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

Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane

Deepak Sehgal

Thesis presented to obtain the degree of Doctor in Science. Program: International Plant Cell and Molecular Biology

Piracicaba
2022

Deepak Sehgal
Molecular Biologist

Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Profª Drª **CLÁUDIA BARROS MONTEIRO VITORELLO**

Thesis presented to obtain the degree of Doctor in Science. Program: International Plant Cell and Molecular Biology

Piracicaba
2022

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

Sehgal, Deepak

Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane / Deepak Sehgal. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2022.

156 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Carvão de cana-de-açúcar 2. Transcrição 3. Redes de co-expressão 4. Via de floração 5. RT-qPCR 6. Ethephon I. Título

With love and gratitude,
Dedicated to my family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the University of Sao Paulo and “Luiz de Queiroz” college of agriculture (ESALQ) for providing me with excellent infrastructure and offering the opportunity to help me realize my dream of pursuing a PhD. Faculty members and the administrative staff were extremely helpful, and provided me with the best support they had to offer.

I would like to sincerely acknowledge the support of funding agency Coordination for the Improvement of Higher Education Personnel- BRAZIL (CAPES) for the doctoral scholarship (88882.378369/2019-01), and São Paulo Research Foundation (FAPESP) for financing the research work.

Thanks, would be too small of a word to offer to my advisor, Dr. Cláudia Barros Monteiro-Vitorello. I have no words to describe how grateful I am that you accepted me as a graduate student. You have not only helped me become a better scientist but a better human being as well. Thank you so much for your unwavering support and belief in me all these years. I’m eternally grateful for your compassion, constant support, guidance, patience, and time. I will always fondly cherish my time at your lab.

I would like to express my sincere gratitude to my advisory committee members Dr. Luis Eduardo Aranha Camargo, Dr. Maria Carolina Quecine Verdi and Dr. Alessandra Alves de Souza for their time and invaluable scientific suggestions to help me navigate my research work in the right direction.

I would like to thank Dr. Luis Eduardo Aranha Camargo for his valuable suggestions over the past three years to help me think better as a scientist, and for critical evaluation of my work.

I would also like to thank Dr. Maria Carolina Quecine Verdi for being a wonderful teacher and a great friend. Thank you for helping me get through a difficult time, your advice has been immensely helpful. I would always cherish our friendship.

I would also like to thank the current and previous members of our lab- Jessica Mendes, Gustavo Crestana, Gustavo Lima, Pedro Beretta, Patricia Schaker, Andressa Bini, Mariana Marrafon, Natalia Teixeira, Jackeline Borba, Juliana Calderan, Lâina Oliveira, Marcella Ferreira, Joyce Ferreti, Luiz Quecine, Larissa Nazato, Natalia Araújo, Mariana Braga, Tiarla Souto, Pedro Vilanova, Hugo Rody, Renato Bombardelli, Clesivan Pereira and Thiago Maia.

Jessica Mendes, Gustavo Crestana, Gustavo Lima, and Pedro Beretta, deserve a special mention for their help and friendship in making my stay at the lab a memorable one. It has been such an honor working with all of you.

I would also like to thank Dr. Elaine Vidotto for kindly providing technical support to our lab. It was a pleasure working with you. I would like to extend my sincere thanks to Dr. Silvana Creste for kindly providing us with Ethephon to conduct our validation experiment.

I would like to specially thank my Brazilian family- Enio Oliveira, Martha, Marcele, Tiago, Adriano Couto, Angela, Leo, Alexandre, Patricia, Gustavo, Nueda, Fernando, João Paulo, Kirstin, Jose Eduardo Guercio, and Juliana Guercio for all the love and emotional support. You will always have a very special place in my heart.

I would like to thank my friends from ESALQ and Piracicaba, Akemi Niitsu, Aparecida Silva, Perla Oliveira, Bruno Trevenzoli, Vinicius Trevenzoli, Luis Serezino, Verusca Rossi, Lucas Capelaro, Bruna Factor, Guilherme Hosaka, Fernando Correr, João Pacheco, Rodolfo Silva, Gabriela Machado, Renata Carvalho, Thaisa Tessutti, Fabio Sanches, Helena Gutierrez, and Gabriel Molina for their friendship and hospitality.

A sincere thanks to all the members of defense committee for kindly accepting the invitation to critically evaluate my research work

I would like to offer my sincere gratitude to Alexandre Santos at the post-graduation office for his unwavering and constant support in the last five years. You are an immensely kind person, and I'm lucky to have a friend like you. Thank you so much for protecting me like an elder brother.

I would also like to thank the support staff and faculty members of the International Program in Plant Cell and Molecular Biology for their help and support in the last five years, especially Rodrigo Pessanha.

This journey would have not been possible without the support of my friend, and elder brother, Vinod Sharma, thank you so much for your unconditional support. I know I can always count on you.

From the bottom of my heart, I would like to offer my sincere gratitude to my family members for their moral and emotional support, and for not letting me give up on my dreams, even though it meant I would be far away for a long time in a foreign country and we won't get a chance to meet and hug each other whenever we want. Maa, thank you so much for everything you have done for me. A special thanks to my younger brother Sunny for all the love and support he has given me all my life.

I would like to thank my wife and my best friend, Tanveer for her unconditional love and support. Thank you so much for encouraging me to pursue my dreams and being my safety harbour.

SUMMARY

RESUMO……………………………………………………………………………….……9
ABSTRACT……………………………………………………………………………...11
1. GENERAL INTRODUCTION…………………………………………………….…13
1.1. Objectives…………………………………………………………….............15
References……………………………………………………………………….……16
2. GENETIC AND EPIGENETIC REGULATION OF FLORAL TRANSITION AND ASSOCIATION OF FLOWERING RELATED GENES IN PLANT IMMUNITY, A REVIEW…………………………………………………………………………19
2.1. Gene regulatory network conferring floral transition across multiple plant species………………………………………………………...…………………19
2.2. Correlation of flowering genes with disease resistance or susceptibility under biotic or abiotic stress…………………………………………….……………………29
2.3. Flowering induction or repression under abiotic stress……………………………29
2.4. Effect of biotic stresses on the flowering pathway……………………………..….31
2.5. Smut fungus and its association with flowering pathway………………………....32
2.6. Conclusion……………………………………………………………………….……34
References………………………………………………………………………........35
3. RNASeq PROFILING UNRAVELED AN ANTAGONISTIC EXPRESSION BEHAVIOR PATTERN OF FLOWERING RELATED GENES IN SUGARCANE SMUT RESISTANT AND SUSCEPTIBLE GENOTYPES IN AN EARLY RESPONSE STUDY……………………………………………………………..45
Abstract……………………………………………………………………….…………45
3.1. Introduction………………………………………………………………………….45
3.2. Material and Methods……………………………………………………………….48
3.2.1. Biological Sample Collection……………………………………………..........48
3.2.2. RNA extraction and RNASeq library preparation…………………………..…49
3.2.3. Transcriptional data preprocessing and mapping………………………………49
3.2.4. Data collection and reference dataset……………………………………………49
3.2.5. Differential expression and Network Analysis……………………………………50
3.2.6. Arabidopsis flowering-related genes database and selection of sugarcane orthologs……………………………………………………………………………..50
3.2.7. RNASeq data validation through quantitative real time PCR…………………..51

3.3. Results……………………………………………………………………………52
3.3.1. Co-expression networks of flowering related genes……………………………54
3.3.2. Pathway wise distribution of flowering differentially expressed genes (FDEGs)..58
3.3.3. Candidate genes expression behavior validation via RT-qPCR………………59
3.4. Discussion……………………………………………………………………...60
3.4.1. Genetic and epigenetic control of flowering related genes in resistant genotype SP80-3280 indicates toward suppression of floral transition…………………………..61
3.4.2. Potential epigenetic regulation related to smut……………….………………62
3.4.3. Genetic regulation of flowering-related genes………………………….……..66
3.4.4. Flowering differentially expressed genes seem to activate flowering pathway in the susceptible genotype IAC66-6……………………………………..………………67
3.5. Conclusion………………………………………………………………………70
References………………………………………………………………….………71
4. INVESTIGATING EFFECTS OF FLOWERING REPRESSOR ETHEPHON ON FLOWERING RELATED GENES EXPRESSION BEHAVIOR IN AN EARLY INFECTION STUDY IN SMUT RESISTANT AND SUSCEPTIBLE GENOTYPES………………………………………………………………………81
Abstract……………………………………………………………………………...81
4.1. Introduction……………………………………………………………………81
4.2. Material and Methods………………………………………………………………84
4.2.1. Sugarcane buds collection…………………………………………………...…84
4.2.2. Ethephon treatment and smut fungus inoculation………………………………84
4.2.3. Sample collection at 48 hours after inoculation…………………….................85
4.2.4. DNA extraction and smut fungal infection verification…………………………..85
4.2.5. Plants cultivation in the greenhouse……………………………………………86
4.2.6. Sample collection at vegetative phase-35DAI…………………………………...86
4.2.7. RNA extraction and cDNA preparation…………………………………………...87
4.2.8. Flowering-related genes expression analysis by RT-qPCR……………………87
4.3. Results……………………………………………………………………………88
4.3.1. Confirmation of smut fungus infection via PCR……………………………….…88
4.3.2. Germination rate calculation and physiological growth assessment………….89
4.3.3. Flowering genes expression behavior analysis by RT-qPCR…………………..90

4.3.3.1. Expression profile of candidate flowering genes-RT-qPCR-48 hai…………90
4.3.3.2. Expression profile of candidate flowering genes-RT-qPCR-35 DAI…………92
4.3.3.3. Relative expression analysis (Livak method) to analyze effects of ethephon-48 hai and 35-DAI………………..……………………………………………………94
4.4. Discussion………………………………………………………………….…....100
4.5. Conclusion……………………………………………………………………....103
References……………………………………………………………………………104
APPENDICES………………………………………………………………………..109

RESUMO

Representação de componentes compartilhados na via de floração e via de desenvolvimento do chicote de carvão em cana-de-açúcar

A cana-de-açúcar é uma das culturas de rendimento industrial e agrícola mais valiosas, sendo cultivada atualmente em mais de 100 países em todo o mundo, e também contribui significativamente para a economia brasileira. A doença do carvão da cana, causada pelo fungo biotrófico basidiomiceto *Sporisorium scitamineum*, é uma ameaça proeminente para a cana-de-açúcar, custando até 80% do rendimento e, por sua vez, tem grande impacto econômico. Portanto, desvendar os componentes do *crosstalk* molecular entre os fungos de carvão e a cana-de-açúcar é imperativo para encontrar uma solução para gerenciar essa doença e proteger os meios de subsistência dos agricultores em todo o mundo. Este trabalho teve como objetivo contribuir para estes estudos, entendendo a correlação dos fungos de carvão e a via de floração da cana-de-açúcar. No primeiro capítulo, a compreensão atual da via de floração na planta modelo *Arabidopsis thaliana* e a correlação dos fungos de carvão com a via de floração foi revisada. Os fungos de carvão são biotróficos por natureza e geralmente infectam espécies de gramíneas, sendo que algumas das espécies modulam as estruturas florais das plantas hospedeiras e interferem na via de floração para sobreviver e se reproduzir. Como a cana-de-açúcar também é uma espécie de gramínea, hipotetizamos a existência de um mecanismo semelhante no patossistema do fungo de carvão da cana-de-açúcar. Desta forma o objetivo geral deste trabalho é entender essa correlação estudando a interação cana-carvão em nível molecular, focando especificamente nos componentes da via de floração. No segundo capítulo, usamos RNASeq para gerar perfis transcricionais de genótipos resistentes (SP80-3280) e suscetíveis ao carvão (IAC66-6) 48 horas após a inoculação com o fungo de carvão e utilizando redes de co-expressão estudamos o efeito da infecção do fungo em níveis de expressão de ortólogos putativos de genes de floração em cana-de-açúcar. Nossos dados revelaram que os fungos de carvão provocam um padrão de expressão antagônico de genes de floração em dois genótipos de cana-de-açúcar com níveis contrastantes de tolerância à doença. O perfil transcricional e as redes de co-expressão do genótipo resistente sugerem a repressão da via de floração e, ao contrário, os dados do genótipo suscetível ao carvão sugerem que o fungo induz a ativação da via de floração. Nossos resultados também indicaram a potencial regulação epigenética no patossistema da cana-de-açúcar. As plantas em geral aceleram o processo de transição floral sob condições de estresse para se reproduzir antes de sucumbirem ao estresse, e nossos resultados especificamente para o genótipo suscetível sugerem que potencialmente, um mecanismo semelhante também existe na cana-de-açúcar. No terceiro capítulo, com base nesses resultados, realizamos um estudo exploratório para avaliar a influência de um conhecido repressor de floração, o etefon, na interação cana-carvão em nível fisiológico e molecular. Nenhum impacto fisiológico na progressão da doença do carvão ou no processo de desenvolvimento do chicote foi observado devido ao tratamento com ethephon nos momentos avaliados. O efeito do ethephon nos genes candidatos à floração varia de acordo com o genótipo e o estádio de desenvolvimento, e o ethephon parece potencialmente ter a capacidade de modular o comportamento de expressão dos genes candidatos à floração independentemente do fungo, já que todos os genes candidatos à floração testados apresentaram menor expressão nas amostras tratadas

com ethephon em comparação com as não tratadas no genótipo suscetível. Este estudo forneceu, pela primeira vez, uma análise aprofundada da correlação de genes relacionados ao florescimento no patossistema cana-de-açúcar e incentiva a ideia de que os fungos de carvão têm o mesmo modus operandi que outras espécies de fungos do carvão para interferir na via de floração para sua própria finalidade de crescimento e reprodução. Este é o primeiro estudo na direção de descrever os componentes compartilhados nas vias de desenvolvimento da floração e do chicote e deve encorajar estudos mais detalhados no futuro para desvendar todos os componentes nessas duas vias genéticas distintas. Essas informações ajudarão a encontrar melhores soluções de manejo para a doença do carvão na cana-de-açúcar.

Palavras-chave: Carvão de cana-de-açúcar, Transcrição, Redes de co-expressão, Via de floração, RT-qPCR, Ethephon

ABSTRACT

Depiction of shared components in flowering pathway and smut whip development pathway in sugarcane

Sugarcane is one of the most valuable industrial and agricultural cash crops, currently being grown in more than 100 countries across the globe, and also contributes significantly to the Brazilian economy. Sugarcane smut disease caused by basidiomycete biotrophic fungi *Sporisorium scitamineum* is a prominent threat for sugarcane, costing up to 80% of the yield, and in turn, have a major economic impact. Therefore, unraveling components of molecular crosstalk between smut fungi and sugarcane is imperative to find a solution to manage this disease and protect the livelihoods of farmers worldwide. This work aimed to contribute to this cause by understanding the correlation of smut fungi and the flowering pathway in sugarcane. In the first chapter, the current understanding of the flowering pathway in the model plant *Arabidopsis thaliana* and the correlation between smut fungi with flowering pathways was reviewed. Smut fungi are biotrophic in nature and generally infect grass species, and some of the smut fungi species have been demonstrated to modulate the floral structures of the host plants and interfere with the flowering pathway to survive and reproduce. Since Sugarcane is also a grass species, we hypothesized the existence of a similar mechanism in the sugarcane smut fungi pathosystem. This is why we embarked on this journey to understand this correlation by studying sugarcane-smut interaction at a molecular level by specifically focusing on the components of the flowering pathway. In the second chapter, we used RNASeq to generate transcriptional profiling of smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) genotypes 48 hours after inoculation with smut fungi and also constructed co-expression networks to study the effect of smut fungus infection on the expression levels of putative orthologs of flowering genes in sugarcane. Our data unraveled that the smut fungus elicits an antagonistic expression pattern of flowering genes in two sugarcane genotypes with contrasting levels of smut tolerance. Transcriptional profiling and co-expression networks of the smut resistant genotype suggest the repression of flowering pathway, and on the contrary, the data from the smut-susceptible genotype suggest that smut fungus induces the activation of flowering pathway. Our results also indicated the potential epigenetic regulation at play in the sugarcane-smut pathosystem. Plants, in general, accelerate the floral transitioning process under stress conditions to reproduce before it succumbs to the stress, and our results specifically from the susceptible genotype suggest that potentially, a similar mechanism exists in sugarcane as well. In the third chapter, based on these results, we conducted an exploratory study to evaluate the influence of a known flowering repressor, ethephon, in sugarcane-smut interaction at a physiological and molecular level. No physiological impact on the smut disease progression or whip developmental process was observed due to ethephon treatment in time-points analyzed. The effect of ethephon on candidate flowering genes varied depending on the genotype and the developmental stage. Ethephon seems to potentially have the ability to modulate the expression behavior of candidate flowering genes independently of the fungus as all the candidate flowering genes tested had lower expression levels in ethephon treated samples compared to the untreated ones in the susceptible genotype. For the first time, this study has provided an in-depth analysis of correlation of flowering-related genes in the sugarcane-smut pathosystem. It encourages the idea that the sugarcane smut fungus has the same

modus operandi as other smut fungi species to interfere with flowering pathways for its own growth and reproduction purposes. This is the first study in the direction of depicting the shared components in flowering and smut whip development pathways and should encourage more detailed studies in future to unravel all the components in these two distinct genetic pathways. This information will help in finding better management solutions for the smut disease in sugarcane.

Keywords: Sugarcane smut, Transcriptomics, Co-expression networks, Flowering pathway, RT-qPCR, Ethephon

1. GENERAL INTRODUCTION

Sugarcane (*Saccharum spp.*), is a monocotyledonous plant of vegetative propagation, belonging to the Poaceae family, and is one of the most robust and economically important agricultural cash crops, currently being grown in more than 100 countries across the globe. Brazil is currently the largest producer of sugarcane, with an equivalent output of 40% of the global stock, and the state of São Paulo is responsible for about 50% of the national production (FAOSTAT, 2021; CONAB, 2021). In addition, some characteristic features like rapid growth cycle, high productivity, and robustness to withstand adverse conditions make sugarcane one of the best crops with bioenergy potential compared to other crops.

Sugarcane is also home to many pathogens including a biotrophic basidiomycete, *Sporisorium scitamineum (Syd.)* [(Piepenbring et al. (2002) *(Syn: Ustilago scitaminea H. and P. Sydow)*], which is responsible for causing the smut disease in sugarcane. Smut fungus in sugarcane grows systematically and alters shoot apical meristem to produce whip-shaped sorus (Sundar et al., 2012; Marques et al., 2017; Bhuiyan et al., 2021). Structurally, whips are composed of both fungal and plant tissues, where a black mass of fungal teliospores covers the parenchymal and vascular tissues of the host plant. Whip-shaped sori are responsible for fungal reproduction, producing billions of teliospores. A silvery outer membrane covers the sporogenesis, which ruptures upon maturation resulting in the dispersion of teliospores in the field (Marques et al., 2017).

After dispersion, diploid teliospores germinate on the host surface to initiate meiosis, resulting in the production of haploid sporidia, which in turn results in plant colonization once the mating compatible sporidial cells generate infective dikaryotic hyphae, a *sine qua non condition* for smut infection in sugarcane (Bakkeren et al., 2008; Marques et al., 2017). Finally, the dikaryotic hyphae differentiate into an appressorium and penetrate the host tissue (Peters et al., 2017). Smut disease results in yield loss due to compromised juice quality, higher fiber content, reduced cane diameter, lower sucrose content, and tillering as well, costing farmers up to 80 % of their crops (Monteiro-Vitorello et al., 2018).

Smuts are biotrophic fungi, generally infecting cereal and grass species such as maize, sugarcane, wheat, sorghum, rice, and barley and they mostly use host floral structures

as their own reproduction site, and indeed have the ability to modulate the flowering pathway positively or negatively depending on the host species (Glassop et al., 2013; Fan et al., 2016; Schmitz et al., 2018). For instance, downregulation of flowering related genes in rice plants infected with smut fungi (*Ustilaginoidea virens*), and upregulation of flowering related genes in maize plants infected with smut fungus (*Ustilago maydis*) suggests that smut fungus can potentially repress or induce the flowering pathway in their corresponding host species (Brefort et al., 2009; Fan et al., 2015; Schmitz et al., 2018).

Sporisorium reilianum, the causal agent of head smut disease in maize, is a close relative of *Sporisorium scitamineum,* and induces the phyllody formation (conversion of floral organs into the vegetative ones) in the inflorescence, and also results in the loss of apical dominance in infected plants via disruption of flowering pathway (Ghareeb et al., 2011). These studies demonstrated a clear correlation of smut fungi with flowering pathways in grass species, and since sugarcane is also a grass species, we speculated a similar relationship in a smut-sugarcane interaction. Indeed, a transcriptomic study from our research group did corroborate this idea- the transcriptional profiling study of smut-sugarcane compatible interaction in the smut intermediate resistant cultivar RB925345 reported the differential expression of essential flowering regulators, such as FT, AP1, AP3, AG, COL6, and VIN3 (Schaker et al., 2016). The differential expression of flowering-related genes and meristematic function genes suggests that smut fungus induces a transcriptional reprogramming right after infection, which may interfere with the flower development pathway, inducing later the whip development. Thus, it seems that potentially smut fungus induces the activation of the flowering pathway, and instead, it takes an alternative distinct genetic pathway, resulting in the whip instead of flowers (Schaker et al., 2016). Results of this study indicate the possibility of a similar mechanism like stress/pathogen-induced flowering at play in sugarcane-smut interaction, albeit with a different result.

Depending on the end product of the plant species, flowering could be a boon or bane (Maize and Sugarcane respectively). Delayed flowering in maize limits the crop yield due to prolonged vegetative developmental stage. Flowering is imperative for genetic improvement in sugarcane; However, flowering is not a desirable feature from a commercial production perspective as it compromises the sucrose content, resulting in yield loss. Flowering time optimization is a crucial factor for maximizing the crop yield

and a deep understanding of flowering genes and their interacting partners at a molecular level is the key to optimize the flowering time. Biochemical and genetic studies in sugarcane might open avenues in identifying key flowering genes to delay or repress flowering and improve the crop yield.

Plants activate the flowering pathway under some stress conditions to reproduce before it succumbs to the injury caused by the stressor, and interestingly, smut fungi negatively or positively regulate flowering pathways depending on the host species. To better understand the correlation between whip development and flower development, we investigated the modulation in expression behavior of flowering-related genes via RNASeq in smut susceptible (IAC66-6) and smut resistant (SP80-3280) genotypes 48 hours after inoculation (48 hai) with smut fungus. In addition, since the flowering pathway is not well determined in sugarcane, we used a well-established model system of floral organ development in Arabidopsis to identify sugarcane orthologs of flowering time. A better understanding of molecular crosstalk between the smut fungus *S. scitamineum* and the components of sugarcane flowering pathways is the key to find a better management solution for the smut disease. This work aims to depict the shared components of flowering and smut whip development pathways in sugarcane and this work focuses on analyzing the impact of smut fungus specifically on flowering related genes. The detailed objectives of this study are as follows.

1.1. Objectives

- Identification of flowering genes orthologs in sugarcane by comparative studies in Arabidopsis
- Comparative transcription profiling of flowering related genes in resistant SP80-3280 and susceptible IAC66-6 genotypes 48 hours after infection (hai)
- Construction of the co-expression networks of the differentially expressed flowering related genes in smut resistant and susceptible genotypes in sugarcane
- Gene expression validation of key flowering candidate genes identified from co-expression networks and comparative transcriptional profiling by RT-qPCR.
- Physiological and Molecular Validation- explore if the flowering repressor, ethephon could have any impact on smut whip development at a physiological or molecular level

References

Bakkeren G, Kämper J, Schirawski J (2008). Sex in smut fungi: structure, function and evolution of mating-type complexes. *Fungal Genet. Biol.* 45, 15–21

Bhuiyan SA, Magarey RC, McNeil MD, Aitken KS (2021). Sugarcane Smut, Caused by *Sporisorium scitamineum*, a Major Disease of Sugarcane: A Contemporary Review. Phytopathology. 2021 Nov;111(11):1905-1917.

Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R (2009). *Ustilago maydis* as a pathogen. Annu Rev Phytopathol 47: 423–445

CONAB (2021). Companhia Nacional de Abastecimento. Available at https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar

Fan J, Guo XY, Li L, Huang F, Sun WX, Li Y, Huang YY, Xu YJ, Shi J, Lei Y, Zheng AP, Wang WM (2015). Infection of *Ustilaginoidea virens* intercepts rice seed formation but activates grain-filling-related genes. J Integr Plant Biol. 2015 Jun;57(6):577-90.

Fan J, Yang J, Wang YQ, Li GB, Li Y, Huang F, Wang WM (2016). Current understanding on *Villosiclava virens*, a unique flower-infecting fungus causing rice false smut disease. Mol Plant Pathol. 2016 Dec;17(9):1321-1330.

FAOSTAT (2021). Food and Agriculture Organization of the United Nations (FAO), 2021. Available at: http://www.fao.org/faostat/en

Ghareeb H, Becker A, Iven T, Feussner I, Schirawski J (2011). *Sporisorium reilianum* infection changes inflorescence and branching architectures of maize. Plant Physiol. 2011 Aug;156(4):2037-52.

Glassop D, Rae AL, Bonnett GD (2013). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech* 16, 235–240.

Marques JPR, Appezzato-da-Glória B, Piepenbring M, Massola NS, Monteiro-Vitorello CB, and Vieira MLC (2017). Sugarcane smut: Shedding light on the development of the whip- shaped sorus. Annals of Botany 119: 815–27.

Monteiro-Vitorello CB, Schacker PDC, Benevenuto J, Teixeira e Silva NS, Almeida SS (2018). Progress in understanding fungal diseases affecting sugarcane: smut. In. Philippe Root (ed). Achieving sustainable cultivation of sugarcane Volume 2. Breeding, pests and diseases. Chapter 10. 2018. ISBN-13: 9781786761484.

Peters LP, Carvalho G, Vilhena MB, Creste S, Azevedo RA, and Monteiro-Vitorello CB (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and susceptible genotypes. Planta 245: 749–64.

Piepenbring M, Stoll M, Oberwinkler F (2002). The generic position of *Ustilago maydis*, *Ustilago scitaminea*, and *Ustilago esculenta* (Ustilaginales). *Mycol Progress* 1, 71–80 (2002).

Schaker PDC, Palhares, AC, Peters, LP, Aitken KS, Creste SA, Kitajima JP, Vieira MLC, Monteiro-Vitorello CB (2016). Depicting sugarcane transcriptional reprogramming during the biotrophic interaction with *Sporisorium scitamineum* in disease establishment and symptoms development. PLoS ONE, 11(9): e0162237, 2016.

Schmitz L, McCotter S, Kretschmer M, et al (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. **F1000Research** 7 (F1000 Faculty Rev):1812.

Sundar A, Barnabas E, Malathi P, Viswanathan R (2012). “A Mini-Review on Smut Disease of Sugarcane Caused by Sporisorium scitamineum,” in Botany, ed. J. Mworia (InTech), 226.

2. GENETIC AND EPIGENETIC REGULATION OF FLORAL TRANSITION AND ASSOCIATION OF FLOWERING RELATED GENES IN PLANT IMMUNITY, A REVIEW

Abstract

The developmental switch from vegetative to reproductive (floral transition), a major event in the plant life cycle could be a “bane or boon”, depending on the plant species. Plants have evolved to accurately time this switch through the regulatory networks involved at a transcriptional, post-transcriptional, and post-translational level to improve the yield and avoid any adverse effects of precocious or delayed flowering. Genes from the multiple pathways like photoperiod, circadian clock, autonomous, vernalization, ambient temperature, gibberellic acid, nutrients constitute the regulatory network. Current understanding about the flowering network is due to the immense work done in a model long-day plant, *Arabidopsis thaliana* and short-day plant, rice. Studies focused on understanding the nuances of flowering networks have unraveled common and unique components of the regulatory networks, suggesting that plant species share some ancestral features and have devised some new crop-specific mechanisms that positively regulate crop productivity. Interestingly, the functional behavior of some of the major regulators is highly conserved across plant species. A tug of war between florigens and anti-florigens decides whether a plant will flower or not. In addition to genetic regulation, chromatin remodeling complexes also contribute to deciding the ultimate fate of plant architecture and flowering. Furthermore, many studies have introduced a new concept of flowering induction, the ‘stress induced flowering’. Some of the important flowering regulators, like Gigantea, Constants, FT, and many others have been demonstrated to be major regulators of stress-induced flowering. Numerous studies have provided evidence of stress-triggered induction in the last couple of decades, though a deep understanding of this mechanism remains elusive at a molecular level. A fine mapping of stress-induced flowering pathway will provide insights into the essential components and, most importantly, the site of regulation, allowing the site-directed manipulation to create novel resistant varieties to minimize the adverse effects rendered by biotic and abiotic stresses in plants. The highly conserved nature of flowering pathways in species like rice, raises the question of whether the functional behavior of these regulators is similarly translated into other grass family members like sugarcane. This review summarizes the current knowledge about the regulation of flowering under natural and stress conditions across multiple species and components shared by these two novel and distinct genetic pathways.

2.1. Gene regulatory network conferring floral transition across multiple plant species

A plant majorly goes through four developmental stages- juvenile, vegetative, reproductive, and finally senescence. A complex network of genes quantitatively regulates the maintenance of these developmental stages and transitioning to the next stage. The Flowering pathway has been extensively studied in Arabidopsis, and serves

as a floral transition model crop for the scientific community. A thorough comparison of the flowering pathway of Arabidopsis with that of the other angiosperms has revealed that though they still use some of the ancestral function to control floral transition, they indeed have evolved and come up with their unique mechanisms to manage floral transition to withstand adverse environmental conditions. Vegetative to reproductive transition, a stage when shoot apical meristem (SAM) starts differentiating into the flowers instead of leaves, is of utmost importance to almost all the annual plants.

Since floral transitioning is a crucial and significantly valuable trait, plants have used all the mechanisms available in their inventory and have indeed come up with quite a few new ones to accurately manage the timing of floral transition (vegetative to reproductive) through a complex network of external and internal cues to ensure their reproductive success (Colasanti and Coneva, 2009). This complex regulatory network comprises the genes from six major pathways: photoperiod, vernalization, ambient temperature, autonomous, circadian clock, and gibberellin (Fornara et al., 2010, snapshot of flowering). Around 306 flowering time genes have been identified thus far through the numerous loss of function mutational and transgenic plants analyses in the model plant *Arabidopsis thaliana* (Bouché et al., 2016), and some of these genes are not temporally and spatially exclusive as they tend to act more than once and at different locations like a leaf or shoot apical meristem. Some of the genes in each path and cross-paths are presented below.

Photoperiod

Arabidopsis is a facultative long-day plant, which means floral transition commences once the leaves start perceiving long summer days and conversely, short-days negatively regulate floral induction in Arabidopsis in winter. The photoperiod pathway manages this perceptive response via the regulatory network involving GI (Gigantea) and CO (Constans) genes in leaves. The CO enhances the transcript levels of FT (Flowering locus T) and TSF (twin sister of FT) genes, which encode small proteins that move to meristem via phloem to initiate the flowering process in Arabidopsis.

Constans, a B-box type zinc finger transcription factor, which expresses in the leaves is one of the most important components of the flowering pathway functioning as a hub for integrating exogenous and endogenous clues to regulate floral transition through FT regulation in long-day plants like Arabidopsis (Searle and Coupland, 2004).

Gigantea has been implicated to be responsible for a multitude of biological functions. However, the main role of GI is the photoperiodic induction of flowering, which could either be through two genetically distinct pathways dependent or independent of constans gene regulation (Brandoli et al., 2020). First, Gigantea can positively regulate flowering by interacting directly with miR172 (MicroRNA 172) in an CO independent interaction, which negatively regulates the expression of TOE1 (Target of Eat 1) and AP2 (Apetala2), known repressors of FT (Jung et al., 2007). Second, GI also positively influences flowering through its interaction with SPY (Spindly) gene, as SPY represses CO and FT and negatively regulates gibberellins, which positively affects floral transition (Swain et al., 2001; Tseng et al., 2004). GI has also been shown to directly bind to a precisely targeted location on the promoter of FT, resulting in enhancement of transcript levels of FT, and also denying access to FT promoter to the repressors of FT, SVP (Short vegetative phase), TEM1 (Tempranillo 1) and TEM2 (Tempranillo 2), which neutralizes their effect on FT (Flowering Locus T) activity (Sawa and Kay, 2011). GI has been functionally characterized extensively in numerous plant species from gymnosperms to monocot and dicot angiosperms, and these studies have established that the functional activity of GI is highly conserved (Wang et al., 2016). For example, Gigantea ortholog in Soybean, GmGIa performs the similar function of positive regulation of flowering through activation of florigen ortholog GmFT2a (Watanabe et al., 2011). Additionally, GmGIa positively regulates *gma-miR172a,* which represses expression of Glyma03g33470, resulting in enhancement of transcript levels of FT, AP2 and LFY (Leafy) to cause precocious flowering in soybean. In wheat, another long-day species, flowering occurs precociously under both long-day and short-day conditions in an overexpression study of GI ortholog in wheat, TaGI1 (Zhao et al., 2005). Notably, the functional conservation of GI in regulating flowering time has been reported in short-day species like maize (ZmGI) and rice (OsGI) as well. However, GI seems to play a role of a negative regulator of flowering in maize as it represses FT like gene, ZCN8 (*Zea mays* centroradialis 8) and orthologue of Constans, CONZ1 (Constans of Zea mays 1) albeit only under long-days (Bendix et al., 2013). A recent study demonstrated the functional conservation of GI also in sorghum as the mutation in SbGI resulted in the lower transcriptional activity of SbCo, SbEhd1 and even the orthologues of FT-like genes in sorghum, SbFT, SbCN8, and SbCN12 (Abdul-Awal et al., 2020). In rice, GI negatively regulates flowering indirectly through positive

regulation of Hd1, which represses Hd3a under long day conditions. Though in rice under short day conditions it positively regulates flowering like another short-day plant species, sweet potato (Hayama et al., 2003; Tang et al., 2017). The Constans orthologue of Sorghum, SbCo acts as a positive regulator of flowering by activating the Ehd1 orthologue of sorghum, SbEhd1, and orthologues of maize FT like genes ZCN8 (SbCN8), ZCN12 (SbCN12) in sorghum (Yang et al., 2014a)

Constans orthologue of rice, Hd1 (orthologue of CO) has contrasting behavior in short-days and long-days conditions. (Shrestha et al., 2014). Interestingly, the functionality of GI-CO-FT module in rice (OsGI-Hd1-Hd3a) is not imperative to induce flowering. The unique module Ehd1-Ghd7 (Early Heading date 1-Grain number, plant height, and heading date 7) can also trigger the activation of florigens in rice (Doi et al., 2004). The functioning of Hd1 under long-days as a florigen repressor has been implicated to be due to the timing variations in phytochrome signaling (Izawa et al., 2002). In addition, Hd1 is dependent on genetic background and under long days. When Ghd7 represses the transcriptional activity of Hd1, this repression activity delays flowering induction in rice under long day conditions (Zhang et al., 2017).

The significance of Ehd1 in flowering induction in rice makes it a primary target of regulation as well, and numerous regulators, repressors (Ghd7, Ghd8, Hd16, OsCOL4) and activators (Ehd2, Ehd3, Ehd4, Hd17, OsMADS50, OsMADS51) modulate expression behavior of Ehd1 (Shrestha et al., 2014). Among these repressors, Ghd7 is the most important expression modulator of Ehd1. The Ehd1 repression stops as the day length shortens because the transcript level of Ghd7 drops to a level when it can no longer repress the activity of Ehd1 (Itoh et al., 2010). The loss of function mutation study determined another layer of flowering regulation in rice. Hd2 (Heading date 2), a QTL encoded by OsPRR37, which has homologues in Arabidopsis (PRR7), barley and wheat (PPD1), exclusively represses only one of the florigens paralogs in rice, Hd3a regardless of day length, however the effect is more prominent during long days (Koo et al., 2013).

Circadian clock

The circadian clock is another flowering pathway working as an internal timekeeping machinery of the plants, which exerts regulatory effects to the photoperiodic induction of flowering in response to external stimuli. It comprises three interconnected transcriptional-translational regulatory feedback loops: morning loop, central loop, and

evening loop. The components of these loops interact with each other at the transcriptional and post-transcriptional level to positively or negatively regulate their gene expression (Pokhilko et al., 2012).

The morning loop is composed of different *PSEUDO-RESPONSE REGULATORs* (*PRRs*), PRR5, PRR7, and PRR9, expressing in a sequential order (Adams et al., 2015), central loop contains PSEUDO-RESPONSE REGULATOR 1/TIMING OF CAB EXPRESSION 1 (PRR1/TOC1), and two partially redundant myb transcriptional factors: CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY). The evening loop is formed of GI, and the EC (evening complex) formed of LUX ARRHYTHMO (LUX), ZEITLUPE (ZTL), EARLY FLOWERING 3 and 4 (ELF3 and ELF4) genes (Más et al., 2003; Inoue et al., 2017). Morning loop components PRR5, PRR7, and PRR9 express in a sequential order and negatively regulate central loop components CCA1, and LHY. CCA1 and LHY represses the transcription of another component of central loop PRR1/TOC1, which positively regulates the expression of its repressors CCA1, and LHY (Liu et al., 2016). The components of the evening loop ZTL and ELF3 enables plants to accurately perceive changes in the day length and also the control over light input signal to the endogenous clock (Doyle et al., 2002; Kim et al., 2007; Adams et al., 2015; Anwer et al., 2020).

One of the most important functions of GI involves its interaction with LOV (Light, Oxygen, Voltage) domains of ZTL, and FLAVIN-BINDING, KELCH REPEAT, F BOX protein 1 (FKF1). The former protein-protein interaction of GI with ZTL at a post translational level ensure the stabilization and maintenance of blue light photoreceptors of ZTL (Cha et al., 2017) and latter interaction of GI with LOV domain of FKF1 results in the formation of enzymatic complex, which degrades the repressor of CO, called CDF1 (Cycling DOF Factor 1). This degradation of CDF1 leads to enhancement in transcript levels of CO, and which in turn results in the activation of Flowering Locus T, FT (Imaizumi et al., 2003; Sawa et al., 2007).

A multilayered transcriptional and post-transcriptional regulatory system comprising of CDFs (Cycling DOF factors, the repressors of CO, Fornara et al., 2009), COP1-SPA1 complex (Constitutive Photomorphogenic 1-Suppressor of PHYA-105 1, Laubinger et al., 2006; Xu et al., 2016), Photoreceptor PHYB (Phytochrome B) and PHYA and CRY2 (Cryptochrome 2) ensures the enhancement of transcriptional activity of constans in the late afternoon to start the floral transitioning process during long days (Tiwari et al.,

2010). In a parallel CO independent interaction, PHYB negatively influences transcription of FT through directly repressing an upstream activator of FT, PFT1/MED25 (Phytochrome and Flowering Time 1/Mediator 25), thereby delays the floral transition commencement in leaves (Kim, 2020). One of the most common features, which seems to be existing in almost all the angiosperms is the presence of a mobile flowering activator protein, called florigen (e.g., FT, in Arabidopsis) which translates in the leaves and transports to the SAM (Shoot apical meristem) from companion cells of the phloem, where the floral transition occurs in all the angiosperms (Liu et al., 2013). In Arabidopsis, once the florigen protein FT arrives at the SAM, it commences the floral development by interacting with a bZIP transcription factor called FD (Flowering locus D) to form FT-FD dimer, and also through enhancing the transcript levels of a MADS box protein, AP1 (Apetala 1), LFY (Leafy) and SOC1 (suppressor of overexpression of constans 1). This mechanism of activation of meristem identity genes to induce flowering through florigen is also conserved in Rice. OsHd3a-OsFD1 module in rice mimics the functional activity of FT-FD module of Arabidopsis to activate a key meristem identity gene, OsMADS15, an orthologue of AP1 gene in Arabidopsis to induce flowering (Taoka et al., 2011)

Flowering integrator genes, such as FT (flowering locus T) and SOC1 (suppressor of overexpression of constans 1) are at the heart of the floral transition as they merge cues from the different pathways to regulate floral timing. Activation of SOC1 through FT-FD dimer further coordinates the gene expression of key flowering genes AGL24 (Agamous like 24), LFY and SPLs. Combinatorial action of such a highly coordinated gene network brings out dramatic morphological changes at the shoot apical meristem which transforms into inflorescence meristem from vegetative meristem, and starts producing flowers instead of leaves.

There are numerous FT-like genes in the plant playing distinct, overlapping or redundant roles in the flower developmental pathway. Though, FT-like genes are predominantly involved in floral transition, there are few species where FT-like genes do not play a role in flower development like Potato and Onion, where it's responsible for storage organ differentiation. In general, FT-like genes act as positive regulators of flowering, e.g., HD3a, RFT1, FTL1 in Rice; FT and TSF in Arabidopsis (Putterill et al., 2016) and SbFT1, SbFT8, SbFT10 (Wolabu et al., 2016). However, not all FT-like genes regulate flowering process positively as several FT-like genes have been reported as anti-florigens e.g. There are two FT-like genes in sugar beet, however

these two genes have antagonistic function, FT1 acts as anti-florigen and FT2 like majority of FT like genes promote flowering (Pin et al., 2010).

This balance between florigens and anti-florigens governs the overall plant architecture and finally the flowering (Lifschitz et al., 2014). Like a diverse array of FT and FT-like genes representation across different short day and long day plants, a similar diversity in regulators of FT and FT-like genes has also been reported (Yano et al., 2000).

Gibberellins

Gibberellins also play a significant role in inducing flowering in addition to their function as plant growth regulator. Any interruption in biosynthetic pathways of gibberellic acid or degradation of bioactive GA negatively regulates the timing of floral transition especially in short day conditions (Davis, 2009; Fornara et al., 2010; Cho et al., 2017). Notably, three enzymes involved in GA pathway are of utmost importance, GA20ox (Gibberellin 20 oxidase), GA3ox (Gibberellin 3 oxidase) and GA2ox (Gibberellin 2 oxidase). Any changes in the activity of these enzymes results in the positive or negative alteration of the flowering time. A hydroxylation reaction catalyzed by GA2ox, results in the degradation of the immediate precursor of bioactive GA or directly bioactive GA (GA4), limiting its availability and therefore delays the floral transitioning (Mutasa-Göttgens et al., 2009). GA20ox and GA3ox are expressed in the leaves, and they catalyze a number of precursors in the GA biosynthetic pathway to get bioactive GA, GA1 or GA4. The importance of these two enzymes makes them primary targets of regulation and interestingly, concentration of bioactive GA achieves an optimal value at the meristem right before the flowering is induced. After its movement to meristem from leaves, it positively regulates the floral integrator genes SOC1 and LFY to induce flowering. Interestingly exogenous application of GA triggers floral induction in Arabidopsis, though this effect is species specific.

Temperature

Ambient temperature pathway has a crucial role in correctly timing the floral transition through regulation of the key floral integrator genes FT, TSF, SOC1 and FUL. SVP (short vegetative phase), a MADS box TF and a key component of the ambient temperature pathway seems to negatively regulate transcription of FT, especially at the lower temperature in Arabidopsis (Fornara et al., 2010). Andres et al., 2014, demonstrated that SVP not only negatively affects floral integrator genes, it also somehow represses the GA biosynthesis pathway to delay the floral transition process and during photoperiodic induction of flowering, the concentration of SVP reduced, which interestingly coincides with the increase in the level of GA20ox at the shoot apex. This growth in gibberellins also positively regulates the concentration of a key component of the aging pathway, SPL (Squamosa Promoter Binding Like).

In Arabidopsis, age of the plant also influences the timing of flower induction, as the plant matures, accumulation in the concentration of SPL (Squamosa Promoter Binding Like) TFs is observed, which helps in floral induction by positively regulating the transcription of FT, LFY (Leafy) and FUL (Fruitfull) genes. However, at the young stage of plant development, miR156 (MicroRNA 156) keeps the concentration of its target genes SPLs in check as miR156 is abundantly present in the leaves when the plant is young, however the concentration of miRNA156 dissipates as the plant matures and so is the negative influence of miR156 on SPLs (Chen et al., 2010, Xu et al., 2016). Overexpression of miR156 in multiple crops (Maize, Rice, Tomato) has established that their role in flowering pathway is highly conserved (Chuck et al., 2007; Xie et al., 2006; Zhang et al., 2011). Interplay of key components of aging (SPLs), ambient temperature (SVP) and gibberellic acid (GA20ox) pathways stabilizes the floral induction process. Signals from multiple pathways converge to positively or negatively regulate the gene expression of key floral integrator genes like FT or SOC1.

FLC (Flowering Locus C) has been established as a major flowering repressor across multiple species, which suppresses flowering induction through negative regulation of key floral integrator genes FT, and SOC1. Flowering time variations exhibited by different varieties of Arabidopsis due to the different activity/concentration levels of FLC (Flowering locus C) demonstrates the importance of FLC in the regulation of the flowering pathway. Prior to flowering induction through vernalization pathway, a higher concentration of FLC is observed in the plants, which is generally regulated epigenetically via a combinatorial action of chromatin remodeling complexes like

SWR1c, PAF1c, COMPASS, HUB-UBC and FRIc (He, 2012). Moreover, FLC isn't a lone repressor, it forms a repressor complex along with a MADS box protein, SVP (Li et al., 2008) to suppress flowering. Epigenetic regulation of florigens and antiflorigens basically governs the fate of floral induction in the plants as the optimal concentrations of florigens (e.g., FT) and anti-florigens (e.g., FLC) decides whether a plant will flower or remain in the vegetative stage. Interestingly, Autonomous pathway components FCA (*FLOWERING CONTROL LOCUS A),* FY (*FLOWERING LOCUS Y), FPA*, *FLD (FLOWERING LOCUS D)*, *FVE/MIS4 (MULTICOPY SUPPRES- SOR OF IRA1 4)*, and REF6 (*RELATIVE OF EARLY FLOWERING 6*,) FLK (*FLOWERING LOCUS K)* repress the activity of floral repressor gene FLC (Flowering Locus C), via chromatin remodeling, RNA processing, and also through the collective effects of two PRCs (Polycomb repressive complex) PRC1 (Polycomb repressive complex 1) and PRC2 (Polycomb repressive complex 2) (Jiang et al., 2008). Plant species like Barley (*Hordeum vulgare*), wheat (*Triticum* spp.), and Arabidopsis, acclimated to a temperate environment needs a prolonged low temperature treatment (Vernalization) to exert a positive effect on flowering induction as it represses the transcriptional activity of FLC. However, Rice, a tropically domesticated crop does not require a low temperature treatment to induce flowering, and spring varieties of barley and wheat are also excused from this requirement of prolonged exposure to low temperature.

Vernalization

In some cases, antisense transcription of FLC initiates due to the low temperature, resulting in the production of a non-coding RNAs COOLAIR, COLDAIR and COLDWRAP which silences the FLC gene, and thus releases FT from the repressive effects of FLC (Fornara et al., 2010; Heo and Sung, 2011; Kim and Sung, 2017). Initial repression effect on FLC rendered by the low temperature treatment turns into stable repressive form when plants are returned to warmer temperatures. Histone modifying enzyme and chromatin remodeling complexes play a vital role in stably repressing FLC, and a gene, VIN3 (Vernalization Insensitive 3) is a significant player in this process which interacts with PRC2 (Polycomb repressive complex 2) to augment the chromatin repressive mark H3K27me3 (trimethylation of lysine 27 on histone 3) on FLC loci after the vernalization period (Heo and Sung, 2011). One another PHD-finger domain protein, VIL1/VRN5 have been implicated to work in a similar fashion, which

interacts with PRC2 like VIN3 to repress the expression of FLC to delay flowering. The components of vernalization pathway of temperate cereals, wheat and barley are different than that of the ones in Arabidopsis, and VRN (Vernalization) loci in these species have been exhibited to control responses from the vernalization pathway (Ream et al., 2012). VRN2, a CCT-domain protein functions as a floral repressor like FLC in Arabidopsis, and Ghd7 in rice, interestingly it shares a high sequence homology with the floral repressor in rice, Ghd7. Prior to vernalization, higher transcript levels of VRN2 are found in the plants which confer repressive effects on the florigen, FT/VRN3 and therefore, downregulation of this gene is imperative to initiate flowering induction process. During vernalization, augmentation in VRN1 (Vernalization 1) expression level coincides with the lower expression of VRN2 which releases VRN3/FT from the repressive effect of VRN2. Additionally, a loss of function mutation study showed that functional activity of VRN1 is required only to stabilize the repressive effect on FLC loci after vernalization (Chen and Dubcovsky, 2012). Additive repressive effects of PRC1 of FLC via chromatin modification further stabilizes the repressive state of FLC after the vernalization (Questa et al., 2016).

Similar to FT, function of FLC is fairly conserved among angiosperms (Ruelens et al., 2013). However, the closest homologue of Arabidopsis FLC in rice, OsMADS51 has an antagonistic function and it acts as an activator of flowering rather than a repressor unlike its counterpart in Arabidopsis (Kim et al., 2007). The instrumental work done in the past few decades has paved the way in unlocking the regulatory networks involved at transcriptional, post-transcriptional and post-translational level to accurately time the floral transition. These studies have also established that most of the components of the flowering pathway are evolutionarily conserved in angiosperms albeit with few exceptions. Flowering could prove to be boon or bane (Maize and Sugarcane respectively) depending on the end product of the plant species. For instance, delayed flowering or flowering repression in maize would result in yield loss due to prolonged vegetative developmental stage. Flowering is imperative for genetic improvement in sugarcane; However, flowering is not a desirable feature from a commercial production perspective as it results in yield loss. Flowering time optimization is a crucial factor for maximizing the crop yield and a deep understanding of flowering genes and their interacting partners at a molecular level is the key to optimize the flowering time. Biochemical and genetic studies in sugarcane might open avenues in identifying key flowering genes to delay or repress flowering and improve the crop yield.

2.2. Correlation of flowering genes with disease resistance or susceptibility under biotic or abiotic stress

Convergence of signals from multiple external and internal cues and a highly systematic regulatory network ensures the flowering induction. However, they are not the only means to induce flowering. Biotic and abiotic stresses can also induce flowering through its interaction with components of the flowering pathways (Wada and Takeno, 2010; Takeno et al., 2016). For instance, the developmental response of Arabidopsis plants susceptible to bacterial pathogens, *Pseudomonas syringae* and *Xanthomonas campestris*, and an oomycete, *Peronospora parasitica,* is to accelerate the floral transition. Initiation/acceleration of the reproductive development process to three phylogenetically different pathogens in Arabidopsis indicates the possibility of a global developmental response to pathogen infection. These changes in flowering time have also been suggested to be linked to the resistance/tolerance level of plants to pathogen infection (Korves and Bergelson, 2003).

Over the past couple of decades, a few reports have helped unravel a few components of this puzzle of stress-induced flowering. However, a complete understanding of how this complex network regulates its components remains elusive.

2.3. Flowering induction or repression under abiotic stress

Strikingly, different paralogs of flowering genes exert different responses under stress and normal conditions. For instance, under the nutrition stress, only one FT ortholog, the PnFT2, in a short-day plant *Pharbitis nil (Japanese morning glory)* expressed and induced flowering in the non-inductive long-day conditions. The other ortholog of FT, PnFT1 did not express, suggesting that not all the orthologs participate in the stress-induced flowering. The authors suggested the involvement of salicylic acid in stress-induced flowering, but that was not proven (Wada and Takeno, 2010). In Arabidopsis under nitrate limiting conditions, a precocious flowering phenotype was observed even in the double (*ft-7 soc1-1*) and triple mutant (*fca co-2 ga1-3*) lines from autonomous, photoperiod, gibberellic acid, and temperature pathways, indicating existence of a novel pathway that operates downstream of other general flowering pathways (Marin et al., 2011). Another short-day plant, *Lemna paucicostata* (synonym *Lemna aequinoctialis*) flowered early under poor-nutrition stress conditions than the control

plants, and like *Pharbitis nil,* stressed plants also exhibited higher levels of salicylic acid (Shimakawa et al., 2012).

