The color scale represents the FDR corrected *P* values (q-value≤0.01) calculated in the enrichment analysis.

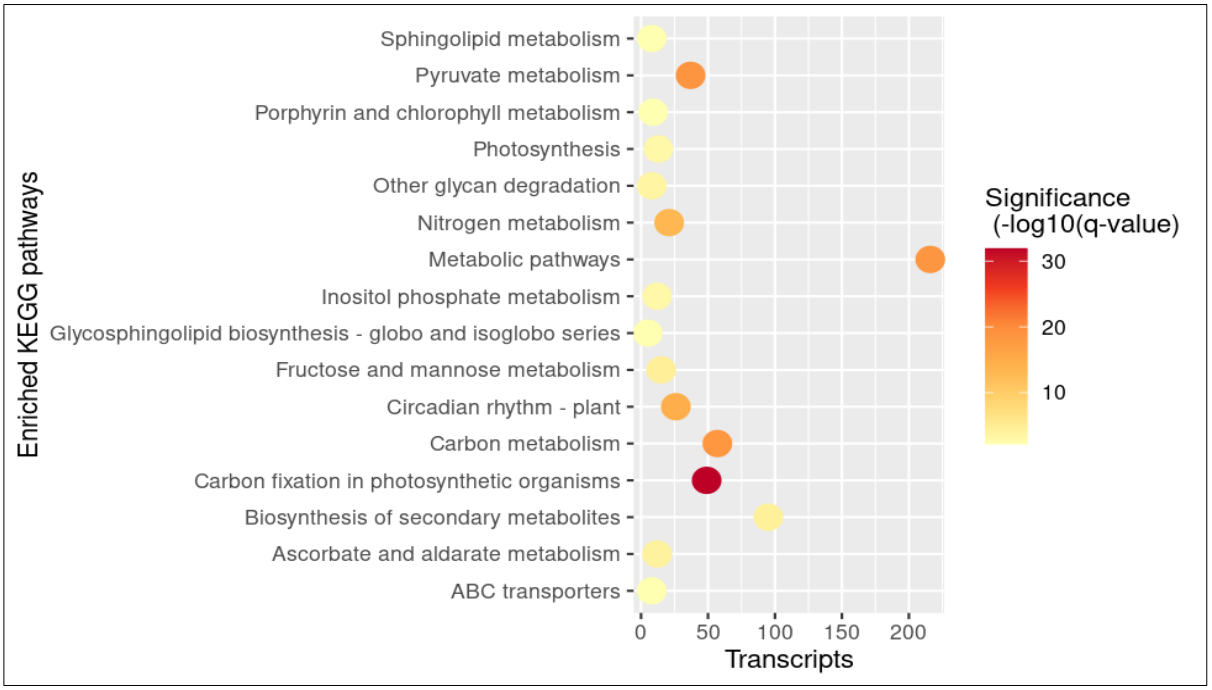


Figure 11 - Dot plot representation of down-regulated KEGG metabolic pathways. The color scale represents the FDR corrected *P* values (q-value≤0.01) calculated in the enrichment analysis.

Other transcriptional studies have demonstrated similar regulation patterns in enriched metabolic pathways in sugarcane submitted to biotic and abiotic stresses, such as bacterial infection (Santa Brigida et al., 2016), drought condition (Belesini et al., 2017; Pereira-Santana et al., 2017; Vital et al., 2017), salinity condition (Vignesh et al., 2021) and fungus infection (Agisha et al., 2022). Specifically, upon *D. saccharalis* herbivory on tolerant and susceptible sugarcane cultivars, and compared to their respective unstressed controls, an overall up-regulation of hormone and triterpenoid biosynthetic processes and a down-regulation of photosynthesis and carbon metabolism was observed (De Mello et al., 2020). Thus, the regulation of these pathways apparently represents a common stress response in sugarcane, and a network of specific gene regulations might play a significant difference in conferring tolerance.

It is noteworthy that, according to the clustering analysis (PCA) of DETs conducted in this study, genotypes classified as susceptible exhibited a more dispersed distribution pattern. This dispersion may reflect inherent differences in their constitutional transcriptional responses, particularly when examining individual gene expression profiles and associated metabolic pathways. Further investigation of genes related to photosynthesis and carbon metabolism revealed a higher number of mapped

reads for a single susceptible genotype (RB047248) (see Supplementary Figures 2 and 3). Consequently, the results obtained from the KEGG pathway analysis regarding photosynthesis and carbon metabolism may not represent a consistent expression pattern across susceptible genotypes.

Based on the findings presented in this section and previously reported pathways associated with pest tolerance, it is important to highlight the involvement of key transcripts within the "Alpha-linolenic acid metabolism" and "Fatty acid metabolism" KEGG pathways. These transcripts contain protein domains associated with enzymes such as 3-hydroxyacyl-CoA dehydrogenase, Acyl-CoA oxidase, Acyl-CoA dehydrogenase/oxidase, and members of the NAD(P)-binding domain superfamily. Functional annotation using DIAMOND (blastx) revealed that many of these transcripts exhibit significant similarity to genes encoding enzymes such as peroxisomal fatty acid β-oxidation multifunctional proteins, 12-oxophytodienoate reductase 7 (OPR7), acyl-coenzyme A oxidase 4 (peroxisomal-like), linoleate 13S-lipoxygenase 11, lipoxygenase 2.3 (chloroplastic-like), and 3-ketoacyl-CoA thiolase 2 (peroxisomal), which are directly or indirectly involved in the biosynthesis of jasmonic acid (JA), a key hormone in plant defense signaling.

In addition, transcripts mapped to the "Steroid biosynthesis" KEGG pathway were found to contain protein domains characteristic of squalene cyclase, terpene synthase, terpenoid cyclase, and members of the NAD(P)-binding domain superfamily. Annotation analysis revealed that these transcripts correspond to genes encoding enzymes such as glutinol/friedelin synthase and cycloartenol synthase, both involved in triterpene biosynthesis, as well as 3β-hydroxysteroid dehydrogenase/decarboxylase, which participates in the synthesis of sterol intermediates critical to brassinosteroid biosynthesis.

The overall pathway enrichment analysis suggests that biosynthesis of Jasmonic acid precursors and defensive-related proteins/metabolites are constitutively up-regulated in tolerant genotypes.

Gene functions were classified using Gene Ontology (GO). DETs were functionally attributed to three GO categories: biological process, molecular function and cellular component. GO terms were assigned (q-value ≤ 0.01) using the online tool PlantRegMap. The enrichment was performed using the group of DETs commonly identified by Kallisto and Salmon software against the available *Sorghum bicolor* GO terms.

The GO enrichment analysis revealed 69 up-regulated and 54 down-regulated GO terms, respectively (Figures 12 and 13). In general, within the biological and molecular function categories, GO terms reported to be involved in plant defense against herbivory were identified. Among these included: “jasmonic-acid biosynthetic process (GO:0009695)”, “regulation of defense response (GO:0031347)” and “terpene synthase activity (GO:0010333)”, all of which were found to be up-regulated; and “photosynthesis (GO:0015979)”, which was found to be down-regulated. However, as previously discussed, the observed down-regulation of photosynthesis-related genes in the tolerant group was influenced by a single susceptible genotype, which introduced bias into the respective metabolic pathway.

Thus, according to KEGG and GO enrichment results, we hypothesize that tolerant genotypes (RB867515, RB047212 and IM76228) presented a broader transcriptional capacity of constitutively activating terpene and hormone-related pathways, which play a central role in plant defense against herbivory.

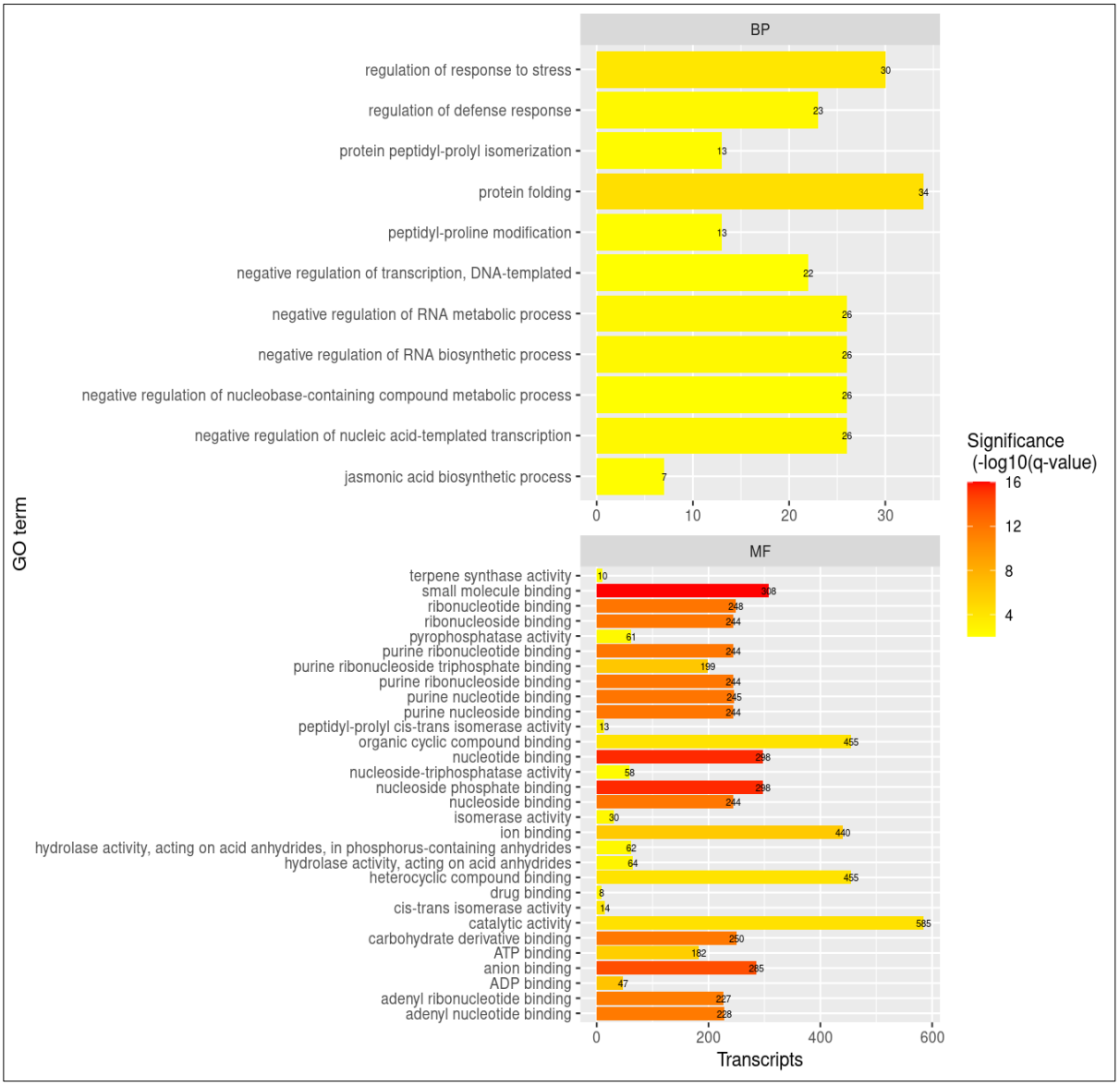


Figure 12 – Bar plot representation of up-regulated GO terms. BP stands for Biological Processes and MF, Molecular Function. The color scale represents the FDR corrected *P* values (q-value≤0.01) calculated in the enrichment analysis.

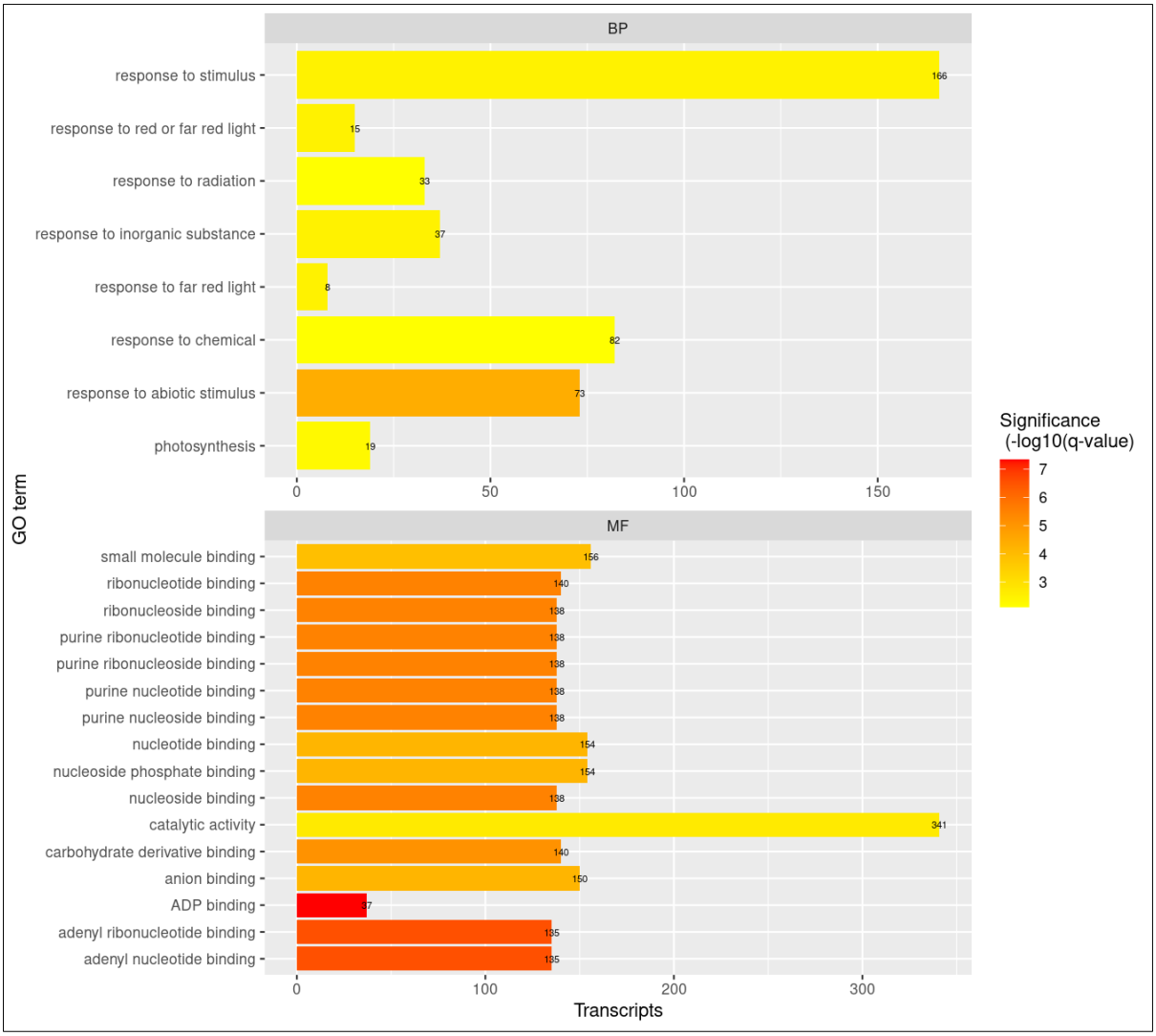


Figure 13 – Bar plot representation of down-regulated GO terms. BP stands for Biological Processes and MF, Molecular Function. The color scale represents the FDR corrected *P* values (q-value≤0.01) calculated in the enrichment analysis.