Interestingly, other stresses, such as high-irradiation stress and low temperature, also triggered activation of the flowering pathway in the short-day plant, *Pharbitis nil* (Shinozaki et al., 1994; Ishimaru et al., 1996). SA-mediated precocious flowering under UV-C light stress was also reported in Arabidopsis. A lower expression level of the flowering repressor, FLC (Flowering locus C) coincided with the SA accumulation in the plants under UV-C light stress. In addition, higher expression levels of FLC in SA deficient plants were also observed, thus suggesting that SA-induced flowering occurs via downregulating the expression of FLC (Martinez et al., 2004). Drought/water scarcity triggers the upregulation of GI, FT, TSF (Twin Sister of FT), and SOC1 to induce early flowering in Arabidopsis under long days, and conversely, delays it under short days through the activation of flowering repressor SVP. This response requires assistance from ABA (abscisic acid), as the early response phenotype disappeared in ABA mutants, which indicates that drought triggers flowering induction in association with the photoperiod and abscisic acid pathways and GI plays a key central role in this response (Riboni et al., 2013).

However, the drought induced flowering response is probably species-specific, as GI does not seem to be involved in drought escape response in rice under short- or long-day conditions. Drought leads to repression of florigens Hd3a and RFT1 in rice, and gi and hd1 mutant lines did not exhibit any difference in the expression levels of florigens Hd3a and RFT1 in comparison to WT under drought. These results indicated that GI and Hd1 were not involved in drought induced flowering response in rice, unlike Arabidopsis (Galbiati et al., 2016). Another abiotic stress, which results in the delayed flowering phenotype is salt stress. Salt stress induces a repressive effect on the transcriptional activities of CO and FT, which affects the downstream action of these genes and delays flowering. Mutation lines of gi and co under salt stress did not exhibit the delayed flowering phenotype, suggesting that salt stress confers its effect through the photoperiod response pathway (Li et al., 2007; Kim et al., 2013). BFT (Brother of FT and TFL1) is a key component of salt stress-triggered repression/delay of flowering as indicated from the bft mutant lines, which exhibited no effect of salt stress. BFT generally interacts with a bZIP transcription factor FD (Flowering locus D), and this interaction of BFT with FD prevents the FT-FD dimer formation, which explains the delayed flowering phenotype under salt stress conditions (Ryu et al., 2014). Even at a

very low dosage, salt sensitive rice cultivars exhibit the same delayed flowering phenotype as Arabidopsis. However, the mechanism behind this strong response remains elusive in rice at a molecular level.

2.4. Effect of biotic stresses on the flowering pathway

Microbial pathogens (Fungi, viruses, and bacteria), insects, nematodes, and arachnids negatively impact the plant growth and development just like the abiotic stresses do. Interestingly, an interaction between Arabidopsis and *Fusarium oxysporum* unraveled a positive correlation between the late flowering phenotype and resistance to *Fusarium oxysporum*. Arabidopsis ecotypes susceptible to *F. oxysporum* exhibited precocious flowering, and conversely, resistant ecotypes flowered late. Furthermore, a gi mutant line infected with *F. oxysporum* showed a higher resistance level to *F. oxysporum* indicating the possibility of GI being a susceptibility factor or susceptibility promoter (Lyons et al., 2015). A couple of other studies have demonstrated a close relationship between plant defense and flowering time (Chin et al., 2013; Yang et al., 2014b). These studies have raised the speculations of the existence of shared components between flowering and defense pathways. Another flowering gene that corroborates this speculation is FLD (Flowering locus D), a putative histone demethylase, acts as a positive regulator of flowering by repressing the floral repressor, FLC. FLD is also an essential component for SAR (Systemic acquired resistance) signaling to increase disease resistance levels (Singh et al., 2013).

Similarly, PIE1 (PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1), a component of the SWR1 complex, also exhibits dual functionality of a negative regulator of flowering, and SAR (Banday and Nandi, 2015). Interestingly, Infection with *Pseudomonas syringae* triggered the transcriptional activity of FLD, which might be the promoting factor of pathogen-induced flowering (Singh et al., 2013).

GRP7 (Glycine rich RNA binding protein 7) plays a dual function: a negative regulator of plant immunity and a positive regulator of flowering pathway (Streitner et al., 2008; Nicaise et al., 2013). HopU1, a mono-ADP-ribosyltransferase is a type III secreted effector of *P. syringae*, which is imperative to acquire complete virulence on Arabidopsis. The loss of function of GRP7 makes it more sensitive to Pto DC3000 infection, and conversely, its overexpression leads to a higher resistance level to Pto

DC3000 infection. Furthermore, GRP7 directly interacts with the mRNAs of pattern recognition receptors, such as FLS2, EFR and HOP1, inhibiting this association and reducing levels of the FLS2 protein. Reduction in FLS2, in turn, results in lower production of ROS and even the callose production, and the significance of GRP7 in regulating expression levels of PRRs, which is the first line of the defense system in plants and makes it an important regulator of plant immune system (Nicaise et al., 2013). Furthermore, GRP7 plays a pivotal role in the flowering pathway whereby it downregulates the transcriptional activity of FLC, resulting in early flowering (Streitner et al., 2008). Moreover, GRP7 also regulates abiotic stress responses as overexpression of GRP7 leads to higher freezing tolerance in Arabidopsis (Kim et al., 2008).

2.5. Smut fungus and its association with flowering pathway

Smut fungi generally targets cereals and grasses species such as Maize, Sugarcane, Wheat, Sorghum, Rice and Barley. Smut fungi across these species mostly use the host floral structures to reproduce, and indeed in some instances, have been demonstrated to modulate the flowering pathway positively or negatively depending on the host species (Glassop et al., 2013; Fan et al., 2016; Schmitz et al., 2018). For instance, downregulation of flowering related genes in rice plants infected with smut fungi (*Ustilaginoidea virens*), and upregulation of flowering related genes in maize plants infected with smut fungus (*Ustilago maydis*) suggests that smut fungi can potentially repress or induce the flowering pathway in their corresponding host species (Fan et al., 2015; Schmitz et al., 2018). Rice false smut fungus, *Ustilaginoidea virens* infects flowers in rice and transforms into false smut balls (Fan et al., 2020) and to do so, it has developed a sophisticated system whereby it mimics the ovary fertilization process to take over the nutrient supply of the host plant (Song et al., 2016).

Ustilago maydis interferes with vegetative and reproductive developmental stages in maize plants by inducing tumor formation in vegetative and floral organs (Brefort et al., 2009). *Sporisorium reilianum,* a close relative of *S. scitamineum,* which causes head smut disease in maize, disrupts the flowering pathway resulting in the formation of phyllody (conversion of floral organs into the vegetative ones) in the inflorescence, and a loss of apical dominance in infected plants as well. Transcriptional profile of maize infected plants demonstrated that head smut fungus modulates the expression

behavior of flowering related genes to induce phyllody formation (Ghareeb et al., 2011).

Sporisorium scitamineum (Syd.) [(Piepenbring et al. (2002) *(Syn: Ustilago scitaminea H. and P. Sydow)*], is responsible for causing smut disease in sugarcane. Smut fungus in sugarcane grows systematically, and alters shoot apical meristem to produce whip-shaped sorus (Marques et al., 2017; Bhuiyan et al., 2021). For a long time, the smut whip in sugarcane was postulated to be a modified floral structure. However, a couple of investigative studies in the last decade (Glassop et al., 2014; Marques et al., 2017), has successfully established that smut whip is instead an elongated internode rather than a modified floral structure (Bhuiyan et al., 2021). Glassop et al. (2014) reported a rare incidence of combination of sugarcane flower and smut in the Ord River region in Australia and similar to the changes induced by floral transition, smut fungus infection also ceases the vegetative phytomers production. Smut infection in grass species occasionally results in phyllody formation or floral reversion (Ghareeb et al., 2011; Bhuiyan et al., 2021), which reflects a potential and perhaps a negative correlation of smut fungus with flowering pathway in the grass species, and since sugarcane is a grass species, speculating the existence of a similar relationship (depicting smut fungus interfering with flowering pathway/reproductive system) in smut-sugarcane interaction would not be a far-fetched idea.

Following study from our research group did provide some insights to support this idea-the transcriptional profiling study of smut-sugarcane compatible interaction in the smut intermediate resistant cultivar RB925345 reported the differential expression of essential flowering regulators, such as FT, AP1, AP3, AG, COL6, and VIN3 (Schaker et al., 2016). The differential expression of flowering-related genes and meristematic function genes suggests that smut fungus induces a transcriptional reprogramming right after infection, which may interfere with the flower development pathway, inducing the whip development later on. Thus, it seems that smut fungus induces the activation of the flowering pathway, and instead, it takes an alternative distinct genetic pathway, resulting in the whip formation instead of the flowers (Schaker et al., 2016). Results of this study indicate the possibility of a similar mechanism like stress/pathogen-induced flowering at play in sugarcane-smut interaction, albeit with a different result.

2.6. Conclusion

The flowering pathway is well mapped in the model crop, *Arabidopsis thaliana*, and a detailed comparison of flowering pathways across multiple species has unraveled the evolutionarily conserved nature of some of its constituents. This information has been instrumental in validating the orthologs of these components in a wide range of species across the plant kingdom. Floral transition, the developmental switch from vegetative to reproductive stage, is a decisive event in the plant life cycle to improve reproductive success. GI-CO-FT is the major regulatory module that precisely times the floral transition by converging signals from multiple pathways. In general, biotic or abiotic stress triggers the activation of the flowering pathway by regulating the transcription of flowering related genes. Under some stress conditions, the plant prematurely activates the flowering pathway to reproduce before succumbing to the injury inflicted by stress conditions. A detailed look at the two distinct genetic pathways that trigger flowering in normal and stress conditions unraveled that these two pathways share some components that are highly conserved. For instance, GI, a master regulator of flowering, is also a susceptibility factor under some stress conditions. A recent study indicated that a novel pathway, similar to the pathogen-induced flowering pathway, exists in the sugarcane like other plant species, generating a different response of developing a modified structure instead of a flower. Elucidation of this alternate and distinct genetic pathway at a molecular level may help control two economically important traits in sugarcane, flowering and smut whip development, both of which harm crop productivity.

References

Abdul-Awal SM, Chen J, Xin Z, Harmon FG (2020). A sorghum *gigantea* mutant attenuates florigen gene expression and delays flowering time. Plant Direct. 2020 Nov 13;4(11): e00281.

Adams S, Manfield I, Stockley P, Carré IA (2015). Revised morning loops of the Arabidopsis circadian clock based on analyses of direct regulatory interactions. *PLoS ONE* 2015, *10*, e0143943.

Andres F, Porri A, Torti S, Mateos J, Romera-Branchat M, Garcia-Martinez JL, Fornara F, Gregis V, Kater MM and Coupland G (2014). SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the Arabidopsis shoot apex to regulate the floral transition. Proc. Natl Acad. Sci. USA, 111, E2760–E2769.

Anwer MU, Davis A, Davis SJ, Quint M (2020). Photoperiod sensing of the circadian clock is controlled by EARLY FLOWERING 3 and GIGANTEA. *The Plant Journal* 101(6): 1397-1410.

Banday ZZ, and Nandi AK (2015). Interconnection between flowering time control and activation of systemic acquired resistance. Front Plant Sci. 2015 Mar 19; 6:174.

Bendix C, Mendoza JM, Stanley DN, Meeley R, Harmon FG (2013). The circadian clock-associated gene gigantea1 affects maize developmental transitions. Plant Cell Environ. 36, 1379e1390.

Bhuiyan SA, Magarey RC, McNeil MD, Aitken KS (2021). Sugarcane Smut, Caused by *Sporisorium scitamineum*, a Major Disease of Sugarcane: A Contemporary Review. Phytopathology. 2021 Nov;111(11):1905-1917.

Bouché F, Lobet G, Tocquin P, Périlleux C (2016). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*, *Nucleic Acids Research*, Volume 44, Issue D1, 4 January 2016, Pages D1167–D1171.

Brandoli C, Petri C, Egea-Cortines M, Weiss J (2020). Gigantea: Uncovering New Functions in Flower Development. Genes (Basel). 2020 Sep 28;11(10):1142.
Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R (2009). *Ustilago maydis* as a pathogen. Annu Rev Phytopathol 47: 423–445

Cha JY, Kim J, Kim TS, Zeng Q, Wang L, Lee SY, Kim WY, Somers DE (2017). GIGANTEA is a co-chaperone which facilitates maturation of ZEITLUPE in the Arabidopsis circadian clock. *Nat. Commun.* 2017, *8*, 3.

Chen A, and Dubcovsky J (2012). Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. PLoS Genetics 8: e1003134.

Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L (2010). SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. J Integr Plant Biol. 2010 Nov;52(11):946-51.

Cho LH, Yoon J, An G (2017). The control of flowering time by environmental factors. Plant J. 90, 708–719.

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007). The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. Nat. Genet. 39, 1517–1521.

Colasanti J and Coneva V (2009). Mechanisms of floral induction in grasses: something borrowed, something new. *Plant Physiol.* 149, 56–62. (doi:10.1104/pp.108.130500).

Davis SJ (2009). Integrating hormones into the floral-transition pathway of Arabidopsis thaliana. Plant, Cell Environ. 32, 1201–1210.

Doi K, Izawa T, Fuse T, et al. (2004). Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes and Development 18: 926 – 936.

Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognár L, Nagy F, Millar AJ, Amasino RM (2002). The ELF4 gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* 2002, *419*, 74–77.

Fan J, Guo XY, Li L, Huang F, Sun WX, Li Y, Huang YY, Xu YJ, Shi J, Lei Y, Zheng AP, Wang WM (2015). Infection of *Ustilaginoidea virens* intercepts rice seed formation but activates grain-filling-related genes. J Integr Plant Biol. 2015 Jun;57(6):577-90.

Fan J, Liu J, Gong ZY, Xu PZ, Hu XH, Wu JL, Li GB, Yang J, Wang YQ, Zhou YF, Li SC, Wang L, Chen XQ, He M, Zhao JQ, Li Y, Huang YY, Hu DW, Wu XJ, Li P, Wang WM (2020). The false smut pathogen *Ustilaginoidea virens* requires rice stamens for false smut ball formation. Environ Microbiol. 2020 Feb;22(2):646-659.

Fan J, Yang J, Wang YQ, Li GB, Li Y, Huang F, Wang WM (2016). Current understanding on *Villosiclava virens*, a unique flower-infecting fungus causing rice false smut disease. Mol Plant Pathol. 2016 Dec;17(9):1321-1330.

Farrona S, Coupland G, Turck F (2008). The impact of chromatin regulation on the floral transition. Semin. Cell Dev. Biol. 19, 560–573.

Fornara F, de Montaigu A, Coupland G (2010). SnapShot: control of flowering in Arabidopsis. Cell 141, 550–550.e2.

Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, Coupland, G (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev. Cell* 2009, *17*, 75–86.

Galbiati F, Chiozzotto R, Locatelli F, Spada A, Genga A, Fornara F (2016). Hd3a, RFT1 and Ehd1 integrate photoperiodic and drought stress signals to delay the floral transition in rice. Plant Cell Environ. 39, 1982–1993.

Ghareeb H, Becker A, Iven T, Feussner I, Schirawski J (2011). *Sporisorium reilianum* infection changes inflorescence and branching architectures of maize. Plant Physiol. 2011 Aug;156(4):2037-52.

Glassop D, Bonnett G, Croft B, Bhuiyan S, Aitken K, Rae A (2014). Flowering-related genes are not involved in the development of smut whip. In ‘Proceedings of the Australian society of sugar cane technologists, 28 April–1 May 2014, Broadbeach, Queensland’. Vol. 36. pp. 244–253. (Curran Associates, Inc.)

Glassop D, Rae AL, Bonnett GD (2013). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech* 16, 235–240 (2014).

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422, 719e722.

He Y (2012). Chromatin regulation of flowering. *Trends Plant Sci.* 17, 556–562.

Heo JB, and Sung S (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 331:76–79.

Imaizumi T, Tran HG, Swartz TE, Briggs WR, Kay SA (2003). FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis. *Nature* 2003, *426*, 302–306.

Inoue K, Araki T, Endo M (2017). Integration of input signals into the gene network in the plant circadian clock. Plant Cell Physiol. 58, 977–982.

Ishimaru A, Ishimaru K, Ishimaru M (1996). Correlation of flowering induced by low temperature and endogenous levels of phenylpropanoids in Pharbitis nil: a study with a secondary-metabolism mutant. J. Plant Physiol. 148, 672–676.

Itoh H, Nonoue Y, Yano M, Izawa T (2010). A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nature Genetics 42: 635 – 638.

Jiang D, Wang Y, Wang Y, He Y (2008). Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. PLoS One 3, e3404.

Jiang S, Xiao L, Huang P, Cheng Z, Chen F, Miao Y, Fu YF, Chen Q, Zhang XM (2019). Nucleoporin *Nup98* participates in flowering regulation in a *CONSTANS*-independent mode. Plant Cell Rep 38, 1263-1271.

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM (2007). The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. Plant Cell. 2007 Sep;19 (9): 2736-48. doi: 10.1105/tpc.107.054528. Epub 2007 Sep 21.

Kim DH (2020). Current understanding of flowering pathways in plants: focusing on the vernalization pathway in Arabidopsis and several vegetable crop plants. *Hortic Environ Biote* 61, 209-227.

Kim DH, and Sung S (2017). Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. Dev Cell 40:302-312.

Kim JS, Jung HJ, Lee HJ, Kim KA, Goh CH, Woo Y, Oh SH, Han YS, Kang H (2008). Glycine-rich RNA-binding protein 7 affects abiotic stress responses by regulating stomata opening and closing in Arabidopsis thaliana. Plant J. 2008 Aug;55(3):455-66.

Kim SL, Lee S, Kim HJ, Nam HG, An G (2007). OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1, OsMADS14, and Hd3a. Plant Physiology 145: 1484–1494.

Kim WY, Ali Z, Park HJ et al. (2013). Release of SOS2 kinase from sequestration with GIGANTEA determines salt tolerance in Arabidopsis. Nat. Commun. 4, 1352.

Kim WY, Fujiwara S, Suh SS, Kim J, Kim Y, Han L, David K, Putterill J, Nam HG, Somers DE (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* 2007, *449*, 356–360.

Koo BH, Yoo SC, Park JW, Kwon CT, Lee BD, An G, Zhang Z, Li J, Li Z, Paek NC (2013). Natural variation in OsPRR37 regulates heading date and contributes to rice cultivation at a wide range of latitudes. *Mol. Plant* 2013, *6*, 1877–1888.

Korves TM, and Bergelson J (2003). A developmental response to pathogen infection in Arabidopsis. Plant Physiology 133, 339–347.

Laubinger S, Marchal V, Gentilhomme J, Wenkel S, Adrian J, Jang S, Kulajta C, Braun H, Coupland G, Hoecker U (2006). *Arabidopsis* SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Develop- ment 133:3213–3222.

Li D, Liu C, Li Shen, Wu Y, Chen H, Robertson M, et al (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Dev Cell 2008; 15:110–20.

Li D, Yang C, Li X, Gan Q, Zhao X, Zhu L (2009). Functional characterization of rice OsDof12. Planta 229: 1159–1169.

Li K, Wang Y, Han C, Zhang W, Jia H, Li X (2007). GA signaling and CO/FT regulatory module mediates salt-induced late flowering in Arabidopsis thaliana. Plant Growth Regul. 53, 195–206.

Lifschitz E, Ayre B, Eshed Y (2014). Florigen and anti-florigen- a systemic mechanism for coordinating growth and termination in flowering plants. Front Plant Sci 2014:5.

Liu L, Zhu Y, Shen L, Yu H (2013). Emerging insights into florigen transport. Curr Opin Plant Biol. 2013 Oct;16(5):607-13.

Liu TL, Newton L, Liu MJ, Shiu SH, Farre EM (2016). A G-box- like motif is necessary for transcriptional regulation by circadian pseudo-response regulators in Arabidopsis. Plant Physiology 170:528–539.

Marın IC, Loef I, Bartetzko L, Searle I, Coupland G, Stitt M, Osuna D (2011). Nitrate regulates floral induction in Arabidopsis, acting independently of light, gibberellin and autonomous pathways. Planta, 233, 539–552.

Martínez C, Pons E, Prats G, León J (2004). Salicylic acid regulates flowering time and links defense responses and reproductive development. Plant J 2004; 37:209- 17.

Más P, Kim WY, Somers DE, Kay SA (2003). Targeted degradation of TOC1 by ZTL modulates circadian function in Arabidopsis thaliana. *Nature* 2003, *426*, 567–570.

Mutasa-Göttgens E, and Hedden P (2009). Gibberellin as a factor in floral regulatory networks. J. Exp. Bot. 60, 1979–1989.

Nicaise V, Joe A, Jeong BR, Korneli C, Boutrot F, Westedt I, Staiger D, Alfano JR, Zipfel C (2013). Pseudomonas HopU1 modulates plant immune receptor levels by blocking the interaction of their mRNAs with GRP7. EMBO J. 2013 Mar 6;32(5):701-12.

Piepenbring M, Stoll M, Oberwinkler F (2002). The generic position of *Ustilago maydis*, *Ustilago scitaminea*, and *Ustilago esculenta* (Ustilaginales). *Mycol Progress* 1, 71–80 (2002). https://doi.org/10.1007/s11557-006-0006-y.

Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O (2010). An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science 2010, 330:1397-1400.

Pokhilko A, Fernández AP, Edwards KD, Southern MM, Halliday KJ, Millar AJ (2012). The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* 2012, *8*, 574.

Putterill J, and Varkonyi-Gasic E (2016) FT and florigen long-distance flowering control in plants. Curr Opin Plant Biol 33: 77–82.

Questa JI, Song J, Geraldo N, An H, Dean C (2016). Arabidopsis transcriptional repressor VAL1 triggers polycomb silencing at FLC during vernalization. Science 353:485–488.

Ream TS, Woods DP, Amasino RM (2012). The molecular basis of vernalization in different plant groups. Cold Spring Harbor Symposia on Quantitative Biology 77: 105–115.

Riboni M, Galbiati M, Tonelli C, Conti L (2013). GIGANTEA enables drought escape response via abscisic acid-dependent activation of the florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1. Plant Physiol. 162, 1706–1719.

Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nature Communications 4: 2280.

Ryu JY, Lee HJ, Seo PJ, Jung JH, Ahn JH, Park CM (2014). The Arabidopsis floral repressor BFT delays flowering by competing with FT for FD binding under high salinity. Mol. Plant, 7, 377–387.

Chin K, DeFalco TA, Moeder W, Yoshioka K (2013). The Arabidopsis cyclic nucleotide-gated ion channels AtCNGC2 and AtCNGC4 work in the same signaling pathway to regulate pathogen defense and floral transition. Plant Physiol. 2013; 163: 611–24.

Sawa M, and Kay SA (2011). GIGANTEA directly activates Flowering Locus T in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 2011, *108*, 11698–11703.

Sawa M, Nusinow DA, Kay SA, Imaizumi T (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science* 2007, *318*, 261–265.

Schaker PDC, Palhares AC, Taniguti LM, Peters LP, Creste S, Aitken KS, Van Sluys MA, Kitajima JP, Vieira MLC, Monteiro-Vitorello CB (2016). RNAseq transcriptional profiling following whip development in sugarcane smut disease. PLoS One E 11 (9), e0162237.

Schaker PDC, Peters LP, Cataldi TR, Labate CA, Caldana C, Monteiro-Vitorello CB (2017). Metabolome Dynamics of Smutted Sugarcane Reveals Mechanisms Involved in Disease Progression and Whip Emission. Front Plant Sci. 2017 May 31; 8:882.

Schmitz L, McCotter S, Kretschmer M, et al (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. **F1000Research** 7 (F1000 Faculty Rev):1812.

Searle I, and Coupland G (2004). Induction of flowering by seasonal changes in photoperiod. EMBO J 23:1217–1222.

Shimakawa A, Shiraya T, Ishizuka Y, Wada KC, Mitsui T, Takeno K (2012). Salicylic acid is involved in the regulation of starvation stress-induced flowering in Lemna paucicostata. J Plant Physiol. 2012 Jul 1;169(10):987-91.

Shinozaki M, Hirai N, Kojima Y, Koshimizu K, Takimoto A (1994). Correlation between level of phenylpropanoids in cotyledons and flowering in Pharbitis seedlings under high-fluence illumination. Plant Cell Physiol. 35, 807–810.

Shrestha R, Gómez-Ariza J, Brambilla V, Fornara F (2014). Molecular control of seasonal flowering in rice, *Arabidopsis* and temperate cereals. *Ann. Bot.* 114, 1445–1458 (2014).

Singh V, Roy S, Giri MK, Chaturvedi R, Chowdhury Z, Shah J, Nandi AK (2013). Arabidopsis thaliana FLOWERING LOCUS D is required for systemic acquired resistance. Mol Plant Microbe Interact. 2013 Sep; 26(9): 1079-88.

Singh V, Roy S, Giri MK, Chaturvedi R, Chowdhury Z, Shah J, Nandi AK (2013). Arabidopsis thaliana FLOWERING LOCUS D is required for systemic acquired resistance. Mol Plant Microbe Interact. 2013 Sep;26(9):1079-88.

Song JH, Wei W, Lv B, Lin Y, Yin WX, Peng YL, Schnabel G, Huang JB, Jiang DH, Luo CX (2016) Rice false smut fungus hijacks the rice nutrients supply by blocking and mimicking the fertilization of rice ovary. Environ Microbiol. 2016 Nov;18(11):3840-3849.

Streitner C, Danisman S, Wehrle F, Schoning JC, Alfano JR, Staiger D (2008). The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. Plant J 56:239–250.

Swain SM, Tseng T, Olszewski NE (2001). Altered expression of SPINDLY affects gibberellin response and plant development. *Plant Physiol.* 2001, *126*, 1174–1185.

Takeno Kiyotoshi (2016). Stress-induced flowering: the third category of flowering response, *Journal of Experimental Botany*, Volume 67, Issue 17, September 2016, Pages 4925–4934.

Tang W, Yan H, Su Z, Park SC, Liu Y, Zhang Y, Wang X, Kou M, Ma D, Kwak SS et al (2017). Cloning and characterization of a novel GIGANTEA gene in sweet potato. *Plant Physiol. Biochem.* 2017, *116*, 27–35.

Taoka K, Ohki I, Tsuji H, Furuita K, Hayashi K, Yanase T, et al (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature 2011; 476:332–5.

Tiwari SB, Shen Y, Chang HC, Hou YL, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ (2010). The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. New Phytol 187:57–66.

Tseng TS, Salomé PA, McClung CR, Olszewski NE (2004). SPINDLY and GIGANTEA interact and act in *Arabidopsis thaliana* pathways involved in light responses, flowering, and rhythms in cotyledon movements. *Plant Cell* 2004, *16*, 1550–1563.

Wada KC, and Takeno K (2010). Stress-induced flowering. Plant Signal Behav. 2010 Aug;5(8):944-7.

Wang T, Sun MY, Wang XS, Li WB, Li YG (2016). Over-expression of GmGIa-regulated soybean miR172a confers early flowering in transgenic Arabidopsis thaliana. *Int. J. Mol. Sci.* 2016, *17*, 645.

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K et al (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 2011, *188*, 395–407.

Wolabu TW, Zhang F, Niu L, Kalve S, Bhatnagar-Mathur P, Muszynski MG, Tadege M (2016). Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. New Phytol 2016; 210:946-59.

Xie K, Wu C, Xiong L (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant Physiol. 142, 280–293.

Xu D, Zhu D, Deng XW (2016). The role of COP1 in repression of photoperiodic flowering. **F1000Res**. 2016 Feb 16;5: F1000 Faculty Rev-178.

Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L and Poethig RS (2016). Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in Arabidopsis thaliana. PLoS Genet. 12: e1006263.

Yang S, Weers BD, Morishige DT, Mullet JE (2014). CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biology 14: 148.

Yang Y, Ma C, Xu Y, Wei Q, Imtiaz M, Lan H, et al (2014a). A zinc finger protein regulates flowering time and abiotic stress tolerance in chrysanthemum by modulating gibberellin biosynthesis. Plant Cell. 2014; 26: 2038–2054.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, et al (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 12: 2473–2483.

Zhang X, Zou Z, Zhang J, Zhang Y, Han Q, Hu T, Xu X, Liu H, Li H and Ye Z (2011). Over-expression of sly-miR156a in tomato results in multiple vegetative and reproductive trait alterations and partial phenocopy of the sft mutant. FEBS Lett. 21, 435–439.

Zhang Z, Hu W, Shen G, Liu H, Hu Y, Zhou X, Liu T, Xin Y (2017). Alternative functions of *Hd1* in repressing or promoting heading are determined by *Ghd7* status under long-day conditions. Sci. Rep. 7: 5388.

Zhao XY, Liu MS, Li JR, Guan CM, Zhang XS (2005). The wheat TaGI1, involved in photoperiodic flowering, encodes an Arabidopsis GI ortholog. *Plant Mol. Biol.* 2005, *58*, 53–64.

3. RNASeq PROFILING UNRAVELED AN ANTAGONISTIC EXPRESSION BEHAVIOR PATTERN OF FLOWERING RELATED GENES IN SUGARCANE SMUT RESISTANT AND SUSCEPTIBLE GENOTYPES IN AN EARLY RESPONSE STUDY

Abstract

Sugarcane smut, caused by a basidiomycete biotrophic fungus (*Sporisorium scitamineum*), is one of the most predominant diseases in sugarcane. A better understanding of molecular crosstalk between smut fungus *S. scitamineum* and its host sugarcane is the key to finding a better management solution for the smut disease. Plants activate the flowering pathway under some stress conditions to reproduce before it succumbs to the injury caused by the stressor. Interestingly, smut fungi generally target floral structure and negatively or positively regulate flowering pathways depending on the host species. In order to better understand the correlation of flowering pathway, and whip development pathway, we generated transcriptional profiling (RNASeq) data for smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) genotypes 48 hours after inoculation. Our data analysis from this early response study unraveled a complex yet distinct association between the floral transition and defense system in the sugarcane-smut interaction. Furthermore, the comparative transcriptional profiling showed that the smut fungus-induced transcriptional reprogramming elicits an antagonistic expression pattern of flowering related genes in smut-resistant and smut-susceptible genotypes. This expression pattern suggests that smut-resistant genotype (SP80-3280) attempts to maintain its meristematic identity and keep the plant in the vegetative stage, whereas smut-susceptible genotype (IAC66-6) seems to positively influence flowering right after infection. We also reported that smut fungus seems to preferentially interact with the genes belonging majorly to autonomous, photoperiod and vernalization pathways 48 hai. This is the first study demonstrating a distinct expression behavior pattern of flowering related genes in two different sugarcane genotypes with contrasting smut resistance levels and warrants a more detailed time course study to depict shared and unshared components of flowering and smut whip development pathways in sugarcane.

3.1. Introduction

Sugarcane (*Saccharum spp.*) is a monocotyledonous plant of vegetative propagation, belonging to the Poaceae family, and is a major agricultural and industrial cash crop grown in more than 100 countries in tropical and subtropical regions of the world. Sugarcane is one of the best crops with bioenergy potential as it has more favorable traits than other crops, like rapid growth, high productivity, and the ability to survive in harsh conditions. Moreover, second-generation bioethanol produced through cellulose depolymerization of sugarcane bagasse is a more sustainable approach than the traditional methods (Mamo et al., 2013; Chami et al., 2020). Brazil is currently the largest producer of sugarcane, with an equivalent output of 40% of the global stock,

and the state of São Paulo is responsible for about 50% of the national production (FAOSTAT, 2021; CONAB, 2021). Despite many environmental issues, the last harvest estimated for 2020/2021 reached 654.5 million tons, representing an increase of 1.8% compared to the previous one and a reduction of 0.6% in the cultivated area (CONAB, 2021).

Sugarcane, one of the most robust and economically important agricultural cash crops, is also a home to many pathogens including a biotrophic fungal pathogen, *Sporisorium scitamineum*, the causal agent of smut disease in sugarcane (Sundar et al., 2012). Smut fungus invasion in the sugarcane defense system results in premature transcriptional reprogramming leading to a whip-like structure emerging from the shoot apical meristem and lateral buds of the infected sugarcane plant (Schaker et al., 2016). Structurally, whips are composed of both fungal and plant tissues, where a black mass of fungal teliospores covers the parenchymal and vascular tissues of the host plant. Whip-shaped sorus are responsible for fungal reproduction (Marques et al., 2017). During the initial developmental stages, whips are covered with a thin silvery membrane and once the teliospores mature, this membrane ruptures, resulting in the dispersion of black powdery mass of teliospores in the field (Sundar et al., 2012). After dispersion, diploid teliospores germinate on the host surface to initiate meiosis, resulting in the production of haploid sporidia, which in turn results in plant colonization once the mating compatible sporidial cells generate infective dikaryotic hyphae, a crucial structure required to infect sugarcane (Bakkeren et al., 2008; Marques et al., 2017). Finally, the dikaryotic hyphae differentiate into an appressorium and penetrate the host tissue (Peters et al., 2017).

Smut disease results in yield loss by compromising juice quality, costing farmers anywhere 10-80% of their crops (Monteiro-Vitorello et al., 2018). Furthermore, it results in higher fiber content, reduced cane diameter, lower sucrose content and tillering as well. However, the gravity of loss is dependent on several factors like environmental conditions and the tolerance level of variety in question (Sundar et al., 2012; Monteiro-Vitorello et al., 2018). Smut has made its presence felt in all the sugarcane growing countries in the world except Fiji (Tom et al., 2017). Modern sugarcane cultivars' complex genetic makeup (2n = 100–130) makes breeding for quantitative resistance extremely difficult (D'Hont et al., 1996; Garcia et al., 2013; de Setta et al., 2014).

Understanding molecular crosstalk between host sugarcane and smut fungi *S. scitamineum* has been the primary research focus of research groups across the globe ever since the idea of a chemical resistance mechanism in sugarcane drifted (James, 1973). In next couple of decades, the discovery of flavonoids inhibiting teliospore germination, the importance of glycosidic substances in the resistance/tolerance level of sugarcane smut disease, and detection of different higher levels of conjugated polyamines to the phenolics in smut infected plants of resistant and susceptible genotypes furthered our understanding of smut resistance mechanism (Lloyd and Pillay, 1980; Lloyd and Naidoo, 1983; Piñon et al*.,* 1999). Several techniques like microarray, cDNA-AFLP (complementary DNA-amplified fragment-length polymorphisms), SSH (suppression-subtractive hybridization) based sequencing, and RNASeq have been employed to analyze the host gene expression after fungal inoculation (Schaker et al., 2016; Que et al., 2014; Wu et al., 2013; Borrás-Hidalgo et al*.,* 2005; Thokoane and Rutherford, 2001).

Published and unpublished data from functional genomics, transcriptomics, proteomics, and metabolomics studies from our lab, collaborators, and other research groups around the globe have contributed significantly in terms of understanding plant defense mechanisms and metabolic changes after pathogen attack (Que et al., 2014; Taniguti et al., 2015; Su et al., 2016; Schaker et al., 2016; Peters et al., 2017; Schaker et al., 2017, Teixeira-Silva et al., 2020; Singh et al., 2021). Interestingly, we observed some of the key floral integrator genes like AP1, COL6, and FT differentially expressed in comparative transcriptional profiling of smut infected and mock plants (Schaker et al., 2016), which navigated our research work to focus on the transcriptome profile of flowering pathway genes in the smut susceptible and resistant genotypes. Flowering is yet another undesirable feature for commercial sugarcane varieties like smut and other diseases as it limits the crop yield as well. Furthermore, flowering and whip bring out anatomical changes in sugarcane via differentiation of shoot apical meristem, resulting in reproductive structures, albeit for sugarcane and smut fungus respectively. Schaker et al. (2016), also proposed that transcriptional reprogramming in meristematic functions probably avert flower development.

To better understand the correlation between whip development and flower development, we investigated the modulation in expression behavior of flowering-related genes via RNASeq in smut susceptible (IAC66-6) and smut resistant (SP80-

3280) genotypes 48 hours after inoculation (48 hai) with smut fungus. In addition, since the flowering pathway is not well determined in sugarcane, we used a well-established model system of floral organ development in Arabidopsis to identify sugarcane orthologs of flowering time.

This work corroborates findings of Schaker et al. (2016) that flowering development and whip development pathways share some components. The main conclusions of this work are: a) transcriptional reprogramming of flowering-related genes in smut infected plants in the resistant genotype SP80-3280 shortly after the infection seems to regulate the flowering pathway negatively; b) transcriptional reprogramming of flowering-related genes in smut infected plants in a the susceptible genotype IAC66-6 shortly after the infection seems to positively regulate the flowering pathway, which corroborates the finding that susceptible Arabidopsis plants accelerate the vegetative-to-reproductive transition in response to pathogens (Korves et al., 2003); c) to our best knowledge, this is the first study reporting the potential role of chromatin remodeling complexes in Sugarcane's defense response to smut fungus.

3.2. Material and Methods

3.2.1. Biological sample collection

Buds were collected from 10 months old plants of smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) genotypes of sugarcane grown at the experimental field of the Department of Genetics, ESALQ, University of São Paulo, Piracicaba. Single-bud setts were prepared and subsequently disinfected by heat and chemical treatments (thoroughly rinsed with water; kept at 52°C in water bath for 30 minutes; submerged in 0.01% sodium hypochlorite solution for 10 minutes and finally rinsed in distilled H_2O thrice). Disinfected single-bud setts were kept in a tray on a moist filter paper bed, covered with Aluminum foil, and transferred to be incubated at 28°C overnight to stimulate germination. Germinated buds were inoculated with the teliospores of sugarcane smut fungus (*S. scitamineum*) strain Ssc04 (viability rate > 80%) suspended in saline solution. 5µl of teliospore (10^9 teliospores mL^{-1}) solution was administered directly over each germinated bud (Peters et al., 2017), kept at room temperature afterwards for 10 minutes for air-drying, and finally transferred the trays to incubate at 28°C for 48 hours in a dark chamber after covering the buds with vermiculite. Three biological replicates, each comprising a pool of 5 buds from inoculated and mock

inoculated (control) samples from each genotype, were collected 48 hours after inoculation (48 hai).

3.2.2. RNA extraction and RNASeq library preparation

RNA was extracted using Trizol® (Invitrogen) with the PureLink™ RNA Mini Kit (Invitrogen) as per kit manual instructions. RNA integrity was verified by running the extracted RNA samples on agarose gel electrophoresis, stained with SYBR safe (Thermo Fisher Scientific, Oregon, USA). RNA quantity (A_{260}/A_{280}) and quality (A_{260}/A_{230}) parameters were assessed by using Nanodrop® 2000 spectrophotometer (Thermo Fisher Scientific Inc.). Following which, RNASeq libraries were prepared with the "TruSeq RNA Sample Prep v2 Low Throughput (LT)" (Illumina) as per kit manual instructions and sent for paired-end sequencing in a HiScanSQ platform (Illumina).

3.2.3. Transcriptomics data preprocessing and mapping

The transcriptomic data were analyzed before (Rody et al., 2019). Shortly, the reads counting table was obtained with the FeatureCounts v1.6.0 tool from the Subread package (Liao et al., 2014). A Gene Transfer Format (GTF) file was prepared for the set of COMPGG transcripts from the fasta file. Only genes with CPM values greater than one in at least three out of six replicates were considered expressed.

3.2.4. Data collection and reference dataset

Paired-end sequenced transcriptomic raw data, for IAC66-6 and SP80-3280 genotypes, were procured from NCBI BioProject PRJNA546134 (Rody et al., 2019). In this study, we referred to the two targeted transcriptome experiments as: SP for the SP80-3280 48 hai, and IAC for the IAC66-6 48 hai. For this study, the reference dataset COMPGG used, was established by Schaker et al., (2017) and comprised two different de novo assembled transcripts sequence datasets. The first dataset, i.e., the COMP dataset, contained 72,269 unique transcripts from six different sugarcane genotypes (Cardoso-Silva et al., 2014). The second dataset, i.e., the GG, comprised 16,219 unique transcript sequences from a previous study in our lab (Schaker et al., 2016).

3.2.5. Differential expression and Network analysis

Differentially Expressed Genes (DEGs) were obtained using the EdgeR v3.30.3 from Bioconductor R (R Core Team, 2020, version 4.0.2) package (McCarthy et al., 2012; Robinson et al., 2010) with default parameters. Only genes with CPM values greater than one in at least three out of six replicates were considered as expressed to prevent RNA-seq artifacts. Genes with $p < 0.05$ were identified as DEGs for the comparison of inoculated/control samples.

In short, the WGCNA software v1.69 (Langfelder and Horvath, 2008) and the pipeline used to construct the genes co-expression networks were according to the step-by-step network construction and module detection tutorial made available with WGCNA (https://github.com/hugorody/RNAseq/blob/master/WGCNA-stepbystep-pipe.R). Finally, the network modules were exported to Cytoscape v3.8 format (Shannon et al., 2003) for visualization and downstream analysis. In addition, the Molecular Complex Detection (MCODE) algorithm (Bader and Hogue, 2003) was used to filter modules densely connected for various functional categories in either the SP or IAC experiments. In this chapter, modules harboring flowering-related genes were depicted considering it had at least one differentially expressed positive or negative regulator of flowering in either of the two sugarcane genotypes, SP80-3280 and IAC66-6. Interacting partners of flowering related genes were identified using STRING database via text mining, and the modules harboring differentially expressed interacting partners were also filtered out. Sequences were manually inspected for orthology.

3.2.6. Arabidopsis flowering-related genes database and selection of sugarcane orthologs

A total of 306 gene identifiers of *A. thaliana* flowering time genes from the FLOR-ID database (Bouché et al., 2016) were procured. Corresponding protein sequences for *A. thaliana* gene identifiers were obtained from the TAIR11 database (Berardini et al., 2015). Sugarcane orthologs of flowering time were predicted through BLASTp searches using the COMPGG largest ORFs sequences as queries against a local assembled BLAST database having the *A. thaliana* Flor-ID sequences, with a cutoff of e^{-05}, minimum of 40% of identity, and minimum of 80% of query coverage. Sugarcane COMPGG sequences that passed BLAST filters were used to assemble the FLOR dataset (Appendix A).

3.2.7. RNA-Seq data validation through quantitative real-time PCR

Differentially expressed flowering-related genes data obtained via RNASeq study was validated through quantitative PCR analysis. Targeted candidate gene sequences were parsed from the COMPGG dataset, and used as input to design primers using Primer3plus online primer designing tool (https://www.bioinformatics.nl) (Untergasser et al., 2007). Default parameters were used for primer size, primer Tm and GC % ratio. Primer specificity was verified again by blasting the primer sequences against the COMPGG dataset. Finally, the quality of forward and reverse primers was verified with NetPrimer (http://www.premierbiosoft.com/NetPrimer/). Because of sugarcane genome polyploidy complexities, we first tested each designed pair of primers by conventional PCR reactions, running in agarose gel stained with SYBR Safe (Thermo Fisher Scientific, Oregon, USA). A total of five candidate genes were tested. Protein arginine methyltransferase 10 (ATPRMT10), Brahma (BRM), Early in short days (ESD7), Splayed (SYD) and Trehalose-6-phosphate synthase (TPS1) were selected for RT-qPCR analysis (Table 1). All RT-qPCRs were performed in the 7300 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA) using GoTaq® qPCR Master Mix kit (Promega, Madison, WI, USA). The GoTaq® 2-step RT-qPCR system kit was used for cDNA synthesis (Promega, Madison, WI, USA). 1 µg of total RNA and Oligo(dT) primers were used for each cDNA reaction as per kit manual instructions. The qPCR master mix with a total volume of 12.5 µl containing 2 µl of 8 times diluted cDNA sample, 6.25 µL of GoTaq® qPCR Master Mix, 0.125 µL CXR reference dye, 0.25 µL of each primer (0.2 µM), and 3.625 µL nuclease-free water was used for each of the three biological replicates and two technical replicates. qPCR cycling conditions were as follows: 95 °C for 02 min; 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and a dissociation curve was obtained for each candidate gene and reference gene primer used in this study to verify the primer specificity (Appendix B).

Sugarcane housekeeping genes encoding for **UBQ1** (Ubiquitin1), **GAPDH** (d-glyceraldehyde-3-phosphate dehydrogenase), **eEF1α** (eukaryotic elongation factor 1**α)** (Iskandar et al., 2004; de Andrade et al., 2017; Huang et al., 2018) and additionally **SRO1** (from this study) were used to normalize the gene expression levels for both the genotypes, SP and IAC. The LinReg PCR program was used to calculate PCR efficiencies and Cq values (Ramakers et al., 2003). Relative changes in the gene

expression levels were calculated with REST software (Pfaffl et al., 2002). Control samples were considered as the calibrators and the student *t*-test was applied to calculate significant changes in the relative expression levels ($p < 0.05$). Expression ratios were calculated by the ΔCt method normalized with the endogenous gene expression and REST software was used to statistically analyze the RT-qPCR results.

Table 1: Primers designed for RT-qPCR validation

Gene	Transcript ID	Forward Primer sequence 5'-3'	Reverse Primer Sequence 5'-3'
UBQ1	Andrade et al 2017	GTCGGGTCCATTCCATCACA	AGCAAAGCATCCTGGCATGT
eEF1a	Huang et al 2018	TTTCACACTTGGAGTGAAGCAGAT	GACTTCCTTCACAATCTCATCATAA
GAPDH	Iskander et al 2004	CACGGCCACTGGAAGCA	TCCTCAGGGTTCCTGATGCC
SRO1	comp202214_c0_seq1	TCAGAGGTGCGGAATACG	CTGCGATGAACGGAAAGG
ATPRMT10	comp184054_c0_seq1	ACTGGCTGGATGGTTTGATG	AATACCTGCTGACCCCAATG
BRM	comp199879_c0_seq1	TAGTGAAGGACGGGAAATCG	AGAAAGTGGTTCAGCGGATG
ESD7	comp184140_c0_seq1	AAGGTGGATGACGAGAAACG	TCCAGCCCTCCAATACAAAG
SYD	comp207886_c1_seq1	TGTTGTGGTGGCAGATAAGC	CATCAATGGCAGCAGAAGTG
TPS1	comp198109_c0_seq1	ACGGCAGTGTTCGGTAAATC	AGATGTTCGCTGATGTCGTG

3.3. Results

In order to evaluate the impact of smut infection on genes involved in floral transition/meristematic functions, we procured all the flowering related genes (306 genes) from FLOR-ID (Bouché et al., 2016), a flowering interactive database of *Arabidopsis thaliana*, and filtered out the expression profile of their respective putative orthologs (with cut-off parameters of minimum 80% of query coverage and 40% of identity percentage) in sugarcane (Appendix C). As a result, sugarcane presented orthologs for 277 flowering related genes, and the average sequence identity was 68 percent. Out of these, 241 flowering orthologs expressed at least in one of the two transcriptome profiles generated for smut-resistant and smut-susceptible genotypes.

FLOR-ID also categorized all the genes in the flowering pathways based on their effects (positive or negative) on flowering. Out of 306 flowering-related genes, 36 (19 in SP80-3280, and 17 in IAC66-6) genes were differentially expressed (FDEGs) in our dataset, which showed a distinct pattern (Figure 1). Only two FDEGs, TEM2/RAV2 and LRB1, were in both resistant and susceptible genotypes. However, they had opposite expression patterns in the resistant (SP80-3280) and susceptible (IAC66-6) genotypes. Therefore, the resistant genotype had 17 and the susceptible genotype had

15 uniquely differentially expressed genes (Figure 1). All the flowering time genes from Arabidopsis regulate flowering pathway either positively or negatively, and hence, are defined as positive and negative regulators respectively on the flowering interactive database (FLOR-ID, Bouché et al., 2016). The expression behavior of their corresponding orthologs in sugarcane was sought in resistant and susceptible genotypes (Table 2 and 3).

The negative regulators of flowering like Brahma/BRM (comp199879_c0_seq1), Early in short days 7/ESD7 (comp207801_c0_seq1), Curly leaf/ Set domain group 1 (comp200190_c1_seq1), Photoperiod Independent early flowering 1/PIE (comp205700_c0_seq1), Ubiquitin-specific protease 12 (comp204750_c2_seq1), Tempranillo 2/TEM2 (gg_06658) were up-regulated and the positive regulators of the flowering Light response BTB1/LRB1 (gg_03849), Protein arginine methyltransferase 10/ATPRMT10 (comp184054_c0_seq1), Trehalose-6-phosphate synthase/TPS 1 (comp198109_c0_seq1), UDP-glycosyltransferase 87A2 (comp202963_c1_seq1) were down-regulated in the resistant genotype SP80-3280 (Table 2).

Differentially gene expression analysis also revealed that positive regulators of the flowering; Glycine Rich RNA binding protein 7/Cold circadian rhythm RNA binding protein 2 (GRP7/CCR2; comp165891_c0_seq1), Light-response BTB1/LRB1 (gg_03849), Gigantea/GI (gg_12613), ADP-Glucose Pyrophosphorylase 1/ADG1 (gg_15551/comp172176_c0_seq1) were up-regulated and the negative regulators of the flowering; Tempranillo 2/TEM2 (gg_06658), Gibberellin 2-oxidase/GA2ox8 (comp202287_c0_seq1) were downregulated in susceptible genotype IAC66-6 (Table 3).

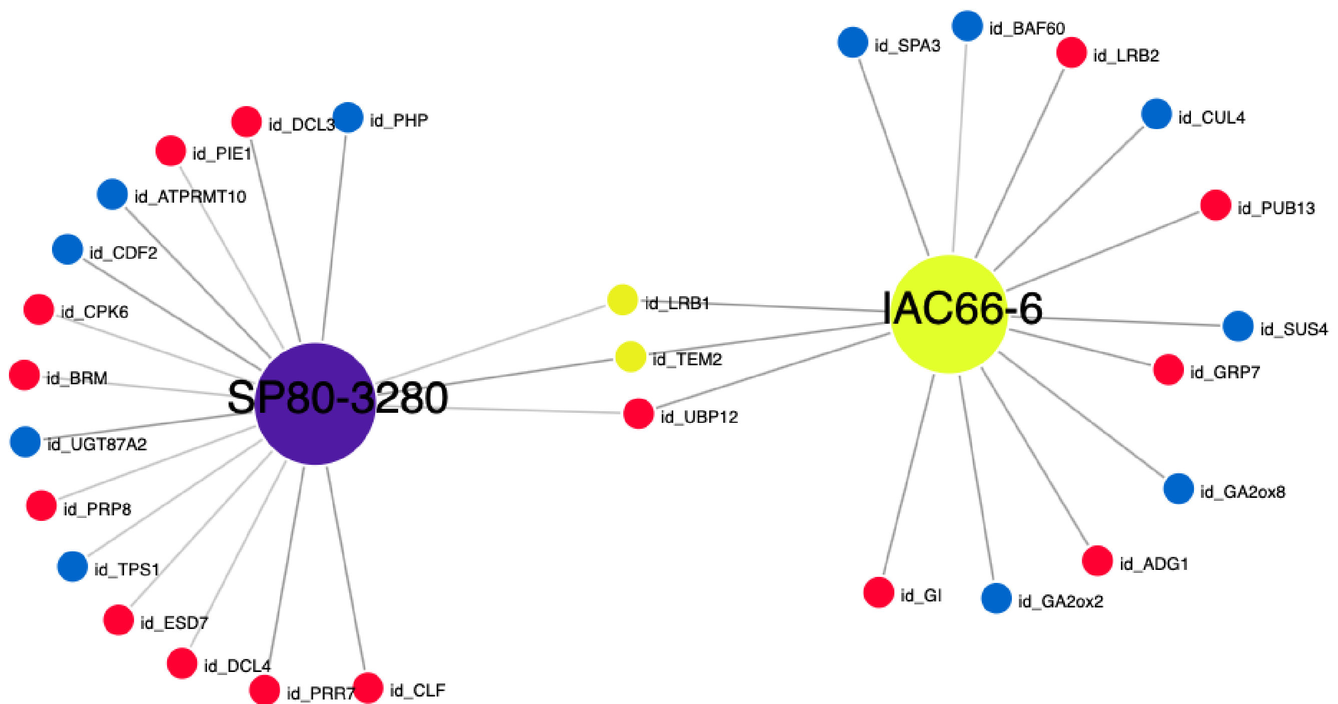


Figure 1: DiVenn (Liang et al., 2019) diagram depicting expression behavior of flowering-related genes from FLOR-ID in resistant genotype (SP80-3280) and susceptible genotype (IAC66-6). Up-regulated genes are highlighted in red, down-regulated in blue, and differentially expressed genes common in two genotypes are highlighted in yellow.

3.3.1. Co-expression networks of flowering-related genes

FDEGs from resistant (19 genes) and susceptible genotypes (17 genes) were clustered into different modules depending on their level of expression similarity. All the FDEGs from the resistant genotype were clubbed in nine different modules (Figure 2) However, only three (*blue2, brown2, lightpink1*) modules included most FDEGs in the resistant genotype SP80-3280.

All the 17 FDEGs from the susceptible genotype were clustered into seven different modules (Figure 2). Like the resistant genotype, most FDEGs were included in three modules (*darkorchid4*, *lightcoral* and *orchid3)* in the susceptible genotype IAC66-6 (Figure 2). Interestingly, the co-expression network modules also demonstrated a clear pattern of FDEGs being assembled into modules depending on whether they regulate flowering pathway positively or negatively. Also, the majority of these genes had the same expression behavior, i.e., up-regulated or downregulated. For instance, the differentially expressed positive regulators (GI, LRB1, PUB13, ADG1/APS1) were co-expressed in the module *darkorchid4* in the susceptible genotype IAC66-6, and all of them were up-regulated (Table 3). Another example from the susceptible genotype is

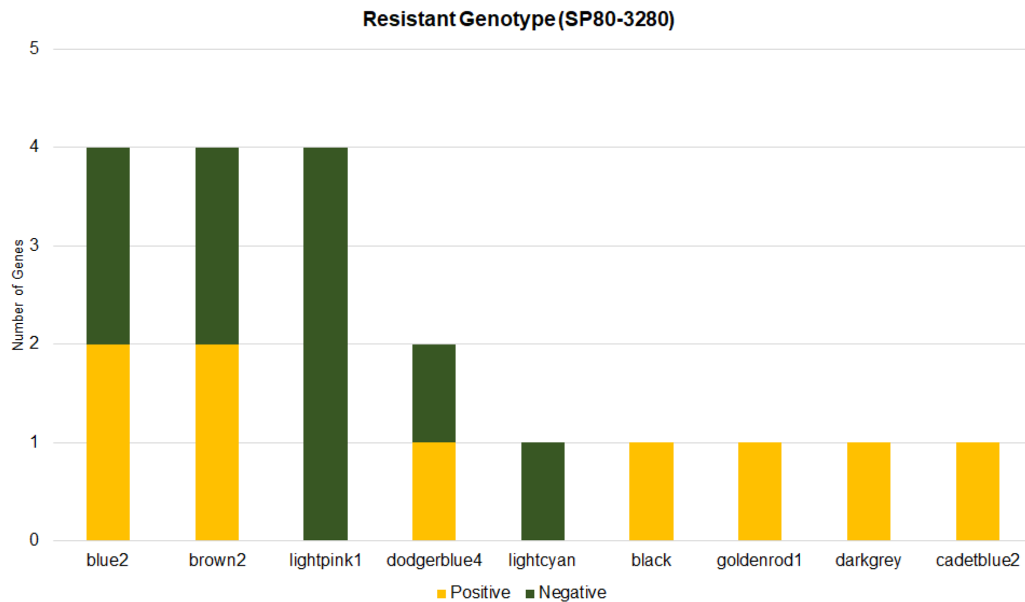
the orchid3 module, where negative regulators GA2ox8, TEM2/RAV2 and GA2ox2 were co-expressed, and all of them were downregulated. One other module which demonstrated a similar pattern was *lightpink1,* where three differentially expressed negative regulators (ESD7, CLF, UBP12) were co-expressed, and up-regulated in the resistant genotype SP80-3280 (Table 2).

Table 2: Co-expression modules from resistant genotype SP80-3280

Module	Comp-gg Seq ID	Gene ID	Gene name	up/down	effect on flowering
blue2	comp205700_c0_seq1	AT3G12810	PIE1	up	Negative
blue2	comp207801_c0_seq1	AT1G08260	ESD7, TIL1, ABO4	up	Negative
blue2	comp206769_c0_seq1	AT5G20320	DCL4	up	Positive
blue2	gg_11571	AT2G17290	CPK6	up	Positive
brown2	comp199879_c0_seq1	AT2G46020	BRM, CHR2	up	Negative
brown2	gg_07573	AT5G39660	CDF2	down	Negative
brown2	comp202464_c0_seq1	AT1G80070	PRP8, EMB33	up	Positive
brown2	comp205075_c0_seq1	AT3G43920	DCL3	up	Positive
lightpink1	comp184140_c0_seq1	AT1G08260	ESD7, TIL1, ABO4	up	Negative
lightpink1	comp200190_c1_seq1	AT2G23380	CLF, SDG1	up	Negative
lightpink1	comp204750_c2_seq1	AT5G06600	UBP12	up	Negative
lightpink1	gg_15862	AT1G08260	ESD7, TIL1, ABO4	up	Negative
dodgerblue4	comp180194_c0_seq1	AT3G22590	PHP, CDC73	down	Negative
dodgerblue4	comp184054_c0_seq1	AT1G04870	ATPRMT10	down	Positive
Lightcyan	gg_06658	AT1G68840	TEM2, RAV2	up	Negative
Black	comp198109_c0_seq1	AT1G78580	TPS1	down	Positive
goldenrod1	comp202963_c1_seq1	AT2G30140	UGT87A2	down	Positive
Darkgrey	gg_03849	AT2G46260	LRB1	down	Positive
cadetblue2	gg_07428	AT5G02810	PRR7	up	Positive

Table 3: Co-expression modules from susceptible genotype IAC66-6

Module	Comp-gg Seq ID	Gene ID	Gene name	up/down	effect on flowering
azure3	gg_15635	AT3G15354	SPA3	down	Negative
azure3	comp198370_c2_seq1	AT3G43190	SUS4	down	Positive
orchid3	comp193746_c0_seq1	AT4G21200	GA2ox8	down	Negative
orchid3	comp202287_c0_seq1	AT1G30040	GA2ox2	down	Negative
orchid3	gg_06658	AT1G68840	TEM2, RAV2	down	Negative
lightpink2	gg_14818	AT5G06600	UBP12	up	Negative
firebrick3	comp199615_c0_seq1	AT5G06600	UBP12	up	Negative
darkorchid4	comp172176_c0_seq1	AT5G48300	ADG1, APS1	up	Positive
darkorchid4	comp197651_c0_seq1	AT3G46510	PUB13	up	Positive
darkorchid4	gg_02988	AT2G46260	LRB1	up	Positive
darkorchid4	gg_12613	AT1G22770	GI	up	Positive
darkorchid4	gg_15551	AT5G48300	ADG1, APS1	up	Positive
Lightcoral	gg_11383	AT5G46210	CUL4	down	Negative
Lightcoral	comp192150_c0_seq1	AT5G14170	BAF60, CHC1	down	Positive
Lightcoral	gg_15361	AT3G43190	SUS4	down	Positive
orangered1	comp165891_c0_seq1	AT2G21660	GRP7, CCR2	up	Positive
orangered1	comp198173_c0_seq1	AT3G61600	LRB2, POB1	up	Positive



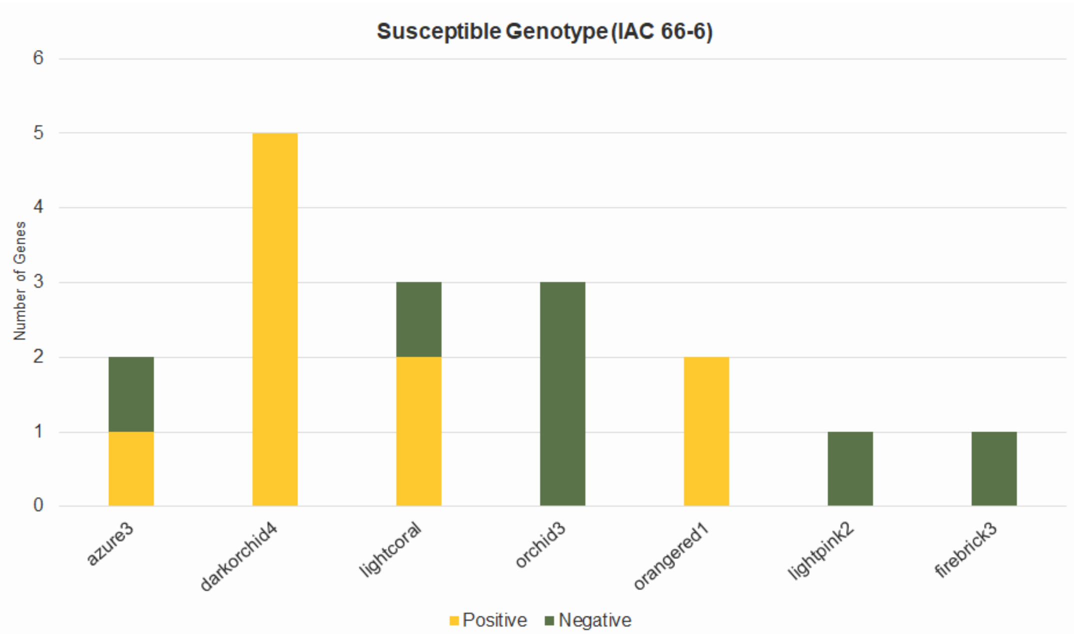


Figure 2: Differentially expressed positive (highlighted in yellow) and negative (highlighted in dark green) regulators of flowering pathway, in different modules of co-expression network of resistant and susceptible genotypes of sugarcane.

Transcriptional profiling and co-expression network data of FDEGs showed that differentially expressed negative regulators were up-regulated and positive regulators were downregulated in the resistant genotype SP80-3280. A similar pattern was observed in the susceptible genotype. However positive and negative regulators had an opposite expression behavior in the susceptible genotype than the resistant one. Despite that, there are a couple of exceptions in the pattern like positive regulators of flowering Dicer-like 3/DCL3 (comp205075_c0_seq1), Dicer-like 4/DCL4 (comp206769_c0_seq1), since they were up-regulated in resistant genotype. However, it could also be due to the higher miRNA/siRNA processing. A few other positive regulators of flowering also had the exceptional behavior pattern like Pre-mRNA-processing-splicing factor/PRP8 (comp202464_c0_seq1; brown2 module), Pseudo response regulator/PRR7 (gg_07428; cadetblue2 module), Calcium-dependent protein kinase 6/CPK6 (gg_11571; blue2 module) in the resistant genotype and BAF60 (comp192150_c0_seq1; lightcoral module) in the susceptible genotype. However, there is no data available about their expression pattern in short day species yet.

3.3.2. Pathway wise distribution of flowering differentially expressed genes (FDEGs)

In order to understand if smut fungus preferentially interacts with the genes from a certain individual pathway (a constituent of flowering pathway) 48 hai, all the differentially expressed genes from smut-resistant and smut-susceptible genotypes were categorized pathway wise (figure 3 and 4). In the smut-resistant genotype (SP80-3280), the majority of the FDEGs belonged to Autonomous (53%), Photoperiod (18%), and Vernalization (17%). Similarly, for the smut-susceptible genotype (IAC66-6), FDEGs belonged to the Autonomous (29%), Photoperiod (22%), and 14% each to Vernalization, Hormones, and Sugar.

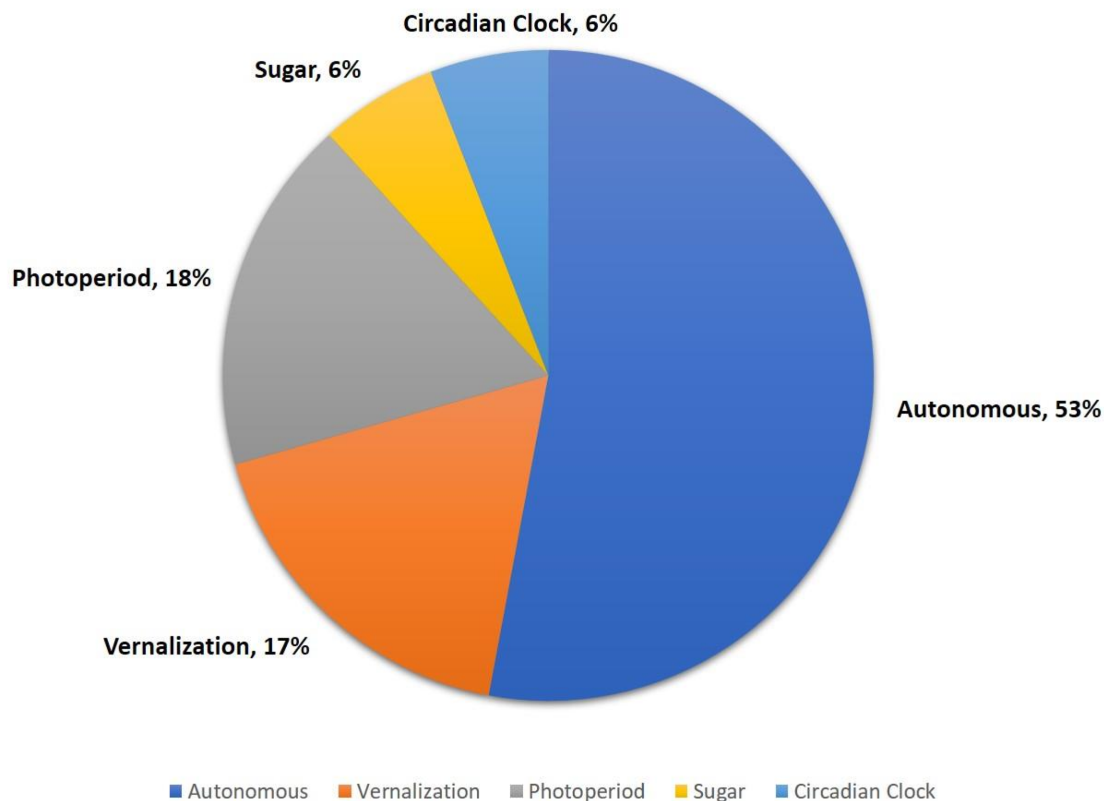


Figure 3: Pathway wise distribution of flowering differentially expressed genes from smut-resistant genotype (SP80-3280) represented in a pie chart.

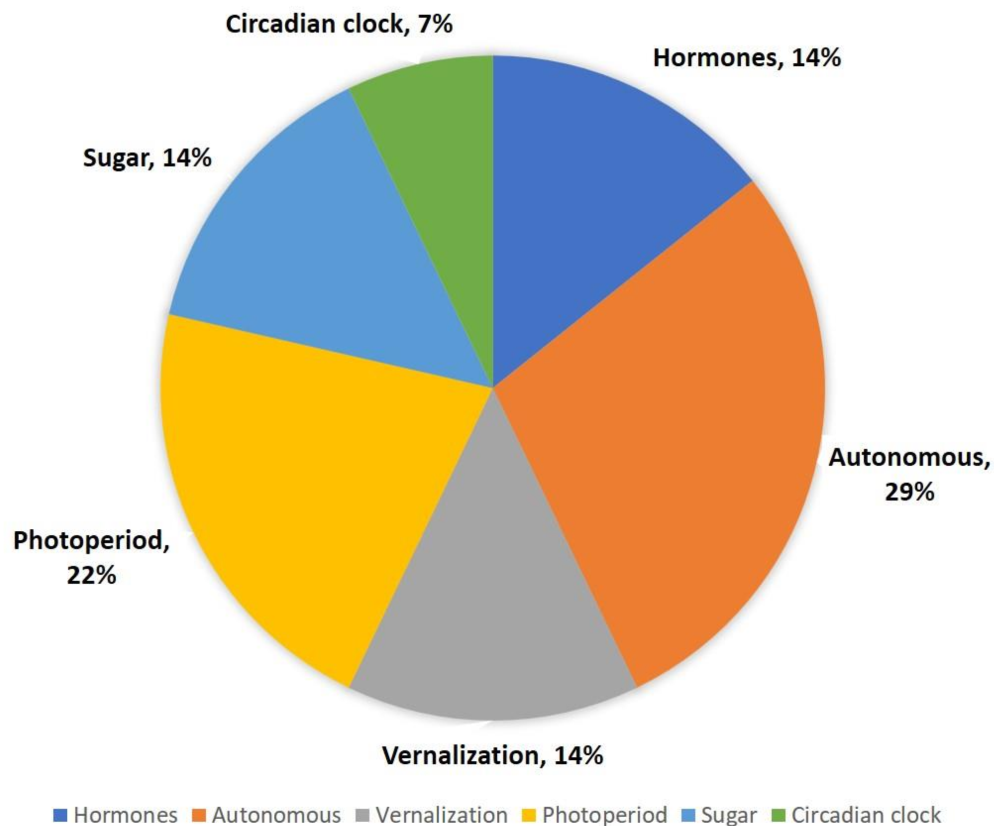


Figure 4: Pathway wise distribution of flowering differentially expressed genes from smut-susceptible genotype (IAC66-6) represented in a pie chart.

3.3.3. Candidate genes expression behavior validation via RT-qPCR

A total of 9 candidate genes (ATPRMT10, BRM, ESD7, GA2ox8, GI, PIE1, TEM2, TPS1, and SYD) were selected for RT-qPCR gene expression validation (flowering related genes and their interacting partners) procured from RNASeq analysis of resistant and susceptible genotypes. However, even after numerous attempts, four genes (GA2ox8, GI, PIE1, and TEM2) could not be amplified with a combination of primers we designed.

To summarize, RT-qPCR analysis validated the RNASeq data for five differentially expressed genes. However, only three genes were statistically significant (p-value <0.05), and the rest of the two genes, though not statistically significant, did have a similar expression behavior pattern as demonstrated in RNASeq data. Interestingly, two candidate genes, BRM and SYD, had a contrasting expression behavior in resistant and susceptible genotypes (Figure 5).

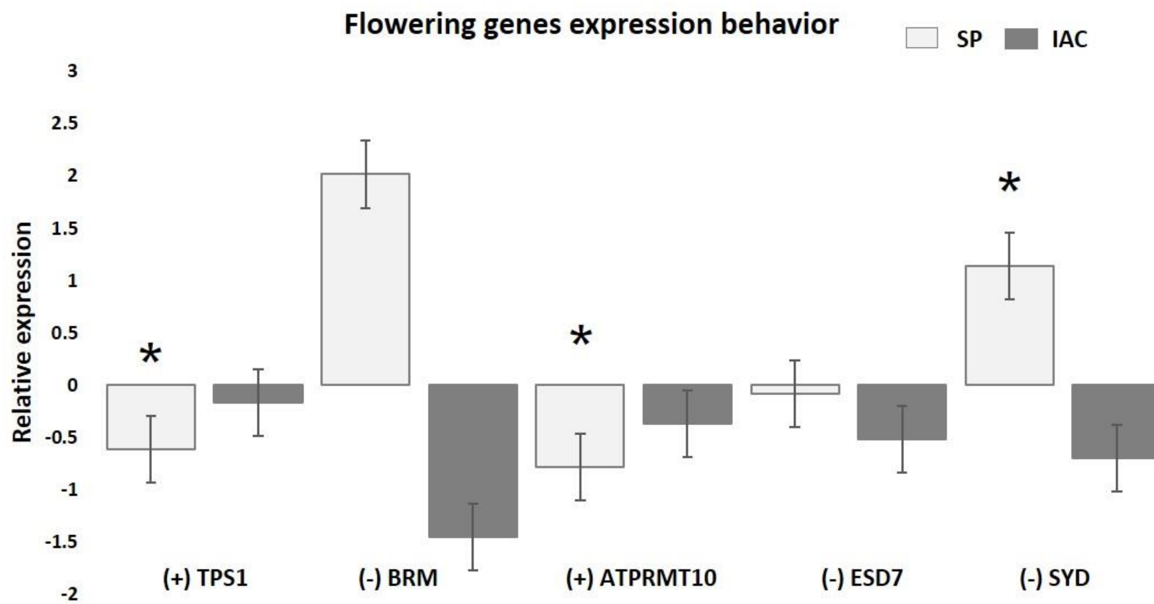


Figure 5: RT-qPCR validation: Flowering differentially expressed genes selected for RT-qPCR validation in smut resistant (SP80-3280) and susceptible (IAC66-6) genotypes: TPS1 (comp198109_c0_seq1), BRM (comp199879_c0_seq1), ATPRMT10 (comp184054_c0_seq1), ESD7 (comp207801_c0_seq1), SYD (comp207886_c1_seq1). REST® software was used to statistically analyze the RT-qPCR data. "*" indicates differentially expressed genes in RT-qPCR reactions (p-value < 0.05. (+) and (-) signs indicate positive and negative regulators of flowering respectively. Standard errors are represented by the bars.

3.4. Discussion

Transcriptional profiling of flowering-related genes in resistant (SP80-3280) and susceptible (IAC66-6) genotypes after 48 hai with *S. scitamineum* revealed that up-regulated genes exceeded the down-regulated genes in both the genotypes. Interestingly, approximately 55% of total DEGs were reported in the resistant genotype, a similar pattern was demonstrated by Que et al. (2014). They also reported a gradual increase in DEGs in susceptible cultivar with an extension in fungal inoculation period. A previous study from our lab (Schaker et al., 2016) demonstrated transcriptional reprogramming in genes related to meristematic functions and reproductive organ development in an intermediate smut resistant genotype (RB92-5345) at two time points: 5 DAI and 200 DAI (after whip emergence). We have observed a similar pattern in resistant and susceptible genotypes in this study. We also observed a distinct pattern where smut fungus infection results in the modulation of expression behavior of flowering related genes majorly from Autonomous, Photoperiod, and Vernalization pathways in smut-resistant genotype, and distribution pattern of FDEGs in smut-susceptible genotype seems to be more fairly distributed,

though more than half of the FDEGs belonged to Autonomous, Photoperiod and Vernalization pathways 48 hai. It is fair to speculate that components of Autonomous, Photoperiod and Vernalization pathways are the preferential interacting partners of smut fungus as differentially expressed genes from these pathways collectively represented 88% and 65% of the FDEGs in resistant and susceptible genotypes respectively (Figure 3 and 4).

3.4.1. Genetic and epigenetic control of flowering-related genes in resistant genotype SP80-3280 indicates suppression of floral transition

Some plant species induce flowering under abiotic and biotic stress conditions (Wada and Takeno, 2010; Takeno et al., 2016). For instance, the developmental response of Arabidopsis plants susceptible to bacterial pathogens, *Pseudomonas syringae* and *Xanthomonas campestris*, and an oomycete, *Peronospora parasitica,* is to accelerate the floral transition. Initiation/acceleration of the reproductive development process to three phylogenetically different pathogens in Arabidopsis indicates the possibility of being a global developmental response to pathogen infection. These changes in flowering time have also been suggested to be linked to the resistance/tolerance level of plants to pathogen infection (Korves and Bergelson, 2003). The differential expression of flowering related genes in sugarcane-smut interaction in our study indicates a similar response. Interestingly, resistant and susceptible genotypes in sugarcane are demonstrating an opposite response to smut infection.

We suggest that smut infection triggers transcriptional reprogramming in both resistant and susceptible genotypes. However, smut elicits an opposite developmental response in both the cultivars indicating a role of flowering-related genes in the disease response of sugarcane. Whereby smut resistant plants seem to suppress flowering pathway (Figure 6), and susceptible plants follow a similar response reported in Arabidopsis (Korves and Bergelson, 2003), and accelerate floral transition (Figure 7). These results seem to be in accordance with the general modus operandi of smut fungus, which interferes with the flowering pathway and generally targets the floral structures of the host plant species (Ghareeb et al., 2011; Glassop et al., 2013; Fan et al., 2016; Schmitz et al., 2018; Bhuiyan et al., 2021) and also suggest that sugarcane smut fungus potentially has the same modus operandi as other smut fungal species.

3.4.2. Potential epigenetic regulation related to smut

Chromatin remodeling factors play a pivotal role in the regulation of flowering pathway through combinatorial action of different histone-modifying enzymes and ATP dependent chromatin remodeling complexes (CRCs) like SWI, SWR1c, PAF1c, COMPASS, HUB-UBC on Flowering locus C and Flowering locus T (He, 2012; Shafiq et al., 2013; Ramirez-Prado et al., 2018). This interplay of different histone-modifying enzymes on FT and FLC loci decides the fate of floral transition.

Additionally, chromatin remodeling has been shown to establish a long and perhaps a transgenerational memory in plants through promoter priming of genes attributed to disease resistance. Although around 40 chromatin remodeling complexes (CRCs) have been reported to play different developmental roles in Arabidopsis, only five (SYD, PIE1, BRHIS1, CHR5, DDM1) have a role in plant defense system (Chen et al., 2017; Ramirez-Prado et al., 2018). To our best knowledge, there are no reports of these CRCs at play in Sugarcane's defense response to smut fungus or in the regulation of the flowering pathway in sugarcane.

Functional interplay between Trithorax group proteins (TrxG) and Polycomb group proteins (PcG) play a crucial role in controlling various developmental processes, including flowering in Arabidopsis (de la paz Sanchez et al., 2015). However, how this complex relationship between these two groups of proteins works in different developmental processes remains elusive. TrxG proteins generally act as transcriptional activators and PcG as the repressors through depositing H3K4me3 and H3k27me3, respectively at a target loci. For instance, Brahma, a TrxG protein, was demonstrated to directly activate a known flowering repressor gene short vegetative phase (SVP) through a gain- and loss-of-function study (Li et al., 2015). In Arabidopsis, SVP upon activation represses the transcription of Flowering locus T, a known florigen, thereby acting as a negative regulator of flowering. The comp184054_c0_seq1 (brown2 module), a putative BRM ortholog in sugarcane, was up-regulated in the resistant genotype SP80-3280, indicating that resistant variety averts/delays the floral transition. Splayed (SYD) is one of the most well studied CRCs in plants along with Brahma, and loss of function mutation analysis generated similar developmental phenotypes for both the genes, indicating that SYD and BRM have overlapping functions (Bezhani et al., 2007). In Arabidopsis, syd mutants exhibited early flowering phenotype through higher transcription levels of a central regulator of floral transition,

which established SYD as a negative regulator of floral transition (Wagner and Meyerowitz, 2002).

A putative SYD ortholog (comp207886_c1_seq1) is also up-regulated in the resistant genotype SP80-3280 and was co-expressed in the module brown2 along with BRM. SYD binds to promoters of jasmonic acid/ethylene-responsive genes and activates their expression. It has also been demonstrated to play a role in disease resistance against a necrotrophic fungus *B. cinerea* infection, in a loss of function analysis, where syd mutants exhibited enhanced susceptibility against *B. cinerea.* Additionally, SYD plays a pivotal role in meristem maintenance, also in Arabidopsis through activating transcription of the Wuschel gene (Walley et al., 2008; Sun et al., 2019). Though the SYD gene has been demonstrated to be a regulator of some defense-related genes, the mechanism behind this regulatory system is not fully unlocked yet.

SWR1 complex (SWR1c) is a highly conserved ATPase chromatin remodeling complex in eukaryotes, facilitating the replacement of canonical histone H2A with histone variant H2A.Z/F primarily at the 5' end of the gene and enhancing the transcriptional competence of the target gene via chromatin remodeling. Recent studies have shed light on the role of SWR1c in different physiological and developmental processes in plants like flowering time, stress response, DNA repair, meiosis, and somatic recombination in plants (Rosa et al., 2013; Aslam et al., 2019). Pleiotropic phenotypic changes in developmental processes and stress response have been observed upon disruption in the functional subunits of SWR1c. For instance, a mutation in PIE1 (Photoperiod-Independent early Flowering 1), a component of SWR1c responsible for mediating the trading of H2A with H2A variant H2A.Z at FLC, results in early flowering. Indeed, FLC is a negative regulator of flowering as it represses the transcription of flowering locus T (Mateos et al., 2015; Madrid et al., 2021).

Similar developmental changes (precocious flowering) were observed in mutants of non-catalytic subunits of SWR1c, ARP6, SWC2 and YAF9A, confirming that their functionality is imperative for swapping of canonical H2A with its variant H2A.Z (Aslam et al., 2019). A recent study involving immunoprecipitation and mass spectrometry has unraveled all the subunits of the SWR1c complex in Arabidopsis (Luo et al., 2020). Besides PIE, which is the catalytic subunit of the complex, SWR1c is comprised of some other non-catalytic subunits like ARP4 (Actin related protein 4), ARP6 (Actin

related protein 6), SWC2 (SWR component 2), SWC4 (SWR1 complex subunit 4), SWC6 (SWR1 complex subunit 6), CHR11/CHR17 (Chromatin-Remodeling protein 11, Chromatin-Remodeling protein 17), MBD9 (Methyl-CpG-binding domain 9), HTA 9 (Histone H2A protein 9), HTA11 (Histone H2A protein 11), GAS41/YAF9A (Gliomas41), RIN1 (Resistance to *Pseudomonas Syringae PV Maculicola* Interactor 1), RIN2 (RVB2A). Moreover, two of these subunits, PIE1 (comp205700_c0_seq1) and SWC2 (gg_08973), were up-regulated in resistant genotype, indicating the possibility of SWR1 complex being employed to delay the floral transition through activation of flowering repressor gene FLC (Choi et al., 2007; Aslam et al., 2019). Interestingly enough, both PIE1 and SWC2 were co-expressed in module blue2, which strongly suggests the possibility that the SWR1 complex has a role in the resistant genotype.

Early in short days (ESD7)/ Tilted 1 (TIL1)/ ABA overly sensitive 4 (ABO4), which encodes the catalytic subunit of DNA polymerase epsilon, not only plays a role in DNA replication or repair but also acts as a chromatin remodeler and a transcriptional repressor as well. ESD7 represses the expression of FT and SOC1 at specific times in a day (Olmo et al., 2010) and thus acts as a negative regulator of flowering. The up-regulated transcript level of sugarcane orthologs {comp207801_c0_seq1 (blue2 module), comp184140_c0_seq1, gg_15862} revealed in this study also indicates the cumulative effect it brings out in combination with other negative regulators of flowering in resistant genotype right after the infection with the smut fungus. ESD7 has been speculated to be a component of a chromatin mediated gene silencing complex along with its genetic and physical interactors TFL2 (Terminal flower2), EBS (Early bolting in short days) and ICU2 (Incurvata2), which were demonstrated to play a similar function of epigenetic repression of important floral integrators such as FT, SOC1, and AGAMOUS (Olmo et al., 2010).

A putative ortholog of a Polycomb group protein, CLF (Curly leaf/ Set domain group 1) (comp200190_c1_seq1; lightpink1 module), is up-regulated in the resistant genotype SP80-3280, following the pattern of other negative flowering regulators in this study. CLF represses the floral homeotic gene Agamous through augmentation of H3K27me3 repressive marks at Agamous locus, resulting in a delay in floral transition (Goodrich et al., 1997). In addition, the interacting partner of the CLF, SDG2 (Short Domain Group2 protein) (comp199888_c1_seq1; brown2 module), was also up-regulated in

the resistant genotype. SDG2 is a histone H3 lysine4 trimethyl transferase, a positive flowering regulator that activates the flowering repressor gene FLC. Guo et al. (2010) demonstrated precocious flowering in SDG2 mutant plants and the reduced levels of the FLC transcripts, which established the role of SDG2 in positively regulating the expression of FLC gene. Interestingly, all the differentially expressed negative regulators here were co-expressed basically in three modules (blue2, brown2 and lightpink1).

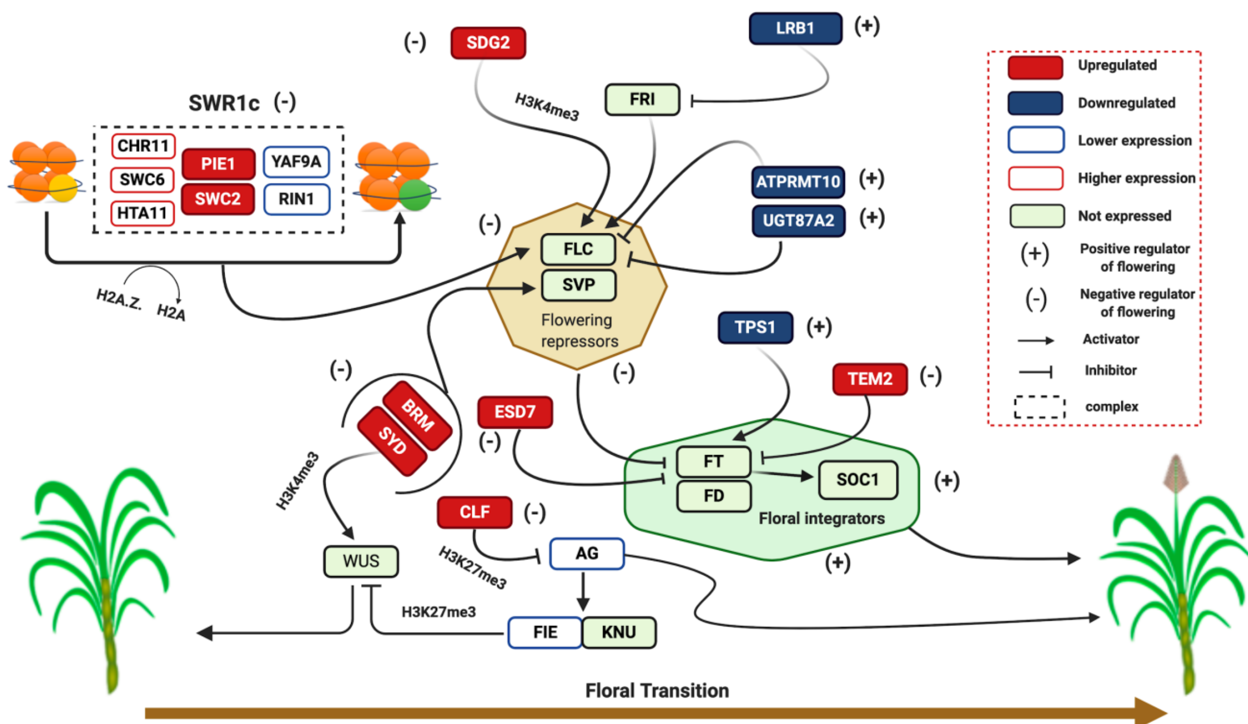


Figure 6: Expression behavior pattern of positive (+) and negative regulators (-) of flowering in an early response to smut infection in the smut resistant genotype SP8032-80. **Red** and **blue** rectangles represent the differentially expressed up-regulated and downregulated genes respectively. Rectangle with border highlighted in red reflects higher expression of the gene in comparison to the control sample (not a DEG). Similarly, a rectangle with border highlighted in blue shows lower expression of the gene in comparison to the control sample (not a DEG), and a rectangle with border highlighted in black shows that the gene didn't express at this stage. Arrows indicate the transcription activation and the inhibitor line represents transcription repression. Rectangle with the dashed black border represents a complex and its constituents enclosed. Brackets with plus sign **(+)** and brackets with dash sign **(-)** adjacent to a gene indicate positive and negative regulators of flowering respectively.

3.4.3. Genetic regulation of flowering-related genes

Wahl et al. 2013, established the crucial role TPS1 gene plays in flowering pathway and characterized it as a positive regulator of flowering, as loss of function mutants displayed highly delayed flowering or no flowering at all in Arabidopsis plants even in the inductive environmental situation (long day condition). Downregulation of the putative ortholog of TPS1 gene (comp198109_c0_seq1; black module) in the resistant genotype is consistent with the trend displayed by the other negative and positive regulators of flowering. For instance, the ATPRMT10 gene (comp184054_c0_seq1; dodgerblue4 module) is a known positive regulator of flowering, which has been postulated to be imperative for suppressing flowering repressor, FLC (Niu et al., 2012), and transcript level of this gene is down-regulated in resistant genotype SP80-3280.

Castillejo and Pelaz (2008) employed mutation and co-immunoprecipitation analyses, to establish Tempranillo genes (TEM1, TEM2), which are redundant in function, as novel direct repressor of florigen, FLT. TEM1/TEM2 negatively regulate the expression of FT by binding to its 5'-UTR region and overexpression of TEM1/TEM2 genes in Arabidopsis did not exhibit floral transition at all. Besides, loss of function of TEM1/TEM2 resulted in precocious flowering in Arabidopsis. Their findings established Tempranillo genes as negative regulators in flowering pathway, and antagonistic behavior of TEM2 (gg_06658) in resistant (up-regulated in SP80-3280; lightcyan module) and susceptible genotypes (downregulated in IAC66-6; orchid3 module) of our study, in response to smut infection, strongly indicate the repression and activation of the flowering pathway in resistant and susceptible genotypes, respectively.

Following the observed pattern, another positive regulator of flowering, UDP-Glucosyl Transferase 87A2 (UGT87A2, comp202963_c1_seq1; goldenrod1 module) was downregulated in resistant genotype SP80-3280. Wang et al. 2012, through loss of function study, showed that UGT87A2 negatively regulates FLC as higher transcript levels of FLC were observed in UGT87A2 mutant plants, and later on, suppression of FT, SOC1 (SUPPRESSOR OF OVEREXPRESSION OF CO 1**),** AP1 (Apetala1), and LFY (LEAFY)) was also observed in mutant plants. Another positive regulator of flowering, the Light-Response BTB1 (LRB1, gg_03849; module darkgrey), was downregulated in resistant plants whereas up-regulated in the susceptible ones (though not the same transcript, gg_02988). LRB1, LRB2, and CUL3A (CULLIN 3A)

are part of a proteasome, where LRB1 and LRB2 directly interact with Frigida (FRI) and degrade it. Frigida positively regulates the flowering repressor, FLC, and FRI degradation downregulates FLC, consequently activating flowering (Hu et al., 2014). The majority of the FDEGs (flowering differentially expressed genes) demonstrated a distinct pattern in this study. However, a few exceptions were also reported, while there is no data for most of these genes with exceptional expression behavior pattern (not following the similar pattern as most of the FDEGs) in SD species, a few of the following genes did. A couple of these genes are Dicer Like 4 gene (DCL4, comp206769_c0_seq1; blue2 module) and Dicer Like 3 gene (DCL3, comp205075_c0_seq1; brown2 module). DCL3 and DCL4 are positive regulators of flowering, and their upregulation in the resistant genotype diverts from the pattern. However, these genes are also involved in post-transcriptional gene silencing machinery through generating siRNAs/miRNAs.

The expression pattern of known interacting partners of these gene Argonaute 1 (AGO1, comp207043_c1_seq1; darkorange3 module), Nuclear RNA polymerase D1A (NRPD1, gg_6180; brown2 module), Nuclear RNA polymerase D1B (NRPD1B, comp198195_c2_seq1; blue2 module) and Suppressor of gene silencing 3 (SGS3, comp204769_c0_seq1; aliceblue module) enabled us to explain this deflection in expression pattern. Up-regulation of these four known interacting partners of DCL3 and DCL4 indicates a transcriptional gene silencing machinery in place and the maintenance of silenced genes through NRPD1A, NRPD1B and SGS3 genes (Chen et al., 2018).

3.4.4. Flowering differentially expressed genes seems to activate flowering pathway in the Susceptible genotype IAC66-6

Numerous endogenous and exogenous cues positively and negatively regulate several developmental processes like floral transition in plants. The role of gibberellins (GA) as a positive regulator of flowering in short-day (SD) and long-day (LD) plants is well established. GA is a potent regulator of flowering in SD plants (Davière and Achard, 2013). Schomburg et al. (2003) reported pleiotropic developmental effects in Arabidopsis by overexpressing the GA2ox8 gene, which catabolizes active GA. Overexpression of the GA2ox8 gene resulted in the depletion of bioactive GA content,

which in turn caused a severe delay in flowering in SD conditions. GA2ox8 and GA2ox7 double mutants resulted in early flowering in both SD and LD conditions.

Here in susceptible genotype IAC 66-6, putative orthologs of two Arabidopsis GA2ox paralogs, GA2ox8 (comp193746_c0_seq1) and GA2ox2 (comp202287_c0_seq1), were downregulated, which could be an indication that susceptible genotype might be activating floral transition after getting infected with smut fungus. Both of the paralogs of the GA2ox gene (GA2ox8 and GA2ox2) co-expressed in the module orchid3 along with another negative regulator of flowering, TEM2/RAV2. The expression behavior of TEM2/RAV2 and LRB1 genes, negative and positive flowering regulators have already been discussed, which established that flowering-related genes in resistant and susceptible genotypes have antagonistic expression behavior in this study. In addition to LRB1, Light-Response BTB2 (LRB2, comp198173_c0_seq1; orangered1 module) gene was also up-regulated in the susceptible genotype.

Gigantea (GI, gg_12613; darkorchid4 module) is one of the master regulators in different signaling pathways, including the circadian clock, photoperiod, sugar, and light signaling pathways. They all positively regulate flowering through activating flowering locus T (FT) directly or indirectly via the regulation of Constans (CO) (Fornara et al., 2010; Sawa and Kaye, 2011). GI also has a role in freezing tolerance as a regulator of oxidative stress through CDF transcriptional repressors and functioning in other several abiotic stresses besides biotic ones (Fornara et al., 2015; Mishra and Panigrahi, 2015). Upregulation of this gene (gg_12613) in the susceptible genotype might indicate a similar master regulator sort of function for Gigantea in sugarcane. Another positive regulator of flowering, GLYCINE-RICH RNA-BINDING PROTEIN 7/COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2 (GRP7/CCR2, comp165891_c0_seq1; orangered1 module), also had elevated transcript levels in the susceptible genotype, following the suit of the majority of differentially expressed positive regulators of flowering in the susceptible genotype. Streitner et al. (2008) showed a flowering-timing function of GRP7. T-DNA mutants of GRP7 exhibited a severe delay in flowering in both short-day and long-day conditions and had elevated transcript levels of FLC.

Ventriglia and colleagues (2008) characterized the ADP GLUCOSE PYROPHOSPHORYLASE 1, which encodes the small subunit of ADP-glucose

pyrophosphorylase. They showed that mutant plants were starchless with flowering delay in long-day and short-day. This phenotype promoted by the ADG1 mutant established it as a positive regulator of flowering. In our data, the upregulation of ADP Glucose Pyrophosphorylase 1 (ADG1/APS1, comp172176_c0_seq1/gg_15551; darkorchid4 module) in the susceptible genotype in response to smut infection is yet another evidence of flowering activation in susceptible plants of sugarcane.

For the CUL4-DDB1-COP1-SPA complex, Chen et al. (2010) reported that CULLIN4 (CUL4)-Damaged DNA Binding Protein1 (DDB1) physically interact with DDB1 binding proteins (DWD) such as COP1 (Constitutive Photomorphogenic Protein 1)-SPA (SUPPRESSOR OF PHYA) protein complexes. CUL4 mutated plants exhibited precocious flowers in short- days, and these cul4 mutated plants also had higher FT transcription levels. However, no changes were reported in the transcriptional level of CO. This complex represses the photomorphogenesis and thus affects the floral transition or prevents precocious flowering. All the subunits of this complex CUL4 (gg_11383), DDB1 (gg_13690), COP1 (comp202116_c2_seq1), SPA3 (gg_15635) were downregulated in the susceptible genotype, which suggests it is no longer suppressing photomorphogenesis and positively regulating the flowering pathway. All the constituents of the complex were differentially expressed (downregulated) except COP1, which though was not differentially expressed, did have a lower expression level in comparison to the control samples.

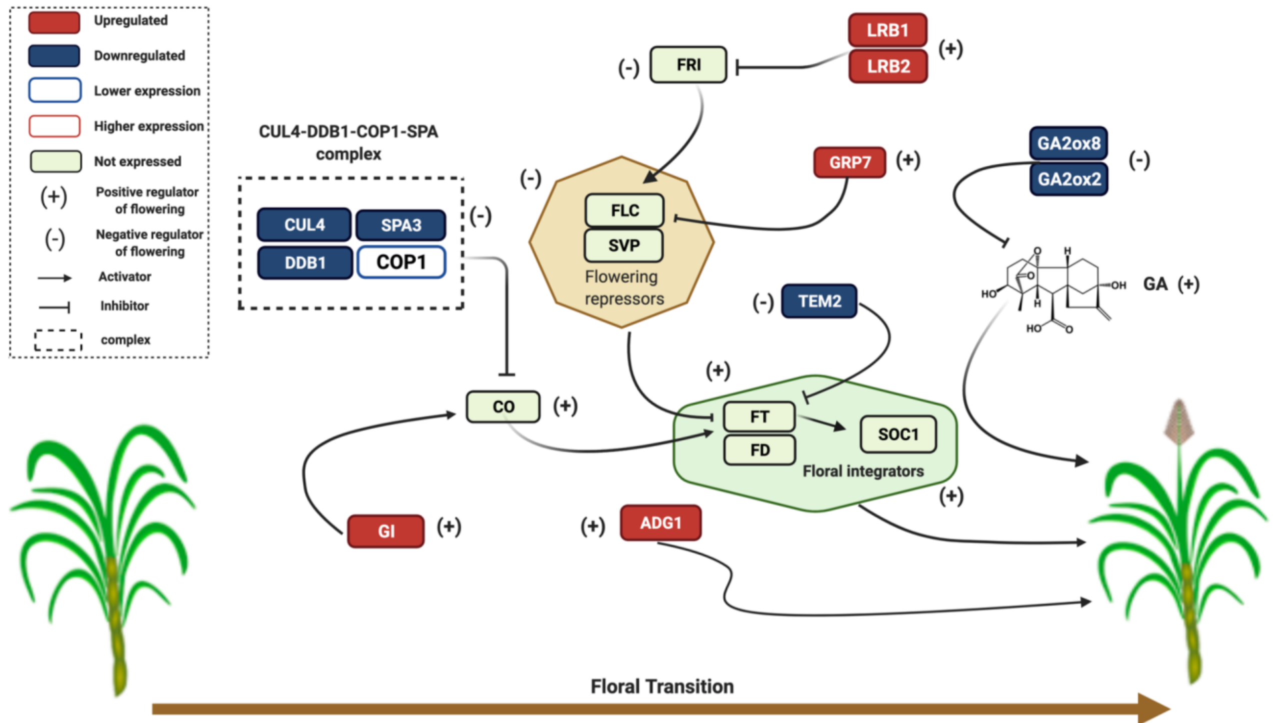


Figure 7: Expression behavior pattern of positive (+) and negative regulators (-) of flowering in an early response to smut infection in the smut susceptible genotype IAC66-6. **Red** and **blue** rectangles represent the differentially expressed up-regulated and downregulated genes respectively. Rectangle with border highlighted in red reflects higher expression of the gene in comparison to the control sample (not a DEG). Similarly, a rectangle with border highlighted in blue shows lower expression of the gene in comparison to the control sample (not a DEG), and a rectangle with border highlighted in black shows that the gene didn't express at this stage. Arrows indicate the transcription activation and the inhibitor line represents transcription repression. Rectangle with the dashed black border represents a complex and its constituents enclosed. Brackets with plus sign **(+)** and brackets with dash sign **(-)** adjacent to a gene indicate positive and negative regulators of flowering respectively.

3.5. Conclusion

Transcriptional profiling of smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) genotypes in this early response study unraveled a complex yet clear association between the floral transition and defense system in the sugarcane-smut interaction. Furthermore, the data suggest that SP80-3280 attempts to maintain its meristematic identity and keep the plant in the vegetative stage, whereas IAC66-6 seems to activate flowering right after infection. Interestingly, the smut fungus seems to interact preferentially with the flowering genes majorly from the autonomous, photoperiod and

vernalization pathways 48 hai. More than half of the flowering differentially expressed genes in smut resistant genotype belonged to the autonomous pathway.

Several plant species initiate floral transition under biotic and abiotic stress has been discussed above; however, detailed analysis of the induction of flowering in a smut-sugarcane compatible interaction has not been reported yet to the best of our knowledge. These preliminary results warrant a thorough study depicting shared and unshared components of flowering and smut whip development pathways in sugarcane. Smut disease and flowering both result in yield loss in sugarcane. A comparative transcriptional study might provide insight into the shared master regulators of these two developmental pathways, which might help design strategies to find better management solutions for two commercially essential issues in sugarcane.

References

Andrews S (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aslam M, Fakher B, Jakada BH, Cao S, Qin Y (2019). SWR1 Chromatin Remodeling Complex: A Key Transcriptional Regulator in Plants. Cells. 2019 Dec 12;8(12):1621.
Bader GD, and Hogue CW (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2 (2003).

Bakkeren G, Kämper J, Schirawski J (2008). Sex in smut fungi: structure, function and evolution of mating-type complexes. *Fungal Genet. Biol.* 45, 15–21.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015 Aug;53(8):474-85.

Bezhani S, Winter C, Hershman S, Wagner JD, Kennedy JF, Kwon CS, et al. (2007). Unique, shared, and redundant roles for the *Arabidopsis* SWI/SNF chromatin remodeling ATPases BRAHMA and SPLAYED. *Plant Cell* 19, 403–416.

Bhuiyan SA, Magarey RC, McNeil MD, Aitken KS (2021). Sugarcane Smut, Caused by *Sporisorium scitamineum*, a Major Disease of Sugarcane: A Contemporary Review. Phytopathology. 2021 Nov;111(11):1905-1917.

Borrás-Hidalgo O, Thomma BP, Carmona E, Borroto CJ, Pujol M, Arencibia A, Lopez J (2005). Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with Ustilago scitaminea or Bipolaris sacchari. Plant Physiol Biochem. 2005 Dec;43(12):1115-21.

Bouché F, Lobet G, Tocquin P, Périlleux C (2016). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*, *Nucleic Acids Research*, Volume 44, Issue D1, 4 January 2016, Pages D1167–D1171.

Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, et al. (2014). De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS One. 2014;9: e 88462.

Castillejo C, and Pelaz S (2008). The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. Current Biology: CB. 2008 Sep;18(17):1338-1343. DOI: 10.1016/j.cub.2008.07.075.

Chami DE, Daccache A, Moujabber ME (2020). What are the impacts of sugarcane production on ecosystem services and human well-being? A review. *Annals of agricultural science, 65*, 188-199.

Chen H, Huang X, Gusmaroli G, Terzaghi W, Lau OS, Yanagawa Y, Zhang Y, Li J, Lee JH, Zhu D, Deng XW (2010). Arabidopsis CULLIN4-damaged DNA binding protein1 interacts with CONSTITUTIVELY PHOTOMORPHOGENIC1-SUPPRESSOR OF PHYA complexes to regulate photomorphogenesis and flowering time. Plant Cell 22, 108–12.