3.5 Expression of defense-related genes

3.5.1 Expression of triterpenoid-related genes

Terpenoids represent a diverse and versatile class of secondary metabolites involved in plant defense against biotic stresses, including herbivory and wounding. Their biosynthesis in plants occurs in various cellular compartments, primarily the cytoplasm, plastids, and endoplasmic reticulum (Vranová; Coman; Gruissem, 2013). Terpenes are derived from the assembly of five-carbon building blocks, known as hemiterpene or isoprene units (C5H8), which originate from two main metabolic pathways: the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP)

pathway (Li et al., 2023). Both pathways contribute to the synthesis of the universal terpene precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Tholl, 2015).

In plants, MVA pathway takes place in the cytoplasm, producing IPP and DMAPP from acetyl-CoA, primarily for the synthesis of sesquiterpenes and triterpenes. Conversely, the MEP pathway, located in plastids, synthesizes IPP and DMAPP from pyruvate and glyceraldehyde-3-phosphate, mainly contributing to the production of all other types of terpenes (Vranová; Coman; Gruissem, 2013). Once synthesized, IPP and DMAPP serve as precursors for a variety of terpene compounds. Terpene synthases (TPSs) catalyze the condensation of IPP and DMAPP to form various terpene skeletons, with TPS specificity determining the type of terpene produced (Chen et al., 2011).

The initial condensation of IPP and DMAPP by geranyl pyrophosphate synthase (GPS) forms geranyl pyrophosphate (GPP), a C10 precursor of monoterpenes. Further condensation of GPP with another IPP molecule, catalyzed by farnesyl pyrophosphate synthase (FPPS), forms farnesyl pyrophosphate (FPP), a C15 precursor of sesquiterpenes and triterpenes (via squalene). Subsequent addition of another IPP to FPP, catalyzed by geranylgeranyl pyrophosphate synthase (GGPS), produces geranylgeranyl pyrophosphate (GGPP), a C20 precursor of diterpenes and other higher terpenes such as carotenoid and tetraterpenes (Li et al., 2023; Tholl, 2015; Vranová; Coman; Gruissem, 2013).

Condensation of two FPP molecules, catalyzed by squalene synthase, forms squalene, the central substrate of triterpene (C30) biosynthesis. Squalene then undergoes epoxidation, catalyzed by squalene epoxidase, to produce 2,3-oxidosqualene, a common precursor for both sterols and triterpenes. The acyclic 2,3-oxidosqualene undergoes cyclization by oxidosqualene cyclase (OSC) enzymes, adopting one of two main conformations: chair-boat-chair (CBC) for sterol biosynthesis (such as cycloartenol and lanosterol), and chair-chair-chair (CCC) for triterpenoid biosynthesis (such as β-amyrin, α-amyrin, lupeol, glutinol, and friedelin) (Thimmappa et al., 2014).

Cyclization is initiated by protonation of the epoxide group, leading to a cascade of carbocation rearrangements, which ultimately yields tetracyclic sterols or tetracyclic/pentacyclic triterpenes, depending on the product specificity of the OSC.

Oxidosqualene cyclases may be monofunctional, producing a single cyclized product, or multifunctional, yielding two or more major cyclic products.

Triterpenoid skeletons can be further modified by redox reactions, methylation, acylation, and glycosylation, mediated by enzymes such as cytochrome P450 monooxygenase (CYP450), UDP-glycosyltransferase (UGT), and glucosidases. These modifications generate a wide variety of bioactive triterpenoids with various structures, complex chemical properties, and unique functional characteristics (Boutanaev et al., 2015; Han et al., 2022).

Cycloartenol, a pentacyclic triterpene skeleton derived from 2,3-oxidosqualene, serves as the principal precursor for sterols and steroid hormones in plants. In contrast, lanosterol, a tetracyclic triterpene, fulfills the same role in fungi and animals (Ghosh, 2016). For decades, it was believed that cycloartenol synthase was the sole precursor for sterol biosynthesis in plants. However, recent studies have demonstrated that some higher plants possess enzymes capable of converting 2,3-oxidosqualene to lanosterol as well (Ohyama et al., 2009).

Plant genomes contain sets of related genes (gene families) that encode enzymes acting on similar substrates and producing similar products, yet these genes have clearly diverged across different plant lineages (Chen et al., 2011). OSCs are pivotal enzymes belonging to such a gene family, playing a central role in the diversification of triterpenes and sterols. Over 150 OSCs from 75 plant species have been functionally characterized through heterologous expression, and comprehensive evolutionary analyses have revealed that several pentacyclic triterpene synthases repeatedly originated in multiple plant lineages (Wang et al., 2021). The expansion of OSC members in higher plants is believed to have occurred primarily through tandem gene duplication, followed by positive selection and diversifying evolution (Xue et al., 2012). Duplication of ancestral cycloartenol and lanosterol synthase genes is thought to have contributed to the amplification of the OSC gene family in monocotyledons and dicotyledons, respectively (Wang; Wei; Feng, 2022).

In this study, we identified two differentially expressed annotated OSCs that use 2,3-oxidosqualene as a substrate to form triterpenoids: cycloartenol synthase and glutinol/fridelin synthase. Both were enriched in the “steroid biosynthesis” KEGG pathway. Glutinol and Friedelin are pentacyclic triterpenes which are formed through a chair-chair-chair conformation by Glutinol synthase (KEGG reaction R09909: (S)-2,3-Epoxysqualene <=> Glutinol) and Fridelin synthase (KEGG reaction R09910: (S)-

2,3-Epoxysqualene <=> Friedelin), respectively. In contrast, cycloartenol is a tetracyclic phytosterol/stigmasterol precursor produced by Cycloartenol synthase in a chair-boat-chair conformation (KEGG reaction R03200: (S)-2,3-Epoxysqualene <=> Cycloartenol).

While cycloartenol synthase-annotated transcripts were up and down-regulated in this study, glutinol/friedelin synthase-annotated transcripts were consistently up-regulated. Figure 14 illustrates the significantly enriched “steroid biosynthesis” KEGG pathway.

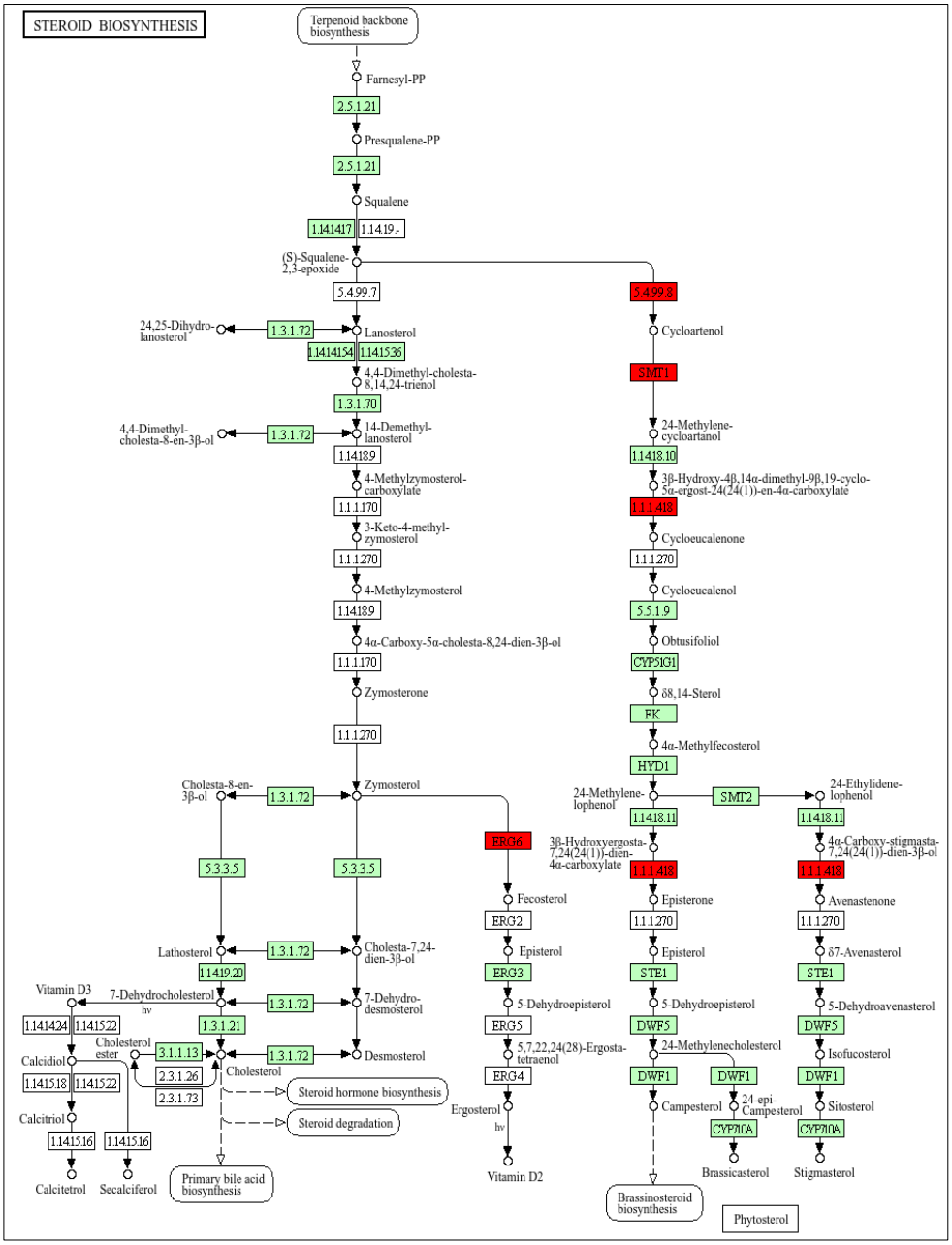


Figure 14 – *Sorghum bicolor* “steroid biosynthesis” KEGG pathway. Red marks indicate enriched steps.

As for plant defenses, terpenoids have been attributed to have direct and indirect effects on herbivores. Directly, terpenoids serve as repellents and reducers of larval feeding and oviposition by herbivores and disturb the nervous system in insects by inhibition of acetylcholinesterase; and indirectly, their synthesis and emission might attract predators or parasites of herbivores, which facilitates the location of injured plants (Sharma; Anand; Kapoor, 2017).

Specifically, the pentacyclic triterpenoid Friedelin has been implicated in imparting negative effects on insects' larval development, feeding, and even leading to mortality. Bioassays have demonstrated the insecticidal potential of friedelin, exhibiting antifeedant, larvicidal, and pupicidal activities against *Helicoverpa armigera* and *Spodoptera litura* (Baskar; Duraipandiyan; Ignacimuthu, 2014). It also presented high mortality of *Aedes aegypti* fourth instar larvae (Famuyiwa; Famuyiwa; Aladesanmi, 2018), adult *Musca domestica,* and *Aedes albopictus* (Skuse) fourth instar larvae (Huang et al., 2009). Furthermore, the phytotoxic and insecticidal potential of friedelin was evaluated in feeding bioassays and oral cannulation, resulting in negative post-ingestive effects, reduced biomass gains, and food consumption on sixth instar *Spodoptera littoralis* larvae (Moiteiro et al., 2006).

An experiment aimed at unraveling the leaf wax composition of sugarcane genotypes resistant to sugarcane borer (IM76228, RB867515, RB047212, RB057249, RB047016, RB047248, among others) was conducted using GC-MS chromatography. This study revealed three triterpene metabolites in sugarcane whorl wax: lanosterol, handianol, and friedelin (Wartha, 2018; Wartha et al., 2022). Analyzing results from *D. saccharalis* untreated (BI – before infestation) and treated plants (AI – after infestation) plants, lanosterol was consistently present in all susceptible genotype samples under both conditions (BI and AI). However, it was not equally screened across tolerant genotypes, with IM76228 showing a lanosterol GC-MS peak in only one replicate. Friedelin was consistently absent in all susceptible genotypes (BI and AI) but present in most tolerant ones. Handianol, a synonym for the triterpenoid cycloartenol (Human Metabolome Database[1]), which is structurally similar to lanosterol, showed an uneven distribution that could not be associated with tolerance groups.

[1] https://hmdb.ca/unearth/q?utf8=%E2%9C%93&query=handianol&searcher=metabolites&button=

Quantitative analysis of friedelin content in wax components and leaf tissues from the same genotypes described above (unpublished data, personal communication) revealed that genotype IM76228 (*S. robustum*) contains significantly higher levels of friedelin in both wax and leaves compared to the sugarcane hybrids. Notably, this genotype presented quantitatively fewer samples with a lanosterol peak. Both sterols and triterpenoids are synthesized as products of one common precursor (2,3-oxidosqualene), and substrate competition might take place. Yet, undetermined mechanisms may favor the production of one metabolite over the other in sugarcane. Apparently, friedelin and lanosterol are associated with sugarcane borer resistance and susceptibility, respectively, and their interplay towards friedelin biosynthesis dictates how plants cope with the pest attack (Wartha, 2018).

As demonstrated, friedelin acts as a post ingestive antifeedant toxin to insect larvae and, presumably, is associated with tolerance against herbivory. To the best of our knowledge, little is known about friedelin complex metabolic pathways in sugarcane and its action on *D. saccharalis* borer larvae, which leads to further investigation and studies to shed light on the interplay between triterpenoid biosynthesis and its importance in conferring tolerance.

Friedelin has been widely studied in plants, primarily due to its pharmacological properties, including anti-inflammatory, antitumor, antioxidant and gastroprotective, neuroprotective, antimicrobial, and antidiabetic effects (Singh et al., 2023). Concerning agricultural applications, in addition to its known natural anti-insect activity, friedelin and other friedelane triterpenoids have also been proposed as a potential herbicide, as they inhibit photosynthesis (Torres-Romero et al., 2010). The discovery of enzymes responsible for friedelin biosynthesis, based on functional expression in *Saccharomyces cerevisiae* and *Escherichia coli,* has been reported mostly in non-grass plant species, such as *Maytenus ilicifolia* (Alves et al., 2018), *Quercus suber* (Busta et al., 2020), *Populus davidiana* (Han et al., 2019), *Tripterygium wilfordi*i (Liu et al., 2020), *Kalanchoe daigremontiana* (Wang et al., 2010b), and *Euphorbia tirucalli* (Kumar et al., 2023).

The first published paper reporting OSC with a friedelin product in the *Poaceae* family was performed using the species *Imperata cylindrica*. The reported OSC (IcOSC3) was characterized as a multifunctional enzyme that produces both glutinol and friedelin as its major products (Naraki et al., 2021). The corresponding mRNA

sequence (NCBI accession number LC476845) is 2,271 bp (base pairs) long, with an *Open Reading frame* (ORF) of 756 aa (amino acids).

By searching into individual transcriptome assemblies generated from each RNA-seq libraries in this study, at least one transcript annotated as glutinol/friedelin synthase (ORF equal to 756 aa) was identified in each genotype of the tolerant group (RB867515, RB047212 and IM76228) and in one susceptible genotype (RB047248). The susceptible genotypes RB057249 and RB047016 did not present any transcripts that matched IcOSC3. All identified sequences showed a high identity percentage with IcOSC3: >95% at the nucleotide level and >94% at the amino acid level (Table 9).

Concerning Cycloartenol synthase (CAS), at least one transcript (ORFs equal to 757 aa) with a CAS annotation was found across all genotypes. These transcripts exhibited >97% identity at both nucleotide and protein level to the NCBI *Sorghum bicolor* accession “cycloartenol synthase isoform X3” (accessions: XM_021459858.1 - nucleotide and XP_021315533.1 - protein) (Table 8).

Table 8 - Nucleotide and protein identity of OSCs found in in this study relative to *I. cylindrica* glutino/friedelin synthase and *Sorghum bicolor* cycloartenol synthase.

Genotype	RB867515	RB047212	IM76228	RB047016	RB057249	RB047248
***Imperata cylindrica* Glutinol/Friedelin synthase sequence**						
Nucl. identity (2271 bp)	96.52	96.52	95.99	-	-	95.99
Prot. identity (756 aa)	96.56	96.43	94.97	-	-	94.58
***Sorghum bicolor* Cycloartenol synthase sequence**						
Nucl. identity (2274bp)	97,89	97,71	98,11	97,58	97,7	97,52
Prot. identity (757 aa)	98,02	97,62	98,15	97,49	98,14	98,01

For homology assessment, the GLU/FRS and CAS putative sequences from RB867515 were aligned using the NCBI blastn tool (*https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome*). This analysis revealed a 64.28% identity and 88% coverage at the nucleotide level. At the protein level, we observed a 53.31% identity (with 71% positives, corresponding to amino acids that are either identical or have

similar chemical properties) and 99% coverage, yielding a blast score of 881. This significant score strongly supports, indicating common ancestry and similar structure/function for these two protein sequences (Pearson, 2013).

To assess the transcriptional activity of genes associated with the above-mentioned triterpenoids, RNA-seq reads from biological samples of each genotype (18 samples) were aligned to the ORF sequences of *Sorghum bicolor* CAS (NCBI ID XM_021459858.1) and *Imperata cylindrica* GLU/FRS (IcOSC3) using *bowtie2* software.

For the CAS reference sequence, read alignment was detected across all genotypes (Figure 15), indicating basal expression in both tolerant and susceptible samples. However, as previously noted, five CAS transcripts were found to be differentially expressed (four up-regulated and one down-regulated), suggesting genotype-specific regulatory patterns.

In contrast, when aligned to the GLU/FRS reference sequence, a distinct transcriptional profile was observed. Tolerant genotypes demonstrated markedly higher expression levels, as evidenced by a substantially greater number of mapped reads (Figure 16). Susceptible genotypes, by comparison, exhibited minimal expression, with only a limited number of reads aligning to the friedelin gene (Figure 16). These findings indicate a strong correlation between elevated friedelin gene expression and pest tolerance in sugarcane, supporting the hypothesis that friedelin plays a functional role in the tolerance phenotype.

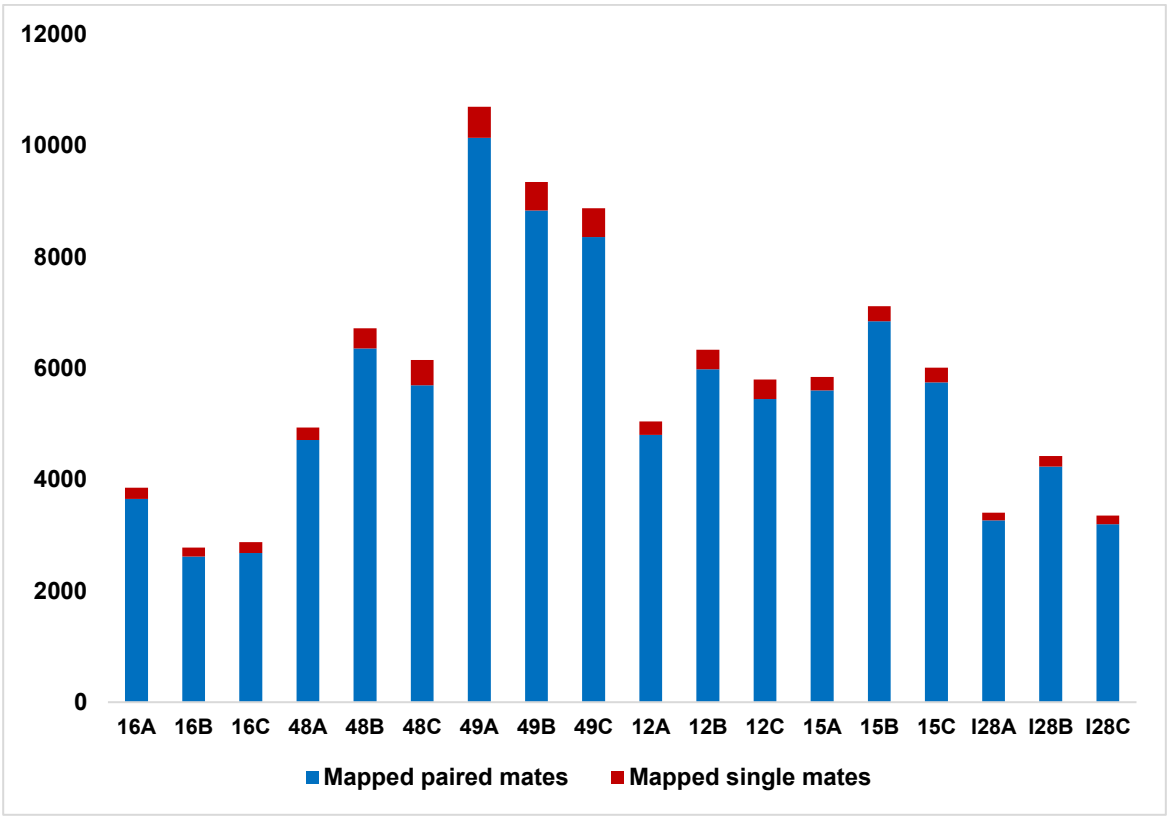


Figure 15 - Bar plot representation of mapped reads of each biological sample to the Cycloartenol synthase coding sequence from *S. bicolor* (NCBI accession XP_002453282) using bowtie2 software. Read alignment was observed across all samples.

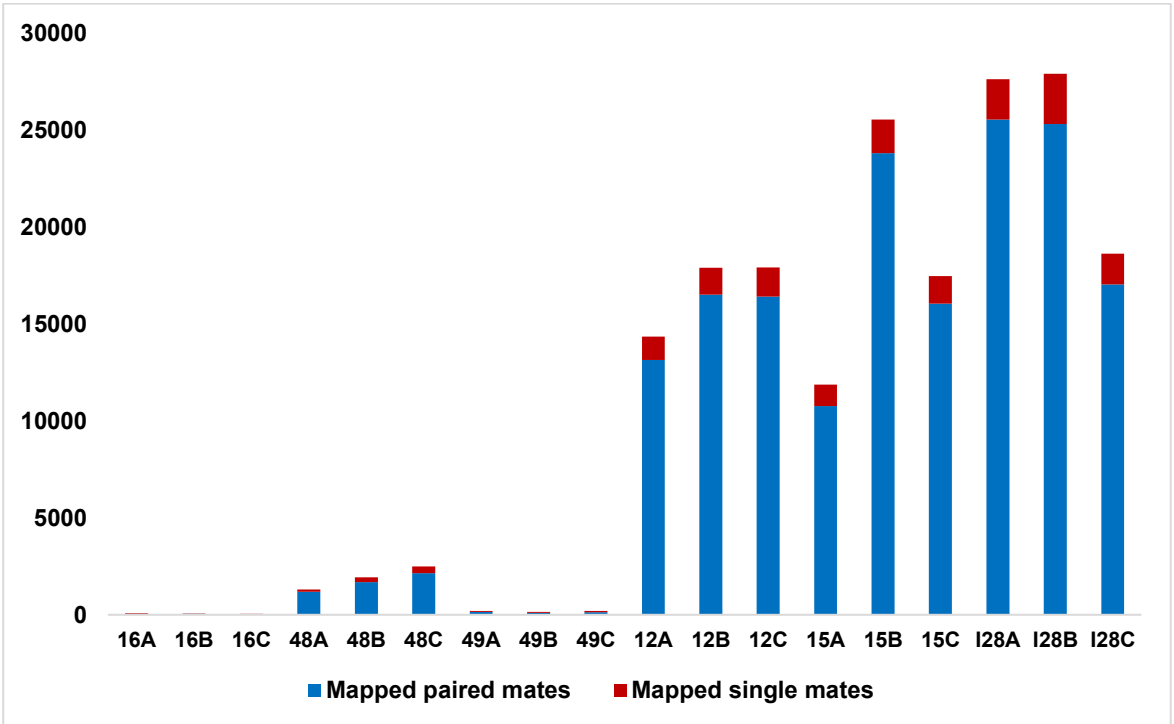


Figure 16 – Bar plot representation of mapped reads of each biological sample to the glutinol/friedelin synthase coding sequence from *Imperata cylindrica* (NCBI accession lsORF3) using bowtie2 software. Noticeable, RNAseq revealed that tolerant genotypes presented a great number of reads aligned to lsOSC3 and, in contrast, fewer reads were present in susceptible genotype libraries.

PCA clustering based on the expression data specific to those transcripts annotated as “glutinol/friedelin synthase” (26 transcripts) revealed 3 distinct groups, separating all 3 susceptible genotypes from tolerant ones, and showing a separation of RB867515 and RB047212 from IM76228 (Figure 16). PC1 and PC2 explained 73% and 19% of the variance, respectively, which combined, they account for 92% of the total variability in the dataset, indicating that these two components effectively summarize the major patterns in the data.

The susceptible genotypes (represented by blue dots) clustered tightly in the upper right quadrant of the plot, while the tolerant genotypes (red dots) were positioned on the left side, with one genotype (IM76228) located on the far right and others (RB867515 and RB047212) on the far left, indicating two distinct transcriptional subgroups within the tolerant category (Figure 17). The strong separation between groups implies distinct transcriptional signatures associated with susceptibility and tolerance. Also, replicates of each genotype cluster very closely together, suggesting high technical reproducibility and biological consistency in gene expression patterns within genotypes. The clustering pattern observed within the tolerant group was expected, as RB867515 and RB047212 are commercial sugarcane hybrids, whereas IM76228 is a wild accession of *Saccharum robustum*. The genetic divergence between these genotypes may lead to the expression of multiple transcript variants of genes involved in glutinol/friedelin synthase biosynthesis, potentially contributing to the pronounced separation among tolerant genotypes in the PCA. A dendrogram based on alignment of protein sequence of friedelin-related transcripts found in this study supports the results observed in the PCA analysis, reinforcing the genetic and transcriptional distinctions among the evaluated genotypes.

The dendrogram in Figure 18 illustrates a phylogenetic alignment tree, which infers evolutionary history, identifies closely related organisms, and cluster organisms with common traits. RB047212 and RB867515 show the highest sequence similarity, clustering together closely. This likely indicates that these two genotypes share conserved glutinol/friedelin synthase-related protein sequences, possibly due to their similar genetic background (both are commercial hybrids). IM76228, a wild *Saccharum robustum* genotype, clusters with hybrids but forms its own distinct branch, suggesting moderate divergence in protein sequences from the hybrids. This reflects both shared evolutionary history and species-level differences. RB047248 is highly divergent from the other three genotypes, branching off earlier. This indicates significant differences

in its protein sequences, which could stem from either distant ancestry or unique adaptations.