Chen W, Zhang X, Fan Y, Li B, Ryabov E, Shi N, Zhao M, Yu Z, Qin C, Zheng Q, Zhang P, Wang H, Jackson S, Cheng Q, Liu Y, Gallusci P, Hong Y (2018). A Genetic Network for Systemic RNA Silencing in Plants. Plant Physiol. 2018 Apr;176(4):2700-2719.

Chen W, Zhu Q, Liu Y, and Zhang Q (2017). Chromatin remodeling and plant immunity. *Adv. Protein Chem. Struct. Biol.* 106, 243–260.

Chen W, Zhu Q, Liu Y, Zhang Q (2017). Chromatin Remodeling and Plant Immunity. *Adv. Protein Chem. Struct. Biol.* 2017, *106*, 243–260.

Choi K, Park C, Lee J, Oh M, Noh B, Lee I (2007). Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. Development 2007, 134, 1931–1941.
CONAB (2021). Companhia Nacional de Abastecimento. Available at https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar

D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet. MGG* 250, 405–413.

Davière JM and Achard P (2013). Gibberellin signaling in plants. Development 140(6): 1147–1151.

de Andrade LM, Dos Santos Brito M, Fávero Peixoto Junior R, Marchiori PER, Nóbile PM, Martins APB, Ribeiro RV, Creste S (2017). Reference genes for normalization of qPCR assays in sugarcane plants under water deficit. Plant Methods. 2017 Apr 17; 13:28.

de la Paz Sanchez M, Aceves-García P, Petrone E, Steckenborn S, Vega-León R, Álvarez-Buylla ER, Garay-Arroyo A, García-Ponce B (2015). The impact of Polycomb group (PcG) and Trithorax group (TrxG) epigenetic factors in plant plasticity. New Phytol. 2015 Nov;208(3):684-94.

de Setta N, Monteiro-Vitorello CB, Metcalfe CJ *et al.* (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15, 540 (2014).

del Olmo I, Lopez-Gonzalez L, Martín-Trillo MM, Martínez-Zapater JM, Pineiro M, and Jarillo JA (2010). *EARLY IN SHORT DAYS 7* (*ESD7*) encodes the catalytic subunit of DNA polymerase epsilon and is required for flowering repression through a mechanism involving epigenetic gene silencing. Plant J. 61: 623–636.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019). The Pfam protein families database in 2019. Nucleic Acids Res. 2019 Jan 8;47(D1): D427-D432.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007). Locating proteins in the cell using Targetp, SignalP and related tools. *Nat Protoc* 2: 953–971. 10.1038/nprot.2007.131.

Fan J, Yang J, Wang YQ, Li GB, Li Y, Huang F, Wang WM (2016). Current understanding on Villosiclava virens, a unique flower-infecting fungus causing rice false smut disease. Mol Plant Pathol. 2016 Dec;17(9):1321-1330.

FAOSTAT (2021). Food and Agriculture Organization of the United Nations (FAO), 2021. Available at: http://www.fao.org/faostat/en

Fornara F, de Montaigu A, Coupland G (2010). SnapShot: Control of flowering in Arabidopsis. Cell 141: 550, e1–e2.

Fornara F, de Montaigu A, Sánchez-Villarreal A, Takahashi Y, van Themaat EV, Huettel B, Davis SJ, Coupland G. (2015). The GI–CDF module of Arabidopsis affects freezing tolerance and growth as well as flowering. The Plant Journal 81, 695–706.

Garcia AA, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira ML, Vicentini R, Costa EA, Mancini MC, Garcia MO, Pastina MM, Gazaffi R, Martins ER, Dahmer N, Sforça DA, Silva CB, Bundock P, Henry RJ, Souza GM, van Sluys MA, Landell MG, Carneiro MS, Vincentz MA, Pinto LR, Vencovsky R, Souza AP (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep. 2013 Dec 2;3:3399.

Ghareeb H, Becker A, Iven T, Feussner I, Schirawski J (2011). *Sporisorium reilianum* infection changes inflorescence and branching architectures of maize. Plant Physiol. 2011 Aug;156(4):2037-52.

Gillis J, Mistry M, Pavlidis P (2010). Gene function analysis in complex data sets using ErmineJ. Nat Protoc. 2010;5(6):1148–59.

Glassop D, Rae AL, Bonnett GD (2013). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech* 16, 235–240 (2014).

Goodrich J, Puangsomlee P, Martin M *et al.* (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386, 44–51 (1997).

Guo L, Yu Y, Law JA, Zhang X (2010). SET DOMAIN GROUP2 is the major histone H3 lysine 4 tri-methyltransferase in *Arabidopsis*. Proc Natl Acad Sci U S A 107: 18557–18562.

Hagberg A, Schult D, Swart P (2008). Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T, Millman J (eds) Proceedings of the 7th Python in science conference (SciPy 2008), pp 11–15.

He Y (2012). Chromatin regulation of flowering. *Trends Plant Sci.* 17, 556–562.

Hu X, Kong X, Wang C, Ma L, Zhao J, Wei J, Zhang X, Loake GJ, Zhang T, Huang J, et al (2014). Proteasome-mediated degradation of FRIGIDA modulates flowering time in Arabidopsis during vernalization. Plant Cell 26: 4763–4781.

Huang N, Ling H, Liu F, Su Y, Su W, Mao H, Zhang X, Wang L, Chen R, Que Y (2018). Identification and evaluation of PCR reference genes for host and pathogen in sugarcane-*Sporisorium scitamineum* interaction system. BMC Genomics. 2018 Jun 19;19(1):479.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. (2019.) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1): D309–D314.

Huerta-Cepas JH, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al (2017). Fast genome- wide functional annotation through orthology assignment by eggNOG-mapper. Molecular Biology and Evolution. 2017; 34 (8): 2115–2122.

Iskandar HM, Simpson RS, Casu RE, Bonnett GD, Maclean DJ, Manners JM (2004). Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. Plant Mol Biol Rep. 2004;22:325–37.

James GL (1973). Smut spore germination in sugarcane internode surfaces. Proceedings of The South African Sugar Technologists' Association 47: 179–80.

Kim D, Langmead B, Salzberg SL (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360.

Korves TM, and Bergelson J (2003). A developmental response to pathogen infection in Arabidopsis. Plant Physiology 133, 339–347.

Langfelder P and Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).

Li C, Chen C, Gao L, Yang S, Nguyen V, Shi X, Siminovitch K et al. (2015). The Arabidopsis SWI2/SNF2 Chromatin Remodeler BRAHMA Regulates Polycomb Function during Vegetative Development and Directly Activates the Flowering Repressor Gene SVP (2015) *PLoS Genetics*, 11 (1), art. no. e1004944.

Liang Sun, Dong S, Ge Y, Fonseca JP, Robinson Z, Mysore KK, Mehta P (2019). "DiVenn: An Interactive and Integrated Web-based Visualization Tool for Comparing Gene Lists." Frontiers in Genetics 10 (2019):421.

Liao Y, Smyth GK, Shi W (2014). featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).

Lloyd HL and Naidoo G (1983). Chemical-assay potentially suitable for determination of smut resistance of sugarcane cultivars. Plant Disease 67: 1103–05.

Lloyd HL, and Pillay M (1980). The development of an improved method for evaluating sugarcane for resistance to smut. Proceedings of The South African Sugar Technologists' Association 54: 168–72.

Luo YX, Hou XM, Zhang CJ, et al. (2020). A plant-specific SWR1 chromatin-remodeling complex couples histone H2A.Z deposition with nucleosome sliding. The EMBO Journal. 2020 Apr;39 (7): e102008.

Madrid E, Chandler JW, Coupland G (2021). Gene regulatory networks controlled by FLOWERING LOCUS C that confer variation in seasonal flowering and life history, *Journal of Experimental Botany*, Volume 72, Issue 1, 20 January 2021, Pages 4–14.
Mamo G, Fayar R, Karlsson EN (2013). Microbial glycoside hydrolases for biomass utilization in biofuel applications. In: Gupta VK, Tuohy MG, eds. Biofuel technologies. Recent developments. Berlin: Springer, 171–188.

Marques JPR, Appezzato-da-Glória B, Piepenbring M, Massola NS, Monteiro-Vitorello CB, and Vieira MLC (2017). Sugarcane smut: Shedding light on the development of the whip- shaped sorus. Annals of Botany 119: 815–27.

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

Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, Fornara F, Schneeberger K, Krajewski P, Coupland G (2015). Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. Genome Biology 16, 31.

McCarthy DJ, Chen Y, Smyth GK (2012). “Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.” *Nucleic Acids Research*, 40 (10), 4288-4297.

Mishra P, and Panigrahi KC (2015). GIGANTEA—an emerging story. Frontiers in Plant Science 6, 8.

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD (2019). InterPro in 2019: improving coverage,classification and access to protein sequence annotations. Nucleic Acids Res. 2019 Jan 8;47(D1): D351-D360.

Monteiro-Vitorello CB, Schacker PDC, Benevenuto J, Teixeira e Silva NS, Almeida SS (2018). Progress in understanding fungal diseases affecting sugarcane: smut. In. Philippe Root (ed). Achieving sustainable cultivation of sugarcane Volume 2. Breeding, pests and diseases. Chapter 10. 2018. ISBN-13: 9781786761484.

Niu L, Lu F, Zhao T, Liu C, Cao X (2012). The enzymatic activity of Arabidopsis protein arginine methyltransferase 10 is essential for flowering time regulation. Protein & Cell. 2012 Jun;3 (6): 450-459.

Park S, Oh S, Ek-Ramos J and Van Nocker S (2010). PLANT HOMOLOGOUS TO PARAFIBROMIN is a component of the PAF1 complex and assists in regulating expression of genes within H3K27ME3-enriched chromatin. Plant Physiol. 153, 821-831.

Peters L, Carvalho G, Vilhena MB, Creste S, Azevedo RA, Monteiro-Vitorello CB (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and susceptible genotypes. Planta 2017, 245, 749–764.

Peters LP (2016). A more detailed view of reactive oxygen species metabolism in the sugarcane and Sporisorium scitamineum interaction. University of São Paulo.

Peters LP, Carvalho G, Vilhena MB, Creste S, Azevedo RA, and Monteiro-Vitorello CB (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. Planta 245: 749–64.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786. 10.1038/nmeth.1701

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30, e36–e36.

Piñon D, de Armas R, Vicente C and Legaz M E (1999). Role of polyamines in the infection of sugarcane buds by Ustilago scitaminea spores. Plant Physiology and Biochemistry 37: 57–64.

Que Y, Su Y, Guo J, Wu Q, Xu L (2014). A global view of transcriptome dynamics during Sporisorium scitamineum challenge in sugarcane by RNA-seq. PLoS One 9(8): e106476.

Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 2003, 339: 62-66. 10.1016/S0304-3940(02)01423-4.

Ramirez-Prado JS, Piquerez SJM, Bendahmane A, Hirt H, Raynaud C, Benhamed M (2018). Modify the histone to win the battle: Chromatin dynamics in plant-pathogen interactions. Front Plant Sci 9: 355.

Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010, 26: 139–40.

Robinson MD, McCarthy DJ, Smyth GK (2010). *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. 26(1): p. 139-40.

Rody HVS, Bombardelli RG, Creste S, Camargo LEA, Van Sluys MA & Monteiro-Vitorello CB (2019). Genome survey of resistance gene analogs in sugarcane: genomic features and differential expression of the innate immune system from a smut-resistant genotype. *BMC genomics*, 20(1), 809.

Rosa M, Von Harder M, Cigliano RA, Schlögelhofer P, Scheid OM (2013). The Arabidopsis SWR1 chromatin-remodeling complex is important for DNA repair, somatic recombination, and meiosis. Plant Cell 25: 1990 – 2001.

Sawa M and Kay SA (2011). GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* 108, 11698–11703.

Schaker PDC, Palhares, AC, Peters, LP, Aitken KS, Creste SA, Kitajima JP, Vieira MLC, Monteiro-Vitorello CB (2016). Depicting sugarcane transcriptional reprogramming during the biotrophic interaction with *Sporisorium scitamineum* in disease establishment and symptoms development. PLoS ONE, 11(9): e0162237, 2016.

Schmitz L, McCotter S, Kretschmer M, et al (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. F1000Research 7 (F1000 Faculty Rev):1812.

Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JA, Amasino RM (2003). Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell.* 2003;15(1):151-163.

Shafiq S, Berr A, Shen WH (2014). Combinatorial functions of diverse histone methylations in *Arabidopsis thaliana* flowering time regulation. *New Phytol.* 201, 312–322.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504.

Singh P, Singh RK, Song QQ, Li HB, Guo DJ, Malviya MK, Verma KK, Song XP, Lakshmanan P, Yang LT, Li YR (2021). Comparative analysis of protein and differential responses of defense-related gene and enzyme activity reveals the long-term molecular responses of sugarcane inoculated with *Sporisorium scitamineum*, Journal of Plant Interactions, 16:1, 12-29.

Streitner C, Danisman S, Wehrle F, Schoning JC, Alfano JR, Staiger D (2008). The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. Plant J 56:239–250.

Su Y, Xu, L, Wang Z, Peng Q, Yang Y, Chen Y, Que Y (2016). Comparative proteomics reveals that central metabolism changes are associated with resistance against *Sporisorium scitamineum* in sugarcane. *BMC Genomics* 17, 800 (2016).

Sun B, Zhou Y, Cai J, Shang E, Yamaguchi N, Xiao J, Looi LS, Wee WY, Gao X, Wagner D, Ito T (2019). Integration of transcriptional repression and polycomb-mediated silencing of WUSCHEL in floral meristems. Plant Cell 31:1488–1505.

Sundar A, Barnabas E, Malathi P, Viswanathan R (2012). “A Mini-Review on Smut Disease of Sugarcane Caused by Sporisorium scitamineum,” in Botany, ed. J. Mworia (InTech), 226.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan; 47: D607-613.

Takeno K (2016). Stress-induced flowering: the third category of flowering response, *Journal of Experimental Botany*, Volume 67, Issue 17, September 2016, Pages 4925–4934.

Taniguti LM, Schaker PDC, Benevenuto J, Peters LP, Carvalho G, Palhares, AC, Quecine MC, Nunes FRS, Kmit MCP, Wai A, Hausner G, Aitken KS, Berkmanc PJ, Moolhujzenc PM, Coutinho LL, Creste SA, Vieira MLC, Kitajima JP, Monteiro-Vitorello CB (2015). Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane. PLoS ONE, 10(6): e0129318.

Teixeira-Silva NS, Schaker PDC, Rody HVS, Maia T, Garner CM, Gassmann W, Monteiro-Vitorello CB (2020). Leaping into the Unknown World of *Sporisorium scitamineum* Candidate Effectors. J Fungi (Basel). 2020 Dec 4;6(4):339.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge**.** Nucleic Acids Res. 47: D506-515 (2019).

Thokoane LN and Rutherford RS (2001). cDNA-AFLP differential display of sugarcane (*Saccharum spp*. hybrids) genes induced by challenge with the fungal pathogen *Ustilago scitaminea* (sugarcane smut). *Proc. South African Sugar Technol. Assoc.* 75, 104–107.

Tom L, Braithwaite K, Kuniata LS (2017). Incursion of sugarcane smut in commercial cana crops at Ramu, Papua New Guinea. Proceedings of the Australian Society of Sugar Cane Technologists 39: 377–84.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, and Jack AM (2007). Leunissen: **Primer3Plus**, an enhanced web interface to Primer3 Nucleic Acids Research 2007 35: W71-W74.

Ventriglia T, Kuhn ML, Ruiz MT, Ribeiro-Pedro M, Valverde F, Ballicora MA, Preiss J, Romero JM (2008). Two Arabidopsis ADP-glucose pyro- phosphorylase large subunits (APL1 and APL2) are catalytic. Plant Physiol 148: 65–76.

Wada KC, and Takeno K (2010). Stress-induced flowering. Plant Signal Behav. 2010 Aug; 5(8): 944-7.

Wagner D and Meyerowitz EM (2002). SPLAYED, a novel SWI/SNF ATPase homolog, controls reproductive development in Arabidopsis. Curr Biol 12:85–94.

Wahl V, Ponnu J, Schlereth A, Arrivault S, Langenecker T, Franke A, Feil R, Lunn JE, Stitt M, Schmid M (2013). Regulation of flowering by trehalose-6-phosphate signaling in Arabidopsis thaliana. Science, 339, 704–707.

Walley JW, Rowe HC, Xiao Y, Chehab EW, Kliebenstein DJ, Wagner D, et al. (2008). The chromatin remodeler SPLAYED regulates specific stress signaling pathways. *PLoS Pathog.* 4: e1000237.

Wang B, Jin SH, Hu HQ, Sun YG, Wang YW, Han P and Hou BK (2012). UGT87A2, an *Arabidopsis* glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. *New Phytol.* 194(3):666–675.

Wu Q, Xu L, Guo J, Su Y, Que Y (2013). Transcriptome Profile Analysis of Sugarcane Responses to *Sporisorium scitaminea* Infection Using Solexa Sequencing Technology, *BioMed Research International*, vol. 2013, Article ID 298920, 9 pages, 2013.

Yang J, Li CY, Wang Y, Zhu Y, Li J, He X, Zhou X, Liu L, Ye Y (2006). Computational analysis of signal peptide dependent secreted proteins in Saccharomyces cerevisiae. Agric Sci China. 2006;5(3):221–7.

4. INVESTIGATING EFFECTS OF FLOWERING REPRESSOR ETHEPHON ON EXPRESSION BEHAVIOR OF FLOWERING RELATED GENES IN A SUGARCANE SMUT COMPATIBLE AND INCOMPATIBLE INTERACTION STUDY

ABSTRACT

Sugarcane (*Saccharum spp.*) is one of the most valuable cash crops in the world and there are many factors which adversely affect the crop yield. *Sporoisorium scitamineum*, the causal agent of smut disease in sugarcane is one of these factors, which results in up to 80% of the yield loss. Transcriptional profiling (RNASeq) of smut-resistant and smut-susceptible genotypes shortly after inoculation with smut fungus unraveled that smut fungus elicits an antagonistic expression behavior of flowering related genes in resistant and susceptible genotypes. Smut fungus generally interferes with the flowering pathway for its growth and development across different plant species. From a commercial point of view, flowering is not a desirable trait as it also compromises the crop yield. Ethephon has the potential to inhibit flowering in sugarcane, and enhance the sucrose levels as well. Here, the objective of this study was to explore if ethephon could have any impact on smut whip development at a physiological or molecular level. We treated buds of smut-resistant and smut-susceptible genotypes with 10 ppm ethephon solution for 10 minutes before inoculating them with smut fungus (*Ssc 04*). We observed that ethephon influenced the germination rate differently in two genotypes. Apparently, no physiological impact on the smut disease progression or whip developmental process was observed due to ethephon treatment. We were able to report that the effect of ethephon on candidate flowering genes varied depending on the genotype and the developmental stage. All the candidate flowering genes (ATPRT10, BRM, ESD7, SYD, and TPS1), in smut susceptible genotype had lower expression levels in ethephon treated samples than the ones not subjected to ethephon treatment at 48 hai, suggesting that ethephon potentially has the ability to modulate the expression behavior of candidate flowering genes even in the presence of fungus.

4.1. Introduction

Ethephon (C2H6ClO3P) or 2-CEPA (2-chloroethyl phosphonic acid) is a versatile plant growth regulator with a multitude of applications, used across a wide array of plant species. Ethephon, an ethylene generating chemical compound, hydrolyses quickly to produce non-toxic chemicals like ethylene gas, phosphoric and hydrochloric acid, which are naturally present in the plants as well. Its application varies across plant species depending on the time of application and the concentration used (Li and Solomon, 2003). Ethylene has been implicated in plant growth inhibition, enhancing germination rate, inducing and repressing flowering, affecting secondary metabolites production, herbicide efficacy augmentation, pollen sterility induction, enhancing fruit quality and fruit ripening, pests eradication, delaying bloom in deciduous fruits,

increasing chlorophyll, proline, peroxidases content and enhancing yield in a wide range of crops (Sun et al., 2015; Cunha et al., 2017; Iqbal et al. 2017; Jain et al., 2018; Islam et al., 2021; Li et al., 2021).

Ethephon was the first phytohormone used to manage post-harvest quality and improve crop management in many plant species (Abeles et al., 1992). Various factors play a critical role in the stimulation of the endogenous ethylene biosynthesis pathway. These include environmental conditions, mode of application, effective chemical absorption by the plant, the coverage area of foliar spray, stress status of the host plant, action site, hormone concentration, and sensitivity. A fine-tuned balance of the aforementioned conditions controls the activation of endogenous ethylene biosynthesis in plants (Abeles et al., 1992; Cunha et al., 2017). The ethephon application in sugarcane has also provided some insights into the benefits of using sugarcane at a commercial scale.

The first commercial-scale ethephon application as a chemical ripener was reported from South Africa (Donaldson and Staden, 1989), and since then numerous studies have exhibited benefits of ethephon application in sugarcane to accelerate ripening process, enhancing the crop yield, flowering inhibition, break apical dominance, augmenting sucrose accumulation and drought resistance as well (Li et al., 2002, Liao et al., 2003; Li and Solomon, 2003, Fong Chong et al., 2010; Silva and Caputo, 2012; Cunha et al., 2017; Jain et al., 2018). Ethephon acts by decomposing into ethylene, which further induces the downstream signaling pathway (Cunha et al., 2017), and the ethylene biosynthesis is transcriptionally regulated by the activities of two enzymes: 1-aminocyclopropane-1-carboxylate synthase (ACS) and ACC oxidase (ACO). First, ACS converts *S*-adenosylmethionine (SAM) into 1-aminocyclopropane-1-carboxylic acid (ACC), and finally, ACO converts ACC into ethylene (Ruduś et al., 2013). Transcriptional activity of ACO is indeed a prominent signal of auto-regulation of ethylene biosynthesis as reported by various research studies (Zhong et al., 2003; De Paepe et al., 2004; Nemhauser et al., 2006; Chang et al., 2013; Ruduś et al., 2013; Li et al., 2015; Cunha et al., 2017).

Though there is still a lot to be unraveled about the ethylene signaling pathway, a canonical genetic model and its components have been described in plants through molecular genetics experiments. Briefly, five different ethylene receptors (ETR1, ETR2, ERS1, ERS2, and EIN4) embedded in the ER membrane (endoplasmic reticulum), and the central units of ethylene signaling- a protein kinase CTR1

(constitutive triple response 1) and a transmembrane protein, EIN2 (ethylene insensitive-2) and transcription factors like EIN3 (ethylene insensitive-3), EIL (ethylene insensitive-3 like), ERFs (ethylene response factors) constitute the canonical ethylene signaling pathway in plants (Binder, 2020). When ethylene is not present, receptors activate the CTR1, which interacts with EIN2, and marks it for proteolysis, rendering it unable to act as a positive regulator of downstream signaling. EIN2 activates the expression of EIN3, EIL, and ERFs and induces the ethylene signaling. When ethylene is present, it inhibits the receptors and represses the activity of CTR1, and thus negates the negative effects it imparts on ethylene pathway (Wang et al., 2003; Binder et al., 2007; Li et al., 2015; Binder, 2020).

From a commercial perspective, flowering is not a desirable feature in sugarcane as it limits the crop yield. Therefore, ethephon application helps in flowering inhibition and enhances the sucrose accumulation in the stalk. (Moore and Osgood, 1989). A recent transcriptomic study shed light on the effect of ethephon on sugarcane at a molecular level by identifying the ethephon's target genes and its action site as well. It also unraveled ethylene as a potential stimulator of other hormonal signaling pathways and a possible hormonal crosstalk as around 18% of differentially expressed genes were responsible for hormone biosynthesis and signaling (Cunha et al., 2017).

Ethylene has also been implied to play a role in plant immunity, and depending on the kind of stress and surrounding environmental conditions, it could either act as a positive or negative regulator of disease resistance (Broekaert et al., 2006; Santa Brigida et al., 2016). For instance, exogenously applied ethylene in rice triggers the activation of pathogenesis-related genes, and ACS2 gene overexpression in rice enhances disease resistance levels against two of the most economically important rice pathogens, sheath blight and blast (Agarwal et al., 2001; Helliwell et al., 2013; Santa brigida et al., 2016). A transcriptional study on sugarcane infected with *Acidovorax avenae subsp avenae* (Aaa), the causal agent of red stripe disease induced the activation of the genes involved in ethylene and jasmonic acid biosynthetic pathways, along with SAR induced genes, pathogenesis-related genes and the genes involved in the oxidative burst as well (Santa Brigida et al., 2016). Ethylene has been demonstrated to be correlated with plant immunity in other crops as well. Knockdown of OsEDR1 (enhanced disease resistance 1) gene in rice results in lower expression of ACS (ethylene biosynthesis gene) and transgenic rice plants from this experiment

demonstrated enhanced resistance levels to bacterial leaf blight disease (Shen et al., 2011).

A recent study provided insights into the effects of ethephon on sugarcane at the physiological and molecular levels. Interestingly, ethephon treatment led to increased levels of flavonoids, peroxidases and proline (Jain et al., 2018), which have been exhibited to be involved in disease resistance in a sugarcane-smut interaction (Schaker et al., 2016; Monteiro-Vitorello et al., 2018). Our transcriptomic study (chapter 2) unraveled the antagonistic effects smut fungus induces on the flowering related genes in smut resistant and smut susceptible genotypes (a potential suppression and activation of flowering pathway, respectively) 48 hours after inoculation. This exploratory study was conducted to determine whether ethephon, which has the potential to negatively regulate flowering in sugarcane, can a) influence sugarcane-smut interaction and whip development process, and b) influence the expression behavior of selected flowering differentially expressed genes (unraveled via RNASeq profiling-chapter 2).

4.2. Material and Methods

4.2.1. Sugarcane buds collection

A total of 240 (120 each from smut resistant genotype, SP80-3280 and smut susceptible genotype, IAC66-6) healthy buds were collected from 10 months old sugarcane plants grown at the experimental field of the department of Genetics, ESALQ, University of São Paulo, Piracicaba. Buds were subsequently disinfected by heat and chemical treatments (thoroughly rinsed with water; kept at 52°C in water bath for 30 minutes; submerged in 0.01% sodium hypochlorite solution for 10 minutes and finally rinsed in distilled H_2O thrice). Disinfected buds were kept in a tray on a moist filter paper bed, covered with aluminum foil and transferred to be incubated at 28°C overnight to stimulate pre-germination.

4.2.2. Ethephon treatment and smut fungus inoculation

To evaluate the effects of flowering repressor (ethephon) on smut whip development pathway, and expression behavior of flowering related genes in different stages of plant development, 10 months old buds of smut resistant and susceptible genotypes were treated with 10 ppm ethephon solution (Jain et al., 2018). Briefly, with two sugarcane genotypes, there were 4 experiments, each constituting 60 germinated

buds. Germinated buds were dipped in 10 ppm solution of Ethephon (2-chloroethylphosphonic acid) for 10 minutes and subsequently inoculated with the teliospores of sugarcane smut fungus (*S. scitamineum*) strain Ssc04 (viability rate > 80%) suspended in saline solution (10^9 teliospores mL−1). 5µl of teliospore solution was administered directly over each germinated bud, kept at room temperature afterwards for 10 minutes for air-drying (Figure 8), and finally transferred the trays to incubate at 28°C for 48 hours in a dark chamber after covering the buds with vermiculite for pre-germination as described by Peters et al. (2017).

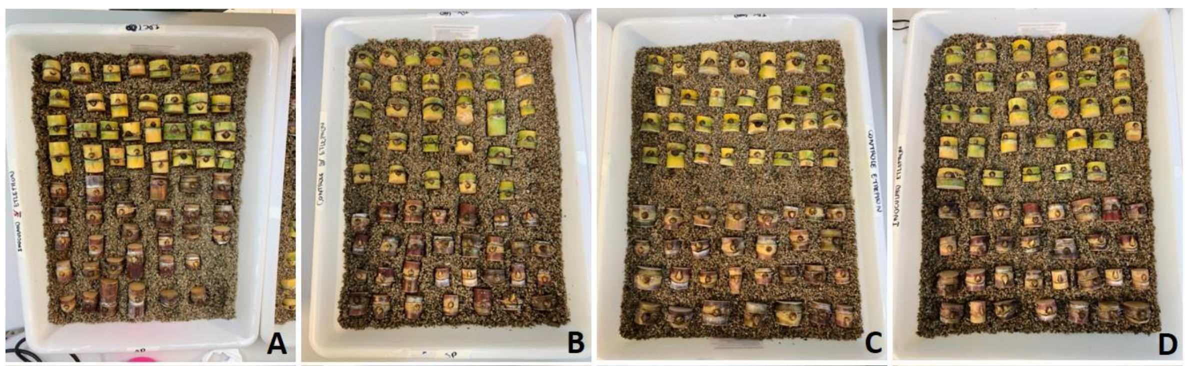


Figure 8: Sugarcane smut resistant (SP80-3280) and susceptible (IAC66-6) buds after ethephon treatment and inoculation with smut fungus: A) Ethephon-Mock inoculated-SP-IAC buds. B) Ethephon inoculated SP-IAC buds. C) No Ethephon-Mock inoculation-SP-IAC buds. D) No Ethephon inoculated SP-IAC buds

4.2.3. Sample collection at 48 hours after inoculation

Three biological replicates, each comprising a pool of 5 buds from inoculated and mock inoculated (control) samples from each genotype and two different treatments were collected 48 hours after inoculation (48 hai). Subsequently, the samples were ground using the liquid N_2 and were transferred to -80°C until further use. A total of 120 buds were collected at this point and rest of the buds were maintained in the dark room for two weeks before transferring to the greenhouse.

4.2.4. DNA extraction and smut fungal infection verification

DNA was extracted with the CTAB method (Doyle and Doyle, 1990) from 48 hai smut resistant and susceptible genotype samples treated with the ethephon (control and inoculated) and not treated with ethephon (control and inoculated). DNA samples were

treated with RNase and subsequently used to verify the presence of fungus in plants after inoculation.

4.2.5. Plants cultivation in the greenhouse

The rest of the buds remained after the 48hai sample collection were placed in trays containing moist vermiculite and kept in the dark chamber for two weeks to measure the germination rate. After two weeks, they were transferred to the greenhouse in the vessels filled with equal parts of topsoil and tropstrat substrate in a randomized design and with daily irrigation (Figure 9).

Figure 9: Experimental setup in a completely randomized design in the greenhouse at 16 days after inoculation.

4.2.6. Sample collection at vegetative phase-35DAI

SAM (shoot apical meristem) samples were collected from the plants being grown in the greenhouse 35 DAI (days after inoculation). Samples at this stage were composed of SAM collected from a single plant per replicate. Samples were immediately frozen in liquid N_2 and were ground using liquid N_2 and transferred into 2 ml eppendorf tubes to prepare for RNA extraction. Rest of the plants were cultivated for five more months

to see if ethephon would bring out any physiological changes in the treated samples in two contrasting sugarcane genotypes.

4.2.7. RNA extraction and cDNA preparation

Total RNA from samples collected at 48hai and 35 DAI (days after inoculation) was extracted using a modified cetyl trimethylammonium bromide (CTAB) method (Yoshida et al., 2010). Briefly, 100 mg of grounded tissue was suspended in 800 µl of CTAB solution preheated at 65°C (2% CTAB, 2% PVP, 25 mM EDTA, 2 M NaCl, 1% **β**-mercaptoethanol, 100 mM Tris-HCl) in a 2 ml eppendorf tube, mixed well by vortexing and inverting the tubes and followed by addition of equal volume (800 µl) of phenol: chloroform (4:1, pH 4.7, Sigma). The samples were subjected to incubation at 55°C for 10 minutes with vortexing every two minutes for 20 seconds and inverting the tubes. Afterwards, samples were centrifuged at 15000 rpm at 4ºC for 15 minutes, and the aqueous phase was extracted with an equal volume of phenol: chloroform (4:1, pH 4.7), and subsequently with chloroform. The RNAs were precipitated by adding 1/4th volumes of 10 M LiCl and kept at 4°C overnight. Following day, samples were centrifuged at 4°C at 15000 rpm, following which the RNA pellet was washed twice with 70% ethanol, air dried at room temperature and finally, dissolved in 35 µl nuclease-free water. RNA integrity was verified by agarose gel electrophoresis, stained with SYBR safe (Thermo Fisher Scientific, Oregon, USA). RNA quantity (A_{260}/A_{280}) and quality (A_{260}/A_{230}) parameters were assessed by using Nanodrop® 2000 spectrophotometer (Thermo Fisher Scientific Inc.). The GoTaq® 2-step RT-qPCR system kit was used for cDNA synthesis (Promega, Madison, WI, USA). 1 µg of total RNA (DNase treated, Sigma Aldrich, St. Louis, MO, USA) and Oligo (dT) primers were used for each cDNA reaction as per kit manual instructions.

4.2.8. Flowering-related genes expression analysis by RT-qPCR

To understand the effects of ethephon treatment on a molecular level, the expression behavior pattern of flowering related genes was analyzed using RT-qPCR. The qPCR master mix with a total volume of 12.5 µl containing 2 µl of 8 times diluted cDNA sample, 6.25 µl of GoTaq® qPCR Master Mix, 0.125 µl CXR reference dye, 0.25 µl of each primer (0.2 µM), and 3.625 µl nuclease-free water was used for each of the three biological replicates and two technical replicates. qPCR cycling conditions were as follows: 95 °C for 02 min; 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and a dissociation

curve was obtained for each candidate gene and reference gene primer used in this study to verify the Primer specificity (Appendix D). Sugarcane housekeeping genes encoding for **GAPDH** (d-glyceraldehyde-3-phosphate dehydrogenase), **eEF1α** (eukaryotic elongation factor 1**α)** (Iskandar et al., 2004; de Andrade et al., 2017; Huang et al., 2018) and additionally, **SRO1** (from transcriptomic data, chapter 2) were used to normalize the gene expression levels for both the genotypes, SP and IAC. Five candidate genes; ATPRMT10, BRM, ESD7, SYD and TPS1 were selected from 48 hai-RNASeq data of sugarcane smut resistant and susceptible genotypes (SP80-3280 and IAC66-6) to observe if ethephon treatment brings out any changes in their expression behavior pattern. The LinReg PCR program was used to calculate PCR efficiencies and Cq values (Ramakers et al., 2003). Relative changes in the gene expression levels were calculated with the $2^{-\Delta\Delta CT}$ method normalized with the endogenous gene expression (Livak and Schmittgen, 2001). One-way ANOVA followed by Tukey's multiple comparisons test was employed to calculate the significant variance (p-value<0.05) through GraphPad prism version 9.3.0 for windows, GraphPad Software, San Diego, California USA (www.graphpad.com).

4.3. Results

4.3.1. Confirmation of smut fungus infection via PCR

A PCR amplicon of ~509 bp generated by targeting the rDNA (ribosomal DNA) internal transcribed spacer region (ITS1, 5.8S & ITS2) of *S. scitamineum* using the following primer set: Hs 5'-AACACGGTTGCATCGGTTGGGTC-3' and Ha 5' - GCTTCTTGCTCATCCTCACCACCAA- 3' (Bueno, 2010) in an conventional PCR reaction confirmed the presence of fungus in the 48 hai samples of the smut resistant and susceptible genotypes infected with Ssc04 (figure 10).

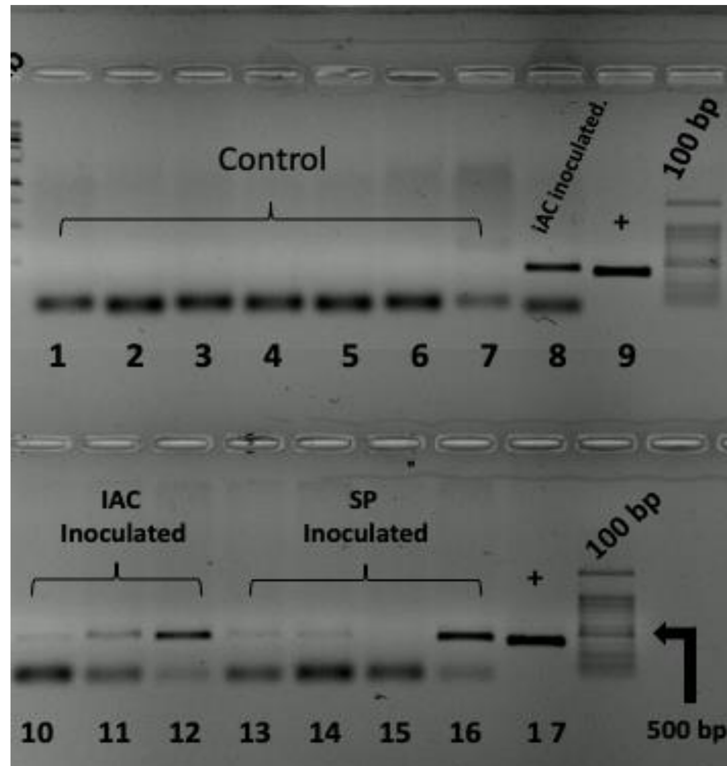


Figure 10: Confirmation of smut fungal infection by conventional PCR. Amplicons (~509 bp) confirms the fungal presence in samples collected 48 hai. 100 bp ladder (Thermo Fisher).

4.3.2. Germination rate calculation and physiological growth assessment

Briefly, smut susceptible genotype IAC66-6 demonstrated a better germination rate (61.66%) than the resistant genotype SP80-3280 (51.66%). Ethephon treated samples from both the genotypes exhibited a contrasting effect on the germination rate in comparison to the ones not subjected to ethephon treatment. Ethephon treated buds from the resistant genotype SP80-3280 exhibited a bit better germination rate (56.66%) than the ones not subjected to ethephon (46.66%), and on the contrary, ethephon treated buds from the smut susceptible genotype IAC66-6 had a lower germination (56.66%) in comparison to the untreated ones (66.66%). Plants were observed for any effect of ethephon on physiological aspects of growth and development and more specifically to see if ethephon would have any effect on the whip development process. Samples were cultivated in a greenhouse for a total of 7 months (215 days) after inoculation and only one whip emerged from a single IAC plant 90 days after inoculation, which was not treated with ethephon. No whip

emergence or disease symptoms were observed in any other plants in the experimental duration of seven months.

4.3.3. Flowering genes expression behavior analysis via RT-qPCR

The selection of candidate flowering genes was based on the previous results obtained in RNASeq experiments and their gene expression validation subsequently via RT-qPCR (Chapter 2). Since the gene expression of these candidate genes have already been analyzed under routine conditions (no exogenous treatment), RT-qPCR analysis of the same genes in this experiment unraveled if the ethephon (sugarcane flowering repressor) could modulate the expression behavior of flowering related genes in question here, and in turn could have any impact on smut whip development process or fungal disease progression in a compatible and incompatible interaction with sugarcane. Indeed, some of the candidate flowering genes did exhibit a different expression profile under the influence of ethephon in this particular experiment than the one we demonstrated earlier as the experimental conditions were slightly different here due to the nature of the control treatment for ethephon (chapter 2).

4.3.3.1. Expression profile of candidate flowering genes-RT-qPCR-48 hai

To analyze the effects of ethephon on a molecular level, the expression behavior pattern of candidate flowering genes in ethephon treated smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) plants 48 hai was compared to the expression behavior of the same genes observed from RNASeq transcriptional profiling (Chapter 2). RNASeq profiling revealed that TPS1 was downregulated (DEG) in 48 hai smut resistant (SP80-3280) plants, and interestingly, ethephon treated samples in the same genotype exhibited a similar expression profile via RT-qPCR, though it was not statistically significant (Figure 11). On the contrary, smut-resistant plants not subjected to ethephon treatment demonstrated higher expression levels of TPS1. Interestingly, TPS1 in smut-susceptible plants exhibited an antagonistic expression profile to that of smut-resistant plants (Figure 11). Another candidate gene, BRM, was upregulated (DEG) and downregulated in the smut-resistant and smut-susceptible genotypes respectively, in the transcriptomic analysis (Chapter 2). However, the expression levels of BRM gene were low (not statistically significant) in all the smut-resistant and smut-susceptible plants subjected to or not subjected to ethephon treatment.

RNASeq analysis and further RT-qPCR analysis (chapter 2, figure 5) revealed that ATPRMT10 gene was downregulated (DEG) in smut-resistant plants and also had lower transcript levels in smut-susceptible plants (not statistically significant) 48 hai. Smut-resistant plants treated with ethephon showed a similar expression profile as reported in RNASeq data analysis. However, the smut-resistant plants not treated with ethephon exhibited higher expression levels, contrary to the expression profile of ethephon treated samples and as well as of the ones from the RNASeq data analysis. Smut-susceptible (IAC) plants not treated with ethephon here had lower expression levels of ATPRMT10 gene, in accordance with the RNASeq data and on the contrary ethephon treated smut-susceptible plants exhibited higher expression levels, not statistically significant though (figure 11).

Transcriptional profiling of smut-resistant and smut-susceptible genotypes 48 hai, exhibited that ESD7, another candidate flowering gene, was upregulated (DEG) in smut-resistant plants and had lower transcript levels in smut-susceptible plants. Ethephon treated plants had lower expression levels of ESD7 in the smut-resistant and smut-susceptible plants. Smut-susceptible (IAC) plants which were not subjected to ethephon treatment exhibited a similar expression pattern, and smut-resistant (SP) plants not treated with ethephon had higher transcript levels of ESD7 as reported in RNASeq study (Chapter 2). Interestingly, SYD gene expression behavior was exactly similar to ESD7 in all the treatments (RNASeq, ethephon treated, not treated with ethephon) in both the genotypes 48 hai, except the IAC plants treated with ethephon, which had the similar lower expression levels as reported in IAC plants not treated with ethephon (figure 11).

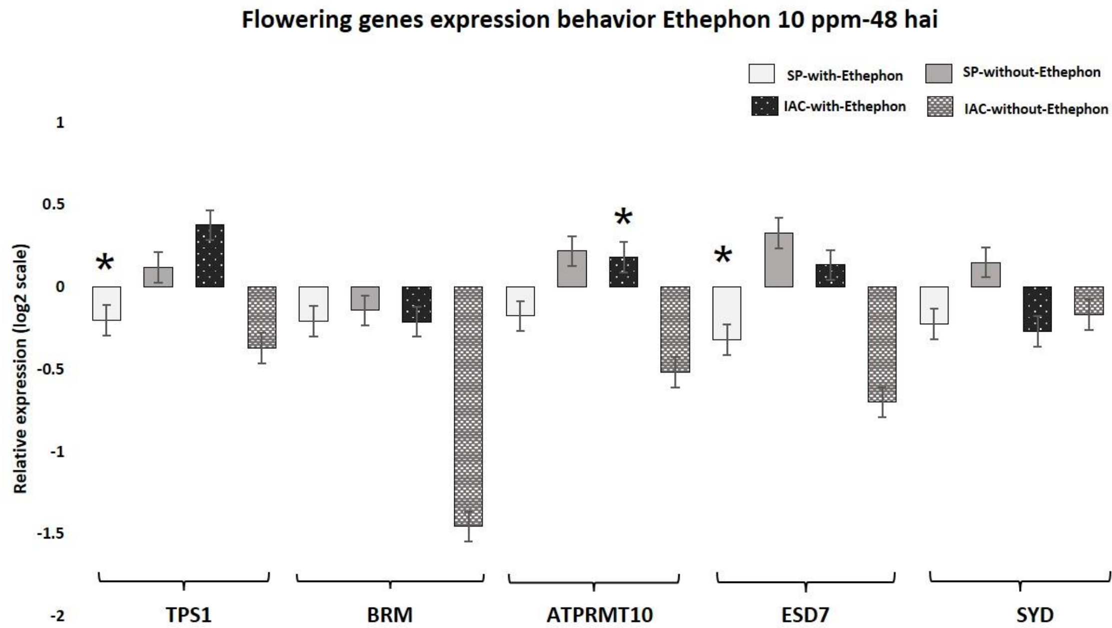


Figure 11: RT-qPCR validation-48 hai: Flowering genes selected for RT-qPCR validation in smut resistant (SP80-3280) and susceptible (IAC66-6) genotypes treated with flowering repressor Ethephon: TPS1 (comp198109_c0_seq1), BRM (comp199879_c0_seq1), ATPRMT10 (comp184054_c0_seq1), ESD7 (comp207801_c0_seq1), SYD (comp207886_c1_seq). REST® software was used to statistically analyze the RT-qPCR data. "✱" indicates differentially expressed genes in RT-qPCR reactions (p-value < 0.05).

4.3.3.2. Expression profile of candidate flowering genes-RT-qPCR-35 DAI

All the candidate flowering genes used for analyzing expression behavior patterns under the influence of ethephon 48 hai were also tested 35 DAI (days after inoculation) via RT-qPCR to analyze changes in expression behavior pattern of these flowering genes if any. Here we summarize how the expression behavior modulated 35 DAI in comparison to the 48 hai expression data (RT-qPCR).

In case of TPS1 gene, the expression behavior at 35 DAI changed completely for IAC plants in both ethephon-treated and untreated samples (higher to lower and lower to higher expression levels, respectively). However, the SP plants did not demonstrate the same pattern. Expression level of ethephon-treated SP plants did not change, and still had lower expression levels, and untreated samples, on the contrary, had their expression level changed from higher to lower expression levels (figure 12).

Interestingly, expression level modulation of the BRM gene on IAC plants was quite clear, changing from lower to higher expression levels in both ethephon-treated and untreated samples. However, expression levels didn't exhibit any modulation in expression behavior pattern in ethephon-treated or untreated SP samples and it

consistently maintained a profile of lower expression levels (figure 12). Expression levels of another candidate gene ATPRMT10, did demonstrate a change in expression levels, though only in ethephon treated SP plants (lower to higher expression levels). Rest of the treatments in both the genotypes had lower expression levels of ATPRMT10 gene as recorded 48 hours after inoculation via RT-qPCR (figure 12).

Expression profile of another candidate flowering gene, ESD7 didn't exhibit any modulation 35 DAI (days after inoculation) in comparison to the one observed at 48 hai via RT-qPCR as well (figure 12). Expression profiling of SYD, a candidate flowering gene analyzed 35 DAI via RT-qPCR exhibited a shift in expression behavior in comparison to the one analyzed at 48 hai in IAC ethephon-treated and untreated samples. However, in case of SP, the modulation in expression behavior was recorded only in the samples not treated with ethephon, and the ethephon treated samples didn't demonstrate any modulation in expression behavior (figure 12) when comparing to the samples tested at 48 hai via RT-qPCR (figure 11).

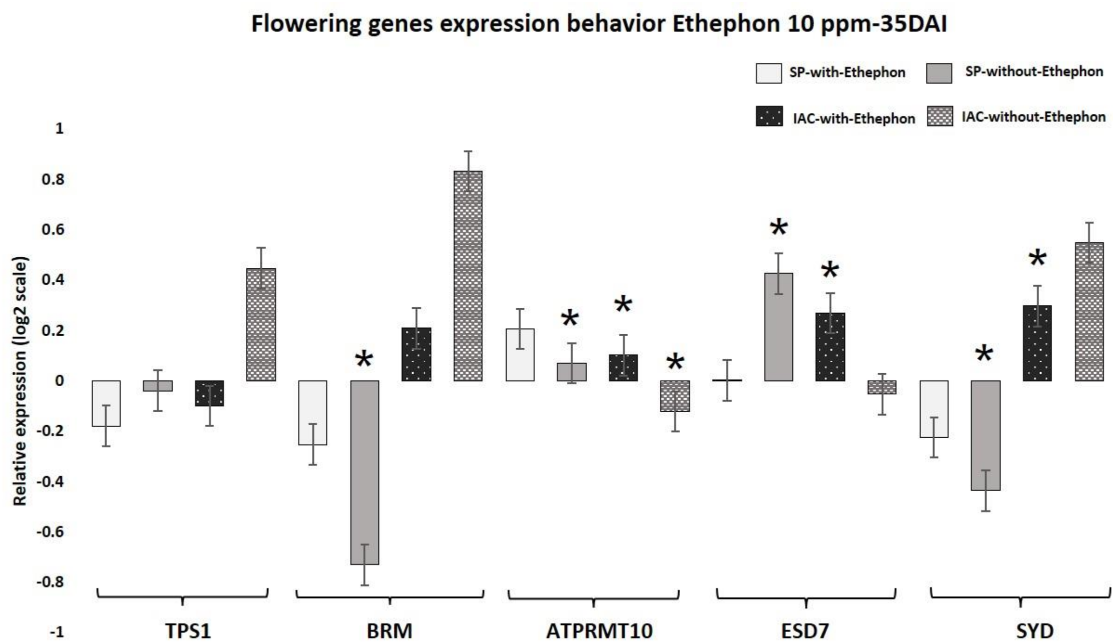


Figure 12: RT-qPCR validation-35 DAI: Flowering genes selected for RT-qPCR validation in smut resistant (SP80-3280) and susceptible (IAC66-6) genotypes treated with flowering repressor Ethephon: TPS1 (comp198109_c0_seq1), BRM (comp199879_c0_seq1), ATPRMT10 (comp184054_c0_seq1), ESD7 (comp207801_c0_seq1), SYD (comp207886_c1_seq1). REST® software was used to statistically analyze the RT-qPCR data. "✱" indicates differentially expressed genes in RT-qPCR reactions (p-value < 0.05).

4.3.3.3. Relative expression analysis (Livak method) to analyze effects of ethephon-48 hai and 35-DAI

Relative expression of the candidate flowering genes was calculated by the $2^{-\Delta\Delta CT}$ method (Livak) in smut-resistant and smut-susceptible plants to determine the effects of ethephon on a molecular level. One-way ANOVA followed by Tukey's multiple comparisons test calculated the significant variance (p-value<0.05) between, ethephon-treated and untreated samples which were either mock-inoculated or inoculated with smut fungus. This analytical approach is better than the one described earlier as it clearly delineates the causal agent (ethephon, smut fungus inoculation or neither) responsible for modulation in gene expression in two sugarcane genotypes with contrasting smut tolerance levels.

ATPRMT10- 48 hai- A comparative expression profile of ATPRMT10 gene generated via RT-qPCR analysis was mapped in smut-resistant and smut-susceptible genotypes (figure 13). To summarize, five of the comparative gene expression analysis for ATPRMT10 gene were statistically significant (p-value<0.05) among ethephon treated or untreated samples tested from two genotypes- i) Lower expression levels were observed in the IAC inoculated (ethephon-untreated) samples in comparison to the IAC-mock inoculated samples (ethephon-untreated). ii) A statistically significant variance in the expression levels of ATPRMT10 gene was observed between IAC mock inoculated (ethephon-treated) and IAC-mock inoculated (ethephon-untreated) samples. iii) Similarly, a statistically significant variance in expression levels were observed between IAC-inoculated (ethephon-treated) and IAC-inoculated (ethephon-untreated) samples. iv) Inter-genotype comparative expression analysis of SP-IAC mock-inoculated samples not subjected to ethephon treatment was statistically significant as well, with IAC mock-inoculated samples exhibiting approximately 7-folds higher expression levels than in SP-mock inoculated samples. v) Similarly, expression levels varied significantly between SP-IAC inoculated samples not subjected to ethephon treatment as IAC inoculated samples had around 6 folds higher expression levels than SP inoculated samples (figure 13).

Interestingly, three other candidate genes BRM, TPS1, and SYD had in general exactly the same expression pattern at 48 hai as described above for ATPRMT10 gene, though with varied significant variance (figure 13). ESD7 gene at 48 hai also had all

the aforementioned statistically significant expression pattern in the same samples, and additionally, had another two inter-genotype statistically significant expression profiles comparing ethephon treated SP and IAC mock inoculated and smut-inoculated samples. Ethephon treated mock inoculated IAC samples had higher expression levels than ethephon treated mock inoculated SP samples, and a similar comparative expression profile was observed in their corresponding smut-inoculated samples as well. It is worth mentioning that mock-inoculated and smut-inoculated samples of the smut-resistant and smut-susceptible genotypes under ethephon had differential expression response. SP samples treated with ethephon had higher expression levels of ESD7 in mock-inoculated samples than the inoculated samples, and on the contrary, comparatively higher expression levels of ESD7 were observed in ethephon treated inoculated samples than in mock-inoculated samples in smut-susceptible genotype, IAC66-6 (figure 13).

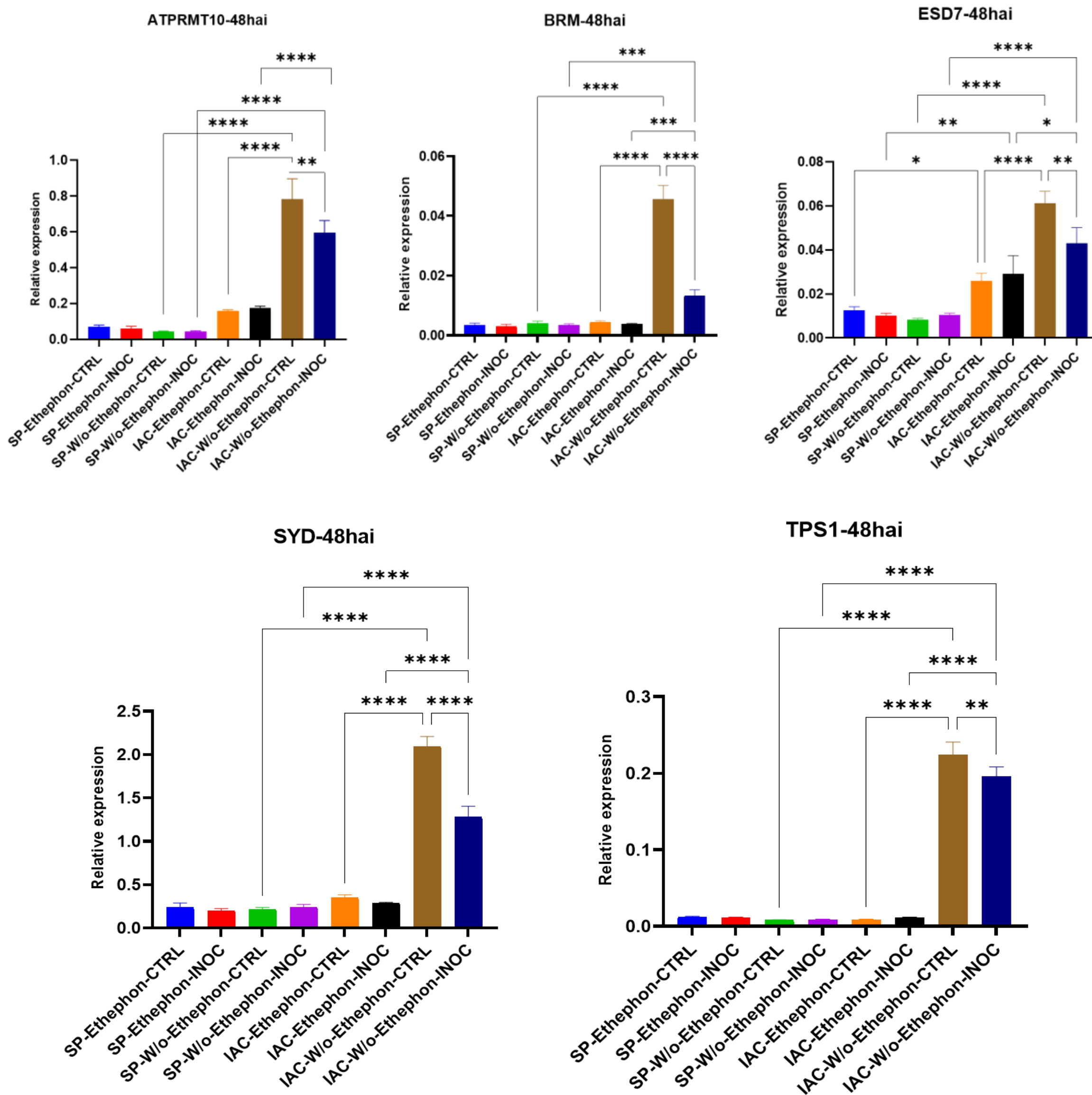


Figure 13: RT-qPCR validation-48 hai: Expression profile of flowering related genes in sugarcane smut resistant (SP80-3280) and susceptible (IAC66-6) genotypes treated with Ethephon (10 ppm): TPS1 (comp198109_c0_seq1), BRM (comp199879_c0_seq1), ATPRMT10 (comp184054_c0_seq1), ESD7 (comp207801_c0_seq1), SYD (comp207886_c1_seq1). The $2^{-\Delta\Delta CT}$ method (Livak) for relative gene expression was employed and bars demonstrate varying degree of significant variance {(p-value <0.0332 (*), <0.0021 (**), <0.0002 (***), <0.0001 (****)}. Standard errors are represented by the bars.

Expression profile of candidate genes- 35 DAI ATPRMT10-35 DAI- A comparative expression profile of ATPRMT10 gene generated via RT-qPCR was analyzed at 35 DAI in smut-resistant and smut-susceptible genotypes treated with ethephon. To summarize, seven of the comparative gene expression analysis for ATPRMT10 gene were statistically significant (p-value<0.05) among ethephon treated or untreated samples tested from two genotypes (figure 14)-

i) Lower expression levels were observed in the IAC inoculated (ethephon-untreated) samples in comparison to the IAC-mock inoculated samples (ethephon-untreated). ii) A statistically significant variance in the expression levels of ATPRMT10 gene was observed between IAC mock inoculated (ethephon-treated) and IAC-mock inoculated (ethephon-untreated) samples. iii) A statistically significant variance in expression levels between IAC-inoculated (ethephon-treated) and IAC-inoculated (ethephon-untreated) samples was observed, implying that ethephon elicits a strong expression response in smut inoculated samples than the ones not subjected to ethephon. Another noteworthy observation is that the ATPRMT10 gene is upregulated (inoculated/mock-inoculated) under the influence of ethephon in the smut-susceptible genotype, though expression was not statistically significant.

On the contrary, it is downregulated (inoculated/mock-inoculated) in samples not subjected to ethephon treatment. An exact response (statistically significant) replica was elicited in the smut-resistant plants as reported for smut-susceptible genotype, suggesting that such response at this stage is not genotype specific, and rather potentially could be a global response. iv) An Inter-genotype comparative expression analysis of SP-IAC mock-inoculated samples (ethephon-treated) was statistically significant as well, with IAC mock-inoculated samples exhibiting higher expression levels than SP-mock inoculated samples. v) Similarly, expression levels varied significantly between SP-IAC inoculated samples (ethephon-treated) as IAC inoculated samples also exhibiting higher expression levels than SP inoculated samples (figure 14). vi) ATPRMT10 gene is upregulated (inoculated/mock-inoculated) under the influence of ethephon in smut-resistant genotype at 35 DAI (figure 14).

BRM-35 DAI- A comparative expression profile of BRM gene generated via RT-qPCR was analyzed at 35 DAI in smut-resistant and smut-susceptible genotypes treated with ethephon. To summarize, nine of the comparative gene expression analysis for BRM gene were statistically significant (p-value<0.05) among ethephon treated or untreated samples tested from two genotypes (figure 14)-

i) BRM gene is upregulated (inoculated/mock-inoculated) in the smut-susceptible plants not subjected to ethephon treatment. Noteworthy to mention here is that it was downregulated at 48 hai ii) Higher expression levels (statistically significant) of BRM gene were observed in IAC mock inoculated (ethephon-treated) in comparison to the

IAC-mock inoculated (ethephon-untreated) samples. However, it had the contradictory expression profile at 48 hai. iii) A statistically significant variance in expression levels between IAC-inoculated (ethephon-treated) and IAC-inoculated (ethephon-untreated) samples was observed, implying that ethephon elicits a much stronger expression response in smut inoculated samples than the ones not subjected to ethephon.
iv) BRM gene is downregulated (inoculated/mock-inoculated) in the smut-resistant plants not subjected to ethephon treatment. v) Lower expression level (statistically significant) of BRM gene was observed in SP-mock inoculated (ethephon-treated) in comparison to the SP-mock inoculated (ethephon-untreated) samples. vi) An Inter-genotype comparative expression analysis of SP-IAC in the mock-inoculated samples (ethephon-treated) was statistically significant as well, with IAC mock-inoculated samples exhibiting higher expression levels than SP-mock inoculated samples. vii) Similarly, expression levels varied significantly between SP-IAC inoculated samples (ethephon-treated) in an Inter-genotype expression analysis as IAC inoculated samples also exhibiting higher expression levels than SP inoculated samples (figure 12). viii) Inter-genotype expression analysis also had another statistically significant expression profile where SP-mock inoculated samples (ethephon-untreated) had higher expression levels of BRM in comparison to the IAC-mock inoculated samples (ethephon-untreated). ix) SP-inoculated samples (ethephon-untreated) had lower expression levels of BRM in comparison to the IAC-inoculated samples (ethephon-untreated).

ESD7-35DAI- Briefly, only three of the comparative gene expression analysis for ESD7 gene were statistically significant (p-value<0.05) among ethephon treated or untreated samples tested from two genotypes (figure 14)- i) SP-mock inoculated samples treated with ethephon exhibited higher expression levels than the SP-mock inoculated samples not subjected to ethephon treatment, suggesting a positive regulation of ESD7 gene under the influence of ethephon in smut-resistant genotype at 35 DAI. ii) An Inter-genotype comparative expression analysis of SP-IAC in the mock-inoculated samples (ethephon-treated) was statistically significant, where SP mock-inoculated samples exhibited higher expression levels than IAC-mock inoculated samples. Antagonistic expression profile was generated for the same at 48 hai iii) Another Inter-genotype comparative expression analysis of SP-IAC in the inoculated samples (ethephon-untreated) was statistically significant, where SP-inoculated samples

exhibited higher expression levels than the IAC-mock inoculated samples and interestingly, this one also had antagonistic expression response at 48 hai (figure 13). SYD-35 DAI- In case of SYD, only 3 statistically significant comparative expression profiles were observed at 35 DAI, and all of them were inter-genotype- i) Inter-genotype comparative expression analysis of SP-IAC in the mock-inoculated samples (ethephon-treated) was statistically significant, where SP mock-inoculated samples exhibited higher expression levels than IAC-mock inoculated samples. ii) Similarly, SP-IAC samples not subjected to ethephon exhibited a similar behavior pattern. However, SP-samples had lower expression levels in comparison to IAC-samples at 48 hai, exhibiting a complete shift in expression behavior at 35 DAI. iii) A similar expression pattern was also reported in inoculated samples (ethephon-treated) of SP-IAC, with higher expression levels in SP samples in comparison to IAC samples.

TPS1-35-DAI- To summarize, four statistically significant comparative expression profiles for smut-resistant and smut-susceptible genotypes were generated 35 DAI by RT-qPCR. SP mock inoculated samples (ethephon-treated) had higher expression levels than the SP mock-inoculated samples not subjected to ethephon treatment, and similarly SP inoculated samples (ethephon-treated) exhibited higher expression levels than the SP mock-inoculated samples not subjected to ethephon treatment. Moreover, SP mock-inoculated and inoculated samples treated with ethephon comparatively exhibited higher expression levels than the IAC mock-inoculated and inoculated samples respectively (figure 14).

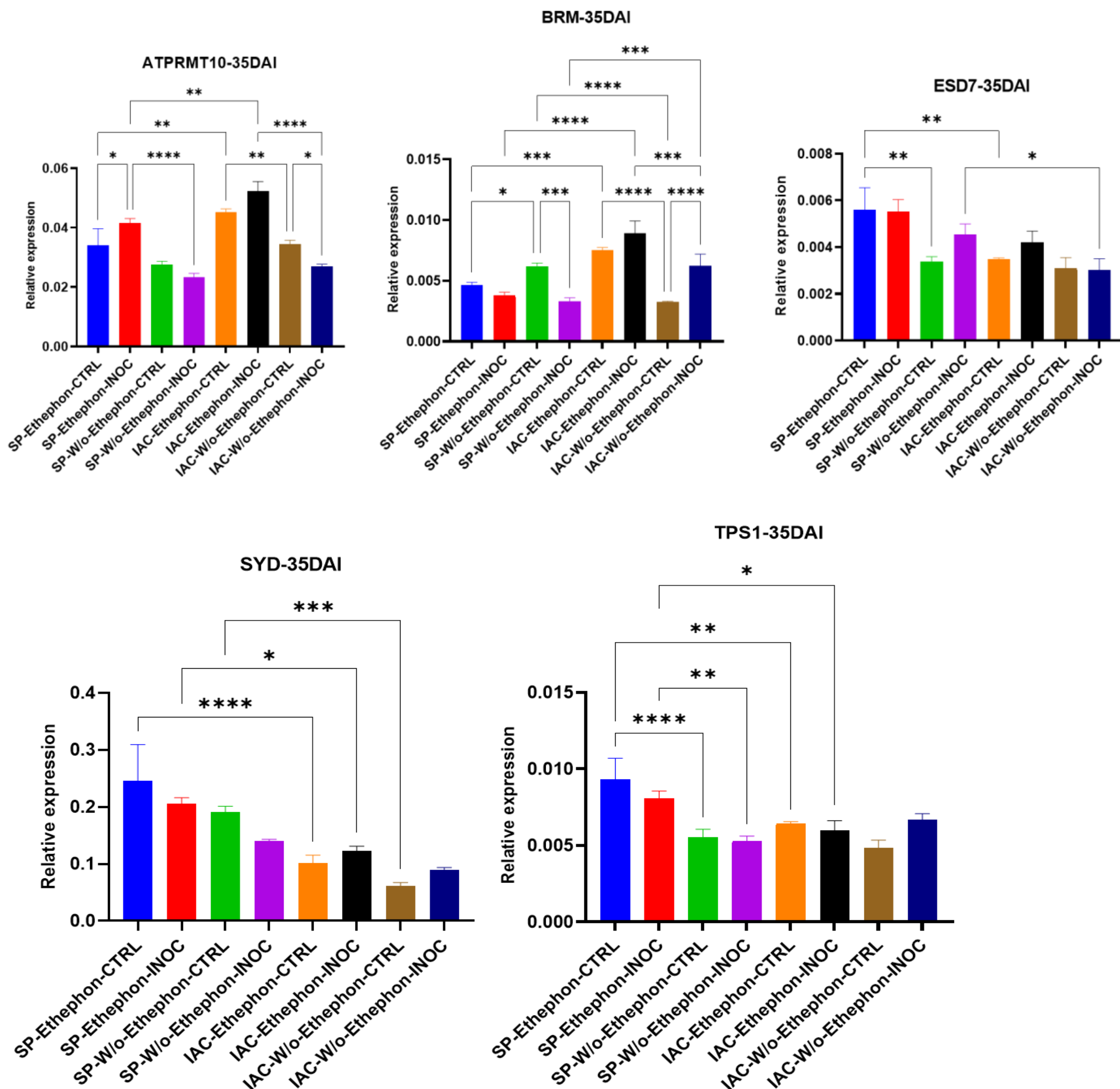


Figure 14: RT-qPCR validation-35 DAI: Expression profile of flowering related genes in sugarcane smut resistant (SP80-3280) and susceptible (IAC66-6) genotypes treated with Ethephon (10 ppm): TPS1 (comp198109_c0_seq1), BRM (comp199879_c0_seq1), ATPRMT10 (comp184054_c0_seq1), ESD7 (comp207801_c0_seq1), SYD (comp207886_c1_seq1). The $2^{-\Delta\Delta CT}$ method (Livak) for relative gene expression was employed and bars demonstrate varying degree of significant variance {(p-value <0.0332 (*), <0.0021 (**), <0.0002 (***), <0.0001 (****)}. Standard errors are represented by the bars.

4.4. Discussion

Flowering and smut disease both are detrimental to sugarcane from a commercial point of view as it limits the crop yield, and as sugarcane is one of the most valuable crops in the world, and a major contributor to Brazilian economy, therefore any factor

adversely affecting sugarcane crop yield has a great economic importance as well (Jain et al., 2018; Monteiro-Vitorello et al., 2018). Our transcriptomic data unraveled how smut fungus differentially regulates flowering related genes in smut-resistant and smut-susceptible genotypes (Chapter 2). Generally, smut fungus targets floral structures to grow and reproduce and in some crops have been shown to positively or negatively influence the flowering pathway (Glassop et al., 2013; Fan et al., 2016; Schmitz et al., 2018). Based on the aforementioned studies, and our data (chapter 2, Schaker et al., 2016) we speculate that sugarcane smut fungus also potentially influences the flowering pathway by regulating the expression of flowering related genes.

Ethephon has a multitude of applications including flowering inhibition, improvement of germination rate, enhancing sucrose levels, and in conferring drought resistance in sugarcane (Moore and Osgood, 1989; Jain et al., 2018). We conducted this exploratory study to understand whether ethephon has the potential to influence the smut disease progression or whip development in sugarcane as both ethephon and smut fungus are associated with flowering pathways. Ethephon improves the germination rate in sugarcane (Jain et al., 2018). However, we observed that this positive influence of ethephon on germination rate is rather genotype specific as smut-susceptible genotype IAC66-6 had a better germination rate than the smut-resistant genotype SP80-3280. Apparently, we did not observe any physiological changes influenced due to ethephon application as only one whip emerged in this experiment in the smut-susceptible (inoculated) plants not subjected to ethephon treatment. Initially our goal was to measure the expression levels of flowering related genes at three different time points- 48 hai (early infection), 35 DAI (vegetative stage), and after whip emergence (reproductive stage). Since only one whip emitted in this experiment, we could not measure the gene expression at the reproductive stage as we did not have enough whips to satiate the minimum number requirement for the experiment to draw a conclusion from a statistical point of view. Gene expression analysis of candidate flowering genes at two different time points provided new insights into the effect of ethephon on candidate genes tested here. To summarize, the effect of ethephon on candidate flowering genes varied depending on the genotype and the developmental stage. Ethephon seems to negatively regulate the expression of ATPRMT10 (a positive flowering regulator) in the smut-susceptible genotype at 48

hai, and this effect was solely governed by ethephon as the samples were mock-inoculated. Interestingly, a similar expression profile was observed for all the candidate flowering genes in IAC in mock inoculated samples at 48 hai. It will be very difficult to draw such an inference from the inoculated samples as the change in expression behavior could either be due to ethephon, smut fungus or a combined effect of both. An inter-genotype gene expression comparison of ethephon untreated samples clearly demonstrates that smut infection elicits a much stronger response of ATPRMT10 gene (positive regulator of flowering) in the smut-susceptible genotype (IAC66-6) than smut-resistant genotype (SP80-3280).

Noteworthy to mention here is that apparently all the candidate genes at 48 hai in smut-susceptible genotype, not subjected to ethephon treatment were downregulated, and apparently, no such expression behavior pattern was observed in smut-resistant genotype (ethephon-untreated) suggesting that smut fungus elicits a much stronger response in the smut-susceptible genotype. Strikingly, one another clear expression pattern was observed for all the candidate genes at 48 hai in smut-susceptible mock-inoculated and inoculated plants, where smut-susceptible plants treated with ethephon had lower expression levels than the ones not subjected to ethephon treatment, suggesting that the ethephon potentially minimizes the effect of smut fungus on the expression level of candidate flowering genes (figure 13).

Higher expression levels of all the candidate genes in ethephon treated mock inoculated samples in comparison to mock inoculated samples not subjected to ethephon treatment demonstrates that ethephon positively regulates the expression of all the candidate flowering genes in smut susceptible genotype and interestingly, ATPRMT10, BRM and ESD7 had the antagonistic expression response at 48 hai. Inter-genotype expression level comparison revealed that SP mock-inoculated and inoculated samples (ethephon-treated) had higher expression levels of ESD7, SYD, and TPS1 than in comparison to IAC mock-inoculated and inoculated samples (ethephon-treated) respectively at 35 DAI. However, ATPRMT10 and BRM expression profiles were exactly opposite for the same samples suggesting that ethephon elicits a different response from the candidate genes at this developmental stage. Comparatively, ethephon had a more uniform and genotype specific effect at 48 hai. Noteworthy to mention here, is that the candidate flowering genes tested here

are the constituents of different pathways, and there is a very limited understanding in relation to the potential crosstalk between ethephon and the pathways these selected candidate genes belong to. We did a thorough search to identify experimental data depicting any direct or indirect interaction of ethylene with these five candidate genes and their respective pathways. However, we could not find any useful information related to this study. Moreover, we detected a limited representation of the genes attributed to hormone biosynthesis in our transcriptomic study (Chapter 2).

Two paralogs of GA2ox gene (GA2ox2 and GA2ox8), which catabolize the active GA content, were the only differentially expressed genes in our transcriptomic data. Interestingly, exogenously applied ethylene has a negative effect on the gibberellin metabolism (Dugardeyn et al., 2008), and since ethephon is an ethylene releasing compound, the ethephon application might positively regulate the expression of two genes downregulated in the susceptible genotype. GA2ox is a negative regulator of flowering and any modulation in the expression behavior of these genes might have a negative impact in a sugarcane-smut compatible interaction. Hormonal crosstalk plays a vital role in plant growth and development, and how ethephon influences this crosstalk needs to be investigated in detail in the sugarcane-smut pathosystem.

4.5. Conclusion

In our investigative study here, we used ethephon, a known flowering repressor in sugarcane to explore if it has the potential to affect smut disease progression or whip development process at a physiological and molecular level. The candidate flowering genes used to measure gene expression level in a time course manner via RT-qPCR analysis were picked from differential expressed flowering genes detected in an early response transcriptomic study of smut-resistant and smut-susceptible genotypes. To our best knowledge, there are no reports of use of ethephon to explore such avenues and though our work was entirely exploratory in nature, we were able to define the effects of ethephon on the gene expression level of flowering-related genes in smut-resistant and smut-susceptible genotypes at 48 hai and 35 DAI. We did not observe any effect of ethephon on smut-whip disease progression and whip development process on a physiological level as only one whip emitted from infected plants in smut-susceptible genotype (IAC66-6) not subjected to ethephon treatment. We also

observed that ethephon influences the germination rate in a genotype specific manner, where it might affect the germination rate either positively or negatively. We were able to report that the effect of ethephon on candidate flowering genes varied depending on the genotype and the developmental stage.

References

Abeles FB, Morgan PW, Saltveit ME (1992). The role of ethylene in agriculture in Ethylene in plant biology (eds Abeles, F. B., Morgan, P. W. & Saltveit, M. E.) 264–296 (San Diego Academic Press, 1992).

Agrawal GK, Rakwal R, Jwa NS (2001). Differential induction of three pathogenesis-related genes, PR10, PR1b and PR5 by the ethylene generator ethephon under light and dark in rice (Oryza sativa L.) seedlings. J Plant Physiol. 2001; 158: 133–137.

Binder BM (2020). Ethylene signaling in plants. *The Journal of biological chemistry*, *295*(22), 7710–7725.

Binder BM, Walker JM, Gagne JM, Emborg TJ, Hemmann G, Bleecker AB, Vierstra RD (2007). The Arabidopsis EIN3-binding F-box proteins, EBF1 and 2 have distinct but overlapping roles in regulating ethylene signaling. Plant Cell 19, 509–523.

Broekaert WF, Delaure SL, De Bolle MFC, Cammue BPA (2006). The role of ethylene in host-pathogen interactions. Annual Review of Phytopathology. 2006; 44: 393–416.

Bueno CRNC. (2010). Infection by *Sporisorium scitamineum* on sugarcane: influence of environmental variables and development of a method for early diagnosis. Universidade de São Paulo.

Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, Nery JR, Qiao H, Yang A, Jamali A, Chen H, Ideker T, Ren B, Bar-Joseph Z, Hughes TR, Ecker JR (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife. 2013 Jun 11;2: e00675.

Cunha C, Roberto G, Vicentini R. *et al.* (2018). Ethylene-induced transcriptional and hormonal responses at the onset of sugarcane ripening. *Sci Rep* 7, 43364 (2017).

De Paepe A, Vuylsteke M, Van Hummelen P, Zabeau M, Van Der Straeten D (2004). Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in Arabidopsis. Plant J. 39, 537–559.
Donaldson RA and Staden JV (1989). A review of chemicals used as ripeners of sugarcane in South Africa. XX Proc. ISSCT, pp, 647-655.

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus (Madison). 12, 13–15.

Dugardeyn J, Vandenbussche F, Straeten DVD (2008). To grow or not to grow: what can we learn on ethylene–gibberellin cross-talk by *in silico* gene expression analysis? *Journal of Experimental Botany*, Volume 59, Issue 1, January 2008, Pages 1–16.

Fan J, Yang J, Wang YQ, Li GB, Li Y, Huang F, Wang WM (2016). Current understanding on *Villosiclava virens*, a unique flower-infecting fungus causing rice false smut disease. Mol Plant Pathol. 2016 Dec;17(9):1321-1330.

Fong Chong B, Mills E, Bonnett GD, Gnanasambandam A (2010). Early exposure to ethylene modifies shoot development and increases sucrose accumulation rate in sugarcane. J. Plant Growth Regul. 29, 149–163 (2010).

Glassop D, Rae AL, Bonnett GD (2013). Sugarcane Flowering Genes and Pathways in Relation to Vegetative Regression. *Sugar Tech* 16, 235–240.

Helliwell EE, Wang Q, Yang Y (2013). Transgenic rice with inducible ethylene production exhibits broad-spectrum disease resistance to the fungal pathogens Magnaporthe oryzae and Rhizoctonia solani. Plant Biotechnol J. 2013 Jan;11(1):33-42.

Iqbal N, Umar S, Per TS, Khan NA (2017). Ethephon increases photosynthetic-nitrogen use efficiency, proline and antioxidant metabolism to alleviate decreased photosynthesis under salinity stress in mustard. Plant Signaling Behavior.

Islam MT, Liu J, Sherif SM (2021). Ethephon-Mediated Bloom Delay in Peach Is Associated with Alterations in Reactive Oxygen Species, Antioxidants, and Carbohydrate Metabolism During Dormancy. Front Plant Sci. 2021 Oct 14; 12:765357.

Jain R, Singh A, Singh S and Chandra A (2018). Influence of bud sett soaking with ethephon and its foliar application on growth, biochemical attributes and gene expression in sugarcane." *Journal of Environmental Biology* 39 (2018): 472-480.

Li C, Wang Y, Ying P, Ma W, Li J (2015). Genome-wide digital transcript analysis of putative fruitlet abscission related genes regulated by ethephon in litchi. Front. Plant Sci. 6, 502.

Li Q, Guo W, Chen B, Pan F, Yang H, Zhou J, Wang G and Li X (2021). Transcriptional and Hormonal Responses in Ethephon-Induced Promotion of Femaleness in Pumpkin. Front. Plant Sci. 12:715487.

Li W, Ma M, Feng Y, Li H, Wang Y, Ma Y, Li M, An F, Guo H (2015). EIN2-directed translational regulation of ethylene signaling in Arabidopsis. Cell 163, 670–683.

Li Y and Solomon S (2003). Ethephon: A versatile growth regulator for the sugarcane industry. *Sugar Tech* 5, 213–223 (2003).

Li YJ, Yang LT, Li YR, Ye YP (2002). Influence of ethephon sprayed at different stages on growth, agronomic traits and drought resistance of sugarcane. Sugarcane, 9(1): 12-18.

Liao F, Lin JZ, Li YR, Yang LT, Ye YP, Xue H (2003). Effects of ethephon on the leaf microstructures related to drought-resistance in sugarcane. Southwest China Journal of Agricultural Sciences, 16(2): 45-48.

Livak KJ, and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta-Delta Ct)) Method. Methods. 2001 Dec;25 (4):402-8.

Monteiro-Vitorello CB, Schacker PDC, Benevenuto J, Teixeira e Silva NS, Almeida SS (2018). Progress in understanding fungal diseases affecting sugarcane: smut. In. Philippe Root (ed). Achieving sustainable cultivation of sugarcane Volume 2. Breeding, pests and diseases. Chapter 10. 2018. ISBN-13: 9781786761484.

Moore PH, and Osgood RV (1989). Prevention of flowering and increasing sugar yield of sugarcane by application of ethephon (2-chloroethylphosphonic acid). *J Plant Growth Regul* 8, 205.

Nemhauser JL, Hong F, Chory J (2006). Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell 126, 467–475

Peters L, Carvalho G, Vilhena MB, Creste S, Azevedo RA, Monteiro-Vitorello CB (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and susceptible genotypes. Planta 2017, 245, 749–764.

Ruduś I, Sasiak M, Kepczyński J (2013). Regulation of ethylene biosynthesis at the level of 1-aminocyclopropane-1-carboxylate oxidase (ACO) gene. Acta Physiologiae Plantarum 35, 295–307

Santa Brigida AB, Rojas CA, Grativol C, de Armas EM, Entenza JO, Thiebaut F, Lima MF, Farrinelli L, Hemerly AS, Lifschitz S, Ferreira PC (2016). Sugarcane transcriptome analysis in response to infection caused by *Acidovorax avenae* subsp. Avenae. PLoS One. 2016 Dec 9;11(12): e0166473.

Schaker PDC, Palhares AC, Taniguti LM, Peters LP, Creste S, Aitken KS, Van Sluys MA, Kitajima JP, Vieira MLC, Monteiro-Vitorello CB (2016). RNAseq transcriptional profiling following whip development in sugarcane smut disease. PLoS One E 11 (9), e0162237.

Schmitz L, McCotter S, Kretschmer M, et al (2018). Transcripts and tumors: regulatory and metabolic programming during biotrophic phytopathogenesis. F1000Research 7 (F1000 Faculty Rev):1812.

Shen X, Liu H, Yuan B, Li X, Xu C, Wang S (2011). OsEDR1 negatively regulates rice bacterial resistance via activation of ethylene biosynthesis. Plant, cell & environment. 2011; 34: 179–91.

Silva MA, and Caputo MM (2012). Ripening and the use of ripeners for better sugarcane management in Crop management: cases and tools for higher yield and sustainability (ed. Marin, F. R.) 1–24 (InTech, 2012).

Sun ZY, Zhang TJ, Su JQ. *et al.* (2015). A novel role of ethephon in controlling the noxious weed *Ipomoea cairica* (Linn.) Sweet. *Sci Rep* 5, 11372.

Wang W, Hall AE, O'Malley R, Bleecker A B (2003). Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from Arabidopsis is not required for signal transmission. Proc. Natl. Acad. Sci. U.S.A. 100, 352–357.

Yoshida S, Ishida JK, Kamal NM *et al.* (2010). A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, *Striga hermonthica*. *BMC Plant Biol* 10, 55 (2010).

Zhong GY and Burns JK (2003). Profiling ethylene-regulated gene expression in Arabidopsis thaliana by microarray analysis. Plant Mol. Biol. 53, 117–131 (2003).

APPENDICES

Appendix A: COMPGG (nucleotide) sequences used to assemble the FLOR dataset

Gene: PIE1
>comp205700_c0_seq1
CTGCCAGATCCCGCCGCGTCACCTGCAGTTGCTGCACTGTTTGTATACCAGTCCCAGTTTGCTGCACATGCTCTAAGATATGAGATAGCCCTG
TGCGAGTGATTGGAAAGATGGCATCAAAAGGTCCCCGGTCAAAGCTAGACCATGAGACAAGAGCTAGGCGTCAAAAGGCTCTTGAAGCTCCA
AGGGAGCCTCGACGGCCAAAAGTTCACTGGGACCATGTCCTTGCTGAGATGGTGTGGCTGGCAAAGGAGTTTGATTCTGAGAGAAAATGGAA
GTTGTCCATGGCCAAAAAGATTGCTCAAAGAGCAAATAAGAGCATAGTTGACCAAGCAACAAAGGGCGAGAGAAAACAGAAGGAGGAAGAAC
ACAGAATGCGAAAAGTTGCCCTTAATATTTCAAAGGATGTGAAGAAGTTCTGGATTAAAATAGAAAAGCTGGTTGTTTATAAGCATCAACTGGA
GGTTGAGGAGAGAAAGAAGAAGGCACTTGATAAGCAACTTGATTTCCTTTTAGGTCAGACTGAGAGGTATTCGACAATGTTGGCGGAAAACCT
TGTGGACATGCCCTATTCCCAAAACCTAGAAAACGGGACTTCGCAGATAAACCAACCATCCCATCAAGAAGAAGTAGCAGAAGAGAACATAAA
TGCAGCCATTCCTGATGATCTTGATAATATGGAAGTCGATGATGATTATGAGAGCTCCTTGGATGAAGAGCCAGAAGATGATGAGCACACAAT
CGATGTGGATGAAGCTCAGATCACCGAAGCTGAGCGGAATGAAGAACTAGCTGCTTTAGAGGCAGAAGCTGACCTACCACTGGATGTTATTCT
CAAGATGTATACCAAAACCAAAGTTAGCAGGGAAAGCAGTCCAGACAGCAACGACATGTTAAGTGACTTAGGTTCGAAGAATTTGATTGTAGAT
TCTTCGAACCAGGCTAATGGCTGTGACCATGAACCTGCTCACTCTTCAAGTGACGATGGCAATTCTTCTGAGGAAGTGGATGATGGTCAATCT
TATGCTGAGTTTGTGAAGAAAAAGTCATGGTAAAAGTAACGGAAATATTTCTTCTATCAATGATCAGGAGGACGAAGATTACATTGCTGCTGAAG
AAGTAAAGGATGACGAAACAACTTTGTCTGAAGAGGAAGAATTGGCAAAGAAAGACGATCCTGATCATCTGGAAGAGATTAAGTTACTGCAGA
AGGAGAGTGAGATACCACTAGAAGAACTTCTCGCGAGGTACCGAAAGGATGGCTGTGCAGATCATGAAACAGCGGAGTTGGAGAATTCACCC
CATTTTGTCGAAGAGGTCAATACTGACATGTCTTTGGATGATCAATCTGTGAACATTTTGGAAGCGAAAAGTGATATATTTGTGGATCACCAATC
CAGGGATGTGCTGGAAACTGAGCACAAGGCTCTACAATCAGAAATCGATGTGGTGGAAACTGAGCAGGAGGCTCTACAATCAGAAATCGCAT
CAGAGCCTTGTGCTCAACAAAATTTTGTGGAAGAGAATAATCTCACTGATGTTAAGGCAGTCCATGGAGATAAGAGTGATGATGTAATTGCTGA
TGCTGCAGCTGCTGCAAGATCAGCACAACCAACTGGCAACACCTTCTTGACTACAAAAGTGCGCACGAAATTCCCATTCCTTCTTAAGCATTCT
CTTCGTGAATACCAGCATATAGGGCTGGATTGGTTGGTTGCTATGTATGAGAAGAGGCTTAATGGAATTCTAGCAGACGAAATGGGTTTAGGC
AAGACAATCATGACTATCTCCTTGCTTGCACACCTTGCGTGTGAGAAGGGGATATGGGGTCCTCATCTTATTGTCGTGCCGACTAGTGTTATG
CTAAATTGGGAGACTGAATTTCTTAAGTGGTGTCCTGCTTTTAAAATACTGACTTATTTTGGAAGTGCAAAGGAGAGAAAGCAGAAACGTCAGG
GCTGGATGAAACCAAATTACTTCCATGTTTGCATCACAACATATAGGCTTGTCATTCAGGACTCTAAAGTGTTTAAGCGAAAAAAGTGGAAGTA
TCTTATTCTTGATGAGGCTCATCTGATAAAGAACTGGAAATCACAGCGTTGGCAGACTTTGCTCAACTTTAATTCAAAGAGACGTATTCTTCTGA
CTGGAACTCCACTGCAAAATGACCTTATGGAACTTTGGTCCCTCATGCATTTTCTGATGCCTCATGTATTTCAGTCTCATCAAGAGTTCAAGGAT
TGGTTCTGCAATCCAATTTCTGGGATGGTGGAAGGCCAAGATAAAGTAAACAAAGAAGTTATAGATCGATTGCACAATGTCCTTCGTCCATTTA
TACTGCGTCGACTGAAAAGGGATGTTGAGAAACAGTTACCAAGGAAGCATGAACATGTCATATATTGCCGACTTTCTAGAAGGCAAAGGAACT
TGTATGAAGATTTTATCGCTAGCTCAGAGACACAAGCAACACTGGCAAGTGGGAATTATTTTGGCATGATAAGCATCATTATGCAACTTAGAAA
GGTCTGTAACCATCCAGATCTTTTTGAAGGTCGCCCAATAATAAGCTCATTTGATATGGCTGGGATTAACATGCAGCTCAGCTATTCAGTCTGC
ATGCTCCTTGATAAGAGTCCATTTTCTCAGGTGGACCTGTCTGATATGAATTTTGTGTTTACTCAAAATGAATTTAGCATGAGTTCCTGGGAAGC
TGACGAGGTCATCTCTGCCTTTCCTGCAAGTATCACCTCCAGGAATTGTGACTTGGATATTTTTTGCTCTAATAAGGATCATCAGGGAAGTAAT
TTGACAAACATTTTTGAAGATATTCAGAAAGCCCTACAGGAGGAGAGAATAAAGGAATCCAGAGAAAGGGCAGCTTCCATTGCATGGTGGAAT
AGAGTTAGATGCCAAAAGAGGCCTGTCTATGGCACAAACACGAGACATGTTTTGACTGTAAATCATCCTGTATCAAATATTCTTGATAAGAGGA
ACAACCCTTTGTGCCACATGGACTATTCATCAAGCCTTGCAGATCTTGTTCTTCCATCTGTGGAACGCTTTCAGAAAATGCTTGATGTTGTTGAA
TCATTTACATTCGCAATTCCTGCTGCTCGAGCTCCTCCACCTGTTTGCTGGTGCAGCAAAGGAAAGTCTCCTGTTTTTATTGATCCGGAATATA
GAGAAAAATGCATGAACAAGTTTTCTCCCACTCTGTCTCCTATAAGGTCTGCTATTGTTCGTCGTCAAGTCTACTTTCCTGATAGGCGTTTGATC
CAGTTTGATTGTGGGAAGTTGCAGGAACTTGCAATTCTGTTAAGGCGTTTGAAGTCAGAAGGGCACAGGGCCTTGATATTTACTCAGATGACT
AAGATGCTTGACGTCTTGGAAGAGTTCATAAATTTATATGGGTATACATATTTACGTTTAGATGGTTCTACTCCCCCAGAAGAGAGGCAGACAC
TGATGCAGAGGTTCAACACAAACCCAAAGTTTTTCCTTTTCATTTTGTCCACTCGTAGTGGTGGTGTGGGAATCAACCTTGTAGGTGCAGACAC
TGTCATCTTCTATGACAGTGACTGGAACCCTGCAATGGACCAACAAGCACAGGACAGATGCCACAGGATTGGTCAGACTCGTGAAGTTCACAT
ATATAGACTCATTAGTGAAAGCACTATAGAGGAGAACATTCTCAAGAAAGCAAATCAGAAACGAGCTCTTGATGATTTAGTGATACAACGTGGT
AGCTACAATACGGAATTCTTCAAGAAACTTGATCCTATGGAGTTCTTTTCCGGGCACACGTCTCTCCATGCGGAAGATCAGGAGAAGAATTGCT
CTACAACCGCGGGAGCTTCAAATGATGTGGATCTGGCTCTGTCAAATGCAGATGTTGAAGCAGCTATTAGACAGGCAGAAGATGAAGCTGACT
ATATGGCTCTCAAGAAGCTGGAGGAGGAAGAAGCCGTGGACAACCAAGAATTCAGTGAGGAGGTTGCTGGCAGACCAGAGGATGATGAATTG
GTTAATGAGGAGGATGGAAAACCTGATGAGCACATTAATGAAGAACACAAATACAACTCTTCTGATGTAGAGAAGGAGAAGCACATTACTTTGT
CTACCAAGCGATTGAATGATGAGAAGGTTCTTACATTAGCTGTTGGTGATGAAGATACTGACATGCTTGCCGATGTGAAACAGATGGCAGCTG
CAGCAGCTGCAGCAGGGCAAGCAAGTTCATCTTTTGAGAACCAGCTCCGTCCAATTGATAGATATGCAATGCGCTTTATGGAACTCTGGGATC
CAGTAATTGACAAAGCTGCTGTGAATCATCAAGTAAATGTTGAGGAGGAAGAATGGGAGCTTGATCGTATTGAAAAACTCAAAGAGGATTTGG
AAGCAGAAATTGATGAAGACCAAGAACCACTGTCTTATGAATCATGGGATGTTGATTTTGCTACAACAGCCTATCGGCAACATGTTGAGGCTTT
AACTCAAAAGCAGTTGTTGGAAGAACAGGAAAGACAGGCTCAGGAAGCAGCAAAAGAGTTGGAGGAGAAGAATGATAATATGAGCAGTCACC
GCAGAAAGTCAAAAAAGAACAAAAAGAAGACAGGCAAATTCAAGTCCCTGAAAAGAGTGCGTTTGTCGTCTGAATCAGAAGTCGTACTGGAGG
AAACCTCTGTAGATACAATGAGCATTGATGACAATGCACCCTCACCTGAGCTTATAAGTGATGAATCGCCACACCATTATTCTAACAAGCGTAA
GAAGATTATGTCTGCTACTGAGGAAGATAGTAACAGCAGAAGTTTAAAGAAGTTTAAGAAAACTACTAAGTCAAGTTTTATTTCTGAAGCCTTGT
CACCTAGGCTGAAGGAGGACCTTAATGATTCAGATCCAAAATCAATGGCTAGAACTAAAAGTGATGGCAGAATTTCCATCCCTTGCATGCCAG
TAAAACGTGTTATTGTAATAAAGCCTGAGAGATTGAAAAGGAAGGGAATATGGTCTCGAGATTGTGCTTCAGACTCATGGACATCTGAGCAAGA
TGCAGTTCTTTGTGGAACTGTGCATGAGTATGGTCCTCTTTGGGAATTGGCAAGTGATTTTCTTCATTCCTTGCCAGGTGGGTCCTATAGGGGA
AGATATCATCATCCTGTGCATTGCTGTGAGAGATACAGAGAACTGTTCTGCAAACATGCAATATCAGCAACAGATAATTCTAACAATGAAAAGG
TTCCTTCTGGGACTGGAAAGGCTATACTTAGAGTATCCGAGGATCAAGCTCAGATGTTGGTGAATGTGACCAGTGAACTTCCTAACAATGAATT
GCTTCTCCAGAAACACTTCATGGCTGTAATTTCGTCTGTCTGGAGATCAAAATGTCGCCGTGATCCCTGCTGTTTTACGAATACTTACTCTAGT
GCATTACATATGTTTTCTCCTGTGAAGAAGCCTGGTGGATCAAGTGAGAACTGGCCCATGGTAAACTTTAGACCAAGCTTCAATCTAGTTAGGA
AAGCCCTTGCAGATGCTCAGGCTAAGTCTACACTAATGGTGATTCCACCACTGTCAAGAAATCAGGAATACCGGCGGAATTATTTAGAATTAGA
GTTAGATTTCTTGAAAGATCAACATGCTTATGAGGAAGACTTCCCATCTGTAATAAATGTGTCCATACTGGAACCAGAACCATCCAAACAGGCT
CCAGAGCCAGTGGAACAATCATTATTGTCTGGACTTTCTTGTAGACAAGCTGAGAACCGACTCAGGATAGCATCAGAAGCTTGTTATGATGGT
GATAGTTCTCATTGGGCATCATCAGCTTTCCATATAAATGATGCCACCCGCCATAAATCTGGTTCAAAGTCCATAGGAAAGCACAAAGCAGCAT
CCGAATCTGGTAGGCCTACCAAATCCAAAATCCAGAAGATTACTGAATCGCATCAGGAAGGGCCAACTGTCATGAGTAATTTTCTCCGTATGC
CTGCTCAACTATTGCCAAGTACAGCTGACTTCCACATTAGCGACTCCCTATCTGAATTTGGTATCAGCGATTCTGAATTCCACTACTCTGAGGA
TCTCTTGCAAGAGGTCAACGATCTCGAGTTCTTCCCAGATCAGGGTGACTCTGGTCTCCTTCCTGGTATTGAAGAGTTAGAACCTCTTTCAGAT

TTCACAGATATCGGATGAACATGGATGGGCAAATTATTCTTGATGGGATGCCCTTCTATACCAAGAGGAGACAAGATTGCCTCTGGCAGGAGC
CCAAGGATATAGGTTGAGTTGCTCAGCCATGCACATATGTGCTGCAAGCTTGATCATGGTTGGCTGGAGCAAAGCTGCTGGCTTGGTGAGGT
CACATTGATCTGACAAGCTTGGAGGATGGGTTTGCACTAATGGCGGCATTGGGTAATTTCCATACATGTATGTATATTGGTAGTCTAGAATCGT
AGGATGTAGAAATGCCTCCCTCATTTGTGTTTGGGGGTAGACCTTGTAACTCGACCACATGGAAATTTTGTTTGAGCAGGTACCAACAGAGAA
GGGAAAAGAGGAGGCCCTGAGCATCACCCATTTGTTGCTGTATCTCGTGTTATGTTCATTCTCTGTTTAACCCAATGTTGTTTGGCTATCTTGC
CATCGAAAACATCTCAGATGTCAAGTTGGTGGTTGCCATCTCTATCAGAAATTTAGGATGTGGCGCCAATGTAGTTGCCAAAGAATGATCAGTC
AAATGCAGCTTATGGGCGCATTATATTTGGTCAGGAATCAATGCATGGTATCAATACTGTCAGAACTGGTGTTCCAGTTTCACTTGAATACAATT
TCTGATTTGATCGAGACCCCGGAGGCAAGTTTTTAGGAGGTAAAGAGCGTGGCCATAGTTATACATGTATGTAATAAAATGATGTAAGGTAGTA
ATCTTCACTCTGGACTTTTGGGCTTATGGTTCTGTGGACTCTCGGTCTCGGTTATGCCAGTTAATAGCCATAACTTCTTGTTGGAGGCGAGGG
GGCTTGTAGAACAGCGGGCTTCTGCCGTGAGCCTGTAACAATGCAATTGATTGTGCGCCACCATACTGGTTTTTTTTGGTAAGTAGACGTTGC
ATCACTGCAAGGCAACGTTGAGAACATGAGATAGGGTAACTGGTCTGTTTCGCTCAAAAATACTCTGAGCTTGTATTCAGTCTTGGCCTGCGAT
AACCTGTCAGCTGAGAATGTTTGGATAATCCCATAATATGTAAGCTGACTAGCTGAGAATGTCATGTTTGGCCAATCAGGTCTAGATGCTGTCT
GTAAGCTGAGAATATTTGGCCAATATCAAGGTCTAGATGCTGTCCCTCGCCAAGCATTTTATGCTTCAAAAAAAA