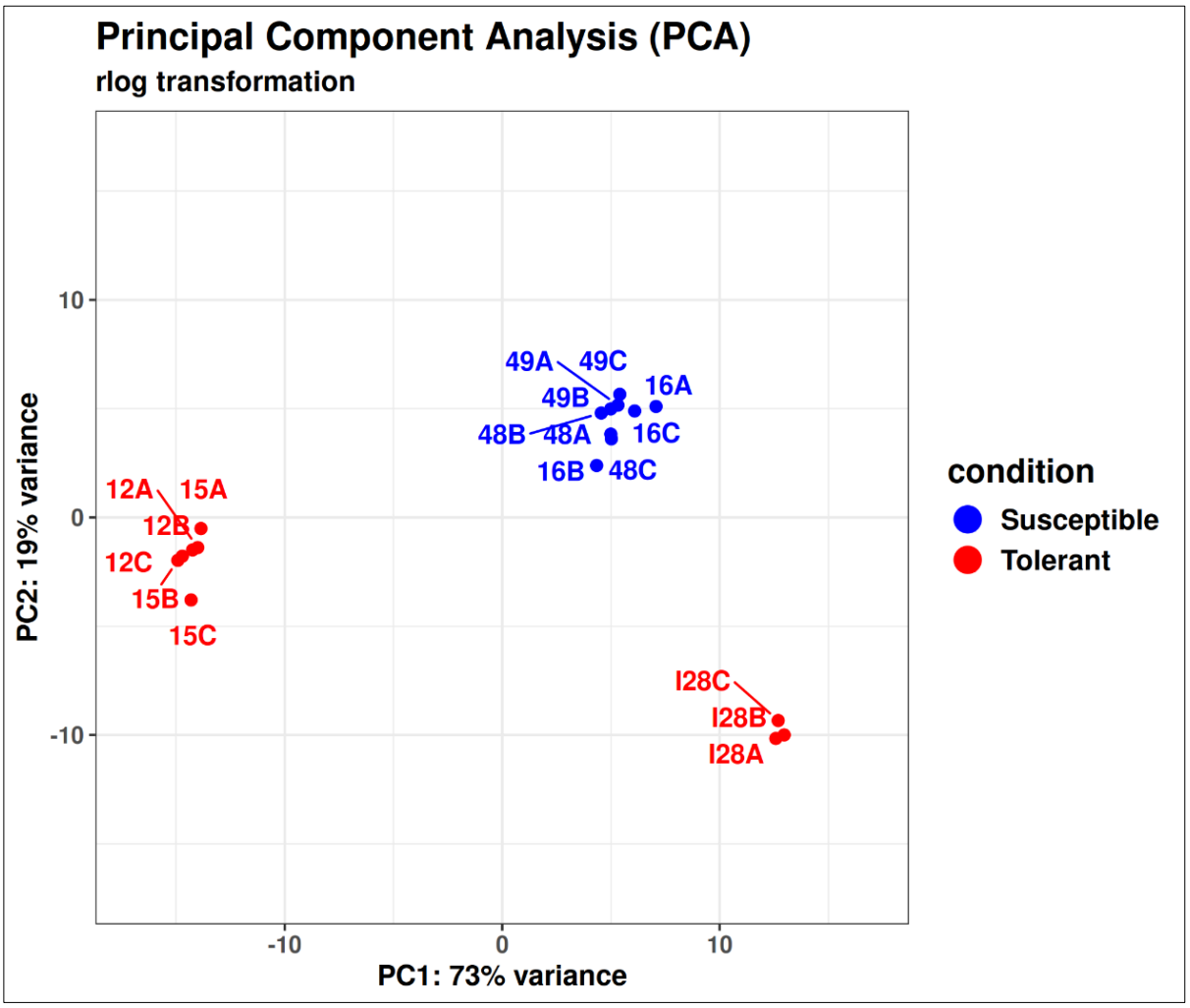


Figure 17 – PCA plot of rlog-transformed data from kallisto filtered expression matrix containing 26 glutinol/friedelin synthase transcripts. Results were very similar to Salmon alignment data. The components clustered samples into 3 distinct groups separating all three susceptible genotypes from tolerant ones and setting apart RB867515 and RB047212 from IM76228. Numbers 12, 15, I28, 16, 48, and 49 stands for RB867515, RB047212, IM76228, RB057249, RB047016 and RB047248. Letters A, B and C represent biological replicates.

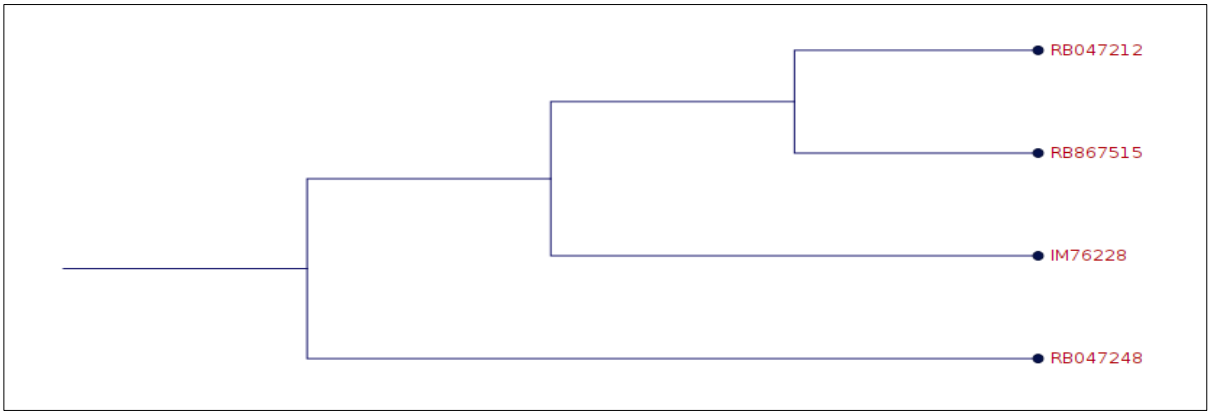


Figure 18 - Phylogenetic alignment tree of four protein sequences derived from the glutinol/friedelin-related transcripts found in this study. A clustering pattern is observed by grouping proteins with similar traits (tolerant samples).

By aligning the putative glutinol/friedelin and cycloartenol synthase ORF sequences from tolerant and susceptible samples found in this study to the R570 and *S. bicolor* v. 5.1 genomes available in Phytozome via blastn, a similar exonic pattern between the ORF sequences was observed. Both sequences possessed 18 exons of similar length, differing only in exons 1, 10, 16, and 18, as illustrated in Figure 19.

Exon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GLU/FRS	201	186	90	195	85	167	189	114	123	120	99	57	47	144	178	75	84	117
CAS	192	186	90	195	85	167	189	114	123	123	99	57	47	144	178	81	84	120

Figure 19 – Exonic structure of glutinol/friedelin and cycloartenol synthases found in *S. bicolor* and R570 genomes.

The alignment analysis in the R570 genome revealed three loci pertaining to the GLU/FRS gene located on two chromosomes within homoeologous group 01: two loci at homolog 01E and one locus at homolog 01B. All loci exhibited varying intronic lengths. Cycloartenol synthase loci were identified on chromosome 03, specifically in the homologs A, B, C, D, E, and F. Furthermore, CAS also presented distinct intronic features.

The very low amount of RNA-seq reads from RB057249 and RB047016 mapping to the reference glutinol/friedelin transcript could be associated with an overall transcription suppression, given that polyploidy often leads to diverse copy numbers of genes (allelic variations) that influence expression. It is worth mentioning that the

OSC gene family shares similar protein domains/motifs, and thus, the low amount of reads aligning to RB057249 and RB047016 susceptible genotypes might originate from genes other than those directly related to friedelin.

As previously mentioned, OSCs are capable of synthesizing one, two, or multiple final cyclic products. Specific mutant amino acid residues in the protein sequence can alter product specificity or dictate the ratio of triterpenoid production. For example, mutagenesis in *Arabidopsis thaliana* cycloartenol synthase revealed that His477 (histidine amino acid at 477th position of the protein) is an essential component for distinguishing the catalytic activity between cycloartenol synthase and lanosterol synthase (Lodeiro; Schulz-Gasch; Matsuda, 2005). A cycloartenol synthase mutant sequence at His477N (original histidine at 477th position mutated to an asparagine) resulted in lanosterol as the major product (88%), whereas mutants at His477Gln (original histidine at 477th position mutated to a glutamine) primarily produced parkeol (73%), another tetracyclic triterpene.

Catalytic activity of enzymes is governed by key amino acids located in active sites that impart protein structure and function. The oxidosqualene cyclase family shares conserved motifs (specific amino acid sequences): DCTAE, QW (aromatic rich motifs of seven residues), and MWCYCR. Specifically, DCTAE is related to substrate binding, with its aspartate residue (A) responsible for substrate connection, cyclization, and protonation. The Q(XXXXX)W motif, located on the exterior of the enzyme, can stabilize carbon cations during cyclization; while MWCYCR is involved in the formation of triterpenoid skeletons (Han et al., 2022; Siedenburg; Jendrossek, 2011; Wang; Wei; Feng, 2022). Although conserved, these motifs have diversified during evolutionary processes of gene duplication and expansion across multiple plant lineages, leading to the wide diversity of triterpenes known today.

All four representative sequences of glutinol/friedelin synthase of RB867515, RB047212, IM76228, and RB047248 genotypes, along with IcOSC3, exhibited four structural Q(XXXXX)W motifs, the catalytic site (DTNAE), and the FWCFCR motif (Figure 20). In comparison, proteins from annotated CAS transcripts displayed different residues for the Q(XXXXX)W motifs, a DCTAE catalytic site, and an MWCHCR motif.

It has been reported that residues surrounding the DCTAE motif strongly influence the B-ring conformation of product specificity. For example, OSCs submitted to mutation at residues near the DCTAE motif (two residues upstream, illustrated in bold: **V**x*DACTAE, where* “x” denotes any possible residue) resulted in loss of function

or altered product specificity (Naraki et al., 2021). However, according to the same author, the upstream residue was not the sole amino acid responsible for inactivating or switching specificity, but rather required context dependency of downstream residues within the DCTAE motif (**V**x*D**CT**AE* – *where bolded* residues illustrate positions that influenced protein function).

In this study, variations in the amino acid composition of residues surrounding the DNTAE motif in putative glutinol/friedelin synthase transcripts were identified. The protein sequences of IsOSC3 and tolerant genotypes exhibited the presence of isoleucine (I481) and alanine (A482) upstream of the catalytic motif (**IA***DNTAE*). In contrast, the susceptible genotype displayed methionine (M481) and Glycine (G482) residues in the corresponding positions (**MG***DNTAE*) (Figure 20). To mitigate potential biases associated with differences in sequencing read coverage, further analysis, including DNA or cDNA sequencing, is necessary to determine whether these amino acid substitutions represent a consistent feature among allelic variants of the GLU/FRS genes.

The second residue upstream from the active site has been attributed to friedelin synthase activity in *Quercus suber*, *Maytenus ilicifolia*, *Populus davidiana*, *Kalonchoe daigremontiana*, all of which possess a leucine residue (**L**S*DCTAE*) (Busta et al., 2020). Glutinol synthase from *Kalanchoe daigremontiana* also features a leucine amino acid at the second upstream residue from the DCTAE motif active site. *Tripterygium wilfordii* is another species that shares this same leucine residue pattern in friedelin synthase genes (Zhou et al., 2019). It has been demonstrated that substituting the non-polar leucine residue with a non-polar isoleucine in Maytenus ilicifolia did not impact friedelin production. However, substitution with a non-polar valine (V) and a polar threonine (T) led to inactivation and the production of *B-amiryin* only, respectively (Souza-Moreira et al., 2016). Nevertheless, other friedelin synthase sequence from *Euphorbia tirucalli,* functionally characterized in yeast expression system (Singh et al., 2023), does not follow the leucine/isoleucine pattern; instead, it presents a valine residue upstream of the catalytic site (**VS***DCTAE*). These findings suggest that amino acid mutations play important roles in diversifying or inactivating triterpene production, and a context dependency of related residues dictates friedelin synthesis in different species.

As the expression of the putative “glutinol/friedelin synthase” gene was undetectable in the other susceptible genotypes analyzed in this study (RB057249 and

RB047016), it was not possible to determine whether the observed amino acid substitution pattern (I482M) is a common characteristic among susceptible genotypes. Therefore, further investigation at the genomic DNA or cDNA level is necessary to clarify whether the differential expression of friedelin synthase in sugarcane is primarily regulated at the transcriptional level, potentially involving transcription factors or promoter activity, or is influenced by structural differences in the protein, where residue substitutions near catalytic sites may impair enzymatic functionality.

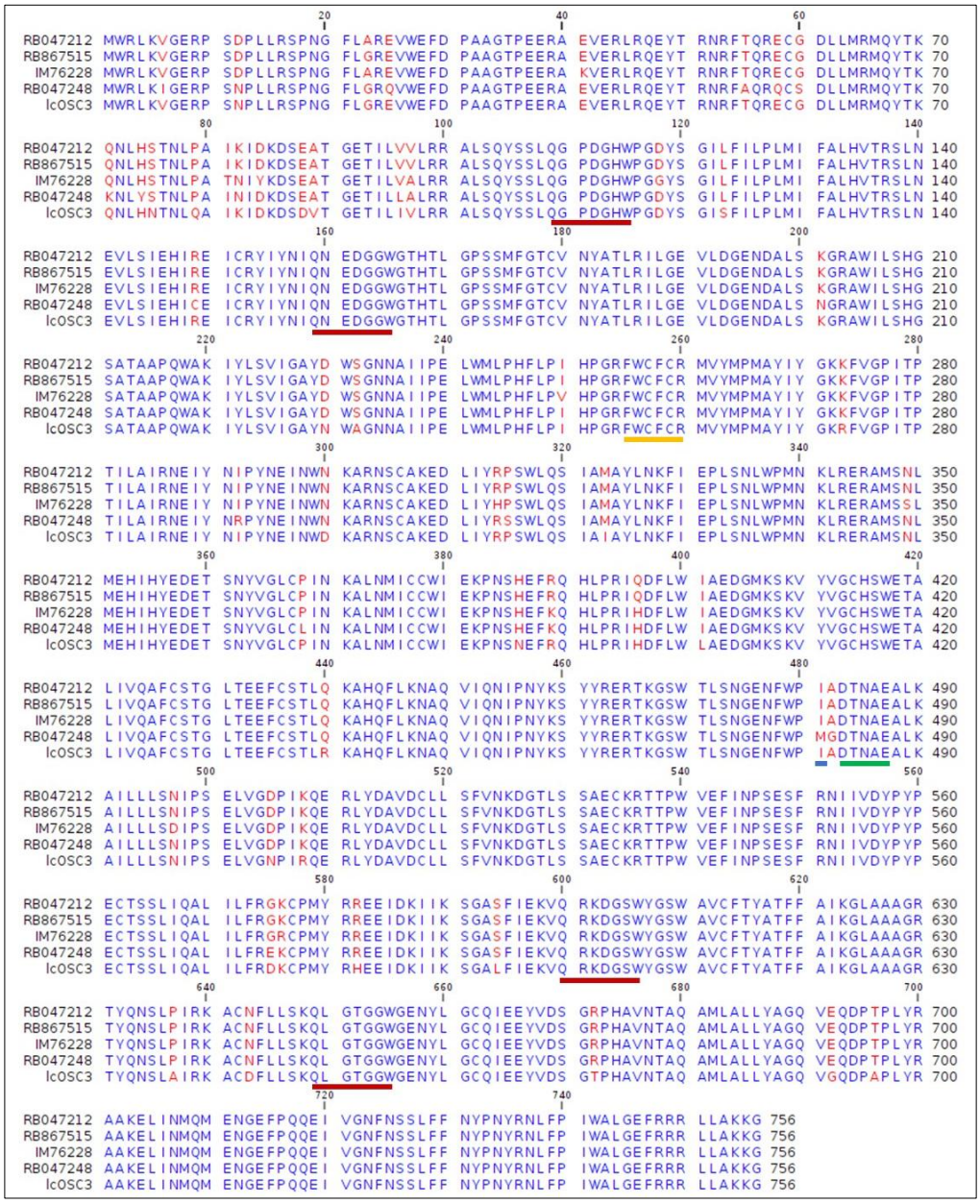


Figure 20 - Amino acid sequence of IcOSC3 and four putative friedelin/glutinol synthases in sugarcane. Four conserved QW motifs, a catalytic site DTNAE, a triterpenoid skeleton FWCFCR motif and a residue involved product specificity are underlined (red, green, yellow and blue, respectively). Residues in red represent divergence among sequences.