Gene: ESD7, TIL1, ABO4

>comp207801_c0_seq1

GGAATAAGAGTGTGGAAGCGGGTGAGTGTCGGGCTAAATTTTCTTGCCATCTTGACTGTTTTGCTTGGGTGAAGCGAGACAGTTACCTTCCAC
AAGGAAGCCAAGGTCTAAAGGCTGTTACAAAAGCCAAGCTAGGTTACGATCCTCTGGAGGTTAATCCAGAAGATATGGTTCGCTTTGCAATGG
AGCAGCCACAGACAATGGCTTCATATTCTGTATCAGATGCTGTTGCGACATACTATCTATACATGACTTATGTCCATCCTTTCATCTTTTCTCTT
GCAACCATAATACCTATGTCACCTGATGAGGTACTGCGGAAAGGAAGTGGGACACTTTGTGAAATGCTGCTAATGGTTCAGGCATTCAAGGCT
AGTGTCATTTGTCCTAACAAGCACCAAGCTGATCTTGAGAAGTTCTACAATAACCGTTTATTGGAAAGTGAAACCTACATTGGTGGTCATGTTG
AGTGTCTTGAAACTGGTGTGTTTAGATCTGATCTCCCCACGAAGTTCCAACTTGAACCATTAGCATATGAGCAACTCATTGGAAATCTTGATCG
TGATCTGCAATATGCCATTGCTGTAGAAGGAAAGTTGGATATTGATTCTGTCACAAATTATGATGAAGTCAAGGATGCCATAGAGCAAAAGCTT
GTTGCATTGCGAGATCACCCAATTCGTGAAGAACGGCCTCTTATCTATCATCTGGATGTTGCTGCTATGTATCCAAATATTATTTTAACTAATAG
ACTCCAGCCACCATCCATAGTTACAGATGTGGACTGCACGGCATGTGATTTTAATCGTCCTGGAAAAAATTGCCTCAGAAAGCTTGAATGGGT
CTGGCGAGGAGAGACCTACATGGCAAAAAAGAATGACTATTATCACATAAAGAGGCAAATTGAGTCAGAGCTGATTCAATCTGGTGGCATCGC
ATCGTCAAAGCCTTTTCTTGACCTCTCAAAACCGGAACATTTACTTAAATTGAAGGACCGTTTGAAGAAATACTGCCAAAAGGCGCACAAAAGA
GTTGTTGATAAACCAATTACTGAAGTCAGAGAAGCTGGAATATGCATGCGTGAAAATTCTTTCTATGTAGACACAGTACGAAGCTTTCGTGATA
GAAGATATGAATACAAAGGTCTTAACAAAACATGGAAAGGAAAATTGGCAGAAGCGAAAGCTAGTGGAAATTCTATGAAAATTCAGGAAGCGC
AGGACATGGTTGTCCTATATGATTCTTTGCAACTTGCTCACAAGTGCATATTGAATTCTTTTTATGGATATGTTATGCGCAAGGGAGCAAGATG
GTACTCTATGGAAATGGCTGGTGTCGTAACATATACTGGTGCAAAGATTATCCAAAATGCTCGATTACTTGTGGAAAAAATTGGAAGGCCACTG
GAACTGGATACAGACGGCATTTGGTGTGTTTTGCCTGGCTCATTTCCAGAAAACTTTACTTTCAAGACAAAAGCTGGGAAGAAGTTAACAATAT
CTTATCCCTGTGTCATGCTTAATGTTGACGTTGCAAGGAACAACACTAATGATCAATATCAAACATTGAAAGACCCAGTAAATAAGACATATACA
ACACACAGTGAGTGTTCCATTGAGTTTGAGGTTGATGGACCTTACAAGGCCATGATCTTACCTGCCTCTAAAGAGGAAGGGATTCTGATAAAG
AAGAGGTATGCTGTCTTCAATGAAGATGGCACCTTGGCGGAACTCAAAGGTTTTGAAATCAAGCGTAGAGGTGAGCTGAAGCTCATAAAAGTT
TTTCAGGCGGAGGTATTTGACAAATTTCTTCATGGTTCAACTTTGGAGGAATGCTATTCTGCTGTTGCCTCTGTTGCTAACCGGTGGTTGGACC
TACTGGATAATCAAGGAATTGATATTGCTGACAGTGAATTGCTTGATTTCATATCGGAGTCGAGCACAATGAGTAAATCTTTAGTTGACTACGG
AGAGCAGAAGTCATGTGCTGTAACTACAGCAAAAAGACTTGCCGAATTCCTTGGAGATTCAATGGTTAAAGACAAAGGGCTACATTGCCAATAT
ATCGTTGCTCGGGAGCCACAAGGCACTCCAGTGAGCGAGCGTGCTGTTCCAGTAGCTATATTCGAGACAGATCCTGAGATTGCAAAGCATTAT
TTAAGGAAATGGTGTAGAATCTCCTCTGATGCAAGCATTCGAACTATTGTTGATTGGTCTTACTACAAACAACGCTTAAGTTCAGCTATTCAGAA
AATCATAACAATTCCTGCAGCAATGCAAAAGATCTCAAATCCTGTTCCTCGAGTTCTTCATCCAGATTGGTTACACAAAAAGGTCAGGGAGAAA
GATGATCGATTTCGGCAGCGTAAACTGCGTGATATCTTCAGTCCCTTGGCTAAAGAAGAGGGGTTGCAGAACTTGAATAGGACTGGAGATATG
GAAGATTTATTAACTTCAAACAAGGATTTGAGAAAAAATCCTTCTCATGGTTTAGACATTGATAAAGAAAATAATCCAAATGGAGCATCAGTGGG
AACTGGTTCGAATAATAGTAAAAAACAACAAAATTGTATAACTGGGTTAAATGTTCCCCATGCTTCACAAATCCAAAATGCTGCTGCTGATGAAA
CAGTTGATAAGAGTACTGACTATCAAGGATGGCTTGATGCTAAGAAAAGAAAATGGAAATATGTCCGTGAACAAAAAAAACGTCGAAGGTTGG
GTGTTGCTGCCACTTTTGACAGTCCTAATGCTCTGCTCTCAGCAAGACATGCCAATCAGTTACCTGGCAACAGTCGGAATAGATCTACATTTTT
CCAGAAACAAGAATTGGCTCTCTTTAGATCACATTGGCAGATAATCCAGCTTGCTTCAAGTACAGTGCCTGGTCACTTTTTTGCATGGGTTGTT
GCTGAAGGGATTATGTTCAAGATTCCTATTAATGTGCCTAGAACTTTTTACCTAAACTCAAAGGCTCCTGTCACAGAAGAGTTTCCAGGAAGGC
GTGTCAAAAAGATTCTCCCACATGGAAGACCTTGTTTCAACCTTATTGAGGTTGTTACAACTGAAGAACAGTTCAGGGCTGAAGGCAAAAAACT
TGCAGCTCATCTAGCAGAGCCAGATGTTGAGGGAATATACGAGACAAAAATTCCACCAGATCTCAATGCTGTTCTCCAAATTGGCTGTGTATG
CAAAGTGGACAAGTCCGCAAAAAAACGAAGTATACAAGATGGATGGGATCTCGCTGAGTTGCAAATGAAAACAACCGCCGAGTGCTCATATCT
GGAACAAACAGTTTCATTTTTCTACTTATACCACAGCTTGAATGAAGGAAGGGCTGTATATGTATTTTACTTCCCTACGTCTCTCCGAATACATG
CAGTGGTTGTCAATCCTTTCCGTAACAAAGAAATATCTCCGTCATTTCTTGAAAAGCAATTTCGAGATGCTTGCCAAGCCCTTGGTTCTGTCCC
TGAGAATCTCTCTTTCCATGTGGACTACCAAACATCAATTGATGCTGGTAGCAAATACGTGCAAAGAATGTTACTTGAGTACAGACAACAACAT
CCTGGACCTGTCATAGGCATAATAGAGTGCCCTAAACTCCAAGCTATAAAGGCAGCTGTACGAGCACTTGATGATTTCCCATGTGTAACCATTC
CTAGCAATGCACGGGACAACAACTATCAGGCTCTCGGTTGGCAAGCAACAGCTGGTCGAACAAGTATGCAGCGTTGTGCTGCATCAACCCAA
TGGTTCAATGAACGGGTTTCACTCGCTAGATATGCACATGTGCCACTGGGTAATTTTGAGCTAGACTGGCTTCTTTTTACTGCTGATGTATTTTT
TTCAAGGGCATTGCGTGATCAACAGCAGGTACTATGGATATCTGATGATGGCATTCCTGATTTAGGAGGCACTTACGAAGGAGAGACATGTTT
TGCTGATGAAGTCATTCAGCCTGCACTTACTTATCCTGGTGCATATAGAAGAGTTTCTGTTGAGCTTAAGATCCACCATTTGGCGGTTAATGCT
CTCCTAAAGAGTAGTCAAGTGGATGAAATGGAAGGAGGATCAATCGGTAGCTTTGGAAACGACATGCCTCCTGGTCAACATGGCACCGAAACT
GATTTTGACGATGCTAGCTTATGTATACCTGCTTTCCAGGTGCTAAAGCAACTTATTCAAAGATGCATTTCAGATGCAGTGTCGTCTGGAAATG
TATTTGCTGATGCAATATTGCAACATCTCTACCGCTGGCTTTGCAGCCCACAATCGAAACTTCATGATCCAGCTCTCCATCGTCTTCTTCATAAT
GTGATGAAGAAAGTCTTTGCTCTATTGTTGGCTGAGTTCCGAAAGCTGGGAGCTAATGTTATCTTTGCAAACTTTTCTAAGATAATCATAGATAC
TGGAAAAGTAGATTTGTCATCTGCACATGCATATTGTGATAGCTTATTAAAGACCTTGCAGACAAGGGATCTTTTCGAGTGGATAGAATTGGAA
CCATTGCACTTCTGGCATTCCTTACTGTTCATGGATCAGTATAACTATGGCGGAATTCAAGCTAAAACACAAAACATGGACTCTTCAGATGGTG
ATAATGATATTGACATTGTCTCAAGCTGGAACATGGCAGAAGATTTGCCCAAAGATACACAGGATCACTTTGTCTTAATTGTATCCGAATTCCTA
TATATTCCATGGAAATTCATGAAGGAACAAGTAGCTACTCAAGCTACCATAAGGGATGATACCTCATGCACTCCATCTGTAACAGTCATGGCTG
CCGAGAATCTTGAGGGCCAGGTGATAGAACACCTCCGTGACAAGATTAGCAATTACTTTGCAGACAAGCTTCTAAGAACTGTAAGTGATATACT
TCACCATTTCAAAGGGAAAAGCAAGCATGAGTCAGATGAACCTGGAAATAGAGAATCTGATCCTCACACGCATAAAGGCGATGCTGCATTGGA
GTTCATTAAGCATATCTGTGCAGTTCTTGCTCTTGATCAGAATGTTCAGCATGATATTTTGAGAATGAGGAAGAACCTGTTAAAGTTGGTTCGTG
TAAAGGAGTTTGCTCCAGAAGCACAATTCCAAGACCCATGTGCCTCATTCACTCTGCCAAATGTGATCTGCAGCTATTGCAATGATTGCCGGG
ACCTGGATCTCTGCCGTGACTCAACACTCCAGGGTCAGGAGTGGAGATGCGCGGTGCCGCAGTGCGGGCAGCCATACCACCGTGAGCAGAT
GGAGAACGCACTTCTACAGGTGGTCCGCCAGCGAGAGAGGCTGTACCACCTGCAGGATCTAGTGTGCGTGCGCTGCAGGCAAGTGAAAGCC
GCCCATTTGTCGGAGCAGTGTTCCTGTGGGGGCTCCTTCCGGTGCAAGGAAGAGTCTTCATATTTTCTCAGCAAAATGCGGGTATTCCTGAAA
GTGGCGGTCAGCCAGAAGTTCCAGCTCCTGCAAGATTGCGTGCAGTGGATTCTTGAAGTCAAATAATGCAAGCTGCTACTTCAACATAGTTTA
GCGATCGATAGGTTGTACAACTTGCATAGGCCCCGAAGAGTGTATATAAACGGCTATATAGCGGTGCCACGTATCACCTTGATCTTTCGTAGA
GTAAAACTGTTCGGACTTCAGAAGATATTAGTTTATCATTGATGAAGACTAACGATGTGTGGTAGCATTGTAGCTGTACCCTGCTCGCAAAATA
TTGTTATTGTATTGTAGATATTCAAAGGTGGATGACGAGAAACGATTGTAACCTTGGTATCGAAACATCTGGATAATTAGCCTTACGGAAAAAAG
AGAGAATTCTAGAAAAAAATCTTCTAGTATGGGACATTGCTGGTGCAAAACTCAACTTTGTATTGGAGGGCTGGAGGCCATGTAGATAACTTTG
GTTATGGCCTATGACTATGAGCAC

Gene: DCL4
>comp206769_c0_seq1
TTTTTGGGAGAAAAAAAAACTATCAAGATCGTATTTAAATTTGATACAAACTTAATAGGACATTGGCATACAATTGCCAACATTAGCTTATATTAT
GGATGTTACGATCATCAGAAAAATAAATTACCGCTTTTCATGAATATGTACCATTCACACCGTTGTTTTTCATGAATATTTGATAATTTTTCCTCA
TTGCAACATACAGGTATATTTGCTAGTTTCTTTTCCTATATGTGAGCTTCGTATGTGTCAATATATCTCTGAACAGAATGCATGTTACAATAACAA
AACTCTTCCCATCTGAGGTCTATGTATGTGACTGAAGGCCACAAACCTAAAATGTTGCTAGACACGATTATCATCTTTTGGTAGGTATCCAAGTT
GCTTAAGATACCACAGGGCTCCCTCTGCCGCATGCTCATGCGCTGCTTTCTTTTGTAGCTTAGGATCGCTATAGCACTCCAAAAGGGTCGCTG
AAGTTCCCGTGATCTGAATAAGGACCTTACAAGTGAACTTTCGAAGGTGGCTTGGTCCTTCATCTTTGCATAACTCAAATTCAGGAGGTTTCCA
ATAGTTTGCAGCACATACTTCATAAAGGAATGACCTTGCTGTTTTGCGTACCATATCACCATGATACTCTTTTTGGTTTATTTTATCAGTTGCCCC
TCTAGGTAGGCAACAATTCTGCATGACCACATTGGACTTATTGCTCCTTCCATTATTAGCATCATTCTTGTCAACCTTGTTATTTCCTGCTGTAC
TCTGAATGGAGGTTTCAGACCTCCCATTCAAAGACACTTCATTTTGAAAAAAGATGTTTCCTTCCATTTCTCCATTTGTCAGTAGATTCTTCATTT
GTACGTCATCCTCAACATTGATCGGTTCTTCATCATAGCCTATCAGTTCTGATTCTTTTTTCGTGGTAGCACGCAAAATTTCCTCCAGTGACTTT
CTTTTATGCTTGTATCCATTATTCTTCAGTTTCAAAAGTGCTTCTTGTGCAGCTACCTTTCTAGCATCTTTTGAATTCCGGTTTGCTGCAGTACAG
CTTATCATCTTGCCGTTTATGTTAACTTCTACCTTGACATGGAACTCTCCATTGCCCTTCGTAGGTTTGGGAAGTCCTAACTCAAAACCATTATA
CTGACAAATTTCTCGGAGTTCTCTCATAGGATCGATGTGCATGCCACAGAAGCTCAATATTGGCTTTAGAAGCATTAGCATTAGCTTCCAAACA
TGGTTCAGGTTGAAGCCGGAATCTAAAAGCACCGCCCCCACACAAGATTCAACAATATCACCAAGAACCTTTGGACATGTTGGTTCTTCTAACA
AGTCTTTCTCTGAACTGGAAAGATTAACATAATTCTCAAATTTATTTACCGCTGCCGTAAGATACTTAGAATCCTTTATGAGATATTTATGGATAG
ATTTCTTAACGGCCACATAAGCAAATGAATTGTTATTGACAGCTAATGATTTCAGATCTGTTAGTTGACCAGGCTTGAGATCAGGGTAAGCTGA
GTAGAGGTATGAGGCCATCAAATATTCCAAAACAGCATCTCCAAGAAACTCCATCCTCTGGTAGCATCCTCCAGAATGCTTATTGAATGAAGGG
TGTACAAATGCTTGGAGAAGAAGACCCTTGCGCTTGAAGGTATAGCCTATCAATTCTTCAAGCTCAGAAATGTTCATGTAATTCATGAGAGACA
AATTGGCGGAGCTTGCATCTAATACTCTATAGAGAGCTGAATTTTCGAAATCAACTTTTATCCCAATCCAATGTAGGAAGGCAAATGCAGCTTT
GAATCCGCACTCGACAATAAAAGCTCCAAGAAGTGACTCAACAACATCTGCAATTGTCTTCCTATGCAACCAATGATGTGATTTTGTACACCTC
AAGTTACAGTTTTCACGTCTATCTGGGTCGATATTCTTCGGGTGTAAAGTCGCTTCTCTGTCAGGATTGCAAACAACTTTACAAGGTCTTCCCA
GTGCAAAGAACTGTGTAGGTTCAAAGTGTTGATCCCGTATGTATACCTGCAAATTTCTTCTAATGGATAACTCATATAAATTCGAATTGCTCACT
ATATCAGAACGTCTCCTGGTCAACTGACCTTCATCAAGTCCTTCATATGAAATAAAGTTATGGCGCCCAACTACATACTTCAAGAAAGCATCAC
CTAGGACTTCTAATCGCTCCAAAGAGATCCTCTCTAAACACCTTTCAGTAGTCAGTGCTTTAAGGATACCCAAGGCACTAATTTGAGAAGCCTC
TGGGAAATAAGATGACATGACATCCTTCAACTCAATAGCCACCAACAAATTCTCCAAGCGACACATTAAAGATGGTAGCAAGGACAAAGAACTA
CCCATATCTTTTGAGAACCCAGTTATCTTCAAAGAGCATAGCTCTGGAGGTAACTCCACAAAGCGCTCCATCAATTCACGACCCTCAGATTCTG
TGCTCTCTAGTTGTCGGTTATGAAGCAGATTACGCAGACAGAAGAGTTGCTTAGCTTTCAAAAGTGGCTGCTCAGGATGGGAAAGCTCGATAT
GTAACCTTTCCTTAAAATATTCTGCATAAGTTGTACCGTTCAACTCACTGTTAGCACTTGTTCCATCCAGAATGGCGTCAACGAAGAAAAACAG
GTCGGTGTGGGGAGTATAGACCAAACTGCCAATCACATCAGCTTTGTTGTATGTTCCACCAAGAAGCTTCAAAGACTCATTTGGCAAATATGAG
TCATGCAGAGACAAACCCATTGGGTCTTGAAATACAGGTGATGATAAACACTTCTTTACTGCTGGCCAATCAATCATATTATCACCATAGAATTT
CTGCTTGATGGGAAGTAAAAGGTAAAATGTTGAATCCATGTCCAATGCAGCATCATTCCCCAACAAAATATAAGATGAAGTGAATTCGGATCTG
TCCAGGAGAACTTTCAAAAACATTTCTTGAAAATTGTGTGCAAGTATCATCTCTTCTTTGTTAAACTCAATCTTTCCCAAATATTTAATTCCTGCTT
TCACAATCCTCGCATGAGCAAGATGCAATTCAACATCCAACTTTTCAGATTCTTCTGGAAGGCGATTGATCACAAAAAGACCAAACATCTGATA
TCGTCTATCTGCTGGTACAGGAATAAATTCTATGTAATAGAAATGCAACTTCAACGAGCAGTCTAGTTTGCATCTTGAAGGTCTCAGAACTGCA
GGGATTAACATCTCATGAAGCTCTTCCCTAAAACTTTCATCCTCATCTTTGTTGCTTTCTGAAATATTTGTTGTTGATACCTTAACCTTTCTAGAG
CCTTGGCCAGGTAGAAGAAAATCTGTCAAAGCACCGAGTTCATGAAGTTTAATGCATGCTTTCAAGCATGCATCTCTCTTAGCTTCATCTTTCG
ATGGACAGGGTTGACCATTCACTTGACGGAAAGCAGCATTTGGTGGAAGAATTAGTCTGCAAACTATTCCATCAATGTCATCAACAAAGAAGAA
TGATGGGGAAGGAACAAAAAACATATCTCTAGGAAGGTTATCACAGTAGCGATGTAATAGAGATACACTGCAAGCAGTGCTAATGGAAGCACC
AGTATCATTGACTCGATAGATGTTCTCTTCAAGGCAATCGAACATATCATTTGATGTTCTCAAGTTAATCTCTTTATCCATAATGCTTTCACCAGT
AATATAATCATCAATTAACTTCTCCTGAGACTGATTTCCCCTCTCCAGGAGAAAAACATATTTAGATTTACTCATCCGGGCACGTCCCCTTGACT
GGATAAAGCTATTAACAGTTTCAGGGAGATCAAACCGCACAACAAGGCAGCAAGTCTGAATGTCAAGTCCCTCCTCACCTACACTAGTAGCGA
CCAAAAGGTTCACCTCACCCGAAGAGAACTTTTCAATGATAGAACCTATCTTGTCCCTTGACATGTTCCTTAATCCCGAGTGGCATCCCACAAG
AAACTCACATTTCCAAAAATCAAGGCACTTCAGATTTTGGAGAATATGTGCTACTACTCTTGCGACAATGGTTCTTTTCACAAAAACAATGCACT
TCATGTTTTCCTCCAGCCTGTATCTCGATAGAACGTCAATAAGAACTGCAAACTTTTTTGAGAAGAAAGGTTCTTCTATTGTCTCTAGGTCAAAT
GAATCAGCATCCGCACCTTCTAAGATGTCGCGGCTTAGAAGAGACGTTGCTTTGTTTACATAGTCATTAATGTCAACCCCTTTTCCATCTAAACT
AACACTATTGGGAGAGAGAAAGGTCCTTGCAGCTTGGAGAGCTCCAAAGAGACCAACTTCTTGCAAACAGAAAATCAAATTTTCATGCAACCT
CCATAAGGACTTCAGTTTCTTCTGTGACTCCTTGAAATTGCATGCACTCTCTCTTAACATGCATTCAGACTGAGACTTATAGCCATCAAGCTCTT
TGATGTAAGTTGTAGTCAAGTTAGAGTGACCAACTGGGCCATAAAAGTACACTTCAATCTTAGGAGAAGCAATCACACTTTCAAGTTCTACATTA
TCAACTGAACAAACCTTTGCATTAAGTAATTCCTCGAGACTGTTGATACATTTAGTGTAGGTAAGCTTGTTAGAACCACCTTTTCCAATAATTGG
TGAAGCAGTCATGCCAAAAACACGAGGATGTTTATCAGCGTTATTGTAGAATTCCTTCATAATTTGTGCATATGGATGCCTTTTTTGTGCTTGTG
CATGATGGCATTCATCAAATATCAAAAGTGCAATTAAGTCCATCTTGATGAAACAATGCCGCAAATTGTGCAATAATATTTGGGGAGTCATTACA
AGGACCTCGTATTCTCCCATCTCCTTCTCCCACGCCTGGTGATCCCTAGAGTTTTTACCACTTCCATAGTAACGTTGAACTTTAAAATTAGTGGA
ATCCGCAATCACCATTGCTTGCTGACGGACAAGGGGTATGGTCGGTGCGAGGAAGACGCAGACCTCACGGCTGGGTTTGCGGATGAGGTGT
CCAAGCTCATACATGAGTAACACGGCAATGTGTGTCTTCCCACATCCCGTTCCAAGGTACACTACGATGTTCTCCTCAACTGCCCTCTTGCAAA
GATCCAGCTGATATTTGCGAGCGATCGTTCTCGGATCCTTGCGTTCGCCCGCGGCCGCGCCGGAGCTCGAGCTCTCAGCCTCGTCCTCGCC
CTCGCCCATGGCGGAAGCTTCGAGCGGAAGCATAAAGGTGTGGGGGAAGGAACAGAACAGAAGGCGGAGGTGTATGTGAAACTGCTGTGCG
CTGTCTGCGTTGAGGTAGCATCGGCCAA

Gene: CPK6
>gg_11571
CCGCTCCATTTTTCTCCTCCTCTCATTTTTGCGTTTAGGCCCCTGGGGAAGCAGGGATTCCGGCGCCGGTCCCTCTCCCCGGCTCTCTCGCC
TCGCCGCTCTGATCGAGGCTGCAGCGTCAGTTGCATTATGGGAAATCAGTGCCAAAATGGAACCTATGGGAGCAATTACAACAACTACAACCA
CTTTCAGAAGGACCCTTTGGCTTCTAGGTACTATGATGGGGCCGGTTCTGAGGATTGCTACTCAGAGCCATCTAAGTCGAGTATAGCGGATCT
TATGCGGCAAGGACTGAGACGAACCTTAACGTCCATCTCTGTCCTTGGTCGAAAGACTCCTAGTGTAACAGAGCATTATACCCTCGGCCGGCA
GCTTGGAGAGGGTCTCACTGGGACAACCTACCTCTGCACTGAGATCAGCACTGGTTGTCAGTATGCATGCAAGTCCATCTTGAAGACAAAGTT
CCGGAACATGCAAGATATCGAAGATGTGCGCCGTGAGATCCAGATAATGCACTACCTTTCAGGCCAAAAGGATACAGTCACCATCAAGGATGT
GTATGAGGATGAG

Gene: BRM, CHR2
>comp199879_c0_seq1
TACAAATCTAATATTGTTTTTAAAAAAAAAACTCCTCAAAATATGAGATGTACATTCTTTTTGGGACGAAAAGAATTTGAGTACTTAAAATACAA
GATGAGATGTACATTCTTTTTGAGACAGAAAGGGTACCTCCTATTCCTATCCATCAATGCTCCAAATGCCGTGTGCCTGTACCAGCACTGTAAC
CCCGTCGGCGACTGAATACTCTGCCAGTTTTGTGCTCATACATTGCACGATTCGAATGCATCATGATGTCAGTCCGGCAACTGATGCACACCG
AATCTTTTCTTCACTCACCAGACCACTCATCGGCCATGTGCCGGCACTACAGAATGATGTAAAACTCGACGCCATGGCCGTCTACAGGACATT
GGAATCCCGGAGCACAATGGTCGCAGTGTAAACAAGCACACGATTCACTCAAGTGTACACAAGCAGGCGGCCAGACTGCAAGAACGATAAAC
ATGTGGAAACAAATAAACTATTATTATCAAGGCCGGCTGCGCCAATGCCTTCATTGCTACTGGATGCATTTTCTAGTTGTTTGTACCACCTACCA
ACATAAGAGTTCCCCTTTACACATTTATGAAGCTAATAACATACAAGAACAATAACCCTACCCCTCCCTCGTTTCTTCATGTCCACAACAACAAA
AGCCCCAGCATGTAGTGTAGGTGCTCATATATGCAAGAACGAACTCAGTTGTGCCAAAATGACAATATAAGCCCCCACCAGTGCAGCCCACAT
TTCATAGGCCAGGCAAATCAAGGTTGCCAGAGAATGTACACCGCAATGCAAAGAATCCTCACATTAGGCTCGGCCGCCTCTTGCCGGTGTCA
GTCCTCAGTCTCTTGGCAGGCTTTGCCCACTGGATGTCTCCAATGCCGGGTGCTGATGAACTTCCACCTTGATCCGAATGCGCGCCCCATCC
AGCGGAATGCATCGACTGCTGAGATGGGTGTGGATCCCTTTGGAACGGCGTCCTGGCACCTCTGGGTGAGGGGACTGGGCCTGGCCGCCC
AGGGTTGGCTGGCGAGAGAGGTCCAGCTCGGCCTGAACTAGAGGGGGACCCAATGCCGCTCCTTGCCCGATCTTGCCTTTTCTTCTTAACGA
TGAACATATCGCTAGGGTGCAACAACCCAGCACCATCCTGGAGCTGATCCCTGCTGCTGGATCCAGAATGTCTGGGATCTCTCTCTGATCTGG
AACTGTGGGGCCGGCTGGGGACCTCGCTGGATTGATTATGCCTGCTATGTCCCGAGCCATGCTGCTCTGCTTCACTGGCACTGGCACTGGCA
CCGGCTCTGCGCTTATGGACTGATGTGGTGTGCTTTGACGATGGGGCAGCAGCACCACTTGCAGCTCCTCCAGGATTCGAAAATGACATCGC
ATTCTTAGCTTCGCTGAAGTCTGAATCTGGGAAGGCAATCTTCATTATGTTGAAAAATAAGTTATGGAGGGTCTCAGCCTCTATCCGGACCTCA
TGCCTATAGCTGAAGTGCTGAACCACACTTTTCAACATCTGTTGCATGTCCGCTATGAACTCAGTTACAGCGCCATACTCGAATCCATCAACTC
GCTGTTCAATTTTTTGTAGGTCCAGGGTGCTACCAGCCGGACCTCTGAATGAGGAATTTTCATTCCTCCGCCACCAAGAAGATATATTAGGTAT
GATTTGATGACCTTCTTTGTCGATCCTCCTCCAAAGCTTGCTTATCACATTTTTGCACTTTCTTTGCATACTGTCAGACATCTTTGTGCCACGAA
ACTCAGGGGTGGCGGAATCAATGGCCTTACTGTTCCAATTTTCCTTGGAACGTTCAGCAGATCCCTCACCAGACCCAGACAAGTAAGTCAATT
TCCCCGACCTAGAAGCTGGTGAAACCTTCCTAGATGGCATGTTGCGCTTCTGTTTCACTATAGGATGAACTGCATCTTGCTGCCTGGCAGCAT
GCTGAGCGAAGACCCCTTCTCCAGATCTATCTTCCTGCTTTACTGCATTTGGTTTTGGACGAAGACGAATACTTCGCTTGCGTTTTATTTTTGGT
TGCAGTACCTGTTCATCCTCACCGTCATCACGTTCATGATTCCAGCTCCCTGACTGCTGGAAGTCCATATGTGAGTCCCCAGACATTGCAATCT
CACCTTCCTCTAGGTCATCCGGCTGCAAATCACCAGTTAAATGCTTGATGATATAGTCAACTGACACAGAAATTCAAGAGAAAAAACAAAACAG
ATAAATAACAACAATGTGTCAACACATACAGTTCTCTTTGACAAAGAACCAGGGCGGGCATCTAATGCAGATAGAGACCGCAATTTTTTAGAAG
ATGAGGAAGGTACAGGTGGTGGTAATCTCCTGCTTCCAGAAGAAGAGCCTGTTGAGCCTGCTTCTTCCATCCTATTAGCTTTCCTGCTTCTTAG
TCCTTCAGTAAATTCATATACATCATCATTCACAGGCTCTTCTTCCTCTGATTCATCCTTGTTATCAGGTATCGAATCATCATTGTCCTCTTCATC
TTCGAATTCCCCTACTTCCCCTTCTTCAGGTAGGGATGCAGTATTCCTCTCCTCAGAATCTTCATCAGATTCTTCAAGATCTTCATCATCGATCT
CCCTGTAGATGGAGTACTTGCCAGTACCCTTTGGCCGCCCCCTTCTTTTCTCAGGTGTATCATTAGTGTCTAAAGCAATGCCCCCAGATGACAT
ATTTCTTGACGGCTTTTTGGACAGACTCGCCACCACAGCATCTACTTCAGTGGAGTTAACACGAAGCCACTTTGGAACCTGATGATGTTTCGTC
ATATCACCTGTCCAATCAAAGTCTTCATCCATCTGATCAAAAAGCTCGACTTCACTTTCAGTCCTAGCAATCATACGGTTTACTTCCTGTAATGA
AGGAACATCATGAACAGAATCTTGATATCTCTCTTCATCATGCAGGAGTGTCTCCAGAGTCATCCGCCTTTCCTCATGGGTTGTTCTTTGATCG
AAACGACCAGCATTAATGACCTCATCTGCCATATCAATTTTGTATTGTTGGATATTGTTGCGGATGAGACTTTCAATTGATCCCATGTATCTGTC
TTTTCCAGCAAGATCATCCTCCAAATCTGCACTCCCTCCATTTCTCAATTCATCCTCTTTCTGATAACTTGATATGTTATCAACAACAGCTTCCAA
GTAAATAACCTTTACCTCCCTAGTCTGCCCTATACGATGGGCCCTAGCAACTGCTTGCTCCTCATTTTGTGGATTTGGATCAGGATCATATATTA
CAACAGTGTCTGCACTCTGAAGATTCAGACCCCTACCAGCAGCACGAATACTAAGCAAGAATATAAAACAATCAGAACCAGGCCTGTTGAAGT
CAACAATTGCTGACTCTCGATCTTCCAAGCTTGTTGTTCCATCAATTCGCCTATAAACAAGTCGTCTCCACTGCAAATAGTCCTCCATGATGTCA
AGAAGCTTTGTCATGGTGCTAAAAAGGAGTACGCGATGACCTGCCTTGTGAAGCTTAATTAAAATTCTATCAAGATTCCACAACTTCCCACAAG
ATCTGATCATAAAATCTTTCCCATGATTTAAGAATGGATATGTCAGCAGAGGATGATTACAGACTTTCCTTAGCTCCATGCACTTGTTATTGAGA
TTCTTGTAAGTCTTGACCTGGTACATGGGATTCCGTTGTGCACGCCTTTCCTCATCTTCAGGATCAACTCTAATGGTACCAGTAGACTTGATCC
AATCATAAATAGCTCCTTGGACGGCAGACATTCTGCATCTCAAAACAATGGAATCCTTCCGTGGAAGTGATCCTTCAACATCTTCCACACGCCT
ACGTAGCATGAAAGGTTCCAAAATCTGATGCAGCCTGTGAATTATTATTACTTTCTTCTCTGTCTCAAGCCAATCATCTTCTTCTTCACTATGTGT
AGGACCATCCCTCTGAAAAGGCTTAGAGAACCAATCTGAAAATGCCTTGCTACTGTCAAATACTTCTGGAAGCAACAAATTCAAGAGGGACCA
AAGCTCCTTGAGATCATTCTGTAGAGGAGTACCGGTGAGAAGGAGGCGCCGCTGGCAACGATAGCGATCAAGATCACGCGCCAGGACAGAG
TCTCTGTCCTTCATCCGCTGTGCCTCGTCAATTATAATGTACTTCCAATCAACCCTTGAAAGCTTGGAACGATCAAACATAACAAATTCATATGT
TGTTACAAGAACATTAAATTTCATGGCCATAACCTCTTGAGAGAACAACTTCTGCCTTTGGTCCTTTGCACCAACATAAAAGATGCAAGATGCA
GATGGTAACCAATTTAGCAGTTCGCTCTTCCAATTGACCAAGACAGCATTTGGCACTATTATGAGATGAGGGCCATAATTCCCTTTAAATTCCAT
CAGGTACGCAATCAATGCCATGACCTGTACAGTCTTGCCAAGACCCATCTCATCAGCCAAAATGCCATTCAACTTATTATTGTACAAAGAAAGC
ATCCACTGTAGGCCCACCAGTTGGTAGTCCCTTAAAGTTCCTGCTCGCAACAGTGACGGCTGCTTCGTAACTCTCTCACTTACAGCATGGGCT
AAACTATAGTACTTGTTAACAGATGTATTATCCCTTGGCGCATTCATCTCAGAGAATGTGTTCCTTATCATAACCTCCTGACCAGCACATTGTGC
AGCAGCCTTCACTTCTTCCTCAGACAGGCCCTGTGCCCGTGCAGCTGCTGCAGCAGCATTTGCTGCCTCTTCTACTTGGTGCTGACTCTTGGC
AGCAGTTATTTTTCCTCCTAGTTTATAAAGGTACTCTTCAGTCTGGGTCAAGAAAGAAGATAGGACATTGTATCTTTGAGCTGCATCCCCAGGA
ACACTAGTCTGCTGTTCCAACAATATTTGACGGTATCTTTCCACATCATTGTTTTTCAATGCCTCCATTCTTTTGTTCCGATCATCATCTTTCTTC
TTTGAGAATTCTCTCAACATCCTTTCATGGTACTTGGCCACCCCTCGATTACGAGTTATTCGAGCATCACGAATCGCCCAATGTGCCTCCAAAA
GCTTCTTGCGCCACTGGAAAATGGATTTCAGCTGCTTCTCTCTTGAAGCTCTCTGCATCTGCTGAACTTGCCTTACAAGCTCAACACGTTGGCG
TTCGCATTGTCTGACAAATTTTCTGTATATCCTATCTGGCATTGCCATAATTTCTTGTTGTTCATGCTCAACTTCATCCCTTAGACGAGCCTGGT
GCTCTAGGAGTTTGAGTTTCTTTTCTTCAATCTGTAGCCTCAAGACAAGATCAGGTCGGATTCTTTTCCTCTCTAGATTAATTGCAAGCAAACCG
CTTATCTTTTTTAAGTTATCCTCACGTTTCTTACCAAGAACTATCATCCCCTCCTGAGCTAATAAATCTTTCACATCATAACCTAGTGACAGATTG
CTGTTGTAATTTGCAGGTCCCATCGAATCGTGTCTCCTAGTGAAGGACGGGAAATCGAATAATGGGCCATGGTATTTTCTGGGAACATCTCTT
GGAGCTGGGGCACTGCTTGTGGAGCCAGTCCTTTTTGCCTGCTCAGCATCCGCTGAACCACTTTCTGCTGGTACAGATTTACCCCTCTCAGCA
TTGTAATCACTTCTCCCAGGTGTTCTCTGGATACTCCGTTCCAGATCCTGCTCAGATTTAATCACAGTAGTGTTACTATGCTCTGGTACAGAAAC
TGGTCCAATTTTAAGAGGCTCTTTCGGTGAAGCTTTCATCACTTGCATGGGACCAGGACCACTGGCAGGACTAGCTTTGTCTTCTGAAGCAGA
AACCTCCACCTTTGGTAAACAGTGTCCCTTCGACAATGCAGGTTTTTCAGGAGCTTTGTCACCACTCTCCATCTGCCTTCCATGTTCATCAGCA
TTGCTTACACCAGGCTTTTCACGATTATTTGTTGCCTGGGGTCCTGAAACTTGTTGCGGTCCTCCTTGTGAATCAGGAGGGCGACCAGATACA
ATTAATTCAAGAACCTCAGGCGGGAGTCGATCACCACGCTTTAAACGCCGGAAAGCTAATATCTGAGCCTTGAGTACATGAAGTTGATTTTTGG
TGAAACCCGTTTGCCCTGTTTGTGGCCGAGCACTTGGAGTTGAGACCTGCGATCCTCCTGCTTCACCAGGGGTAGATGTGGGGGCAGCAACC
TGGTTAAGCTGTTGCATCTGTCTGACATGCTGCATCTGCATTGTGTCTGAATTTGCAAGCACATTTTTTGGATTGGATTGCTCAGGAACCTTACT
CACATGTCCAGAACTTTGGGGGATATGCATCGTTTGCGCACCATTGCCCCCTGTCATAGGTGACCTCACAGCCCTTTCATTTGAACCATCCCT
GTTGTGCACAGATAACTGCTGTTGCAACTGCAAGTTGCTCGAATTCACCACCTTAGCCTCTTCCCCACCAGAAACTGAGCTCGGTGTTAGGGG
TTGACGAGGCTTCAGTAGGGCACCCTGGCTTGGAGCATTACCATGCCCTGGGGTATCACTGTTCACCTGAGAGGGCATTGCTTGCTGCTGCG
CAGCCATGTTTGCCTCGTTCTGCTTCTGCATGGCTGCCATCCTGTTGTTCTGCCAGATCGGCAGGAGTTGAGAAATTGCACTCATGTTTGCAG
GATTGGACAGGTCAAAGTTGTGCTCTTTTGCCCAGGCTTGGATGGCTTGAAACTGCGCTGGTGTTATCGGGCCGCCACCGGCACTGCCCATG
CCCACCTGGCCTTGCATTGGCTGCATGGGCCTGATCATACCCATTGGCGGCAACTGCTGTCCAGGGACTCCTTGAGGAGGCATAGGAGGCC

TCATATCACCACTTCTTTGCTCACTACTGCTTGATTGCCCCTGCTCTTTCTGCTCACACTGCCGCTTATACATCTGGGCCTGGGCCTGATGAAG
GGCCATAAGCTCCTGCATCTTGGCAGGGTTGTTAAGCATGTCCTGGTCTCTTGAGGATGAACCCGCCATATTCATCTTGGCCTGCTGCTGCTG
GAGGAGCATCCCCTGGGCCGCCTTCTGCTGCTGCTGCTGCATCGCAAGCTTCAGGAACTGTTGCTGCATTGCAACCTGGTTGTGCACCTGAC
CTCCAACCATGCCCTGCGGTCCACCTGACATTGGCAGGTTCCTCTGACCCTGGAACGGTGACACTGGACCAGACGACTGAGGGAAGCTCAC
GCCCCCTGGTCCCATCATGGCATGCGGTGCACCAGATGGGAAGGCTTGCTGCTGTTCATGGTGATGCATCATTCCCTGGCCTCTAAACCCTA
GCTGCTGCGCCTGCTGCTGGGCGGACGCCGGGGTCGGCGGCTGCTGCTGGGGTTGCA

Gene: CDF2
>gg_07573
TTTTTTTCTAGAAAAATCAGAACGATTTATAATTTAATTTAGAACGGAGGGAGTACATTAGAACTTTAACCGGTGCCAGCTAAAAGACTATCGAC
ATCAATTCACAGGCCTACTGTATTTACATGCGCTCCTTGAGCAGTCAGAATCTTCAGACCAACCTAGCTATCGTCAGAAGAAAAGAAAAACTGT
AAATCTTGGCCAGGTCTGAGCTGATTGCCAAACTGCTGCACTTAGCGAAAGATCAATTCAAGAAAAGCTTAAATAGGAACATTGACCATCCTCT
GGTATCCTTTGTTTTGAGGACACTTTATTTACGGTATGATGTCACCGATGGGAGCATCTATAAGTTTCTCAGGTCCTCTCCTGGAATGTCTGAG
AGCGGGAAAACGCTGCTGGATTGGCCTGCAGGGATTCTGGCGTCTTAGTCTCACTATTTTTCAGATCCTTGGACTGGAAAGACTTGAATATGA
TTCTCTCATCAGGCTTGATACCAAGAGAGGCCCAGATTGAACTCTTTGCTGCTTCTTCTGGGTTGTCAATTCGGAGAGCTTTAGGAACCCACAA
CTTTTTTTCTGCCTTATGTTCTTCCTGCAAATTAGCTACCCTTGTATGCTTTCCTAACACCCTAGAGTTGCTTCCTGAAAAAGCTAAACTGTTTGG
AGGAAG

Gene: PRP8, EMB33

>comp202464_c0_seq1

CAGAAAGCACAAAGCTCGAGCCATTAATGCAACTAACAATTTATGAAGCCGCTTGAACGTGGAAGTTTGGGATGAAACAGTCTTAGAATAACAT
AGCTACTAAGCAAAATCTTTACTGTTACCATGGCGTGCACTTGTACAATTTTAACCCCTCTCTCCAGGGCTGTCTTACAACAAAAAGGTAGTATA
ATATATGTTTAAGCCACAGGCGAACAACAGGTGACACCCGTCGAGTCAGGTTACAGCTTCACACATGGCCGTTCAAAACAGAGCTAAGTGCGT
TTGAGATACAGTTTCTAACAGACGCCTACGTGAATGTGTCCTCCCTATCACCTTCAGCGACCTCTCCCTCCTCGATGTTGCTGAACTCAAGGAA
GTGCGTCGGCCTGTGGTCCTCGTGATAGAAGTCCCGGGGCATGCCCAGCTTCATGTTGTACTTCATCTGTGGGTCATGCTTCACTCCCATGA
AGTTGAAGTTCCAAGGGGCGTTATCCGGAATCATGTAGAAACCAAGGAAGCGGTCGCTCAGAAGCATCTGGACCTTCTCGTAATGAGTTGGA
AGGTATCCATGCGGGTTGCTTCCGTTGTCCTTGTTGCTGCGTCCCCACTCGTAACCGCTTGGTGTCAACTTGTATGCAGTCAACGAGCAAGAT
CCTGGGGTGAAGCTGCATGTCAGAATGATGCACTTCTCACCATCCCATTGTTTGTTGTTCTCCAGAATCTTGGCATGTGATGTCAGGTCCTGT
GGCGATAGCTGTGGAGCTTCATTTGGCTGTGTATGCATCCATCCCAAGGGTTCGAGATCATTAAGGAACTCATGCTCAGGAAGATTTGCTGGC
AGAGTCACCATCTGGTGTGTTCCATGCTGCGGAGGAATGGCTATACACCTAATCTCTTTGACTTGAGGATTGTCCTGTGGACTCAGCCCATAC
AAGAATCCAGCAATTTGTGTCCGCAGATCTGCTATGCATATGAATTTCTTCAATATGTTCTTTGGCATAATGTAAGTGTAGCCAGTCTCCTTTAT
ATCATCAGAATTGACATAAATGTGATTTACACGAAGATACAAGTTTGTGGCAGAGATGGCTCTGACGCGCCAATCAGTCTTTGATGCGAATGCT
TGTTGCTCGTATGGACTTGTGGTTGTGATAATGAGTTCATCTCCGTGAACATTTGTCGTCCTTGTGGTTACAGCAGTTAACTGAGTCGTCTCTC
TTGACTGTTTCTCAATTTCAGCAATCTGTTGCCTCTGCTGTGATGGTGGAGCAATTTCAGCACCAAGGATGATATCTCGCATTTCAGATTGTGT
GAGTGCAGAAGTATTAACATTGTTTTTCTTTGCATAATCAGAAAGAATGAGGTCCCTTAGTGCACACTCAACCTTCAACCACTGTTCATCATTCA
GAGTTGGCCATATATGATGTGGTTCAGTAACAATTGTTTTGTCAGGTTTGAGCAACATCTTTGCCTTCTCATTATTCACATGAAGCGCACGTAGA
ATAAGCACAATGCGAGAGAATGCTGTGTATGATGAGATACTTTTCAACCAATCATCATATATATTATAAAGAACCATCTGTGGTTCAGTTGCCTT
CAGTATAAGATCACCAAACTTCTCAATCTTCAAGCAAGCCTGGAAAGGAAGCTGCAGCTCGCTGCCCTTGATAACAATGTTAGGGAAGTCAAG
TAGATGGACCTCCAAGGGATCCAACATGCCCTTCCTTGTCACAATAATTTGCTTTGGCTGCTCTTCAACAGGAAGAGATCGCACTAATGCAGC
AACCTCCTCAGCTGTTTTCCACTTGGCCAGCTGCCCAAGACGTTTTTGCCCCGCCCATACACTTGTGTGGATAACCTTAAGAAAAAGTTGACC
CGTTCTTGGATTGAAAATGAAAATAGCACCATTGATTGGTTTGGTGGTCAAGTTACCCTCAAATGTTTTGTGAATAGTGACTCGATACACATTTG
TATCATCCACAAACCATATGATTTGATTGCTGAATATCTCTCCATAGTTTTGTGAAGACAGGTATGGTTCAGTGGGTTCAGAAGAGTACAACTG
CAGACCCTTCCTTATTCTCTCCCTCAGGACATACAGAGCAGGATTTGACTTCATGATCTTGTTAATTGCCTGCTGAATTAGTGGTTTTGATCCG
GGAAACCAGTTGCCAAAAGCAGAGTGCAAATTATACGCTAGATCGATTCCAATCATCACACCAGTTGGCGATGGGTATATGGACATATTGTCT
GTTGTGTAATCCATGAACTTGGCTCTTGTATAACGTTCTATGTCATGTGAATCATAGTCACCCCAACGCAGTTGCACATCAATCCAGTACTTGTT
ACTTGCTTTCTGATCAAAGAGATCCTTCGACTCAGAAACTAAGCTTGGTTTGGACATCGTCCATCTGTGTGCAGCAAAGAGAAGGATATCTGCA
CAGGAGCTGTTCATCTTGTAACTCTTCCTGGGATGTATTGTCTCCTTCTGCACAGTCTCAATCTCAAGTGCATCCAACTCCTGGTCTAAAACTT
GGCAAAGATCCATAACAACACTTTCATGGATCTTTTGCCACAGATGTGCACGGAAGATCTGGATCAGAGAAATCTTCAATGTCGGGATTTTCCC
ATGCATGAATATCCCCGTAAGATCTAGCTGGACCTGGAAACCAACGTACACATTTGCACGATTGATGGTTGGTGACCACCACAGGGTAAATCT
TCGATTAGGGATTTGATTGAGACCAGAACGCTGTGCATTTGTCAGCTTCTTGTATTTCATTGACTCTTCGAAACCTGATGCCTTCTCCCAAAACA
GACCCTCCCAAGTAGGGAAATAAGTCCCTTTAAATAATGTATGCTCCAGGATGCCTTCTACACCACCAAGTGCTTGAATGACATCAGTTCTGTA
GTTATTTAGATTCCAAAGCTTCCCATCATGCCGCTGATGTGTCCACCAGAATGGGTTCTGTTTCAGCACTTGATATTGCTTAAAATCCGTTCTCA
CTCTCCATCCTTTGTCATAGGCCAGCGTATGGCGGTCTTTTTGGAAAAGAGTGTTGATACGAGGTATACCTCGATCCCATGAATCTTCAAGATC
CTCAAGAGTTAATCGTCTGTTCTGTGATTGTGCTTCCTGCCTCTTCAAAGCATACTCAGCCCAAACACGCTGTGAATCAATAAACTCACTCTCC
CATGGCTGGATGTAGCGGTACAAGTTTGGAATAAGCTGGTCCTCTTCATGGCTCATACCACTTCTAAAATGTGTTACACCAACATCTGTCTGTT
TGCTATACCTAAGATCGCTCTGTGGTATAAGAATATGGCCCATTGATAACATTCCAAGACCTCCAATCTCCTTGGGAGTATAGAAGATAACTGG
AGGAAACCTGCTAGGCATCTTTGAATTCAGGCCAATCTTTATACGAGTCTGGATCTTGTTTTCACATTTGACTAGCAAATCCAAAAGCTCCTGG
GTATGGACAGTTGCTTCACGAAAATATGTCATAAGGCCAATGAGAGCCGTATTCCACTTATTGACTATCTTCGTGAATGTTGTTGACCCTGATG
ACATTAGAATCTGCCGGACACGGTTCTCGAACACTTTCATATGTTCATCATCAACCCTCAGGAAGGCAATTGCTGTCCTTTCTTTTGTCTGTTCA
TTCTGCAAATTCCATACTCCATCTTTTGTGTTGCTGAACGCCTCCTGAGTCATCCGTATCTTTGGCAGTATCCGGACTTCAAAGCCACACATGC
TAAAAAGCAGGTTTGGATTGTCCTTGCTGTAAACAGAGACAAAGCTGTTCTCCCATTCTAAAGTTGTAATGCTCCTCGGAAGGCGGTTCTTCAT
ATCCCAGAACACACTTCTTCCAAGATTCACATCATGTTTCATAAGCCTCATTCTTGCATCTCTGGGCCAACATTTCTTGTTATTGTAACCAACCA
TGTTCTCATTGTTAGGATCAGGATGTTCTGTCAAGTAACGCTGAATTAGATCTCGTGCTTCCTCATGGGTGAACCGGAACAATATATGCACCTT
GTCAATATATCGAGAATACAATCTGATGGGATGCCTTGTCTCAACTTTTGTGTCAGCGTATGTAAGGAACTCATTTGGCATTTGAGGTGGCCCA
GCAATCTCACTTGCACGAGTTAGACCAAGTATCAGGAGATCTAGCACAAGGCCATAATACTGAACAACAAAAGATGCAAATTGAAGCCCTCGA
ATGAGACCATAAGAATTTGTGTGACTCATGTCCTTGTAAGACAACACTACATTGTTCTTTGCAGTGACATAGTCAGCTATGTTATGATCCAAGAC
CAACCGCAGAAGCCTGTTCAACAGAGTCAGATCAATTTTTTCAAAGAACTTCTCAAATTTTGTCTGGAGCATCACAACACACTGTCCATCACTT
GTGTCCCATATATCCAGTAAATTATTTATGCCCTGGCACCATTTATAAACAAGAAGAGGTGGTGGTTCTGAATCAGCAGGCTTAACCCAATTTG
GGAAGAGGTGACGTTTGTCCCCCTCGTACCACAGGTACTGGTCAAGGTATGCATCAGTGATCTTTTCAAGAGGTTCAATTTCATATACCGGAA
TCAAGTAGCTGTACAGATCCATGAATTCTATACCGACTTCTTTGAAAGCTCGCTGAGTAAGGAGATGCCTCTTTATCCTTGACAGTGCTTCATG
TGGATTATCATACGCTTGTTCAATCAACCCAAGCTCCTCTCGTTGCAATTGGTTCAATCTAACTGCTACACTGTATGACTCCTTCAGCCTTTCCA
GTGCAAGTATAAGAAGTTTAGTGTCATGCTTGTATGACAACGGCGGAAAAGGAATTGGCGAGAACTTTCTTGATTCCAGCCAATGAACAGTTGT
TGTATAGATAGCAACAGCTTCTTCAGGAGTTACATATGGCCCATCTTTCAAGTAATTGTGTTGCCTCTCTTGCTCTGCCTTCAGCCACAGACGA
GTTAACCTTCCAAGATTTTTCCGACAAACAGTTTTGTCAACTGTAGCACCACGCCTGATACGCTCACGATTGTAGTGGGCAACATTTGTCCACC
AGTCTGCCTTCGACTTCACATAACGAAGAATCATGTTCTCAATAGGCACTGGCAGACCAGGGACTTTCCATGGGATATTTGCCTTCCAGCATC
GCCATGCCTCACTAAGATGCTGTAATATAGTTCGAGCCTTGTTTTGCTTGATGCCTTCTGGCATAGCATCTAGAACATCATGCATAACTGCAGC
ACGAAGTTCTAAATCAAAGTGAGATTCTACACGTTGCTTTGTTACAGTTTTGGCCACACCCTTGGAATGACGACCCTCAAACTGCCGTGCTAAA
AGATTACCCAACCAGCGCTCCAATAGTGGGACAATACCTCGAAGGAAGAAGAGCCACACTCTCCACATTGGAGCCCAAAACCCACAACCAGG
TCCTTTACCAACAGGACCAGTGTTAAAACGGTAATATATCAAGTGCTTCAGATCTTTGCACATCCTAATCTGCCTCATCAATCGGTACTTGTACC
TATACATACCAGTTAACTGCCCAACATGAGAAAATATATATTGCAGACCATCTGCCAACTGGAAGGCATCGACATTTCCCAGACGGAACTGTAT
GTTAGCATCAACAACCAACTTTGTAAGTCGTAAAATCTCACGACACAGATGAAATGCATTTCCAAAACGAGATTTCTTCCTTTCCTTTGTTGTAA
GAGTCTTCACAGGCTTCAAATTGAAATTGTAATCTAAATGAAGATAGTTAAGATTCTTCCTGTGAATCAACAAGTTTAGCATGTTGTACCCCTGC
TTGCAAACCTGCAAGCCTGCTTCTGCCCAGTCAAGCTCGGTAGTTTGGAAAAACTTAGTTGCTTGGAGTGAACGGAACAGATGCTTCTTTTTCT
GAGCCTTGGGAGGCCTATGATGTAGCTCATTGAGCACATAACACTTCAATAGCTTCTGATAGCTCACACGAACTTTCACAGGATAAGCTGGAG
GGCAATGTTCCTTGTACCACTCAGATACAAGAGGAATATCCTCTGCACGACGTGTCCGACCAGATCTCATGTTAAAAGGCTTTGGAGCAAAAA
GTAATGATATACCAGCAGCTGTTGTATCAGTGTAGAGTTCAGTCTGTTTCAAGAGAGGCTCTACATCTTCTGGGAGACAAAAATCTTCTTCATC
TTCCTCTTCAGTTGTTTTCCTCTCCCGACGATCAACCTTGTTGGTTGAAGTGATAGGGTTTATCAGAGGATCATAGTAGAACGCTGGCAAGTCT
GGATCTTCAGTCTTGATATACATTATCATGGGGGTATGATAGATACCAAGTTTCACTTTCCTTGGCCTGTTGTTGTATAGGTGAGGGAAAGCAA
TCCTGTACTCAGTTCGAAGTGGTTGACGTATGATGAGCTTATTTATGTCATTAAACTCATTCCAGTCCTCATCCCCCTTCTCCATATCACGGTAT
AATGGCTCAAACTTCGGACCCCCTGGAATGCACATATTAAGGGCCTTGGCAGTAAAGAAGGACTCCATATCAAATAGGTAGAAATAATTCCTAT
CGATCAGATCTGACAACAGCTGGCCAGCAAGGCGATATAGTGTTGCCATTATGGGAAGAGATAGATGCCATTTCCTGTAACTGGGGCCATTTA
TAAGTTTTGTCTTCATCAGAGGTTTGTGATCATAGAACCATTCATAGACAGCTGCATCTTCCTCCTCATCCAACTCCAACTGTATAGCTTCAAGT