3.5.2 Expression of genes involved in the biosynthetic and signaling pathways of ethylene (ET) and jasmonic acid (JA)

Phytohormones are signaling molecules that play a crucial role in plant defense mechanisms against herbivory. Hormones participate in signal transduction pathways, promoting cellular crosstalk communication and coordinating transcriptome changes (Erb; Meldau; Howe, 2012; Hu et al., 2024). It has been proposed that the main defense mechanisms against insect attack in sugarcane are regulated by genes involved in the jasmonic acid (JA) and ethylene (ET) biosynthetic pathways (Falco et al., 2001). Analysis of differentially expressed transcripts in this study revealed that the signaling and biosynthetic pathways of ET and JA were differentially and significantly regulated.

A key enzyme of the ET biosynthetic pathway, 1-aminocyclopropane-1-carboxylate oxidase (ACO), was found to be up- and down-regulated. In general, four DETs were screened: one up regulated and three down regulated. Two down-regulated DETs were significantly enriched in the “Metabolic metabolism” and “Biosynthesis of secondary metabolites” KEGG pathways.

In the overall ethylene biosynthetic process, 1-aminocyclopropane-1-carboxylate synthase (ACS) is an enzyme that catalyzes the synthesis of 1-aminocyclopropane-1-carboxylic acid from S-adenosyl methionine (SAM), and ACO catalyzes the last step of the ET biosynthetic pathway by converting 1-aminocyclopropane-1-carboxylic acid to ethylene (Booker; DeLong, 2015).

ET signaling transcripts were also significantly regulated, and among the annotated DETs are ethylene-responsive factor (ERF) and ethylene insensitive protein 3 (EIN3). ERF transcripts were triggered in both contrasting groups, while EIN3 was triggered only in tolerant genotypes. Overall, EIN3 and ERF transcription factors are involved in positively regulating ethylene responses (Chen; Etheridge; Schaller, 2005). These genes are reviewed to be responsible for triggering defenses against abiotic factors and insect herbivory (Nguyen et al., 2016), but little is known about their interplay in sugarcane defense mechanisms against insect herbivory.

JA and its derivatives have been recognized as key players in plant defense responses against herbivores. The first steps in JA biosynthesis involve enzymes that participate in the “linoleic and alpha-linoleic acid metabolism” pathway. The “*alpha*-

linoleic acid metabolism" was enriched in this study as well as the "Jasmonic acid biosynthetic process" GO term. DETs belonging to the JA biosynthetic genes were annotated, such as lipoxygenase (LOX) and 12-oxophytodienoate reductase (12-OPR). These enzymes participate in the oxylipin biosynthetic process, and more precisely they mediate the biosynthesis of JA and its derivatives from *alpha*-linoleic acid precursor via the octadecanoid pathway (Creelman; Mulpuri, 2002; Wasternack; Feussner, 2018). The octadecanoid pathway has been reported to be involved in defense mechanisms against biotic stresses (Santino et al., 2013).

Generally, JA biosynthesis involves three compartments within the cell: chloroplast, peroxisome and cytosol. The *alpha*-linoleic acid is converted into 12-oxophytodienoic acid (OPDA) inside the chloroplast with the activity of lipoxygenase, allene oxide synthase (AOS) and allene oxide cyclase (AOC). OPDA, is then transported into peroxisome and is reduced by the activity of the peroxisomal enzyme 12-oxo-phytodienoic reductase (12-OPR) forming 3-oxo-2(2′[Z]-pentenyl)-cyclopentane-1-octanoate, which subsequently undergoes *beta*-oxidation processes for the final formation of JA. Upon transport to the cytoplasm, JA undergoes further metabolism to form JA conjugates (Devoto; Turner, 2005; Wasternack; Strnad, 2018), such as jasmonoyl-isoleucine (JA-Ile) and methyl-jasmonate (MeJA), which are conjugated by JA amido synthetases and JA methyl transferases, respectively (Acosta; Farmer, 2010). JA-Ile is considered the most important active form of the hormone (Acosta; Farmer, 2010).

Here, LOX and 12-OPR transcripts exhibited higher expressions levels in the tolerant group, suggesting that the jasmonic acid (JA) biosynthesis pathway is more active both at its initiation in plastids and at its completion in peroxisomes in tolerant plants. Conversely, transcripts encoding JA amido synthetase, which is involved in JA diversification within the cytosol, were downregulated. Figure 21 illustrates the significantly enriched "*alpha*- linoleic acid metabolism" KEGG pathway associated with these findings.

Transcriptional changes of JA and ET biosynthesis and signaling pathways related to the plant's defense responses under caterpillar infestation have been reported. For instance, using transcriptome sequencing and analysis, similar hormonal regulation results were identified in sugarcane subjected to infection with the bacterium *Acidovorax avenae* subsp. avenae (Santa Brigida et al., 2016), in corn infested with the borer *Ostrinia furcanalis* (Wang et al., 2017; Yang et al., 2015), and in cotton

subjected to *Hellicoverpa armigera* larvae infestation (Huang et al., 2015). Additionally, it has been demonstrated that LOX isoforms can play different signaling roles in the regulation of plant growth and activation of stress-induced defense responses (Chen et al., 2004), and confer tolerance against wounding and insect attack (Wang et al., 2008). Thus, it is speculated whether different genes of LOX identified in this study, or possible isoforms, could regulate defense-related genes and impair different levels of tolerance in sugarcane.

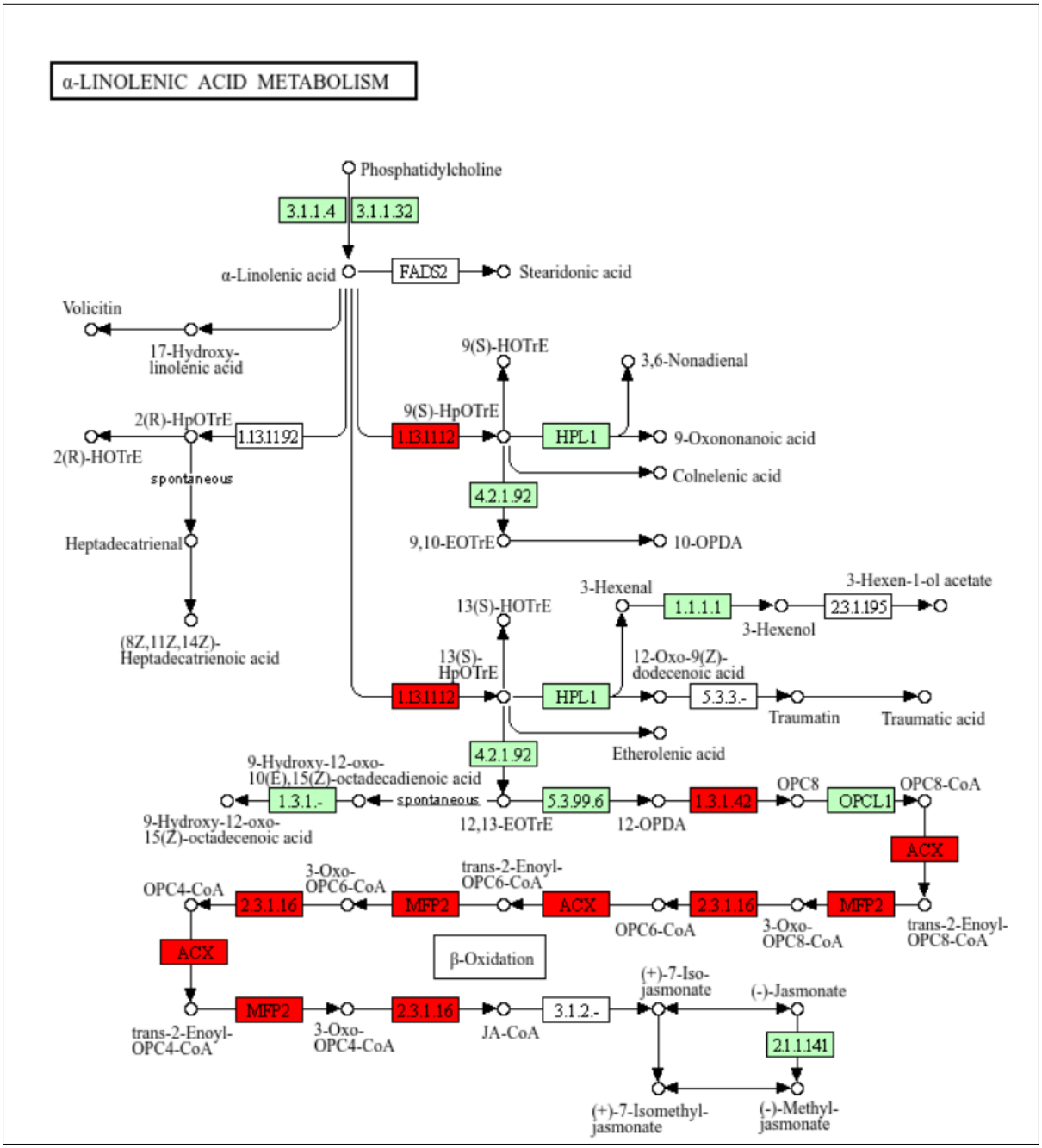


Figure 21 – *Sorghum bicolor* "*alpha*- linoleic acid metabolism" KEGG pathway. Red marks indicate enriched steps.

Hormones are important signaling molecules in terpenoid biosynthesis. It has been demonstrated that expression of terpene synthases was induced by addition of jasmonic acid and ethylene precursor 1-amino-cyclopropyl-1-carboxylic acid, and it has been proposed that ethylene contributes to the herbivory-induced terpenoid biosynthesis by modulating both early signaling events such as cytoplasmic Ca2+-influx and the downstream JA-dependent biosynthesis of terpenoids (Arimura et al., 2008). Jasmonic acid was found to be a very effective elicitor of triterpenoid production, while inhibiting biosynthesis of sterols (Alsoufi et al., 2019; Kim et al., 2011; Lee et al., 2004). On the other hand, ethylene presented a slight effect on triterpenoid metabolism and strongly influenced steroid metabolism (Markowski et al., 2022). In particular, MeJA treatment enhanced accumulation of friedelin synthase mRNA and its respective triterpenoid metabolite in leaves of *Populus davidiana* (Han et al., 2019).

Therefore, experiments revealed that both phytohormones (jasmonic acid and ethylene) influence steroid and triterpenoid biosynthetic pathways and their antagonistic interplay is associated with a metabolic competition which determines the ratio of gene expression and metabolite production.

3.5.3 Expression of transcription factor-related genes

Gene expression is regulated through a wide network of transcriptional and hormonal crosstalk, which includes transcription factor regulation. The mevalonate pathway (MVA), responsible for triterpenoids synthesis, is regulated by a variety of key enzymes and transcription factors (Li; Jia; Qin, 2023). Transcriptional factors (TFs) consist of several modular domains that reflect their core functions in gene expression by finding appropriate DNA elements to bind and recruiting additional protein machinery to that site to affect transcription events (Strader; Weijers; Wagner, 2022). TFs interact specifically with cis elements located in the promoter region of genes. In plants, up to 10% of the genome is estimated to potentially encode TFs (Franco-Zorrilla et al., 2014).

TFs are classified into different families based on specific motifs present in their protein structures. There are about 58 families of TFs in plants, and among them, six major families are found to be involved in innate immune responses to biotic and abiotic stress responses: AP2/ERF (APETALA2/ethylene responsive factor), bHLH

(basic helix-loop-helix), MYB (myeloblastosis related), NAC, WRKY (WRKY conserved motif), and bZIP (basic leucine zipper) (Hoang et al., 2017; Ng; Abeysinghe; Kamali, 2018; Strader; Weijers; Wagner, 2022). Alterations in the expression of TFs change the accumulation patterns of several defensive proteins and numerous enzymes involved in the biosynthesis of secondary metabolites such as terpenoids which are regulated, among others, by the MYC and MYB TF families (Garcia et al., 2021).

In order to identify enriched TF families, enrichment analysis was performed using the online tool PlantRegMap. A total of 26 TF families were identified as enriched among the differentially expressed transcripts (DETs) in the tolerant group relative to the susceptible group. Of these, nine TF families were common to both up- and downregulated DETs, eleven were uniquely associated with the tolerant group, and seven were exclusive to the susceptible group (Table 9).

Table 9 – List of up- and down-enriched Transcription Factor protein families. Bold formats represent common TF family proteins in up and down enriched samples.

TF protein families	
Up regulated	***Down regulated***
MYB family protein	**G2-like family protein**
TCP family protein	**bZIP family protein**
bZIP family protein	**TALE family protein**
bHLH family protein	**MYB family protein**
TALE family protein	**bHLH family protein**
CPP family protein	**BBR-BPC family protein**
BBR-BPC family protein	**CPP family protein**
G2-like family protein	**TCP family protein**
NAC family protein	**NAC family protein**
Dof family protein	YABBY family protein
GRAS family protein	SBP family protein
MIKC_MADS family protein	ERF family protein
C2H2 family protein	MYB_related family protein
HD-ZIP family protein	GATA family protein
B3 family protein	EIL family protein
WRKY family protein	LFY family protein
C3H family protein	
ARF family protein	
E2F/DP family protein	
HSF family protein	

Although transcription factors within the same family typically share similar DNA-binding domains, individual TFs can regulate distinct sets of target genes and thereby exert different biological effects. The table below (Table 10) summarizes the TF families with different individual transcription factors enriched within the WRKY,

MYB, and bHLH TF families, which, in general, are known to coordinate the expression of both defense-related proteins and enzymes involved in secondary metabolite biosynthesis (for more details, see Supplementary Table S2 and S3).

Table 10 – Annotated transcription factors enriched within TF families.

Up-regulated	Down-regulated
WRKY family	
WRKY transcription factor WRKY76 [S. bicolor]	
Probable WRKY transcription factor 4 [S. viridis]	
Probable WRKY transcription factor 70 [S. viridis]	
WRKY transcription factor SUSIBA2-like isoform X1 [S. bicolor]	
Probable WRKY transcription factor 3 [S. bicolor]	
Probable WRKY transcription factor 58 [S. viridis]	
BHLH family	
Transcription factor bHLH35 isoform X1 [S.bicolor]	Transcription factor bHLH49-like [S. bicolor]
Transcription factor bHLH77-like isoform X1 [P. virgatum]	Transcription factor bHLH47-like [S. bicolor]
Transcription factor bHLH168 isoform X1 [S. bicolor]	Transcription factor bHLH96 [S. bicolor]
MYB family	
Myb-related protein Zm38 [S. bicolor]	Transcription factor MYB61 [S. bicolor]
Transcription factor MYBS3 [S. bicolor]	
Myb-related protein 308-like [S. bicolor]	
Target of Myb protein 1-like [S. bicolor]	

The identification of specific transcription factors enriched in tolerant samples might provide deeper insight into the regulatory mechanisms underlying constitutional responses associated with pest tolerance. Notably, the exclusive enrichment of WRKY TFs in tolerant genotypes suggests a pivotal role for this family in mediating defense responses. As transcription factors are key players linking upstream signaling events to downstream gene expression, understanding which TFs are more active offers valuable information on the broader transcriptional reprogramming associated with tolerance.

WRKY transcription factor family is one of the largest groups of transcription factors in plants, and genes pertaining to this family have been implicated in conferring tolerance against multiple abiotic and biotic stresses (Banerjee; Roychoudhury, 2015; Phukan; Jeena; Shukla, 2016). WRKY genes are associated with increased accumulation of triterpenoid compounds (Singh et al., 2017; Sun et al., 2013), their expression is induced with methyl jasmonate (Li; Jia; Qin, 2023), and can bind to W-box sequences in promoters of genes encoding squalene synthase and squalene epoxidases, enzymes directly involved in triterpenoid pathway (Singh et al., 2017).

The plant hormone jasmonate (JA), the main signal responsible for the activation of inducible defenses against arthropods, has an intrinsic protein-protein

interplay with transcription factors within a cascade of signaling and transduction pathways. In the absence of the endogenous bioactive JA-Ile, TIFY/JAZ (ZIM domain-containing protein), known as a key negative regulator of jasmonate-related TFs responses, interacts with transcription factors and prevents them from interacting and promoting JA responsive genes, whilst when JA-Ile is present it mediates the binding of JAZ proteins to the F-box Coronatine Insensitive (COI1) complex, which then undergoes ubiquitination and subsequent degradation of JAZ proteins via the 26S proteasome, releasing transcription factors and allowing them to activate their respective downstream responses (Pauwels; Goossens, 2011; Schweizer et al., 2013).

Besides the basic-loop-helix MYC TF protein family, largely important for expression of defense genes, other TF families have been identified as direct target of JAZ proteins, including R2R3 MYB, EIN (Pauwels; Goossens, 2011) and WRKY (Chen et al., 2021). De-repressed TFs after JAZ suppression interact with a multi-subunit mediator complex (MEDIATOR) which function as a transcriptional coactivator linking DNA-bound TFs to RNA Polymerase II transcription machinery (Kazan, 2017). MEDIATOR25 (MED25) is required for insect/wound and pathogen response branches of the JA signaling pathway (Wang et al., 2015). MED25 interacts with a range of TFs belonging to the MYC family, AP2/ERF, ARF, MYB, WRKY and bZIP (Kazan, 2017; Takaoka et al., 2022).

Ethylene Responsive Factor-related TF has also been identified to play a role in triterpenoid biosynthesis. ERF-type TF can bind specifically to promoters of the genes encoding beta-amyrin synthase, Cycloartenol synthase and Squalene epoxidase, and regulate their expression levels (Chen et al., 2021). Another transcription factor associated with triterpenoid biosynthesis and known to be JA responsive is a bHLH member. It has been shown that a specific bHLH TF (GlbHLH5) significantly responded to methyl-jasmonate (MeJA) induction and upon overexpression significantly increased triterpenoid accumulation, while silencing it demonstrated the opposite effect (Xu et al., 2022). MYB TFs have also been implicated in positive regulation of MeJA responses and subsequently up-regulation of triterpenoid coding enzymes Squalene epoxidase and Cycloartenol synthase (Yin et al., 2020).

4 CONCLUSIONS

Differential expression analysis showed that tolerant genotypes presented a wide array of stress/defense-related genes. Among these were genes involved in the JA and the triterpenoid biosynthesis, especially a putative Glutinol/Friedelin synthase which exhibited high expression in the tolerant genotypes. Also, transcription factor protein families, such as WRKY and bHLH, reported to be responsive to JA and induce OSC biosynthesis and triterpenoid accumulation were found to be significantly enriched in up-regulated transcripts.

We hypothesize that JA, ET and TFs related genes are involved in the regulation of defense-related metabolites (triterpenoids) found in this study. Some differences regarding the regulation levels of common genes between the studied genotypes and the presence of unique DETs in these pathways deserve further investigation. Thus, a more detailed investigation of nucleotide polymorphisms among putative glutinol/friedelin DETs and detailed knowledge of respective gene structures could give insight into how gene regulation is mediated between the tolerance groups analyzed in this study and provide useful information for the development of molecular markers related to tolerance against *D. saccharalis* in sugarcane.

5 DATA AVAILABILITY

The data generated will be available upon publication of the manuscript.

6 REFERENCES

ACOSTA, I. F.; FARMER, E. E. Jasmonates. **The Arabidopsis Book**, v. 8, 2010.

AGISHA, V. N. et al. 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**, v. 13, 2022.

ALSOUFI, A. S. M. et al. Effect of jasmonic acid and chitosan on triterpenoid production in Calendula officinalis hairy root cultures. **Phytochemistry Letters**, v. 31, p. 5–11, 2019.

ALVES, T. B. et al. Friedelin in Maytenus ilicifolia is produced by friedelin synthase isoforms. **Molecules**, v. 23, n. 3, 2018.

ARIMURA, G. I. et al. Herbivore-induced terpenoid emission in Medicago truncatula: Concerted action of jasmonate, ethylene and calcium signaling. **Planta**, v. 227, n. 2, p. 453–464, 2008.

ARRIGONI, E. B. Broca da cana-de-açúcar – Importância econômica e situação atual. Em: ARRIGONI, E.B.; DINARDO-MIRANDA, L.L.; ROSSETO, R. (Ed.). **Pragas da cana-de-açúcar - Importância econômica e enfoques atuais**. Piracicaba: STAB, p. 1–4, 2002.

BANERJEE, A.; ROYCHOUDHURY, A. WRKY proteins: Signaling and regulation of expression during abiotic stress responses. **Scientific World Journal,** v. 2015, 2015.

BASKAR, K.; DURAIPANDIYAN, V.; IGNACIMUTHU, S. Bioefficacy of the triterpenoid friedelin against Helicoverpa armigera (Hub.) and Spodoptera litura (Fab.) (Lepidoptera: Noctuidae). **Pest Management Science**, v. 70, n. 12, p. 1877–1883, 2014.

BELESINI, A. A. et al. De novo transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. **Genetics and Molecular Research**, v. 16, n. 2, 2017.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.

BOOKER, M. A.; DELONG, A. Producing the ethylene signal: regulation and diversification of ethylene biosynthetic enzymes. **Plant Physiology**, v. 169, n. 1, p. 42–50, 2015.

BOUTANAEV, A. M. et al. Investigation of terpene diversification across multiple sequenced plant genomes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 1, p. E81–E88, 2015.

BRAY, N. L. et al. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology**, v. 34, p. 525–527, 2016.

BUCCHINI, F. et al. TRAPID 2.0: A web application for taxonomic and functional analysis of de novo transcriptomes. **Nucleic Acids Research**, v. 49, n. 17, 2021.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, p. 59–60, 2014.

BUSHMANOVA, E. et al. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. **GigaScience**, v. 8, n. 9, 2019.

BUSTA, L. et al. Oxidosqualene cyclases involved in the biosynthesis of triterpenoids in Quercus suber cork. **Scientific Reports**, v. 10, 2020.

CHEN, F. et al. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. **Plant Journal**, v. 66, n. 1, p. 212–229, 2011.

CHEN, G. et al. Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. **Plant physiology**, v. 136, n. 1, p. 2641–2651, 2004.

CHEN, L. et al. The transcription factor WRKY75 positively regulates jasmonate-mediated plant defense to necrotrophic fungal pathogens. **Journal of Experimental Botany**, v. 72, n. 4, p. 1473–1489, 2021.

CHEN, Y. F.; ETHERIDGE, N.; SCHALLER, G. E. Ethylene signal transduction. **Annals of Botany**, v. 95, n. 6, p. 901–915, 2005.

CONAB. Companhia Nacional de Abastecimento - Segundo levantamento - Novembro/2023. 2023.

CREELMAN, R. A.; MULPURI, R. The Oxylipin Pathway in Arabidopsis. **The Arabidopsis Book**, v. 1, 2002.

DE MELLO, U. S. et al. An overview of the transcriptional responses of two tolerant and susceptible sugarcane cultivars to borer (Diatraea saccharalis) infestation. **Functional and Integrative Genomics**, v. 20, p. 839–855, 2020.

DE SETTA, N. et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. **BMC Genomics**, v. 15, 2014.

DEMETRIO, P. A.; ZONETTI, P. DA C.; MUNHOZ, R. E. F. Avaliação de Clones de Cana-de-Açúcar Promissores RBs Quanto à Resistência á Broca-da-Cana (Diatraea Saccharalis) na Região Noroeste do Paraná. **Iniciação Científica Cesumar**, v. 10, n. 1, p. 13–16, 2008.

DEVOTO, A.; TURNER, J. G. Jasmonate-regulated Arabidopsis stress signalling network. **Physiologia Plantarum**, v. 123, n. 2, p. 161–172, 2005.

D'HONT, A. et al. Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. **Molecular and General Genetics**, v. 250, p. 405–413, 1996.

DILLEWIJN, C. VAN. Botany of Sugarcane. **Chronica Botanica Company**, 1952.

DINARDO-MIRANDA, L. L. et al. Resistance of sugarcane cultivars to Diatraea saccharalis. **Pesquisa Agropecuaria Brasileira**, v. 47, n. 1, 2012a.

DINARDO-MIRANDA, L. L. et al. Influência da infestação de Diatraea saccharalis (Fabr.) sobre parâmetros tecnológicos da cana-de-açúcar. **Bragantia**, v. 71, n. 3, 2012b.

DOBIN, A. et al. STAR: Ultrafast universal RNA-seq aligner. **Bioinformatics**, v. 29, n. 1, p. 15–21, 2013.

ERB, M.; MELDAU, S.; HOWE, G. A. Role of phytohormones in insect-specific plant reactions. **Trends in Plant Science**, v. 17, n. 5, 2012.

EWELS, P. et al. MultiQC: Summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, v. 32, n. 19, p. 3047–3048, 2016.

FALCO, M. C. et al. Mechanisms of sugarcane response to herbivory. **Genetics and Molecular Biology**, v. 24, p. 113–122, 2001.

FAMUYIWA, F. G.; FAMUYIWA, S. O.; ALADESANMI, A. J. Activity of the compounds isolated from *Blighia sapida* (sapindaceae) stem bark against *Aedes aegypti* larvae. **Ife Journal of Science**, v. 20, n. 3, p. 601, 2018.

FAOSTAT. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, 2018.

FRANCO-ZORRILLA, J. M. et al. DNA-binding specificities of plant transcription factors and their potential to define target genes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 6, p. 2367–2372, 2014.

GAIDATZIS, D. et al. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. **Nature Biotechnology**, v. 33, n. 7, p. 722–729, 2015.

GARCIA, A. et al. Disentangling transcriptional responses in plant defense against arthropod herbivores. **Scientific Reports**, v. 11, n. 1, 2021.

GARCÍA-ALCALDE, F. et al. Qualimap: Evaluating next-generation sequencing alignment data. **Bioinformatics**, v. 28, n. 20, p. 2678–2679, 2012.

GARSMEUR, O. et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. **Nature Communications**, v. 9, 2018.

GHOSH, S. Biosynthesis of structurally diverse triterpenes in plants: The role of oxidosqualene cyclases. **Proceedings of the Indian National Science Academy**, v. 82, n. 4, 2016.

HAN, J. Y. et al. Functional characterization of an oxidosqualene cyclase (PdFRS) encoding a monofunctional friedelin synthase in Populus davidiana. **Planta**, v. 249, n. 1, p. 95–111, 2019.

HAN, Y. et al. Genome-wide identification of osc gene family and potential function in the synthesis of ursane-and oleanane-type triterpene in Momordica charantia. **International Journal of Molecular Sciences**, v. 23, n. 1, p. 196, 2022.

HEALEY, A. L. et al. The complex polyploid genome architecture of sugarcane. **Nature**, v. 628, p. 804–810, 2024.

HOANG, X. L. T. et al. Transcription Factors and Their Roles in Signal Transduction in Plants under Abiotic Stresses. **Current Genomics**, v. 18, n. 6, p. 483–497, 2017.

HOTTA, C. T. et al. The biotechnology roadmap for sugarcane improvement. **Tropical Plant Biology**, v. 3, n. 2, p. 75–87, 2010.

HU, C. et al. Molecular interaction network of plant-herbivorous insects. **Advanced Agrochem**, v. 3, n. 1, p. 74–82, 2024.

HUANG, J. G. et al. Insecticidal and cytotoxic activities of extracts of Cacalia tangutica and its two active ingredients against Musca domestica and Aedes albopictus. **Journal of Economic Entomology**, v. 102, n. 4, p. 1444–1447, 2009.

HUANG, X. Z. et al. Dynamic transcriptome analysis and volatile profiling of Gossypium hirsutum in response to the cotton bollworm Helicoverpa armigera. **Scientific Reports**, v. 5, 2015.

HUNTER, S. et al. InterPro: The integrative protein signature database. **Nucleic Acids Research**, v. 37, p. D211–D215, 2009.

JIN, J. et al. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. **Nucleic Acids Research**, v. 45, p. D1040–D1045, 2017.

KAPRANOV, P. et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. **BMC Biology**, v. 8, 2010.