GGTTCAACATCCAACAAATTATCAGCATAGTCCAGTGGGGGCTCCTCATCATCAAATGGAGGAAAGCGCATTCGTTTAAAATGACGTCTATCCC
TCTTCTCTCGACGCATCATAATCCACATTGTACCCCACTGTGCGAGGTATATGGGTTCAACAACCCAAGGGATTTCATTAACGAAAGTGATAGC
CCCAGTGATATGATACAAGATTTTCACATGGCGAACCTGCTCCCAAGGCATCGGCATGTTCTCCAGCAGTTTGTACACAGCGTGGGGCACAAA
TTTGAGTGCTCCAAGATAAACACGCTTGTCGTGCCTGTATTTCTTGGAAGACATGTCGCCATGGTCCCTAATGATTTTGCGGACGTGCTCAGG
GGGCATGTCCTCCTTCTGCGCCTCGACGAACCCGAACTTGCGCTTGTCGCCGTAGCGCTTGGAGTTGAGCTGGTGCCACTTGCGGGCCTTCT
CCACCAGCTGCGCCTCCAGCTCCGCCGGCGTCAACGGCTTGCCGCTCTGCGGCGCCCCGGCCGGCG

Gene: DCL3
>comp205075_c0_seq1
CATGCAAGTATTTCTCTCTTTCAGTGACTCTGATAACGTCCGATGCCTTATTTTTGTGGATCGAAAGATTAGTGCTAGAGTCATTGAGCGGACG
ATGAAGAAAATAGGCCGACTCTCCTGTTTCAGAGTCTCTTTTCTAACTGGAGGGAGTTCTTCAGTGGATGCCCTCACCCCTAAAATGCAAAAG
GATACGCTGGATTCGTTCCGCTCTGGAAAGGTCAACTTATTGTTTACTACAGATGTTGCAGAAGAGGGTATACATATTCCAGACTGCTCGTGTG
TAATAAGATTTGATTTGCCAAGGACTACCTGCAGCTATGTGCAATCACGTGGACGAGCCCGACAGAGTGACTCTCAGTACATACTAATGATTGA
ACAGGGGAATGTGAAACAAAATGATTTAATATCTGGGATTATGAGAAGTAAGGCCTCGATGGCTGAGATTGCATCAAACCGAGAGCCTGAGGA
TTCACACCCCAATTTCTTCCCCACTGAAGAAATAAATGAATACCATATAAGCTCAACAGGAGCCAAAATAACTACTGATTCTAGCATCAGTGTTC
TCAACCAGTACTGTGACAAGCTCCCAAAGGACAAGTACTACACCCCAAGACCCACGTTTCAGTTCACCCATTATGGTGATGGTTATGAGTGTA
CTGTAACACTACCATCTAGTGCTGTGTTTCAACTTTTGGTAGGTCCAAAAGCAAGAAACATGCAGAAAGCAAAGCAGCTTGTTTGCCTTGATGC
ATGTAAGAGGCTGCATCAGCTCGGAGCACTTGATGACCATCTTTCTCCATCAGTTCAAGAGCCGCCTCTGGAAATTTCAAGCAAGTCTAGTAT
CTCGACATCCGGTGCAGGTTTAGGTACAACCAAACGAAAAGAACTGCATAGTACAACCAAAGTTCTTTCTATGTCTGGCTCTTGGGGTTCAGAT
AGAAGTGTGACTAAGCTTCAAGGCTACAAGTTAAATTTTGTTTGTGATCAAGTTGGTCAAAAATACTCTGACTTTGTCCTTTTAATTGATGCAAAT
ATAGCAAAAGAAGCTGCTACTTTAGATATTGATCTATATTTGCATGACAAGATGGTAAAAGCTTCAGTTTCTCCTTGTGGGCCTCTTGAGTTGGA
TGCTCGGCAGATGGAACAAACAAAATTATTTCAAGAACTTCTCTTTAACGGTTTGTATGGAAAACTGTTTATTGGATCAAAAGCATCAAAAACCC
CAAGGGAGTTCATTCTCAAGAAAGATGACGCGTTCCTTTGGAATAAAGAAAATATGTATTTGCTTTTACCTGTGGATTCTACTTTGGATTCTCAT
AGAAGTGTTTGCATCAATTGGAGTATAATTGATGCAGCGGCTACAACTGTTGGACTCATGAGAAGCATTTATTCAGATGGCAAACGTAATTTAA
TTGGCAAACTTAATCATGAAAAAAATGACGAAGATTTTATTCGTTTGGCCAACAAGTCATGCAAGTCTAATGATCTTGGAAATATGGTGGTTCTA
GCAATTCACACTGGAAAAATATATACTGTTGTTAAGGTATCTGATTTATATGCCAATAGTAAATTTGATGGTACATCTGACAAGAAAGAAGCAAA
ATTTCGGACATTTGCAGAATATTTTGAGAAGAAGTATAATATATTTCTTTGTCATCCTTTACAACCATTGCTAGTGTTGAAACCCACCCATAATCC
ACACAATCTTCTTTCCTCAAAGAATAGAGATGAAGGCAATGTTGTGGAAAATAAAAATAGGGCGAATAGCCTTGTTCACATGCCTCCAGAGTTG
TTGATTCCCCTTGATTTACCTGTTGATGTTCTGAGAGTATTCTATTTGTTTCCATCTCTAATGCATCGTCTTGAATCACTAATGCTAGCCAGCCAA
CTAAGAAGTAAAATTGCATACACAGATTCTGATATATCAAGCTTTGTGATTTTGGAAGCTATTACAACACTTAGATGCTGCGAGGACTTCTCTAT
GGAGCGTTTAGAGTTATTGGGGGACTCTGTTCTTAAGTATGCAGTGAGTGCCCATCTTTTCATGACTTTTCTTAATAAGCATGAGGGGCAGTTA
TCGTCCAGAAGGCAAGAAACGATATGTAATGCCACACTTTATAGGTTTGGTATTGAACGCAGAATACAGGGTTACATACGTGATGCTGCATTTG
ATCCTCGTCGATGGCTTGCTCCTGGACAGCTGTCCAGCCGTCCCTGTCCTTGTGAATGCCCAGTAAAATCTGAGGTTATAACTGAGGATATTC
ATAGGATTGATGACAAATCTATAATTATAGGCAAGGCGTGTGACAAGGGACACAGATGGATATGTTCCAAAACCATCTCTGATTGTGTTGAGGC
TATAATTGGGGCCTATTATGTGGGAGGGGGATTAAAAGCAGCTTTTGCTGTTCTCAAATGGTTGCAGATTGAGACTGAAATTGAGGAAGAGCT
AATTATGGAAGCCCTGTCGAGTGCTTCAGTTCGAAATTATCTTCCAAAAGTTGATGTAGTTGAATTGCTTGAAGCAAAACTAGGCTACACCTTTC
AAGTGAAAGGTTTGCTAATAGAGGCTCTCACCCACCCATCACAGCAAGAATCTGGAGCAACATACTGCTACCAGCGCTTGGAGTTCCTTGGTG
ATGCAGTCTTGGATATTTTAATAACTCGCCATTTGTTCCTTAGTCATAAAGACATAGATGAGGGGGAGCTGACAGATTTACGGTCTGCATCAGT
AAATAATGAGAATTTTGCACAAGTTGCGGTAAAGCACAATCTCCATCAGTTTCTACAACATTCTTCTGGGTTCCTGCAAGACCAAATCACTGAAT
ATGTGAATAGTCTGGAAGGTTCATCCATGGACAGAACCGGCCTTTTATCCAGTGGATCATCTAGGGGGCCAAAAGTTCTAGGTGATATTGTAG
AAAGTATTGCAGGTGCTATTCTTATTGATAGCAAACTTGATTTGGATTTAGTTTGGCGGGTTTTCAAATCTCTTCTTTCCCCAATTGTCACACCT
GAGAATTTGGAGTTACCCCCATTCAGAGAACTTAATGAGTGGTGTGACAAGAATGGGTACTTTATTGGAATTAAATGTGAAAATCGAGGAGACA
ACACAATGGCTATTCTTAACTTACAACTCAAGGACTTGCTTCTTGTGAGGCAAGGTTGTGGGAAGAACAAAGTAGATGCAAAGGCACACGCAG
CTTCCTTATTGCTCAGGGATCTAGAGGAAAAAGGACTTATAATTCCAAAGAATGCAAGCAGAGTGGAACAATTTGAAAAGAATTCTGGTACTAC
AAAACATTGCAACAACTTGCTTGATGTTATGGACACACAAAATGTAGCACCACCTAGGCAAAAGGAACTAACTGTGTCAAGTACCGCTCCTGG
CTCTGTTCAACCATTGGTTGTGAAGGTTAAGTTGAGTAAAGGAGGACCTCGTGTATCCTTGTACGAGTCGTGTAAAAAGCTTCAGTGGCCAAT
GCCTACGTTTGAATTTGTGAAAGTCGAGCCAAGTGTGTGCTCTTCCGGTGATTCCTCACAAAAGGTTGCACCCCAAGGATTCGCATTTGCTTC
GACAATAACATTGCATATACCGAACGGCGATGTCATCAGCCTCACAGGAGACGGGCGTGTGGATAAGAAGAGCTCACAGGATTCTGCTGCGC
TGCTCATGCTCTATGAGCTGCAGCGGCAAGGTAGATTCCAAGTACAAGAAGTATGAGGAATCGCCACAAATTCGTGGCTCCATTATGAAAGGT
GTAATTTTACCGGCGATTATTGCTTAGGCATCTTGGAATAATGTCATCACCAAACCAGTTGATCTTAGATGGCCGCTGCGCCGGTGTTGATGTA
CCCGTGTAGGTTACTCTTGTTCCTGCCTTGTGGATACAGACATTTCATCTGTATATGTAGGTGCATGCCGATGAGAGCCTGGGAGGGCACTGC
CGTGGTCGTGATGTGTGTACACATGCAGTAATAAGCATCACATGGTTTCCAGACTCGGGAAAAATCTATGAGCTGAAATTGTCACCCAAGGCA
ATATGCTTACATGCTAATTTTGACAGGAAATTGCAAATATTCACGAGATACTCTTATCCATATATTTGGCCGTGCTCTGCAAACAGCTTTTCCTAT
GTCATTATTCTCCCCTTCCATACTATCCTTGTGACGTAAGAACGAAATCGTCATACAAAATAGCACTGAGCACGGGGGGAACAGAGGTTATTGC
AAGCATTTGTATAACTCGTCATAAGAAAAAGAAGTGCCCCGTTCGATCGTTGGAAAAAGAAGTGCCAGGTTCGATCGTTGGTCTGCTGTCTGG
ACTCTGGAGGCTGGAGCATCACCCTTCGCCTGGCCGACATACATGGTGCTGCGCCCATTCAAAATTCTCCAAAGGGATGCGGCCGGACATAC
AAGAGTACATGATCCCATTCAAAATTCTCCAAAGCGATGCGGCCGGACACACATGAGTACATGGTACAATATCTGTTCCAAATTATAATTCATTT
GACTTTTTTACCCTAAATTTGAC

Gene: ESD7, TIL1, ABO4
>comp184140_c0_seq1
ACTAAGGACAAGATGGAGTTGGAAGTTGAGGCATATATTAGGCGACGTTATGAAGGGGAAATTTCTGACATAGAAATAATTGAGAAGGAGGAT
TTAGATCTGAAAAATCATCTTTCTGGCTTGAAGAGAAAATACCTGAAGCTTCAGTTTGATACTGTACAGCAATTAATGCGTGTAAGGAATGACCT
ATTACATGTTGTTGAAAAGAATGAAGAAAAATTAGATGCCGTGGAGGCATTTGAATCAATTCATGGTGTAAAAAGGGTAGAAAGGCCACAAGAT
TACATAAGTTACATCATTGATTTGCGTGAATATGATGTTCCATATCATGTTCGTTTTGCTATAGATAATGATGTTAGAAGTGGACAATGGTACAAT
GTTAGTGTATCTGGTTCTGATGTTTTGCTTCAGAGAAGGGAAGACCTGCTTCAGCGTGCAGAAGTTCATGTATGTGCATTTGATATTGAGACCA
CAAAGCTGCCCCTGAAGTTTCCAGATGCTGAATACGACACCATAATGATGATTTCATACATGATTGATGGACAAGGGTACCTGATAATCAACAG
AGAGTGCGTTGGAGAAGACATTGAAGATTTGGAGTACACACCTAAACCTGAATTTGAGGGGTATTTTAGAGTTAAGAATGTTCAAACTGAGTTG
GATCTACTTAAAGCATGGTTTTCTCATATGCAAGAGGTTAAACCTGGCATTTATGTTACATACAATGGTGACTTTTTTTGACTGGCCATTTTTAGA
GAAGAGGGCTGCACACCATGGAATAAAAATGAATGAGGAGATTGGTTTCCAATGTG

Gene: CLF, SDG1
>comp200190_c1_seq1
TTTTTTTTAAGTTCCAAACAGGGCTATGAACTCATGGCATAATTTCTTTTGGGTAGTACAGATTAGGTATTAGAGACGGCATTTTGTACTCCAAG
TAAGCCTACTTTAATGCAACCCTCCTGGCACAACTAGCCAATAAATTGCACTAAACCTAGCTTAACATGCTACTTTTGCGGCAACAAAAACTGG
AGCAGCATTTCGACTAGTTGCTTCAGAAACGCTCTGACCGTACCCCCCATATGAATCAGGCTCTAAAGGGGCACTCGGTACCTATCAGCAGCT
CATGTTACATCAGATCAGTCGCCCGCCAGCTAATTCTACGGATGGCAGGACCAGCTGCCATGGAGAATTCAAGTACAAGAAACATTCGAGTTA
CAATGACAAGTTCAAATAAATTGGGGGGCCTATGAATGACAGGTTCCTGTCTTGGTTGAATGGCTGCCAAATTTGTGGCCCAACAATGAAAGA
ATCAGAATCAGATTCAGCCTCTGGTGTTTTGGGCGAGCTTCTTTGCACGCCCATTGGACGGTTGCCCGTCATCCTTTCCTCCCGAAGCCTCAG
GCTTCCGGGCCCAAGCAGGGGCTCTGTCAGGCTCATAGCGGTAATCATAGAATAACTCTTCGCCAGCGAGGATTCTTTCTTTGGCGAATATGC
CCACTCTATGATCACCAGCTACCATGATAACCTTGGCATAGCAATTCGGGTCAGGGGCATGGTTGGCAAATTTCAGCTTGTCACCCATTCTGT
ATGCATCAAGAACATACTCATTGTTCAGGTTGAAAAGAAACGATGAGTTCTCACGATCATATATCTTTCCACGCTTATCTGCTTCTTTGTGTGAG
ATTAGTTCCCCAGTGTACTCACCAAGATATTCATGTTTGCTAACACTATTCTTGAGGAATGCTCCCCAGCCAGAGACATCTGATCTTCCAAGTA
AGACCCTTTGTTGTTGTTTAAGAAGCAGTTTCATGTTCCTGCATTCATAATTATCTCCTCTCTGGTTTGGAACTCCCAATGTACCATCACCACAC
CCAACCCAGCAGTTTCTGCAAACATCCGGATCGCATTCCCTGTCAGCAGCAAAACATGGACATTGGCGGCTGCGACACTGGCTCTTTGCACA
GTGACATCCTCGAAAACGGTTCTTGCAAATTTTTGGACACCCGCAGTATTTCTCACAGCATGTCCCATTTGTAAGACATGGACACTGCTTTCCA
CATGCAGATTGACAACCACAAGGATTATATTGTCGACAAGGCTGATCCTTCCTTTCAGTAATCCTTTTCCTTATGAAATGGTAACCTGCAGATTT
CCAGGTGTACTTCAAACGACGGACTTTACCTCGCCTTCTAAAATACCTTGACCTTGTGCGCAACTCAGTACCCTTAATATATCCTTTAACAAGA
GAATCGACACCACTAAGAGCTCCAGATGCACTGCTGTTTTCAATATAATTCATATATTGAAAAACATCTCTGCATGTCTTCATTCCAGCAAGAAG
GTTCCGAGCAATTAAACAACTGTTCCTTCCAAAAATCTCTAATCCTTTCACCAGAAGTCCCTGCTCAATTGCCTTCCAGGATCTCAAGTAACCCT
CCTGCTTGCATATATTCTCATCGACGAATTCTTCTTTCCTTGAACTATCATAACCATCTGTCGCTGAATGACGGTTATTTGCTTCTTCTGTTGAAT
CATTATGAACTTCAGCAGAAGCTGAGTTATTCCCAATTTGAGGTGTACTCTTCTTTTTGGAAATTCTTGTGCTTGGAGAACTCTGTTGTGAGGAT
ACAATTAATTCCTTATTTCCATTTCGTGTATCAGATCTAAGCTTCATGTCCCTTGCCAAAGAGGATCCACTAACAGAATTAGAATCAGATGATGC
CATTTCCCTTTGTCCTTTCTTCACACTCATGAGAATTCTCTTGGCAATTCGTCTATTGGTAATCTTTCTGATTCCACCTTTCGGCCCAATTTTAAC
TTTTGAGGGACTAGGGGAGTGTTGTGGAGATTTATTGCTTATTGGATGTACTTCTGAGTCACTGCTTTCTGAGATAACCCTTGCAGTTGAAGAG
CTTTCACTTTGTTGAGATTTAGTTTTCCTTCCACAGGAGCCGTTCTTTTTCCTATTTGGGCCTGGCTGGTTCATCACATTGTCTGATGAGTGAGC
TGGCTCCTCAACATCAATAAGCATGTGATCAACTCCAGCTACAGAATCTGGTTCAGAGGCCAGTTTATAGCAATGAATACCACAGGGTACACC
ATCATCAACGCCACTCCAAGCTGTCTGTTTTTCTGTAGGAAATACTAAATCTTGAGAACACCCGTGTAGCTTGCAATCAAACACTAGACATCGT
CGACAGAAGAGATTGTCAAAGGAATCCAATGCTGCATCCAAATCTTTGTCACGGTACAAATCTTCCACTTCGGCATTATGCTCAGTCCCTTTCT
TGCAAGAATCCTTAGTTTTCTCACCATGCAGGATTTCATACCTGGCCTTTATGTCATCAGTAGCCCGCTCCATGTATCGAGCTAGCGTTTGCAG
TACAGCATCAGACATGCCACATTCTTGAATTGTCATCCGAATAATGCGATCTTCAGAATGTTTAAATTCCTTTTTTTCCTCCTCATCCTCAATGG
CTTCATCTTCACTATCGCTGCAAATTAGAGCTTCGCCACAACTGGTATCATAGTAAATCCTCCGTCGACCAAGTACAGATTGGTCTTCTGTCAT
CCTCTGGTTCCTGTCCAAAAATATCCATGTTGTATAAGGTGGAAGCTTTGGCACTTCTGGTAATTTAATTGGTCTAATAGCATTCTTGGGACCAA
GATTTCCACCAAAAATAGCATTCGATGAGGAGAAAGGACTTTCATCTTGAAAATTGCCACCATCTTTGTCAACCGGAATAATATCATAACTATGC
AGGGTGCATAGCGCATCATCTTGCCTTTTTGTGAGCAAATTTGATGCTAAATCAGTACCCTTTAATGTATTAATTTGCCTATTTTTTGATAAATTG
TGAGTCCTCTGTGTATAGCTGCTGAGATTAGTCTTATTCTCCCCTATCCTATTCTTAATGTAAGTCAAACGATCTGCGGTAATCCTCTTTTTTAAC
GAGTCAATAACAGATAAAACATAGGGGGAGGCAGCATTTTCTTCTCCAGCTTGCACAGCTGTAATAGTGGTGTCCTGGGCGGCGCTGGTAGA
TGGGCGCGACCGGCCCGCGGAAGCCGAGGCAGATGCGGACGACGCCAC

Gene: UBP12
>comp204750_c2_seq1
TAAGATGAAGGGAACTGTTGTGGAAGGAACAATTGAACAATTATTTGAAGGCCACCACATTAATTACATCGAGTGCATTAACGTAGACTATAAA
TCTAGCAGAAAAGAATCATTTTACGATCTACAACTTGATGTGAAAGGTTGTCGTGATGTTTATGCATCATTTGATAAATATGTGGAAGTGGAGCG
TCTAGAGGGTGATAACAAATACCATGCGGAGCGATATGGCTTACAGGATGCAAGGAAAGGTGTGCTCTTCCTTGATTTCCCTCCTGTTTTGCA
GCTTCAACTGAAGCGTTTTGAATATGATTACATGCGAGATACAATGGTTAAGATTAATGACCGCTATGAGTTCCCTCTTCAACTTGATCTTGATA
GAGACAATGGAAAATATCTTTCTCCAGATGCAGATCGAAGTATTAGAAACCTCTATACGCTTCACAGTGTTCTTGTTCATAGTGGGGGAGTACA
TGGTGGTCACTACTATGCTTTCATTCGGCCGACACTCTCGGATCAGTGGTACAAATTCGATGATGAGCGGGTTACAAAAGAAGACACTAAAAA
GGCACTTGAAGAGCAGTATGGGGGTGAGGAAGAGTTGCCTCAAATAAACCCTGGTTTCAATAACACACCATTCAAATTCACAAAGTATTCAAAC
GCCTATATGCTTGTCTATATACGTGAGAGTGACAAGGAGAAAATAATGTGTAATGTTGATGAGAAGGATATCGCTGAGCATTTGCGGATAAGGT
TGAAGAAAGAACAAGAAGAGAAAGAGCACAAGAAGAAGGAGAAGGCTGAAGCTCATCTCTACACTATCATAAAGATTGCTCGAGACGAGGATT
TAAAGGAGCAGATTGGTAAGAATATATATTTTGACTTGGTGGACCATGAAAAAGTTCGCAGCTTCCGCATTCAGAAGCAGTTGCCTTTTACTTC
ATTCAAGGAGGAAGTTGCAAAGGAATTTGGCATCCCTGTACAGTTTCAGCGCTTCTGGTTGTGGGCTAAGAGGCAGAACCATACATACCGGCC
GAATCGTCCATTGACCCCCCATGAGGAAGCACAATCAGTTGGGCAGCTTAGGGAAGTCTCAAATAAGGCACACAATGCTGAGTTGAAGCTATT
TTTGGAAGTAGAACTTGGGCTGGATTTGCGGCCTCTCCCTCCTCCTGAGAAGGGCAAGGAAGATATTCTTCTTTTCTTCAAACTCTACAACCCA
GAAAAGGAAGAACTTCGTTTTATGGGGAGGCTCTTCGTGAAGGCCTTGGGCAAACCTTCAGAAATCCGGGCAAAACTAAACGAAATGGCTGG
ATTTTCCCCTGATCAAGAAATTGAGCTTTATGAGGAAATTAAGTTTGAACCAAACGTGATGTGCGAAATTATCGACAAGAAACTCACTTTTCGTT
CTAGTCAGCTTGAAGATGGGGACATTATCTGTTTCCAAAAATCACCAGGAGCAGAATATGATGCACAAGTGCGATATCCAGATGTTCCTTCATT
TTTGGAGTATGTCCATAATAGGCAGGTTGTGCATTTCCGGTCTTTGGAGAAACCAAAGGATGATGATTTTTCTTTGGAATTGTCAAAGCTTCATA
CCTATGATGATGTTGTTGAGAGAGTTGCTCGCCAACTTGAATTGGATGATCCAGCGAAAATTCGGCTTACATCCCACAATTGCTACTCTCAGCA
ACCTAAACCACAACCCGTCAAATATCGAGGAGTGGAGCATCTGTTGGACATGCTCATCCATTACAATCAGACGTCTGACATCTTGTATTATGAA
GTGTTGGATATACCACTGCCAGAATTGCAGTTCCTAAAAACCCTGAAAGTCGCATTTCACCATCCTATGAAGGATGAGGTTGTTATTCATAGCA
TCAGGCTTCCAAAAAATAGCACCATTGCTGATGTAATAAATGACTTGAAAACTAAGGTTGAACTGTCCAGTCCTAATGCTGAACTACGTGTGCT
AGAAGTCTTTTACCACAAGATCTATAAGATCTTTCCACTACAAGAGAAGATAGAGAATATAAATGATCAATATTGGACACTACGTGCTGAGGAG
ATTCCAGAAGAGGAGAAAAATATAGGTCCAAATGATCGGCTTATTCATGTTTATCACTTCATGAAAGATATTAATCAGACTCAGCAAATTCAGAA
CTTTGGAGACCCTTTCTTTCTGCTCATTCATGAAGGTGAGACTTTAGCAGAAGTCAAGAAGCGCATACAGAGTAAACTGCAAGTCTCTGCCGAT
GAATTTTCCAAGTGGAAATTTGCGTTTATATCGATGAATCGGCCAGACTACCTCCAGGATTCAGATGTTATATCTACTCGCTTTCAGAGGAGAG
AGGTCTATGGAGCTTGGGAGCAATATCTTGGGATGGAGCACACTGATACTGCACCAAAAAGGGCTTATACAGTGAATCAGAACCGTCACTCAT
ACGAGAAGCCGGTCAAAATTTACAATTGAAATTTCTTACTGGTCTGGCCTCATGGTGAAGTACTCCATATGAGTGGAGCAGCCAAAAAAAGGG
AGGATAAGGAATTTATACTGTCAGGGGGAAACCTCATAGTTGTTAACCATTCGGTGTGCCATGACCATCTATTCCAGTGCTACAGGCCGGCTC
CTGTCATCTCGTCTGTAGCAGCCAGCTCCCATGGGCCATACCAATGATTTTTTTCAAGATTTTCCTTCTTTGCCCTCGGGTGATGGTGCTAGGG
CATTGATGTCTGTTTCGGTTGGGTACAAGGGCAGATATTGCTTCATTCTTAGAGCAAAGTTGGGTACATAAGATTTTGGTAGGAGGCGTGTATA
CAGGCAGTCGGTATAGTTGTGTCTCTCTCGGCTTGAACCCGCTTATGTTGTCTTGTACTATTGTTAATGTAGTGACAATGGTTGACTGTAAATT
GGTTCCAATTTCCAGCTTCCATCGCATGTCTGAAAAAGGAGAGTAATTTTGGGGCCTTTTGGAAACATGCAATATTTGACATCTTGTTGCAAGA
AACTACAGCTTGCTCAAGGACCTACTGCTTCGATCATCACTAAAGCAAGAGAAAAATTGTACTTTTTACTTTTGTTTATTTTTGCAACGTGATCAT
GATGTGTCAAAATATTTGTATGCTTCTTAAATGGACCCAAAAGTTGTTTCCAAAAAAA

Gene: ESD7, TIL1, ABO4
>gg_15862
GGCCCCACCGCCTTCAAGTTTTGGCGCTCATTTCTCCCCTCCGCAAAACCCAAATCTAGGGTTCACTCTTCCCCTTTCCCGGCGTGCAACGGC
AATCAAAAGTATCCAAGCGGCCACTGGGCCAGCGAGTCTTCCCTTCCCGCTCAGCGGAGCGCACTTGCCAGCATGAGCGGGCGGCGGGGTC
CAGCTGCCAGAGGAGGCGGAGGGGGAGGCGGCGGGAACTGGCGGCGAAGCGGGTCGGCGAAGGCGAAGGAGCAGCGGCTGCGCCTGGG
CGCGGAGGAGCTCCTCGAGAGCCGGTTGGGATTCGCGCCATACACTGACGGGGAGCGTCGCCTCGGGTGGCTTCTCACCTTCTCGCCGTCA
TCATGGGAAGACGAGGACACCGGAAAAATCTACAGCTGTGTTGACTTATACTTTGTGTCACAGGATGGGTCTACATTTAAAGTGAAATACAAGT
TTCGGCCTTACTTTTATGCTGCGACTAAGGACAAGATGGAGTTGGAAGTTGAGGCATATATTAGGCGACGTTATGAAGGGGAAATTTCTGACA
TAGAAATAATTGAGAAGGAGGATTTAGATCTGAAAAATCATCTTTCTGGCTTGAAGAGAAAATACCTGAAGCTTCAGTTTGATACTGTACAGCAA
TTAATGCGTG

Gene: PHP, CDC73
>comp180194_c0_seq1
CGGCCGCCGAGGAGGCCACCACGGGGAAGCAAATCCGCGCGCTCGAGCGGCCCTTCAAGGACCGCAACGCGATCCTCGACGCCCGCGGC
CGCGACTTCCTCGCCGTCTTCCAGGCCGCCGTGCGCCGCCAGGACGAGCAGCGCAAGGCCGGCGGCAAGGACGCGGCGCCCTCTTCCCG
CCCCGACTCTGGCTCCGCCGCTGCAGCCCTGGCCAAGCCCAAGGTGCTGGACAGGGCCCTGGGCGACGGCGTCGTCCCCATCATCCTCGT
GCCCAGCGCCTCGCAGACGCTGATTACGATCTACAATGTCAAGGAGTTCCTCGAGGACGGGGTGTTCGTGCCCAGCGAGGAGAGGATGCGG
GCGACCAAAGGGGGGAAGCCGGAGAGCGTCACGGTGCAGAAGAAGCTGATTCGTACGGAGAGAGCTGGCGGCGCTGGGGGTGCCGTCTC
CTTCGAGGTTAGGGACAAGCCGGCTTCGCTCAAGTCAGACGATTGGGGGCGGGTGGTGGCCGTGTTTGTGCTCGGCAAGGAGTGGCAGTTC
AAGGACTGGCCCTTCAAGGACCATGTAGAGATCTTTAACAGAGTTATTGGTTTCTATGTGCGCTTCGAAGATGATAGTGTGGAGGCAGCCAAG
GTGGTGAAACAGTGGAACGTGAAAATCATATCTATAAGCAAAAATAAAAGGCATCAAGACAGAACAGCTGCTCTTGAAGTATGGGAGCGCTTG
GAGGAATTTATGCGGGCGCGTACATAATTGATGTGCAGTCATTGTGACTGTTAAACACCCAAATGTTATGTTGAGCTGGAGGCAGACTTTGTG
CTTGTTGAATTACTGATCGGCAGAATGACTTTATTTTTCTTAAAGGGAAAGGATAAAAGACTACCATTTTCTATTCAATCTCATATTATTGATCTA
TGACAGTAGATGGCAGAATGGAATGGCTGGAAGCCCTGAAGACTGCAATCTTTGTACGCTAGGATGTATTTGTAATGTTGTACTTGCAAACTTC
TAACTTCTTATTTCTAGGCAAGTATATGAGATAATTTTGCAAGTTGCACCTGAAAGTATATTTTGTTTCTACTTCTCTATGTTGCGGGGTCGCTG
GCTGTCAATCTTTTCGTGTTCCTGCTGCTATTATATGAGGTTCTAAGTTTTACATTTTTTTCTGTATACTAATCTGCTTCACTTTGGC

Gene: ATPRMT10
>comp184054_c0_seq1
GCCGGCGCCGCCGCCTCTGGTTCCTCCGGCGCGCGCCCAGTGGACAAGGAGGTGGACTTCGCCAACTACTTCTGCACCTACGCCTACCTCT
ACCACCAGAAGGAGATGCTCTGCGACCGCGTCCGCATGGACGCCTACCACTCCGCCGTCTTCCGCAACGCGCCGCACTTCCAGGGCAAGGT
GGTCCTTGACGTGGGCACTGGAAGCGGCATCCTTGCCATCTGGAGCGCGCAGGCCGGGGCCAGGAAGGTGTACGCCGTCGAGGCGACTAA
CATGGCGGAGCACGCGCGCGAGCTTGCTCGCGCCAACGGCGTCGCCGACATCGTCGAGGTGATACAGGGCACCGTGGAAGACGTTGAGCT
TCCAGAGAAAGTTGATGTCATTATATCAGAGTGGATGGGCTACTTTCTTTTGAGAGAGTCCATGTTTGATTCCGTCATTTGTGCACGTGACCGC
TGGTTGAAGCCAGATGGTGTAATGTATCCTAGTCATGCTAGGATGTGGCTAGCACCCATAAGAACTGGTTTGGGTGACAAAAAGAGGGAAGAT
TTTGATATAGCTATGGATGACTGGAGTCTGTTTGTTCAGGACACTCAAACATATTATGGGGTCAACATGAATGCTTTGACAAAGGCATATCGTG
CAGAACATGAAAAATACTATTTGAAGTCTTCAATATGGAACAATCTTCATCCAAACCAAGTTATCGGTCAACCTGCAGCTATCAAGGAAATTGAT
TGCTTGACAGCCACTGTTGATGAGATTCGAGAAGTCAGAGCACAAGTTACAATGTCAGTCAGAATGGAAGCTAGATTGTCAGCACTGGCTGGA
TGGTTTGATGTTCATTTTCGAGGTAGTGCTCAAAACCCTGGGGTGGAGGAAATTGAATTAACTACAGCTCCAGATGAGCATGGTGGAACTCAT
TGGGGTCAGCAGGTATTCTTGCTGACTCCACCTCTGGGTGTGACTAAAGGAGACAATGTTAATGTTTCCTTCTCGATGGTTCGCTCAAAGGAA
AATCATCGACTTATGGATATGGAGTTCACCTATGAATTGCATGAGTTATCTGGCAGGAAGCACCCAGCAGTCACAACCAAGATGTACTTAGAGT
AGACAAGAGTTGAGTAGGCCTGATTCATCATCGTTAGATGAAAAGACGGAACCAGTTACCCTGTGCCCTCGCATCGCTGTTACAGAAACCTAT
GCTCAGAAGCTGAGCATATGTAAGTGCCTTAGGGCCCCTGTGTGTTTGTTATACCCATGATGTGTACATTGAGACCATGAGGCATGATGACGT
GCCAAAGATTTGGTTGTATTGGTAATATCTTATTGTTTGATTGCTCAAGCAACAGAGGAGCTTATTGCTTCACAACATTAATGTATTAGCAGATA
TTGACCCACATTACAGTTTTAAGCAACATTCAGTGCAGAATTTTGTTGAATTCAAAAAAAA

Gene: TEM2, RAV2
>gg_06658
GCCACCACCGGCAAGGGCGTGCTCCTCAATTTCGAGGACGGCGAGGGCAAGGTGTGGCGGTTCCGGTACTCGTACTGGAACAGCAGCCAG
AGCTACGTGCTCACCAAGGGCTGGAGCCGCTTCGTCAGGGAGAAGGGCCTCCGAGCCGGGGACACCATCGTCTTCTCCCACTCCACGTACG
CCTCGGAGAAGCAGCTCTTCATCGACTGCAAGAAGACGACCAAGACGACGACGGCCGCCACCACCGACGGGGCGCCGGCGCCGGTGCCGG
CGCCAGTGGAGAAGAAACCAAGTACTGAAGCCCGTGTAGTGAGGCTGTTCGGCGTCGACATCGCCGGAGACGGGTGCCAGAAGCGCGCAC
GGCCGGTGGAGATTGCGTTCGAGCATGGGCCGCAGGAGTTGTTGATGAAGAAGCAATGCGTGGCTCATCACCGCTCTCCTGCCCTAGGTGC
CTTCTTGTTATAGCAGCGATCGTACGTATTATACACATAGTAGTATTCTAGTCCTCCCCCTTTTTGTTCTTCCTGTCGCCTGCTTCATCTCGAAT
TCAATTACCACTAGCTATAGTGTTTGTTCTTCGTCTTGCAAAGAATAGACTGCAAAACGAGCTGTATGTATGTACACCAAGTCTTCAGATGCTAG
TAGAAGTAGCAGTCCTGTAAGATCATGCAAGTTTGTGCCATCATCGAGTGCAGCAATAAGTTGATTCGATCTGCTGTAGACTAAACATGTGTGC
TTGCTTCTTGTTCTGGTGTTGTTGGATATTCATCACTCTTGGCAACTCAGTCTAGCTTTGATTAGCTTTGCTGTAGCACGTCGACATGCATGAAT
AAGTCGATAGATGTTGCG

Gene: TPS1
>comp198109_c0_seq1
CGACTGGAGGAGTTGGGTGAGATTATCAAGTGACGGAAGCGAGCACATAAGAAATGCCAAAACCGCATTGCTGAAATGGATATGCAATCATG
AAGGCCCGTGGCTTGTCTGACCAAGTAATGCAGGAATGCGGTTGGGATTTGGAGAGGAGCCCCACGGATCAAGCCACCCATTCCAAGCAGG
AGCGCTCATGATTGAGATGTCACATTGCTTGCATGAGCTGCATTAGTATATTGGTAGCTTTCTCCGATCCCGTGCAAGTTGATTCCGTCAGGG
CCTGCAATCCCTGCCCAGCCTCACGTTGCAACAAACCTCCTTTCAACATCACCTTCTCTAAGCATGCATCCAGACAAGATTTCAGTCTCTAATA
ACTAGTAGTAATAAGTAGGTAAGACTAAAAAAAAGCATGAGATGTGGCCTCATTGCGCAGGAGTATTTCAGAAGAACTGTAGCAAACCAGTTGA
AAAACGCAATCTCAAAAATTCGTAACGAGAAATCAAAAGCCGGCGAACGATTGAATATCATAGCAGAAGCAATGAAACTCTCTAAACTCCTTTT
TATCCTTTCCCATTTGCTCCCGTCTCGCAGTTTTATCACGAGCATAAATTTTGTACAGGATTGTACAGCACGCTGACTAAGCTACGGGTGTACT
CCCTCTGTCCCTAAATGTGTCTCTAGAAAACCGTGTTTCCCTCGTAAACACGCTTCCCGGACGACGTACATATAGGGACGGAGGGAGTAACTC
TGTCGCGCATATGAGGAATGAATATTTTTTTCCCAAGGTGCTTCTCTGGTAAAAGCCCGGAGGCTCCACCAGTTTATTTACTGCCTATCCATGA
ACATGTAATCAGCACAGCTGGATTGGAAGCCAGCTGTTGCCGATGCCAATTCCTTGAGGAAAGAGACAACCTCCTCCGACGAACTCAGCAGG
TAACGGGCGTTGGACCGCTTCCTTCCGACGGCGCAGGAGAAGTAGTTCTCGCCCTTGAGATCAAGGACCGAGGACCCTTCGCGCCAGTCGT
GGTTGCTGCTTGCGCTGCTGTGGCTTGATGAGGACGACCTCTCGGATGCAGCCTGCTGCGCCTTCTGTGTCCTGGACTGTGGGTTCCTGGA
GCTGCTCCTGCCATTCGATGGTGGCCTTCCGTTCGGCCTCCGGTCAAGTGAAGCTGAGCCACCCTCTGGTTTTATTTTGGATTCAGAAGGGTA
CTCGGGATCAAAAAAGACATAGATGTCCTCATCCTTCCCAAGGAAATGCCCTATACACAGGACATAGTCAATTGGAGTAACCATGTTTTCGCTG
TGAACTATCTCCCCTAAAATACGATCAATTGCAGCACCCTTTGTAACTCCAACAGACCGAACTTCAACTGACCGACTCCCTTGAACAACATCAA
CAGCTGCATTTGAGATCGGACCTGTCCACAAGTGCTGCAGCATATCTCTTGCTTGTAGCCTTCCAAATTCAACATCAGCATACTTGTAGTTCCA
CACAAATGATGTTTCACGATGTTCGAAATGGGATCTTGGGGTTCTTTCTGTAAAGTATTCAAAAACATGCTGGAGAAAAAAAAATGATATCAAAA
GCTTGATATTAGTTTGACGCAACAGAACATTATTCATGAAACAGACACACCTTTACGCTGTCAACCCAATCCATGTTCAGATGCTCAGGCATTG
TTGTCATCCATTCTCCGTAAGTCGGGCGTAAAAACATCCCATGCTCTGCCGCCAACCACATTTTAAATTCTCCAAAATTCTGTACACATGGAAA
GATGTGTTGGTTCTAATGGAAAATATAAAGGTCAATCGATCAAACCTAAAATAACAGACAGGTCTGCTTTGATATTACTTCATCAAGAACACTCC
TGTCACTGCCACTGAGAACAATAACTGTAGTGTGCTCATCCTCACAGAGGGCTCTCAGAGGACCCTTTAAGTCAGGATGCAACTTGAGTTCCA
TTTCCTTGATTTGGTCACCACCCCTTCTCCCAGAGGATTCGACTGGCTCAGTCAATGTTGAATTGAAACCCAATATGAGCAGACGATTTTTAGA
GCGCAAATATTGCTGGATTGCCGTTTGACTAGGAAGACCAGGAGGGACTTGTCTTGTCCTCAGTAGTGCCTCAGCAACCGTGTCATTTAGCTC
AAATACAAAAGTTTCAGCCCAATCTTGAGCCGTGTGAGTTGTGACATGAGCATAGTTGTGCCTGTGTCGTTTCTCTCTCTCATCGGATGGCATC
GTTAAAGCATGCCGTATTGAGTCTGCAACTTCTGTAATATTCCAAGGGTTTACTAGAATGGCGCCAGCTCCAAGTGATTGTGCTGCCCCAGCA
AACTCACTAAGTATCAGAACTCCTTTCTTAGACCCTTGGCATGCAACGTACTCATAGCTCACAAGGTTCATCCCATCTCTCAGTGATGTTACAA
GAGCAACATCAGTGACTGCATAAAGAGCACACAAGGCATGGAAATCAAGAGATCGGTCCAGATGATGAATAGGGACAGCAGTCAACGTTCCG
AATCGACCATTTATGCGCCCAACAATTTCATGCACTTGGCTCGTTAGCTTTTGATACTCAGGGACGTCAGTTCTTGTTGGCACAGCAATCTGCA
GTAGAACAACTTTGTTGTTCCAGTCTTGGTTTTCCTCAAGAAACTTTTCAAAGGCCAAAATCTTTTGTGGAATTCCCTTAATCATGTCAAGTCGA
TCAACACCAAGCATTACCTTTCGACCAGCAAAACGTTCTGTCAATTCACTGATGTGCCTTTTTACTGCTGGAAGCTCCAATGCTCGCTTGAAAC
GATCAGAGTCTATCCCAATAGGAAACGCTGCAACCCTGGTTAGTCTTCCTTGATCTTCCACACCCTCAGGGGTACCCTCAAGTCCAAGTATTCT
AGTGCAAGCACTCACAAAATGCCTCGCATAGTCGTAAGTATGAAATCCGACTAAATCAGCACACAACACCGAGCGAAGCAGCTCCAAGCGGG
ACGGCAGTGTTCGGTAAATCTCTGATGATGGGAATGGCGTGTGCAGGAACCACCCGACCTTCATATTGATGTCATGGTCCTTGAGGCACTTG
GGCAGGAACATGAGGTGGTAGTCATGGCACCAGATTACATCCCCCTCCTGGTAGTGCTGGTACACGACATCAGCGAACATCTGGTTAGCACG
CTTGTACGCGTCGAACTGTGACTCAAAGTTCCTCGTTGTTGCCAGCCTGTCCTCCTGTGGTAGTCCTAGGTAGTGGAACAGCGGCCACAGGA
TGTTGTTGCAGTACCCATTGTAGTACTGGTGCACAATCTCCTCGTCCAGGAACACTGGTATGCATCTCTTCTCGGCAAGAGCTTTGGTGAGGG
CTCGCTGGCCAACCTCATCTGGTACGTTGACGCCCGCCCAGCCAATCCATTTCGCATCGACGTCCTTCACGCCAAGCAAGGCGCTCACGAGG
CCGCCGGCGCTGATCTCAAGAGACCATTGGTCCTCGCCGCGGCGGTTGGCGGAAACAGGGAGGCGGTTGGCCACGACGAGGAGGCGCTG
CTTCATGCCACGGAGGCCGAGGCGGCGGAAACCGCTGCGGGAGCCGGCGAGCGGCGGTCGCGCGGCG

Gene: UGT87A2
>comp202963_c1_seq1
GGGGCGCGGGTCACCGTCTCCACGGCCGTGTCCGCGCTCCGCAAGATGTTCCCGGACGCGGCCGTTGGCGGCGATCAGGCGGATGCGGA
GGGCCACCGCGACGCCGCGGGGGTGTGGTACGTGCCTTACTCCGACGGCTACGACGACGGCTTCGACAAAGCCGTGCACGACGCCACGCA
CTACATGGAGCAGATCAAGCTCGTGGGGTCCCGCACGCTGGGTGACGTCCTGGCCCGCCTCCGCGGCGCGGGCAGGCCCGTCACGCTCGT
CGTGTACACGCTGCTGCTGTCGTGGGTGGCCGACGTGGCGCGCGCCCACGCCGTGCCCGCCGCGCTCTACTGGATCCAGCCGGCCACCGT
GCTCGCCGCCTACCTCCACTTCTTCCGCGCCACGGACGGCGTCGACGCGGCCATCGCCGCGGCGGGCGGCGACCCGTGGGCGACCGTCC
GGTTCCCGGGCCTCCCGCCGCTCCGCGTCCGCGACCTGCCGTCGTTCATCGTCTCCACCTCCGAGAACGACCCCTACGCGTTCGTCGTCGA
CGCGTTCCGCCAGCTCATCGAGCTGCTGGACGGCGAGGACAGCCCCAGCGTGCTCGCCAACACGTTCGACGCCGTGGAGCCCGAGGCGGT
GGCGTCGCTGCGGGACCACGGCGTGGACGTGGTCCCCGTCGGACCCGTCCTCTCCTTCCTCGACGACGACGCGGCGGCCGGCGGCGCCA
AGAACGGCGGCAACGACCTGTTCAGTCAGGACGGCAAGGGCTACCTCGACTGGCTGGACGCGCAGGCGCCGGGCTCCGTGGTGTACATCT
CGTTCGGGAGCCTGTCGGTGATGAGCGAGCGGCAGATCGAGGAGGTGGCCCGGGGCATGTCCGAGAGCGGCCGCCCGTTCCTGTGGGTG
CTGCGGGAGGACAACCGCAGCAGCGACGCCGCGCCGCTGGGCGGCGAGCGCGGCATGGTGGTGGGGTGGTGCGACCAGGTGCGGGTGC
TGTCGCACCCGGCCGTGGGCTGCTTCGTCACGCACTGCGGGTGGAACTCCACGCTGGAGAGCATGGCGTGCGGCGTCCCCGTCGTGTGCG
TGCCGCAGTGGACGGACCAGGGCACCAACGCGTGGCTGGTGGAGAGGGTCGGCACCGGCATCAGGGCCGCCGTCAGCGACAAGGACGGC
GTGCTGGAGGCTGACGAGCTGCGGAGGTGCCTCGACTTCGCCACCTCCGAGATGGTGCGCGCCAAGGCCGCCGTGTGGAGGGAGAAGGC
GCGCGCCGCCGCGTCGGAGGGCGGCTCGTCGGAGATGAACCTCAAGGCGTTTGTCGCTAAGCAGATCGCCGGCGGCAACAACTAGGCCAA
CTCGCTACATACTAGCTCTCGCGCGCGTCGTCGCATGATTATTATTGCTCTATATGCCCACATCAGCTGGCATAGATACTATACAAGACTTAGG
CTCTACAACTGTACTTACATTGTTGCTCTTGCTGGATTTGGTTCGATTCGAATTAATAACTAAGCGTCTATCATGTATATGTATGTTCATGTCTTC
ATGATGTGTGTTTTTTTATAAAAAAAATCAGTGTTCTTCATGAAAAAAAA

Gene: LRB1
>gg_03849
GACAATTTGAAAACGAGAACTGAATTCTCTCTTGTTTATTTAAGTGCATCAGTAAAAAGTAAAAACAGCACGCACAGAAGATTTCATTTTCTACA
CTTCAGATCATAATTGCATGGATTTGGACACCGCGGCGCAAACGCAGTTAGTAGTACACAGCTGCGTCTAGGGTTGTTTTATGGTCAGCTCAG
CTCTTAGATGCAATACCCCGTCAATGAAGAAGAGACTGTCATCAGCCATGAATGATGGCCACGGGATCGCAAAGAGATTCCGGTACCCAACC
GCCTTGCCACCAGTGAAGGTATAGTAGCCCTTATATTTGCTGACAAACTCACCTGACGGCCTTGTCCTTGCAGCAAACTCATAATCCACGGTA
ACACTTGTCGACCCTTTCTCTTGCATCCCTAGGAAGAGACCAAAACAATAGAATGCACTTTGCTGGTCCATGTTGCAGTGTGCTGAGAGGAAG
AAGCCCTGTCCAGCTAGATGGAAAGCTTGTGAGTATATCCGTCCGGATGGAAAAAGTCGGCTGCACTCCTCACGCTTCAGGTCCAAGTATGCT
ATGCACTGTGGATATGGTCTATCAAACTCTACAACCTTAAGCGGACGATATTTGTAAGCTCGCTCAGCATATTTCCTACATGTCATCACATCTG
CTGCAAGGGCACGCTGCCTATGTGGTGCATCAGCTTTGTAAAGAAGTGCCTCTGTGACACATTTAGTCGCCTGCTCATGATCCAGATCACTGC
AAGCAAGAACCTTCCGCAACTTCCTGCAGGTCATATGGCAGAACCGAACAAGGGGCAGCAAGCGTGTGCCCAAGATCTCACGTCTTTCCTCC
AATTTTGGGTATTGAGCACGTGCCCACTTTATCACAAAGTCGTAGATAGCATCCTCTGATGCGACCTGAAGGTCATTACTCCATAGGATGGCTT
CAATTCCAGCAAGAGGAATGTTCATTGCTTCATCTTGGAACTTGGTCAAATCCTTGTATTTATTGGCAAGGAATTCCTTGGCAGTGTCTGTCAG
TGGCTGAACTGCAGCAGCCATTGAAATGCTTGAGGGTAGATCCAGATAGAGCAGTGCAGACTCTGTGGTCATAGGCAAACTTCTA