KAZAN, K. The multitalented MEDIATOR25. **Frontiers in Plant**, v. 8, 2017.

KIM, Y. S. et al. Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in Bupleurum falcatum. **Planta**, v. 233, n. 2, p. 343–355, 2011.

KOPYLOVA, E.; NOÉ, L.; TOUZET, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. **Bioinformatics**, v. 28, n. 24, p. 3211–3217, 2012.

KUMAR, A. et al. Functional characterization of five triterpene synthases through De-novo assembly and transcriptome analysis of Euphorbia grantii and Euphorbia tirucalli. **bioRxiv**, 2023.

LA CAMERA, S. et al. Metabolic reprogramming in plant innate immunity: The contributions of phenylpropanoid and oxylipin pathways. **Immunological Reviews**, v. 198, n. 1, p. 267–284, 2004.

LAURENT, G. S. et al. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. **BMC Genomics**, v. 13, 2012.

LE CUNFF, L. et al. Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (Bru1) in highly polyploid sugarcane (2n ~ 12x ~ 115). **Genetics**, v. 180, n. 1, p. 649–660, 2008.

LEBOT, V. Biomolecular evidence for plant domestication in Sahu**l. Genetic Resources and Crop Evolution**, v. 46, p. 619–628, 1999.

LEE, M-H. et al. Enhanced Triterpene and Phytosterol Biosynthesis in Panax ginseng Overexpressing Squalene Synthase Gene. **Plant and Cell Physiology**, v. 45, n. 8, 2004.

LI, C. et al. Advances in the Biosynthesis of Terpenoids and Their Ecological Functions in Plant Resistance. **International Journal of Molecular Sciences**, v. 24, n. 14, 2023.

LI, L.; JIA, X.; QIN, L. Advances in the study of key genes and transcription factors regulating the mevalonate synthesis pathway in Edible and medicinal plants. **Food Science and Technology**, 2023.

LI, S. W. et al. Transcriptome and Gene Expression Analysis of the Rice Leaf Folder, Cnaphalocrocis medinalis. **PLoS ONE**, v. 7, n. 11, 2012.

LIU, Y. et al. Identification and functional characterization of squalene epoxidases and oxidosqualene cyclases from Tripterygium wilfordii. **Plant Cell Reports**, v. 39, n. 3, p. 409–418, 2020.

LODEIRO, S.; SCHULZ-GASCH, T.; MATSUDA, S. P. T. Enzyme redesign: Two mutations cooperate to convert cycloartenol synthase into an accurate lanosterol synthase. **Journal of the American Chemical Society**, v. 127, n. 41, p. 14132–14133, 2005.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, p. 550, 2014.

MA, Y. et al. Transcriptome analysis of sugarcane reveals rapid defense response of SES208 to Xanthomonas albilineans in early infection. **BMC Plant Biology**, v. 23, 2023.

MARKOWSKI, M. et al. Effect of ethylene and abscisic acid on steroid and triterpenoid synthesis in Calendula officinalis hairy roots and saponin release to the culture medium. **Plants**, v. 11, n. 3, p. 303, 2022.

MITCHELL, C. et al. Plant defense against herbivorous pests: Exploiting resistance and tolerance traits for sustainable crop protection. **Frontiers in Plant Science**, v. 7, 2016.

MOITEIRO, C. et al. Biovalorization of friedelane triterpenes derived from cork processing industry byproducts. **Journal of Agricultural and Food Chemistry**, v. 54, n. 10, p. 3566–3571, 2006.

MORLAN, J. D.; QU, K.; SINICROPI, D. V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. **PLoS ONE**, v. 7, n. 8, 2012.

NARAKI, S. et al. Two triterpene synthases from Imperata cylindrica catalyzing the formation of a pair of diastereoisomers through boat or chair cyclization. **ChemBioChem**, v. 22, n. 11, p. 1992–2001, 2021.

NG, D. W. K.; ABEYSINGHE, J. K.; KAMALI, M. Regulating the regulators: The control of transcription factors in plant defense signaling. **International Journal of Molecular Sciences**, v. 19, n. 12, p. 3737, 2018.

NGUYEN, D. et al. How plants handle multiple stresses: hormonal interactions underlying responses to abiotic stress and insect herbivory. **Plant Molecular Biology**, v. 91, n. 6, p. 727–740, 2016.

OHYAMA, K. et al. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. **PNAS**, v. 106, n. 3, p. 725–730, 2009.

PATRO, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. **Nature Methods**, v. 14, p. 417–419, 2017.

PAUWELS, L.; GOOSSENS, A. The JAZ proteins: A crucial interface in the jasmonate signaling cascade. **The Plant Cell**, v. 23, n. 9, p. 3089–3100, 2011.

PEARSON, W. R. An introduction to sequence similarity (“homology”) searching. **Current Protocols in Bioinformatics**, v. 3, 2013.

PEREIRA-SANTANA, A. et al. Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. **PLoS ONE**, 2017.

PHUKAN, U. J.; JEENA, G. S.; SHUKLA, R. K. WRKY transcription factors: Molecular regulation and stress responses in plants. **Frontiers in Plant Science,** v. 13, 2022.

PIMENTEL, G. V. et al. Preferência de oviposição e desempenho larval da broca-da-cana em oito genótipos de cana-de-açúcar. **Ciência e Agrotecnologia**, v. 41, n. 4, p. 439–446, 2017.

R DEVELOPMENT CORETEAM (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2024.

ROBERTSON, G. et al. De novo assembly and analysis of RNA-seq data. **Nature Methods**, v. 7, p. 909–912, 2010.

ROSSATO, J. A. DE S. et al. Characterization and impact of the sugarcane borer on sugarcane yield and quality. **Agronomy Journal**, v. 105, n. 3, p. 643–648, 2013.

ROSSATO, J. A. S. et al. Photosynthesis, yield and raw material quality of sugarcane injured by multiple pests. **PeerJ**, v. 2019, n. 1, 2019.

SANTA BRIGIDA, A. B. et al. Sugarcane transcriptome analysis in response to infection caused by Acidovorax avenae subsp. avenae. **PLoS ONE**, v. 11, n. 12, 2016.

SANTINO, A. et al. Jasmonate signaling in plant development and defense response to multiple (a)biotic stresses. **Plant Cell Reports**, v. 32, n. 7, p. 1085–1098, 2013.

SCHWEIZER, F. et al. Differential contribution of transcription factors to Arabidopsis thaliana defense against Spodoptera littoralis. **Frontiers in Plant Science**, v. 4, 2013.

SHARMA, E.; ANAND, G.; KAPOOR, R. Terpenoids in plant and arbuscular mycorrhiza-reinforced defence against herbivorous insects. **Annals of Botany**, v. 119, n. 5, p. 791–801, 2017.

SHEN, W. et al. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. **PLoS ONE**, 2016.

SIEDENBURG, G.; JENDROSSEK, D. Squalene-hopene cyclases. **Applied and Environmental Microbiology**, v. 77, n. 12, p. 3905–3915, 2011.

SIMÃO, F. A. et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, v. 31, n. 19, p. 3210–3212, 2015.

SINGH, A. K. et al. A WRKY transcription factor from Withania somnifera regulates triterpenoid withanolide accumulation and biotic stress tolerance through modulation of phytosterol and defense pathways. **New Phytologist**, v. 215, n. 3, p. 1115–1131, 2017.

SINGH, S. K. et al. Friedelin: Structure, Biosynthesis, Extraction, and Its Potential Health Impact. **Molecules**, v. 28, n. 23, p. 7760, 2023.

SMITH, C. M.; CLEMENT, S. L. Molecular bases of plant resistance to arthropods. **Annual Review of Entomology**, v. 57, p. 309–328, 2012.

SONESON, C.; LOVE, M. I.; ROBINSON, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. **F1000Research**, v. 4, p. 1521, 2015.
SOUZA-MOREIRA, T. M. et al. Friedelin synthase from Maytenus ilicifolia: leucine 482 plays an essential role in the production of the most rearranged pentacyclic triterpene. **Scientific Reports**, v. 6, 2016.

STEINEGGER, M.; SÖDING, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. **Nature Biotechnology**, v. 35, p. 1026–1028, 2017.

STRADER, L.; WEIJERS, D.; WAGNER, D. Plant transcription factors — being in the right place with the right company. **Current Opinion in Plant Biology**, v. 65, 2022.

SUN, Y. et al. Discovery of WRKY transcription factors through transcriptome analysis and characterization of a novel methyl jasmonate-inducible PqWRKY1 gene from Panax quinquefolius. **Plant Cell, Tissue and Organ Culture**, v. 114, p. 269–277, 2013.

TAKAOKA, Y. et al. Protein-protein interactions between jasmonate-related master regulator MYC and transcriptional mediator MED25 depend on a short binding domain. **Journal of Biological Chemistry**, v. 298, n. 1, 2022.

THIMMAPPA, R. et al. Triterpene biosynthesis in plants. **Annual Review of Plant Biology**, v. 65, p. 225–257, 2014.

THIRUGNANASAMBANDAM, P. P.; HOANG, N. V.; HENRY, R. J. The challenge of analyzing the sugarcane genome. **Frontiers in Plant Science**, v. 9, 2018.

THOLL, D. Biosynthesis and biological functions of terpenoids in plants. **Advances in Biochemical Engineering/Biotechnology**, v. 148, p. 63–106, 2015.

TOMAZ, A. C. et al. Assessing resistance of sugarcane varieties to sugarcane borer Diatraea saccharalis Fab. (Lepidoptera: Crambidae). **Bulletin of Entomological Research**, v. 108, n. 4, p. 547–555, 2017.

TOMAZ, A. C. et al. Genetic parameters and selection of sugarcane in early selection stages for resistance to sugarcane borer Diatraea saccharalis. **Crop Breeding and Applied Biotechnology**, v. 19, p. 208–216, 2019.

TOMAZ, Adriano Cirino et al. Genetic parameters and selection of sugarcane for borer resistance in the advanced selection stage. **Crop Breeding and Applied Biotechnology**, v. 20, n. 1, 2020.

TORRES-ROMERO, D. et al. Friedelane triterpenes from Celastrus vulcanicola as photosynthetic inhibitors. **Journal of Agricultural and Food Chemistry**, v. 58, n. 20, p. 10847–10854, 2010.

VENTURINI, L. et al. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. **GigaScience**, v. 7, n. 8, 2018.

VIGNESH, P. et al. Comparative de novo transcriptome analysis identifies salinity stress responsive genes and metabolic pathways in sugarcane and its wild relative Erianthus arundinaceus [Retzius] Jeswiet. **Scientific Reports**, v. 11, 2021.

VITAL, C. E. et al. An integrative overview of the molecular and physiological responses of sugarcane under drought conditions. **Plant Molecular Biology**, v. 94, n. 6, p. 577–594, 2017.

VRANOVÁ, E.; COMAN, D.; GRUISSEM, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. **Annual Review of Plant Biology**, v. 64, p. 665–700, 2013.

WANG, C. et al. The Arabidopsis Mediator complex subunit16 is a key component of basal resistance against the necrotrophic fungal pathogen Sclerotinia sclerotiorum. **Plant Physiology**, v. 169, n. 1, p. 856–872, 2015.

WANG, H. et al. Transcriptome profiling revealed novel transcriptional regulators in maize responses to Ostrinia furnacalis and jasmonic acid. **PLoS ONE,** v. 12, n. 5, 2017.

WANG, J. et al. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. **BMC Genomics**, v. 11, 2010.

WANG, J. et al. Diverse triterpene skeletons are derived from the expansion and divergent evolution of 2,3-oxidosqualene cyclases in plants. **Critical Reviews in Biochemistry and Molecular Biology**, v. 57, n. 2, p. 113–132, 2021.

WANG, P.; WEI, G.; FENG, L. Research Advances in Oxidosqualene Cyclase in Plants. **Forests**, v. 13, n. 9, p. 1382, 2022.

WANG, R. et al. A novel lipoxygenase gene from developing rice seeds confers dual position specificity and responds to wounding and insect attack. **Plant Molecular Biology**, v. 66, n. 4, p. 401–414, 2008.

WANG, Z. et al. Cloning and characterization of oxidosqualene cyclases from Kalanchoe daigremontiana: Enzymes catalyzing up to 10 rearrangement steps yielding friedelin and other triterpenoids. **Journal of Biological Chemistry**, v. 285, n. 39, p. 29703–29712, 2010b.

WAR, A. R. et al. Mechanisms of plant defense against insect herbivores. **Plant Signaling and Behavior**, v. 7, n. 10, p. 1306–1320, 2012.

WARTHA, C. A. Resistance mechanism of sugarcane genotypes to the sugarcane borer. Dissertação—Viçosa: Universidade Federal de Viçosa, 2018.

WARTHA, C. A. et al. Classification of sugarcane genotypes susceptible and resistant to the initial attack of sugarcane borer Diatraea saccharalis using epicuticular wax composition. **Phytochemistry**, v. 199, 2022.

WASTERNACK, C.; FEUSSNER, I. The Oxylipin Pathways: Biochemistry and Function. **Annual Review of Plant Biology**, v. 69, 2018.

WASTERNACK, C.; STRNAD, M. Jasmonates: News on occurrence, biosynthesis, metabolism and action of an ancient group of signaling compounds. **International Journal of Molecular Sciences,** v. 19, n. 9, p. 2539, 2018.

WHITE, W. H. et al. Re-evaluation of sugarcane borer (Lepidoptera: Crambidae) bioeconomics in Louisiana. **Crop Protection**, v. 27, n. 9, p. 1256–1261, 2008.

XIE, C. et al. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. **Nucleic Acids Research**, v. 39, p. 316–322, 2011.

XU, J. et al. Identification of a Novel Metabolic Target for Bioactive Triterpenoids Biosynthesis in Ganoderma lucidum. **Frontiers in Microbiology**, v. 13, 2022.

XUE, Z. et al. Divergent evolution of oxidosqualene cyclases in plants. **New Phytologist**, v. 193, n. 4, p. 1022–1038, 2012.

YANG, F. et al. Analysis of key genes of jasmonic acid mediated signal pathway for defense against insect damages by comparative transcriptome sequencing. **Scientific Reports**, v. 5, 2015.

YIN, J. et al. Functional identification of BpMYB21 and BpMYB61 transcription factors responding to MeJA and SA in birch triterpenoid synthesis. **BMC Plant Biology**, v. 20, 2020.

ZEBELO, S. A.; MAFFEI, M. E. Role of early signalling events in plant-insect interactions. **Journal of Experimental Botany**, v. 66, n. 2, p. 435–448, 2015.

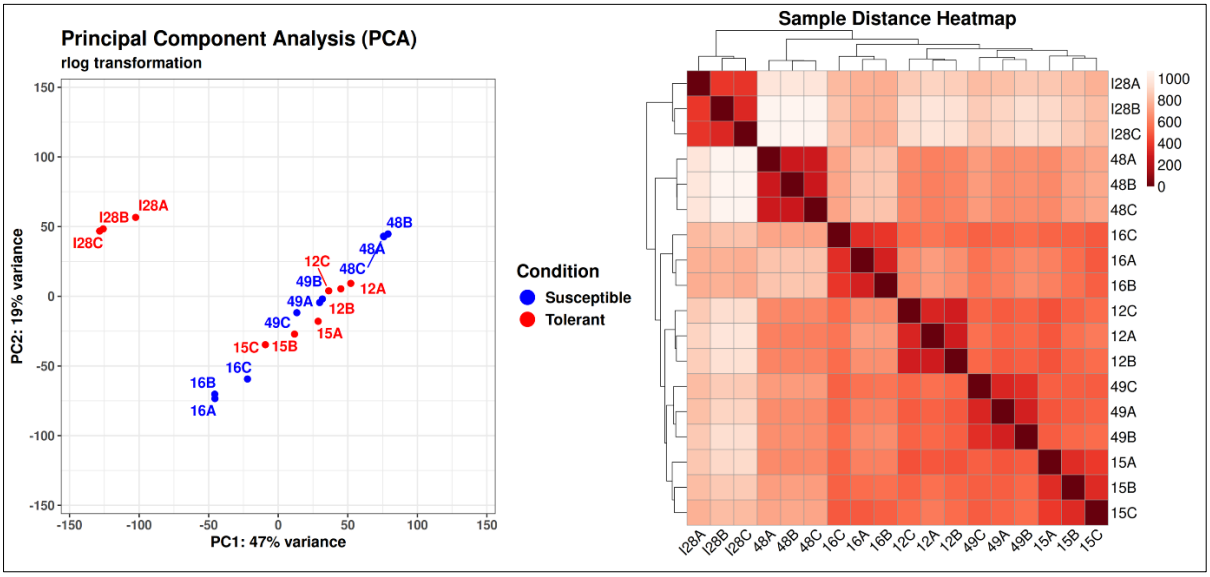
ZHAO, S. et al. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: PolyA+ selection versus rRNA depletion. **Scientific Reports**, v. 8, 2018.

ZHOU, J. et al. Friedelane-type triterpene cyclase in celastrol biosynthesis from Tripterygium wilfordii and its application for triterpenes biosynthesis in yeast. **New Phytologist**, v. 223, n. 2, p. 722–735, 2019.

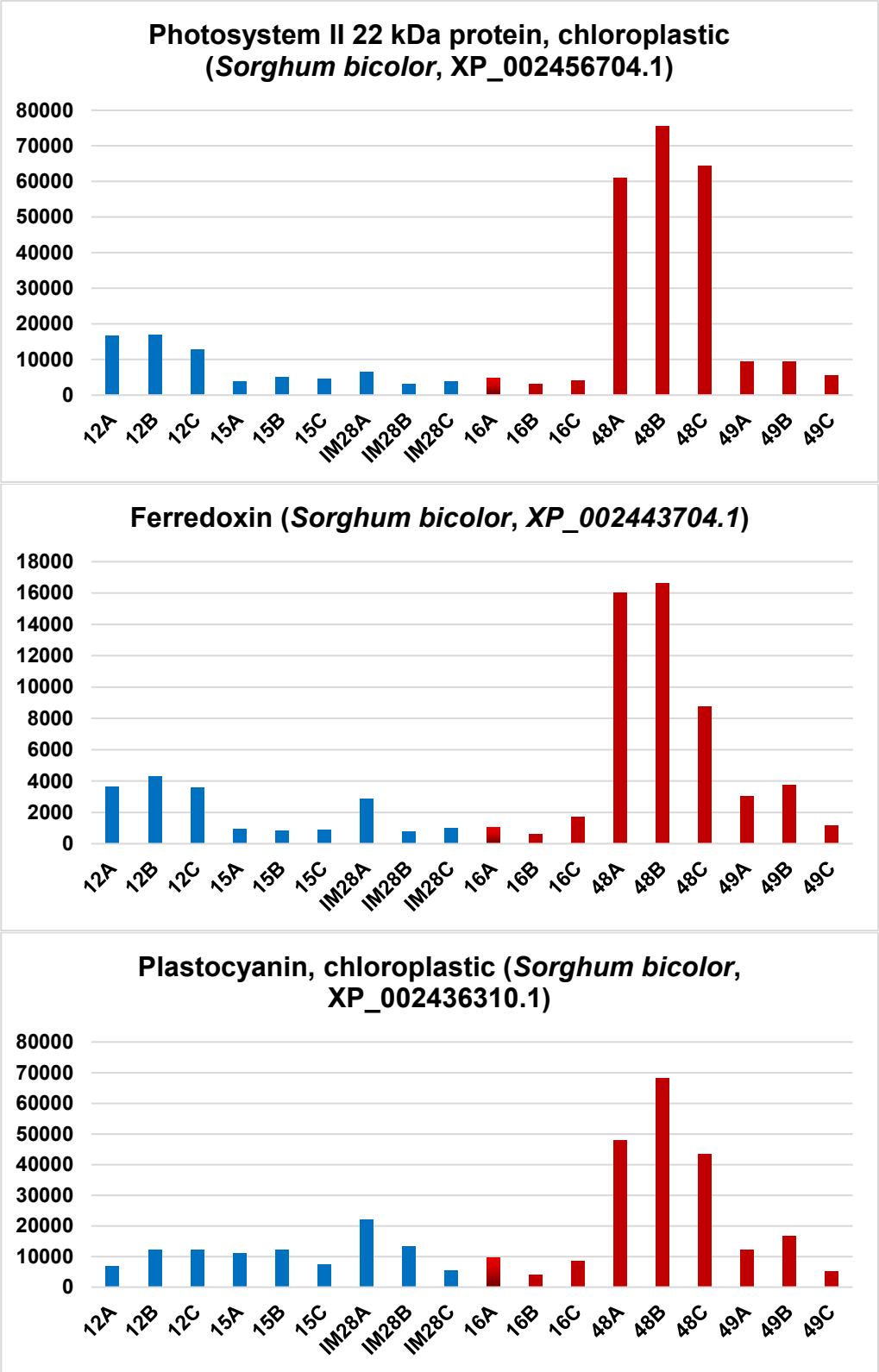
Supplementary Table 1. Classification of genotypes used in this study based on infestation index by sugarcane borer (II, %) at four locations and three agricultural seasons with a natural infestation (reported by (Wartha et al., 2022)

Genotype	Class[1]	II
RB047212	Resistant	6.01
RB867515	Resistant	7.46
IM76-228[2]	Resistant	
RB047016	Susceptible	8.72
RB047248	Susceptible	8.26
RB057249	Susceptible	9.78

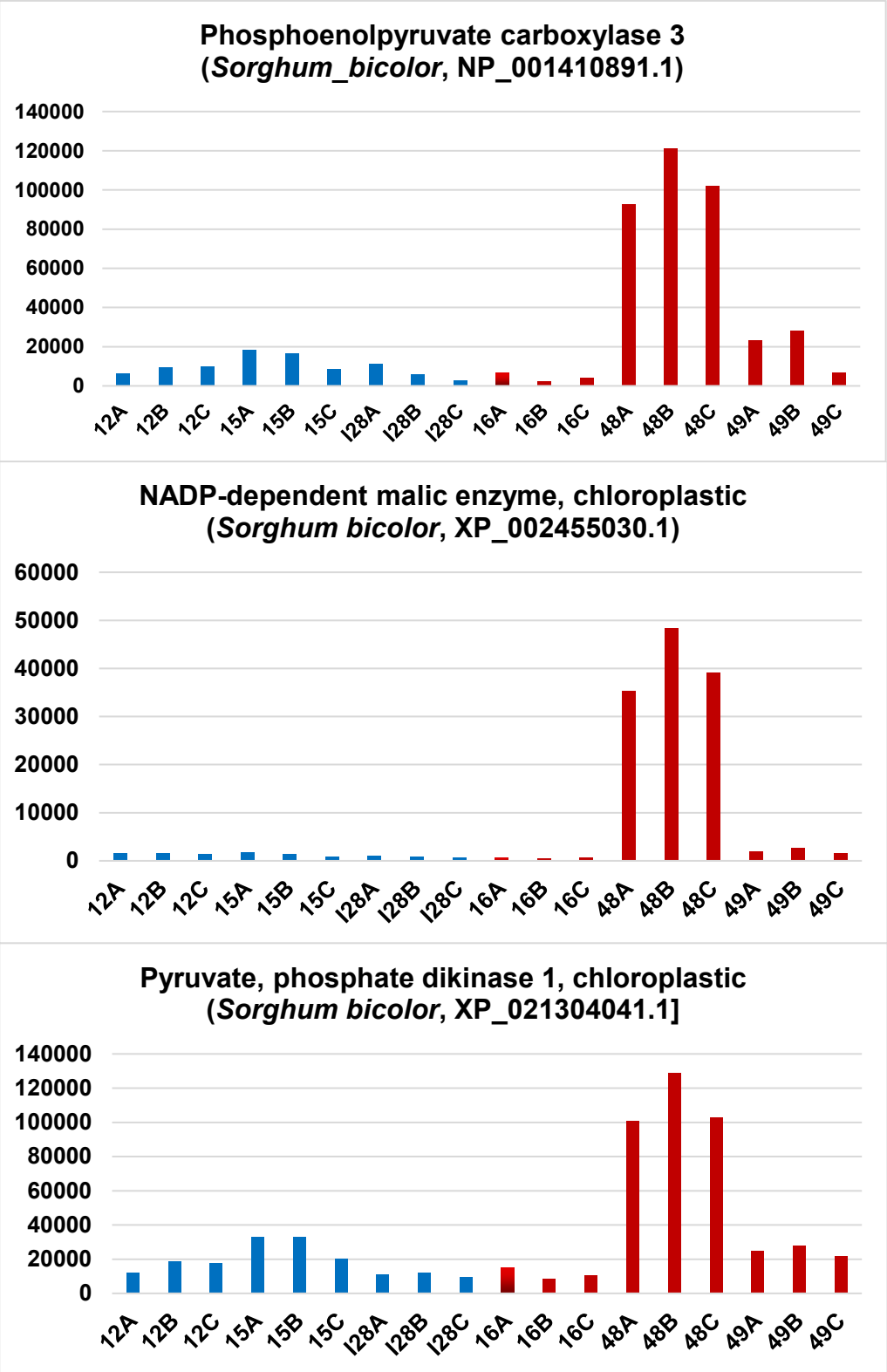
[1] Resistance classes were defined based on the infestation index by sugarcane borer (II, the number of bored internodes divided by the total number of internodes in percentage).
[2] Biometric data are unavailable. The genotype was visually selected in the field because it exhibited low intensity of natural borer infestation. Pimentel et al. (2017) reports that IM76228 presented high larvae mortality rate in the initial stages of leaf feeding.



Supplementary Figure 1 - rlog-transformation PCA plot and heatmap clustering generated using Salmon filtered transcript expression matrix. Numbers 12, 15, I28, 16, 48, and 49 stands for RB867515, RB047212, IM76228, RB057249, RB047016 and RB047248. Letters A, B and C represent biological replicates.



Supplementary Figure 2 – Examples of genes involved in Photosynthesis that presented a higher number of reads mapping to a single susceptible genotype (RB047248). Numbers 12, 15, I28, 16, 48, and 49 represent the genotypes RB867515, RB047212, IM76228, RB057249, RB047016, and RB047248. Letters A, B and C denote biological replicates.



Supplementary Figure 3 – Examples of genes involved in Carbon fixation metabolism that presented a higher number of reads mapping to a single susceptible genotype (RB047248). Numbers 12, 15, I28, 16, 48, and 49 represent the genotypes RB867515, RB047212, IM76228, RB057249, RB047016, and RB047248. Letters A, B and C denote biological replicates.

Supplementary Table 2 – Up-regulated transcription factors enriched in TF families.

Accession	Annotation	Species	Protein domain
	WRKY family		
XP_002462384.1	WRKY transcription factor WRKY76	Sorghum bicolor	WRKY domain
XP_034593926.1	Probable WRKY transcription factor 4	Setaria viridis	WRKY transcription factor, plant
XP_034583912.1	Probable WRKY transcription factor 70	Setaria viridis	WRKY domain superfamily
XP_021310435.1	WRKY transcription factor SUSIBA2-like isoform X1	Sorghum bicolor	-
XP_002444812.2	Probable WRKY transcription factor 3	Sorghum bicolor	WRKY domain superfamily
XP_034602487.1	Probable WRKY transcription factor 58	Sorghum bicolor	WRKY domain
	BHLH family		
XP_002452037.1	Transcription factor bHLH35 isoform X1	Sorghum bicolor	Myc-type, basic helix-loop-helix (bHLH) domain
XP_039850485.1	Transcription factor bHLH77-like isoform X1	Panicum virgatum	Basic helix-loop-helix leucine zipper transcription factor
XP_002465183.2	Transcription factor bHLH168 isoform X1	Sorghum bicolor	Helix-loop-helix DNA-binding domain superfamily
	MYB family		
XP_002444373.2	Myb-related protein Zm38	Sorghum bicolor	-
XP_002455021.1	Transcription factor MYBS3	Sorghum bicolor	Zinc finger, CCHC-type
XP_021321622.1	Myb-related protein 308-like	Sorghum bicolor	SANT/Myb domain
XP_021303049.1	Target of Myb protein 1-like	Sorghum bicolor	ENTH/VHS

Complementary Table 3 – Down-regulated transcription factors enriched in TF families.

Accession	Annotation	Species	Protein domain
	BHLH family		
XP_021310511.1	Transcription_factor_bHLH49-like	Sorghum bicolor	Helix-loop-helix DNA-binding domain superfamily
XP_021307582.1	Transcription_factor_bHLH47-like	Sorghum bicolor	Helix-loop-helix DNA-binding domain superfamily
XP_002489234.2	Transcription_factor_bHLH96	Sorghum bicolor	Helix-loop-helix DNA-binding domain superfamily
	MYB family		
XP_002461087.1	Transcription_factor_MYB61	Sorghum bicolor	Myb-like transcription factor