Gene: PRR7
>gg_07428
ATTCTTTGTACTGGAGCCCACATCGTTGTTGCTACTGCCATTGGATCCCTGCTTTATTGGAGCAGCATCTGAATTTGACTTCATCTTGCATGTA
GAGTCCGTTTTCACAGCCTCAGAACTGTTATCTTGTGGTGAAGAGCTTTCCCCACATCTTGCTCCACCTTGATTGGAAGCCACAGATGTATGGT
ACCTTGCAAAAATCATAACGAGTTAGCATGCAACTTACTGTTATGTTAATAACCTAGTAATGACACTGATATTATGTTTTGCACCTGGTGAAGGC
TGACTGATCTGATCTTTTCAGAATATTTCGCTCATCATGGATTTCGGCTGTAGCTCCATCTGTTCTCAATCTCTTCAAACCCAACTCATGGGGTG
TCACCTCGATATGAGAATCATGATTCTTATCATTTCCATCGGGCATCTTGGAAGGGCCATCAGGTGCATCAACGGCTCGCACTGCTTGTTGGC
CTTCTGTGTTTCTGGCTATTGAGCTAATTAAATCAGCAGTTTCAGTATTTGTTTCACTAGGTATGTCTGTACAATCATCTTCTGCCATCATTTTCT
TTTTGGATTTGCATTGGGGCATGTGAGTCTCATGTCGTTTTTCTGTTGGATTAACAGATACGTCATTGAGAGATGATTGATATTCTGCACTTGAG
TTCCTAGGAGCACCTATTTCTAGGTATCTTCCCATAGATTCATCTTTATTCTCCTTATGTTTCTTGCCATTCCTTTTGTTTGCTGTCGGTAGCCAC
TTGTTACTGCATATCTCTGATTTGGGGTGAATTACTTGTGCACATGTACTATCAGGTGGATCTGCTAGTTGATCAGGAGACATAGGTTGTGGAC
TGTCAATCTCCACAGCACGCTTTGTCCATGAGCTTTGAGTACCACTGCCATTATCACTTCCATCCCTAGCATTGAGTCCGACACTGAAGTCGTC
ATCGTCGTCGTCATCATTTTCTTCATCATCATGATTGCTGTCACTGTCGTTCTCATACTCGTCACCAGTATTTGGTTTGGCACACTTCTGTGTCT
GGATGCCACTTTCACTTCCACTGCCACTGGAACTGTGGCATCGCCTCCAAACGTGCTGCCAAAGGTTCTTAAGCTCATTCTTACGTAGTGGCT
TTACCAAGAAGTCAACTGCTCCCTTCGACAAACACTTAAACACCATACTCATAGAGTCATTTGAAGACATCATAATCACAGGAATGTCCTTGCAA
ATTTTGTGACTAGTGATTTTGCTAAGCAGACCGATGCCAGATAGACAAGGCATGAAAACCTCAGTCAATACAAGGTCAATGTTGTTCTGCAGAT
TTTCAAGATATCGCCATGCATGTGAACCATTTTCAGCAGGGATAACTTCATAGCAGCACTTACGGAGCAGGGCACTGACCACCTGACGAGTAG
AGTCATCATTCTCCACCAGCAAGACTCTCAGTGTCTTCACAGGTAGGAACCTCTCCCAGTGAATGCTTTGTCCTTGTTGCTCCTTCTGTTTATC
CACCTGCTGCATGCCCGGTGGCGCACTGCGCCCATTGGCTAAGTCATCTTCCTTTTCCCTCCATTCATCTGCATCAGCTTCAGCCTCATGGTG
GCCATTATCTAAGGTGACATTCCCTATCCCCAACACATCCTTGCGGGAAGACCCGTCCGTGCCAGCTTGGCAAGCGCTACCCATTAGCCTGAT
TCGAAGAGAGAGAGAGCACTGCAGTGGTTGATTTGTGCAACTAGAGGGAGTTCACAGATGGTGGAGCACATCCTATGTCAGCATAAAAAATG
CATATGCTGAATCCAACTACGAATCTGTAGGGAAAACGCACGGGGATCAAACACCAATCTGTCATGAGCATAACAGATGCTCAGAATAAGCAC
TCCATTCTGCAGTATTGGTCAGAGCCAGAGAGGATTACATGTAGCCGGGCTGGAAGAAATTATTCTTAAGCCATGTGCTCTGCGACGCTCAAA
ATCAGTTGCTTCCGACTCCGTCTCCCTTCAAAATCCAACTAATCCGCGGTCCGAGCTCCACCCAAGGTAACAGCAGCGCCTGGGGAATCGGA
CTCTCGAGATCCACTGGAGGCGACCCTGACAAACTCTACGCGTGGGGTGTTTACGATGACGAGGAGGAGATGAGGCCGCGAGGGGTTCCGG
GCCAAATGGCGGGGAGAGGCGATGCGCTGGTGACTGGTGTGATGGCGAGGAGGCCGCGGAGGAGGCGATAGATAACGGGGAGAGGGGGA
AGGGGAGGCGCTCGCCGTGACGTGGAGCAAGTGAAGGTGGGCAGGGAGTGGGGGCG

Gene: SPA3
>gg_15635
CATACATACAGTATTAGTGTTTCATTGCAATGGTACAACTTTCGAAATTAAACTAGCAAAGTATGTCCATCAGATAAAACTAGACACTATCACTT
GTAGAGCCATATTATGAAAAGAGAAGTAGAGAACAAGTATGTGTGAAAGTTTTATCAAACCTTTGTGTTTGTATGCCCCGTAAATGTTTGAAGTG
GACTGTCAATTACCCTTCCTTGACTCATCGACAAGTCCCAAAGCTTCAATGAGTTGTCAGTAGACGCAGATACTATTGTTGATGCATCCAGATA
CTTAACATAGCTTACAGTCTTTGTGTGCCCAACTAGTGTACAATAAGGAGCTCGTATGTTACGAAGATCATAGCAGTAAATTTTGTGGTCTGCA
GAGCCGATGGCAATGGAGCGAGCAGAATCAGGTTGAAATTGGACAGAACACACGTTTGCCCTTGTCCTAATAGTGCCAATACTCCCAGCCTG
CACAGTTCATAAAAATGTGTTATAACAATGGAGAGAAGCAATTCTACATTCACAGACAGGTTCGATAACCACGCCATCGTTAGTTCAAAAGCTG
ACATACAGTAGGTGCAAGAATAAAATTGCCTGATTCATATCCCACAGCTTCACAGATCCATCATCACTCCCACTGACCAATTTTGTTGGGTCCA
CAATTGAGAAGTCCACCGACCACACACGTCGCTCATGCTCCCTCATCTCAACAAAAACTTGACTCCTAGTAACATCCCAAACCTGTACTATACC
CTCAAAATCACTAGATGCTATATGGCTCTTCATATAACTGTTCCAGCAAATACAACTAAGTTTCGATCTATTAGACATCTCTACCACAGGATAGT
GAATATCACGGTGTTCATTTACAATCATATTATACTCAAATACTTTTATCTTCTTATTTACACCAGCAGTTGCAAAAAATTCTCTATCCCGGTCAA
ACCCTACGGAGCATACTAAGTTTGATGAGTTCAGCAACTCACAATGCTTCAGTTCTGCCCGAACTTTCAACTTACTGAATGACAAGTACTTGCA
CAACCCCTCAAGAAAAGAGTTCACCCAGCCTCTTTGCCTTCTACCATATTGCCTTTCCAAAGAAAAATCATCTATTGAACTTCCCTCAGTCCCAA
TAGCTGAACCAGTAGCCCTCTTAACAACTGATAATGATCACACAGAACCACTTTTGCCTTCTTCCTCCGAATCAAACCTTGGTAATCTTGAAGAA
GATGATTCCTTTGACATTGATATTCTCTTAGCTGAGGGTGAAGGTGCTATAAAGCCGAAGAGAAGAAGACCAAGA

Gene: CUL4
>gg_11383
AGACACGTTGCAACAGCTACAACTCCCTCCCAGAGAACAAAAACAAGGAACTGAAAAATAAATTAATTTACAATGTCTTTTCCGATAATAAAGTG
AGCACGGTGTCTTTAATGGTCCGCCAAAGCGCTGTCAAGCCAGATAGTTGTAGATCTGGGGGTTACTCCTGTCCCGCTCCAGGTACTCCCGG
TCGATCAAGCTCTCTATTCTTTTCTTGATATCAGCTGGCTTAACAGGGAACTTGAGCTGCTGGTAAAGCTCCGTTATTAGAAGGGTGTGGCTGA
GGACCTTCCGCGTCTTCATTATTCGAACTATGGCCGCATCAACCTGATACTGCCGATCCTGGAACACTCTCTCAGTGGTGCTCGTGTTTTCTTC
TACCGTCTCCTTCATTTGAATTGCATTTACCTGCAAGTGCACAACAACAATTCAAGCTTAATGACCATATCAAATATAAAGTGTATGTGTACATAT
ATTACCTTTATGCGATAGAGGGGAGCACTAAAATCTTCATTGAATACAAATTCATCCTTATCTTCTACATCTCTCCCTTTTGGTATCTTTTGGAGA
ACCCGAACTTTACCACATGCAAGTGATTGCAGTGTCCGTCTCAATTCTTTATCCTCGATACCAGTGGAATCTGAAACAATGACACTGCGAGTTC
CTTTTTGCCTTTGGGAAAGTCCACTTTCAAGACACAGTGTCCCAAAGAATTCTGCCACATTAAACGCCTTCCACTGTACTTGCTCAAATAGAACT
CTTTAAATATATCCTGATACACATTCAGTTCGTGTGGGAGTTTCACATCCATTGGTGGATATGTAGGCCAATAGCCTGTTGTAAGCACATGGAC
ACTCATTTCAATGCCAGAGGGAAGCTTTGTCCTAGCTTGAGATGATTGCCTGAATGAATCATTTATTTCTTTGGATAATTCAATGTCCTTGAACA
TTCCCTCCAGCTTGTTGGTAAACTGACTCCCACACTCAGTTTTGAGTTTTGGTCCTGTTTCGGTTTATACAAGGTAAAGATGTATTTGAGGCATT
CTACAAGAAGGATCTAGCTAAGAGGTT

Gene: GA2ox8
>comp193746_c0_seq1
GGCCTGTTTCCGTTTCATATAGTACATTAAGCTTTTAATCAATTGTTTGTTGTAGCATAGTACACAAGTATAATACTGGTGGACCATCATTTATAC
ATTAAGCTTTGTAGCATAGTACACAAAGATTGGTCGAGTGGTGGACCATCATTCATTTATACATACATTAACTGAGACAGAGAATAACTTGAGC
CATATGTACATTCTTTCCTATACACTTCCCCATTTTCAAATATATACGGCCAATACTCTAATTGATTAAAAAACTAAGTGGCAGTGAGCTAAGGT
GGTTGGTTTATAGTGTTTTTTTTTAACTTCTTGTGGGAGTGACTAACAGTACATACACCAAATGATAAAGGGGTCTATATTTCTAGAGGCCCATT
TAGACTACTGGGCTCATACTTACTTATATGTTATGTCACTTCTAGCTATAATATCTATCAAATATATTCAAGCCCAACGTAGTGAGTAGTAATAGT
AGCCCACTACGACCAGGATAGGAGAAGAACGGGCCAGGAAACTTGTTTGGGCCTACTCTGGTTTGCGAAAGCGTTGCAGGCCGAGCTTTCG
GCCCGTGAGCCTGACGTCCTCCCTGATCTGGTTCCGGTACTCCCCGAAGGTGAAGCTCTTGTAGCGAGGAGATCTGCCGGCGCCGCAGCGT
GGCTGGATGAGCGTGTCGTAGCTGGGGCAGAGGAAGAAGGCGACGGAGAACCGCTCCCGCGCGTCGCTGGCCATGACCCGGTGCTCGAC
GCTCCGGTAGCGGTCGTTGCTCCACGCCTGCAGAAGGTCGCCCACGTTGACGATGAGCGCGCCGGGGTTGGGTTCCACGGCGACCCACCG
CCCGGCCTTGACGATCTGGAGGCCGCCGACGCCGTCCTGGTGCAGGATGGTGAGGAAGTCGCTGTCCGTGTGCGGGCACAGCCCGTAGGC
GCCGCTGGGCGCCGTGCACGGCGGGTACCGGTTCAGCCGCAGGAAGCAGGTGCTCCGCGTGCACCGCGACACCACCGCCGTGTCGTCGT
CCTCGCCGTCGCCCACGTCCGCGACGAGGATCCCCGCCAGCCGCTGCGCCAGCTTCGACATCGCCGTGGACACCTGCTCGACCACTAGCC
TGGTCTTGTCATCATCGTCACCGGCGGTGGCTGCCGGCGCCGTCGTATTGGTCGTGGGGATGTGGTAGGCCTCGGACCACGAGAGCTGCTC
CAGGCACGTGGCGGTGGGCGTGCCCCAGCGGTAGCTCTCCGGCGAGAAGTCCAGCAGCGGCTGGCTCGCCTTGAGGTGGAAGGGCCGCC
GGAAGACGGCCACCTGCGCCTGGTGCAGCTCGTCCAGCAGCGCCTGCGGCACGCCGTGGTTGGTCACCTGGAAGAAGCCCCACTCGGAGG
CGGCGCGCACGATGGACGCCCGGCACTGCTCTGTCGCTGCGGCCTCGGCGCCTGATGACTGCTTGTGCTGATGCTGCTGCAGCAGGCTGG
AAAGGTCGATCACCGGGAGGAAGCACTCCTCCTCGTCGGCGTCGTCCAGTCGGTGCGGGTGACTGTCCAAGAGATGCTTGTACGTGGCCAT
CAACGGCGGCTCGTAGTCGTCGTCGTCGTCGGGCGCGGCGCCCTTGCCCCCCTTCTTGCCCTTCTTGCCGCCCTTCTTGGGCTTCTCCTTGT
CGGCGGCGGCGGGCGCAGGCGGGGCTGGGTCTTCCGCTTCGTCCTGCGGGAAGGAGTACTCGTCGTCGTCGATGGCGACGTTCTTCTTGC
GCCCCATCGGCGGCGAGCTCGGAGACCCTAGGGTTTGGGCTGGGGGCGGGAGAGAATGCGGCGAGCGGGCGGGCGGACGGACGGAGAT
CGACGCGGCGGGGGCGAG

Gene: GA2ox2
>comp202287_c0_seq1
TCGTCAACCACGGCGTGCCCGCGGCCACCATGGGCAGGGCCGAGTCCGAGGCCGTCAGGTTCTTCGCGCAGGCGCAGGCCGACAAGGAC
CGCGCGGGGCCGGCGTACCCGTTCGGGTACGGGTACGGCAGCAAGGGGATCGGGCTCAATGGCGACATGGGGTGGCTCGAGTACCTCCTC
CTCGCCGTCGACTCCGCGTCGCTCTCCGACGCCTGCCCCGTCCCCTCGACCGCCGCCTTCCGGAGCGCGCTGAACGAGTACGTCGCGGCC
GTGCGGAAGGTGGCGGTGCGTGTGCTGGAGGCGATGGCGGAGGGCCTGGGCATTGCGGACGCGGACGCGCTGAGCTCGATGGTGGCCGG
CGCCGGGAGCGACCAGGTGTTCCGCGTGAACCACTACCCGCCCTGCCCGGCGCTGCAGGGCCTGGGCTGCAGCGCCACGGGCTTCGGCG
AGCACACCGACCCGCAGCTCATCTCCGTGCTCCGCTCCAACGGCACGTCCGGCCTGCAGATCGCGCTCCGCGACGGCGCGCAGTGGGTGT
CCGTGCCCTCCGACCGCGACGCTTTCTTCGTTAACGTCGGCGACTCGTTGCAGGTGGGTGGGCGTTTTAATTAATTTGGCACGTGACGTGAC
ACCAATTTCTTCTTTCTTATTTTGTGGCCCGCGCTCGAGTGTGTTTTTGTATCCGCCTTTCCTCTACAGTGCTGCAACCAGCAAGCTAGCTGCT
AGATTCTGCATGCTATTAAAAATCTAGTGCGCGCTTTCGTGAACAAGTCAAACTGTGGCCACACACGGAATGCCGGGCGGCCAGTCAACTCAT
CAGTCACCACGCATTAGATAGCATTAACATAATCAAGTGAACCACATCACACGACGGCAAGGACGCGTTAGATTAGATAGCTGCAGAGCAACA
CACCTTGATCTGCAGTGCATGTCTCACAGTGGTGCGGTGCGTGCTGGCTTTGTGTGCAGGTGCTGACCAACGGGCGGTTCAAGAGCGTGAA
GCACCGGGTGGTGACCAACAGCCTCAAGTCTAGGGTTTCCTTCATCTACTTCGCGGGACCGCCGCTGGAGCAGCGGATCGTGCCGCTGCCG
GAGCTGCTGGCGGAGGGCGAGGAGAGCCTGTACAAGGAGTTCACGTGGGGCGAGTACAAGAAGGCCGCGTACAAGACGAGGCTCGGCGAC
AACAGGCTGGCCCAGTTTGAGAAGCATAGTAGTAGCATCTAGCTAGCAGGGCGGCCCGGCCGGTTAAGCACACCAGCTAGCTTTACAGGAC
GACAACAGGCAACCATGACGAACGAACGAAGACGAGCGAGTTACTTGCAAGAAATTAAATGATGGCGAAGAAGAAGGTAAAGACGATGATGA
CATGATCGATGTGTACAGCTGGGGGACGGGAACGGGAACGGCGACTACTCCAGTAGCTAGATTTTTGAGGTGGGTGTTTCCGCAAGTTGCTC
TAATGGCCTGTTTTGGGGGCGTGCGGTGGGGTTACTTGCCTATTTCGGCATGCTGGAGCCGGACGGGCGATAAGCCGCCGCCGCCGCCGG
TAGTTAGCGTTAACCAAAATACTAGCATGGCAGCGAAGGGACGATCCTTCCTCTCTGGCTCGCTAGCTTCCCACGTTAGCGTGCTTCTTGCTT
TGCTTTCCTTTTTTTTTTTGTGCTTTTAAGCATCAATTCGCGTGTATGTACGTATTGTCGTGTAAATGGAGGTTTTGTGTGTGCGCCTATGTGGA
CCTCCTGTTTACCTCTTACTCGATCGGCTGCCAATGTTCCATCCACTGATCCACAGCCATTCCTCTCTCTGTCGAGCGATCGCTTTCGGGCACT
GGATTTCCTAGCGCACTTTTTTGCCGTTAAAATGCACTTGATGGCAAGATCTCTGTATTCTTTTTTTTTTGTGAAGTATTCAATGGACGCCTTTAT
TGAAAGGGCCGGCCGGCTCCCTTCTTTTTCACTGGGTTGTTTGGTTCCCT

Gene: TEM2, RAV2
>gg_06658
GCCACCACCGGCAAGGGCGTGCTCCTCAATTTCGAGGACGGCGAGGGCAAGGTGTGGCGGTTCCGGTACTCGTACTGGAACAGCAGCCAG
AGCTACGTGCTCACCAAGGGCTGGAGCCGCTTCGTCAGGGAGAAGGGCCTCCGAGCCGGGGACACCATCGTCTTCTCCCACTCCACGTACG
CCTCGGAGAAGCAGCTCTTCATCGACTGCAAGAAGACGACCAAGACGACGACGGCCGCCACCACCGACGGGGCGCCGGCGCCGGTGCCGG
CGCCAGTGGAGAAGAAACCAAGTACTGAAGCCCGTGTAGTGAGGCTGTTCGGCGTCGACATCGCCGGAGACGGGTGCCAGAAGCGCGCAC
GGCCGGTGGAGATTGCGTTCGAGCATGGGCCGCAGGAGTTGTTGATGAAGAAGCAATGCGTGGCTCATCACCGCTCTCCTGCCCTAGGTGC
CTTCTTGTTATAGCAGCGATCGTACGTATTATACACATAGTAGTATTCTAGTCCTCCCCCTTTTTGTTCTTCCTGTCGCCTGCTTCATCTCGAAT
TCAATTACCACTAGCTATAGTGTTTGTTCTTCGTCTTGCAAAGAATAGACTGCAAAACGAGCTGTATGTATGTACACCAAGTCTTCAGATGCTAG
TAGAAGTAGCAGTCCTGTAAGATCATGCAAGTTTGTGCCATCATCGAGTGCAGCAATAAGTTGATTCGATCTGCTGTAGACTAAACATGTGTGC
TTGCTTCTTGTTCTGGTGTTGTTGGATATTCATCACTCTTGGCAACTCAGTCTAGCTTTGATTAGCTTTGCTGTAGCACGTCGACATGCATGAAT
AAGTCGATAGATGTTGCG

Gene: UBP12
>gg_14818
GGCTTTGCAAAGTCTTTTTTTATAAACTTCAGCATTGTGATAACAGTGTTGCTACAAAAGAGCTCACCAAATCTTTTGGATGGGATAGCTATGAT
TCATTCATGCAGCATGATGTTCAAGAGTTGAATAGGGTCCTATGTGAAAAGTTGGAGAACAAGATGAAGGGAACCACCGTGGAAGGAACAATA
CAAAAATTATTCGAGGGTCACCATATGAATTATATCGAGTGCATTGATGTTGACTCAAAATCTACCAGGAAAGAGTCATTCTATGATCTTGCACT
CGATGTCAAGGGATGTTCTGATGTCTATGCATCATTTGATAAGTATGTTGCAGTGGAGAGGTTAGAAGGCGATAATAAGTATCAATCTGAGGAA
CATGGTCTACAGGATGCCAAGAAAGGAATGCTTTTTATTGACTTCCCTCCAGTTTTGCAACTTCAGTTAAAACGGTTTGAATATGATTTTGTGCG
GGATACAATGCTCAAGATAAATGACCGTTATGAGTTCCCACTTCAATTGGATCTTGATAGAGATGATGGAAAATATCTTTCTCCAGACGCGGAT
AGAAGTGTGCGTAACCTATATACTCTTCACAGCGTGCTGGTTCATAGCGGCGGAGTTCATGGAGGACACTATTATGCCTTTATTCGCCCGAAG
CTGTCTGATCAATGGTACAAGTTCGAGGATGAAAGAGTGACGAAAGAAGACATGAAACGGGCATTGGAGGAGCAATATGGTGGTGAAGAAGA
GGTTAGGTTGAGAAAGGAACATGAAGAGAAGGAAAACAAGAAAAAGGAGAAGGCTGAGGCTCATATGTTCACTGCATTGAAGGTGGCCCGAG
ACTCTGATTTCAAAGAGCAAATTGGAAGACATGTGCATTTTGATCTTGTGGACTTTGACAAAGTTAATAGCTTCCGTGCACCTAAGAACATGTCA
ATCAATGATGTCAAGGTGGAG

Gene: UBP12
>comp199615_c0_seq1
GTCGCTCACTCAGCTTTTATAGCGGAAAAACCCCTTCTTTTTTTCTTTCGTCCTTGTCTCTTCCCTTCTCCTTTTCGCCTTCCTCCATCCTAAACC
CTAGCCACACTGTCCTAAATACGCTCTTCTCCGCTGCGATCGCTGCAGCGATCCGATCTTTCGCCGGTCACCGGCGCCGGAGACGCCGCGC
TCTATAATCATGGTTGTCTCGGCCGCCCCCGAGCTGCCACAGCAGGAGCAGCCACAGCAGGACCAGGATGAGGAGATGCTGGTCTCGCATC
AGGATGGCATCGAGGGGCCGCAGCAGGATGTTGTTGAGGGTCCGCAGCCGATGGAAGAATCTGCATCCGCTGTTGAGAACCAACTTGTGCC
AGATACCTCCACATCCAGATTTACATGGTGTATTGAAAATTTCTCCAAGCGGAATGTTAGAAAGCACTACTCTGATGATTTTATTGTTGGGGGAT
ACAAATGGCGAGTGCTCGTCTTCCCCAGAGGAAATAATGTGGAATACCTTTCAATGTACCTGGATGTTGCTGATTCAAACTTGCTGCCTCCTG
GTTGGAGTAGGAATGCACAATTCAGCCTTGCTGTGGTAAACCAATTAGACAGCAAAGCATCATTAAGAAAAGAAGCGCAACACCTATTCAATTC
CCGAGAAAGTGATTGGGGTTTTTCATCTTTTATGCCTTTGTTGGATCTGTATGATTCAAGTAAAGGATATGTTGTGAATGATAAATGTATCATAG
AAGCTGAGGTTGCCGTGCGTAAAACTTTTGATTTCTGGAACTATGACTCCAAAAGGATGACTGGTTATGTTGGCCTGAAGAATCAAGGGGCTA
CCTGTTACATGAACTCTCTTCTTCAGACTTTATACCACATTCCCTACTTTCGGAAGGCTGTATATCATATGCCTACTACAGAGAATGATACACCT
TCAGGGAACATTCCATTGGCTTTGCAAAGTCTTTTTTTATAAACTTCAGCATTGTGATAACAGTGTTGCTACAAAAGAGCTCACCAAATCTTTTG
GATGGGATAGCTATGATTCATTCATGCAGCATGATGTTCAAGAGTTGAATAGGGTCCTATGTGAAAAGTTGGAGAACAAGATGAAGGGAACCA
CCGTGGAAGGAACAATACAAAAATTATTCGAGGGTCACCATATGAATTATATCGAGTGCATTGATGTTGACTCAAAATCTACCAGGAAAGAGTC
ATTCTATGATCTTGCACTCGATGTCAAGGGATGTTCTGATGTCTATGCATCATTTGATAAGTATGTTGCAGTGGAGAGGTTAGAAGGCGATAAT
AAGTATCAATCTGAGGAACATGGTCTACAGGATGCCAAGAAAGGAATGCTTTTTATTGACTTCCCTCCAGTTTTGCAACTTCAGTTAAAACGGT
TTGAATATGATTTTGTGCGGGATACAATGCTCAAGATAAATGACCGTTATGAGTTTCCACTTCAATTGGATCTTGATAGA

Gene: SUS4
>comp198370_c2_seq1
GGACAACGCGGTACAGACCAGGCATTGTGAATGCCATGTGTGACTCGTACTGACCGACGGTGTCCTTGTTTCCAGCAATCTCTTGGAAGGTA
CTTGTGATAATGAAGTCAGCATGGTTCATAGCAATCAAGTCAGTGGTGAACTGGCACGAGAAGTGGTAGTGGTCCTCAAACTTCTTCCAGTAG
AGGTCAGAGTTAGGGTACTTAGTTTTCTCAAGGGCGTGGGCAATGGTACAGTGAGTAACACCCATCTTGTGCGCAAGCAAACACGCAACAAG
GTTTCCGTCACTGTAGTTTCCGATGATCAGATCAGGATTGGCCTGAAGCTCTCCAGCAATCTCATGTGCCACATCATCAGTGTAAGTCTCTAGG
TACGGCCAGACTTCAAAACGTGAGATCCACTTGCGAACAATTCCATTTTCTGTTCTGAATGGCACACGAAGGATATGGCAGTGCTCAGTGCCA
AGGACCTTCTCAAGCCGCTGACCACAGGTGGTGCCAGTTGCATCAGGGAGCAACCTGGTAACAATAAGGATCTTTGGTGTGATGTCAAGACC
ACACTGCTTGATCCTCAGCAGCATTTCATTCTCCATAGCACGGACTTGGTCCAAAATGTAGACAACCTGGCCTCCAGTGTCAGGGTAACCCAA
GACATTAGCTTGAGCAAAGTAACCATGAGGGGAGAGGATAACAACATTGAACACCATGGGGATCGTTCCAAGGAACTTCTCCAGGGTGGACG
GATCTGGGGCCTCAAGAAGGTCCAAGAGGAGGTGAATAGTCTCCTGCGCGCGCTTAGCGCAGTCACCCCAACCCTTCTCCAGACCAAGTTCC
TGGAACCTGTGGTGAAATTCAGAGTATGGGGTATCAGCTTGGAGGGTGGACAGGTGCTCCTCAGCCTTCCTCAGAGCGCCTTGCAGAGCACT
GAGACTGCGGATTCTGTCATTCAACATCATGGTCATCCCCTTGTAGTTGTGGGCGCGAAGGAAGTTGAGCAAGGGGTACATGCTCTCCTTGTC
ATGGAAGAGCTTTGATGACAGGTGCCTGTTGAGGAACTGCACACCATTGCCAATGGACTTTGACAGAGAAGGACGGGGGAAGGATGCATTGA
ATGGCTCAAAGTCCAGCTCAAGAACAAAGTTGTTGTTGGGGCCTTCTTCCACAAGCTGTTCCTTGAACTGCAGGTACTCTGGGACTCTCAGCT
CCTCAACAGCGAGCTCGCTGACGTTGACCCTCACATACTCCCAGACACCAGGCCTAGGGCGGATGGCAAGTGCAACCCATGGGGGGATAAC
AATTGCTTCCTGAGCTGCCCTCAGGACATCCTCAAAGGCGCCATCCTTCAACTTCTCACGCTCAGCCTCAGGGATTGCACTGTTGTACTCGGC
AATGATCTGGTGGGGCTGCAGCATACCCTTTCCAAGGTTTTTCAGCCTGGTGAAGACGGCGACAAGCTCATTGGGGTGGGCAGAGAGTGAAT
CGCCAATGCGCTCCCTGACGCTGTGGAGGCGGCTCAGGACACGGTCGCCGGCAGCTTCCCCCATTGCTATCCTCTTGTTGGGTCCTCCTCA
AGCCTGCACAGTGAAACAGAAACGGAAGCTTCAATCAGCATTGAAAACTGAACTGAAAAAGGGTTCAGATCCGTTTACATACAAGCTGCAAAT
GTCATCCTAAAACCCTCCCAGCATGACTCATACATAATCCCAACTGCCAACAACCACACCTTTGACCCTTCACCAAGAAACTATTGCAGGCAAG
AAATCAACAGAACCTACTTCTGGAAGGGGAGAAACAGCAAACAACAGGTGCAAGACCAAAAATCGACTCCTCAAATCCTGTCCTAGTCCTAGA
CTAGAAACAAGCAACAATCAATACAACAAGAGGCATACAAGAATCGATGTAATGATGAAACGCTTCAGAAGACCGAGGAATGGCGGAGGGTG
GGAGGCGCCAATTTATACAGATCTGACGAGAGAAGCCACCAAAAACCACATCGATGGATCCCATGAAGACAAGGCTGGGCCGTATGGCACCA
ATGGCCTCCGGCTCCAAAAACCGTTGAAACCAAAGCAGGCGAGGACGAAGAGTGACGCGGCAGGCTACTTCTAGTTCTAAGCACGAGATCAG
CATGGGGACTTGGGGTTCCTTGCAGGCAATCGCAAGGTGACTAGGATACACAGAAGTAATTACTCCTAGATGGTAATAAGACACAGGATTAGT
CTAAATTAACCATGGCTATTAGCACAAGAACCCAGAACCGTCAGCCTTGTCTCCGAGTCCCGAGACAGGGGGCACATGGCGGGGCAGAGGA
ATGGATAAAGCCAACCACTTGTCCGAGAAAAATCATCACCAAAATATGTATTCCAAAAATAAAAAATGGTTGAGGAACAACAAGCACCGGATCT
AGCGTTCTACAAACTTGGGTATGGTCTGGAAAGAAAAAAAACATATGAAACCTTAAACCCAGAGAGCAAGAGAATGATGGTCAGTAAACCCCA
GATCCATCATCCATCCAGAAGCAACACCATGTTTCCCTGCGCGGATAAGCAGGGTCCCTCCGCTCCACTAGATCTCCTCAAGTGGCAAGTTGA
CCCAGGTTGTTATTACTAAAGTTAAAACCCCAATAATTCTCTGGATCTCACGCACTGGTGCAAGAGCCATAAA

Gene: ADG1, APS1
>comp172176_c0_seq1
ATATCTTCCCAGTAACCATCATATAAGTAAGCTTGCACCCTTAATCCAATTTCTGTTGCGCCAGGAATAACCTCACTTCCGAAGTCATTTGCTGC
AGGAAAGTTTTCTCTGAGAAGCCGAAGCATCACATCTTTGCTAAAAACATAGATTCCCATACTAGCAATATAAGGCAGTTCCTTGGCCCTCTCA
GGATCAAGGCCCAATATAGTGGTGTCAACCATCATTGATCTCAGCTTCTCTCCTTTTGGTTTTTCTGCAAACTCAACTATTCTCCCTTCATCATC
AATTTTCATAAGACCAAATGCAGTTGCACGTTGTTCATCCATTGGCAGGGCTGCAACAGTTATATCAGCATCTGTTTCTCTATGGGCTTGAATG
AACTTTTGGTAGTCCATACGGTACAGGTGATCTCCAGCCAGAATAAGAAACTCCATGACATTGTGCTCCTCAAATAGCCACATATATTGACGAA
CGGCATCTGCAGTACCCTGAAACCA

Gene: PUB13
>comp197651_c0_seq1
CTAGGCACTAGCTGGGGCCGCGCGGGAGGACGCGGTCCGGTTCACGCCTCGCCTCCCCTCTCGCCATGGCCGGCGACCGCGGCCGCGGG
GCCGCGGCGGAGAGGGTGGCCGCCGCCGTGGAGGCGGTGGGATCCGGTGGCTGGGAGTTCCGGAACGCGTACAGGCGCCAGCTGCTGG
CGCTGTCGCGGCGGATCCGGCTCCTCGGCCCCTTTGCCGAGGAGCTCCGGGAGGCGCGCCGGGGGCCCGCGGAGGAGGAGGAGCAGGA
GCGGGCGCTGGTGCCGCTCGCCGACGCGCTGGAGGAGGCGCTTGACCTGCTCCAGCTCGGCCGCGACGGCAGCAGGATCTTCCTGGTCCT
TGATAGGGATAAAGTAATGAAGAAGTTTCATGAATCAATTGCTCAGCTGGAGCAAGCCTTGTGTGATTTTCCATACGATAAATTGGATATATCTG
ATGAAGTTAGAGAGCAGGTTGAGTTAGTGCATGTACAGCTCAAAAGAGCCAAAGAACGGGTTGACATGCCTGATGATGAGTTCTACAATGAAT
TACTTTCGCTGTATAACAAGAGCCATGACCCAAGTGCTGAATTGGATATCCTTGGAAGATTGTCAAAAAAGTTACACCTGATGACCATCACTGA
TCTCACACAAGAATCCCTTGCTTTGCATGAGATGGTGGCATCTGGTGGAGGTAACGATCCTGGAGAACACGTTGAAAAGATGTCAATGCTATT
GAAGAAGATCAAGGACTTTGTGCAAACTCAGAACCCTGAAATGGGCCCGCCGGTGACTACCAAGCTAATGGATTCCAACGGGCAGCCAAGAC
CTGTAAACATTCCTGATGAGTTCCGCTGTCCAATTTCTCTTGAGCTGATGAAAGACCCCGTTATTGTTGCTACTGGCCAGGTGCCTACATCTGA
TTCTACTTTTTCATACTCAAGTACAAGGCAACTTCAATACATCATATGAGATACTTGTTTGTTTGTGTTTTTTCAGACATATGAGCGGACATTCAT
CGAGAAATGGCTAGCTTCTGGGCATCATACTTGTCCAAATACTCAACAGAGGATGCCAAACACAACACTGACACCAAACTATGTTCTTAGAAGT
CTTATTGCGCAGTGGTGCGAGGCCAACGGGATTGATCCACCGAAGCGCTCAACGCAGCCTGACAAACCCACATCTTCATGCTCTCCGAGTGA
ACGTGCTATCATCGATGCTCTTCTCTCCAAGTTATGCTCTGCAGATCCTGAGGAACAGAGATCATCTGCTGCAGAGCTCCGGCTCCTAGCAAA
GCGGAATGCAAATAATCGGATCTGCATTGCTGAGGCTGGTGCCATCCCCTTACTATTGAGTCTCCTGTCGTCATCTGATCTGCAGACTCAGGA
ACATGCTGTGACTGCACTTCTGAACCTCTCTATACATGAGGATAACAAGTCTAGCATCATATTATCTGGTGCTGTCCCGGGCATAGTTCATGTT
CTGAAGAATGGAAGCATGCAGGCACGGGAGAATGCTGCAGCCACCCTTTTCAGCCTGTCGGTGGTTGATGAATACAAAGTAACAATTGGGGG
AACAGGAGCTATCCCTGCCCTTGTAGTTTTGCTAAGTGAGGGCAGCCCACGTGGGATGAAGGATGCGGCAGCTGCTCTATTCAACTTGTGTAT
TTACCAAGGGAACAAGGGCCGTGCTATACGAGCTGGCCTTGTGCCACTCATCATGGGGCTCGTGACAAACCCCACTGGTGCTCTGTTGGATG
AGGCTATGGCGATACTCTCAATACTGTCCAGCCATCCTGAGGGTAAGGCTGCGATCGGGGCAGCAGAGCCTGTTCCCGTGCTTGTGGAGATG
ATTGGAAGTGGGTCACCGAGGAATAGGGAGAATGCTGCAGCTGTGATGTTGCACCTGTCTGTGCATAACGTGCATTTAGCTCGGGCACAGGA
GTGCGGCATCATGGTTCCGCTACGGGAACTGGCCCTGAATGGCACAGAGAGGGGGAAAAGAAAGGCGGTGCAGCTCCTTGAGCGAATGAGC
AGATTCCTGGTTATGCAACATGAGGATATTGCTTCCCAACAGTTGGACGAGATCGAGCGTCAGCTTGTTGCCTTAGGACCTAATGTTACAGAG
GCTGATCTTGACCAACTGGGACTTCTTTAACCTCCATATCCTACGGTTGCATGATGATGATGAATCTATCTTTGCCACAATACTACTAAAAGGGT
TTGTTTTTGGATAAGTACTACCAAAAGGATGAGCAAGCTGTATCCGAGTCACTTTGCTTCCGGATCCATGATTTATGCTGAACGAGAGTTCTTG
TATGGCAACGTTCTATCGGTACTGAAATCTGTGATGAATCCGTCTGGCTTGGCTTCTTCCATGCTTGTGTGCTGAAGCCCCCATCGCTCCATCT
GTTTCTGGGTATGTAAGGTAGCCCAGAATGTTGAGATGTGTACAGTCTACTTAGTATGCCGCTCCAAATGTCATGTATGGCTATGTTTTGACTA
GAATATGTACATTTTCACTATGGTACTAGATAAATGAGAAGTTCTAGCGGCGGACTTAAATCAGTAAATCTTATTTCATGCTACTGGATGTTGGC
GTGAGTTTCAGATTTCAGAATTTTTATTATGGTGACGCGTATGCATTTGTGAATCCAACATTGCCACTACGTATTGTAGTTTACGGCAGTGCTAG
TTGCAACCAGAATGCCATTTCATGCCGTCGAAATGCAG

Gene: LRB1
>gg_02988
CCTCCTACCCTTCCCTTCTGTCTCCTCCTTCCCCTCCCGGCCACACGCCCTCCCCCCTCGAGCCCTCCGCCGGCTAGGGTTACGGCCACCGC
CTTCCACGCCTCCCCCACTTGCCGTAGGCGGAGGAGGAGGCTGGTCGGTGTGCGCGCGACCGCGGCGGCGGAGATGGACCCGGACTTCTC
GCCCGGGGGCGGGGGGCCCAGCTTTGAGTTCGCCTTCAACGAGGTCAACTTCTCCGACCGGGAATTGCGGATCGAGGTCGTCGCCGGGGA
CGACGACGCGCCGGGCTCCAGCGGCGGCGGTCCCGGAGGAGGTGGCCTCGCCGACTGGGCGCGCCACCGCAAGCGCCGCCGCGAGGAG
CTCCTCAAGGAGAAAGGTGATGATGGGCAAGGTATTGATTCATCTTGGACTGTGGTGGGTACGCCAGTTTTACGAGTCAAGACAATTTATATC
AGTTCGGCGATTCTTGCTGCGAAGAGTCCTTTCTTTTTCAAGCTTTTCTCAAACGGCATGAAAGAATCGGATCAGAGACATGCAACCCTTAGAA
TTACTGATTCAGAGGAAACTGCCCTCATGGAGCTTTTAAGCTTTATGTATAGTGGAAAGTTGACAACTACTGAGCCCACTCTTCTGCTGGATAT
CTTGATGGCTGCCGACAAATTTGAGGTTGTTTCGTGCATGAGGTACTGCAGCCAGTTGCTCACAAGCTTGCCCATGACCACAGAATCTGCACT
TCTCTACCTAGATCTCCCCTGCTCAATTTCAATGGCAGCTGCAGTTCAACCTCTGACAGATGCAGCTAAGGATTTCCTTGCTGTTAAATACAAG
GATTTGACTAAGTTCCAAGATGAAGTGATGAACATCCCTCTTGCTGGAATTGAAGCCATCTTGTCAAGCAATGACCTCCAGGTGGCATCTGAA
GATACCATCTATGACTTCTTGCTCAGATGGGCCCGTGCACAATACCCGAAATCAGAAGAGAGACGTGAGATCCTTAGTTCTCGGCTGCTCCCA
CTTGTGCGATTCAGTCACATGACATGTAGGAAGCTGCGGAAGGTCCTAACGTGCACGGATATAGACCACGAGCAAGCAACCAAGTGTGTCAC
CGAGGCTCTTCTGTACAAAGCTGATGCACCACACCGGCAACGAGCTCTTGCAGCGGACACAATAACCTGTCGGAAATTCGCTGAGCGGGCTT
ACAAGTACAGACCACTCAAAGTCGTTGAGTTTGATCG

Gene: GI
>gg_12613
AGCTGAATGGGGGACCATGCCTATCATAGCACAGCGTCCCATCTATGATGCGTGAAAGAATTGGATGTACAACTGCATGACCATGCTCCGGG
TGATGAAGAACAAAAGTTGCCAACACTTCATCTACCAAGCGACTTTCTTTTGATGGATAGGAACTTTGAATGAGCTGAGCTATATCTTCAGGGA
ATTGCTCGCTGTCAGCTGTAAACTGGCCAAAGTACTCGACATAGGCCAAAATTTGTGCCTTCTTCTGCTCCACATCCTGGGGTGGAGGCCAGT
ACAATGATGAGAATTGGAGCCCATCGATCCACTTCACATTTGAATCTGACATGCTTGGTAGCTCAAGAGCCTGTAATCAACACCGTGTTGAAGT
TGCCAGTCAGTTCATCCATCTATGCCACCAAATTCCTACTCTTAGTAATGCTACTTGCTTCGAAACCAAGAAACTTTCAATAAGACACAAGTTCT
CCTAAGTGTACTGTTAACTTCTTCAATCAACCCCCACAGAGGGGGGGTCTGCATCCCCCAACGTGAACAAGTATCATATCAACCAATCCCAAA
CATAGTTGCTTGAAAACCTCTCGACGGGGCGCGCTCTAATCCGAGGCTCTCCGCGGGCAGCCGGCGGCGTCCGTGCGCAGATTTCGCGGC
GGGGCGGCGGGTGAGGCGGAGGGGAGGAGAGAGAAATGGCGAGGAGCGAGGAGAGGCGAGAAAAAGG

Gene: ADG1, APS1
>gg_15551
AGACGCGAAAGCAAAACGTCAAACAGGCTCGTCTGAAACTCCCAGAGCCGGTAGAGGACACGCCACTCCACAGTGTCAGTCACTGCCCCCAT
CACGCCACCCTCCCTCCCTTCTAGGCCGTTGCCGCGAGCTCGCTCCAACCCACCGCCGATGGCGACGGCTATGGGCGCGATGGCCGCCAC
CTCCCTGGCCCCGATCCTGGCCGCTGCTACGTTCCCCGGTGATCTCGGCCTCGGACGCCGCCGGGCGGCTGTGTCAGGGTGGCGCGCCGG
CGGGAGACGGCTGCGTGCGTCGCCGCCTGCCCAGAGGCCGTTCCTGTTCTCGCCGAGGGGCGTTTCGGACTCGCGGAGCTCGCAAACGTG
CCTTGATCCGGACGCCAGCACGAGTGTTCTTGGGATCATCCTTGGAGGTGGTGCTGGGACAAGGCTGTATCCACTGACGAAGAAGAGGGCG
AAACCAGCAGTGCCGTTGGGCGCCAACTACAGGCTCATAGATATCCCTGTCAGCAACTGTCTGAACAGTAACGTCTCCAAGATATATGTGCTA
ACACAGTTCAACTCTGCTTCGCTCAACCGCCACCTCTCCAGGGCCTATGGGAACAACATTGCCGGGTACAAGAATGAGGGATTCGTTGAGGT
CCTTGCAGCACAACAGAGTCCAGAGAATCCCAACTGGTTTCAGGGTACTGCAGATGCTGTGCGTCAATATATGTGGCTATTTGAGGAGCACAA
TGTCATGGAGTTCCTTATTCTTGCTGGAGATCACCTGTACCGTATGGACTACCAAAAGTTCATTCAAGCCCATAGAGAAACAGATGCTGATATA
ACTGTTGCAGCCCTGCCAATGGATGAACAACGTGCAACTGCATTTGGTCTTATGAAAATTGATGATGAAGGGAGAATAGTTGAGTTTGCAGAA
AAACCAAAAGGAGAGAAGCTGAGATCAATGATGGTTGACACCACTATTTTGGGCCTTGATCCTGAGAGGGCCAAGGAACTGCCTTATATTGCT
AGTATGGGAATCTATGTTTTTAGCAAAGATGTGATGCTTCGGCTTCTCAGAGAAAACTTTCCTGCAGCAAATGACTTCGGAAGTGAGGTTATTC
CTGGCGCAACAGAAATTGGATTAAGGGTGCAAGCTTACTTATATGATGGTTACTGGGAAGATATTGGTACTATTGAAGCATTTTATAATGCAAA
CTTGGGAATAACCAAGAAGCCTGTTCCGGATTTTAGCTTCTATGACCGTTCTGCTCCAATTTATACGCAACCTAGATACTTGCCTCCTTCAAAG
GTTCTTGATGCCGATGTGACAGACAGTGTTATTGGTGAAGGTTGTGTTATTAAACATTGCACAATCAACCATTCCGTAGTTGGACTCCGTTCCT
GCATTTCTGAAGGTGCAGTTATAGAGGACTCTTTGCTAATGGGTGCAGACTATTACGAGACTGAGGATGATAAGAAAGTCCTTTCTGAGACTG
GTGGCATTCCCATTGGTATTGGGAAGAATGCACATATCAGAAAAGCAATAATCGACAAAAATGCTCGTATTGGAGAAAATGTGAAGATAATCAA
TTTTGATAATGTCCAAGAAGCAGTAAGGGAGACGGAAGGATACTTTATCAAAAGTGGCATTGTCACAGTGATTAAAGATGCCTTAATCCCTAGT
GGAACAATCATATAAAGTCAGCCAACACCATGTATATAAGGTTGGTTTGCAGCAGGTGGTGATCACCAATGCACATCAAGTTCTGTCAGCCCTT
TCCAGACCATGCTTATTCCCATGTTGCAAAAATCACCAAGCTCTGCATCAACTTGCAGAGGATAAAAATCAAATAATAAAATGGTGCCATTTCGA
AGGCAACATTTTTGTCACCAGTATGAGGCACTGTTCTAAATGTACTGTTGAGTGGAACTTGGATAATGTGTGCTTTTGTAATCGCACACCTAAA
ACACCCTGTGCTTATACTCTTTCTTACCTTAATATGCTCCCTGCATTGCCCAGGGAATGAGGATTAAAAATGCCATACATAACATGATTTTTGTT
AGACTGCTGTAGAGCTTGCATAAGTTTCTTCTGGCCTTATAGAGATGTTCAGAGTGAAACTATCATCTTGGATGTGCGAAGAGTTGCCTGTCTT
TTCTGAGCTTGTAATGACACAAGTTTCTGTGCCTTTGTGCATGATGTTGTTGAACCTATATTAAGTCTATATAATATGGTGTATGCTCTTATATGA
TTTGGCAAAAAT

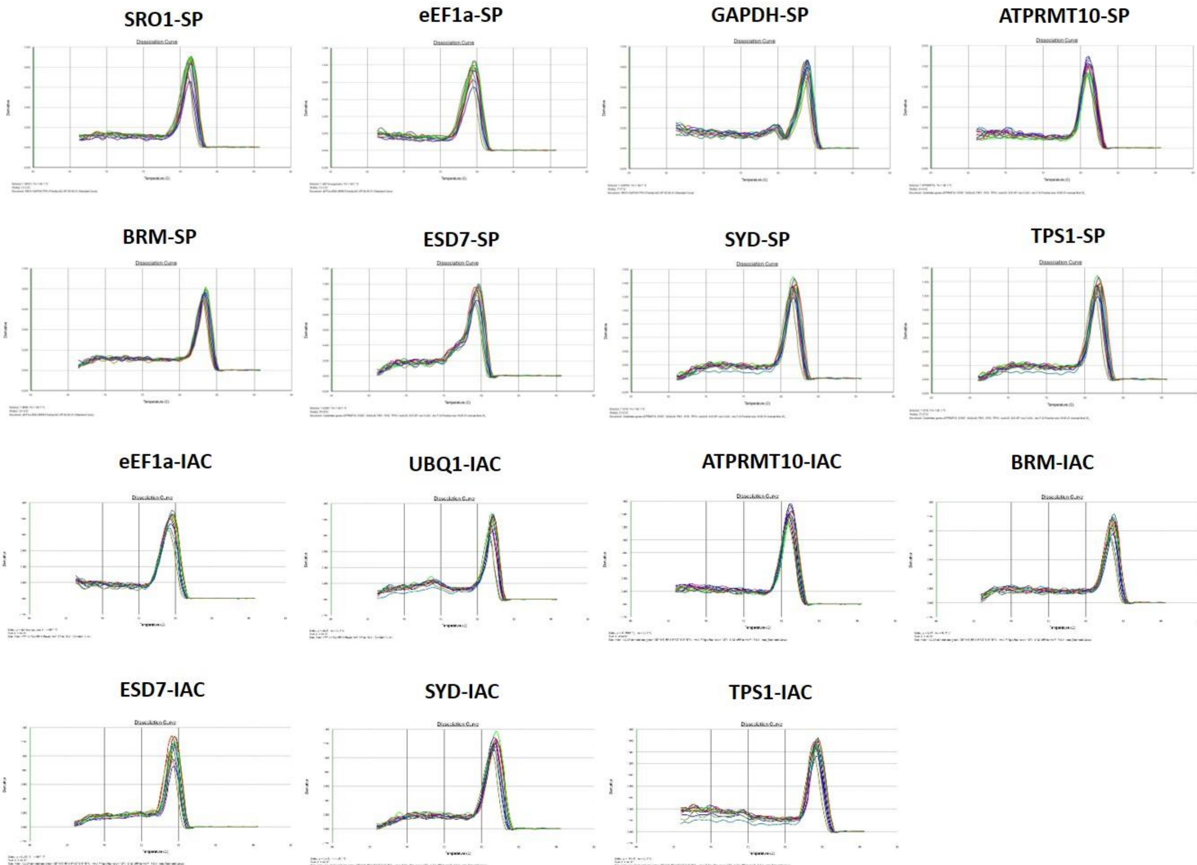
Gene: BAF60,CHC1
>comp192150_c0_seq1
GACCAGGTCATCATAGAACTACTATCTGAAAACAAAAGGTTCTTAAAGAAAAGAGTAATATGGCTTATAAGAAGATCCTGGAATAGATAGAAAAT
TTCCTTGACAAAAGGCAACCACACCTCAAAGCTTCTTCCAGCTCAAGGAATCATAGAAAAGCTAGACTGTCATACTCCTGCATTCCAGTTTTTTT
ATAATGCAATTTAGGCTTACTATTTTTAGGGCCACACCAAGCAACATTATGGTTCAGTTCAGCACTTGGCATTATACTTCCACAATAGTTAGACA
ATATAGAACACAACATACAAGTATACAACGACTAGCCTGAACATCCTTTCACAACACTGAACAAATCTCTCGCTAATACGTCTGAATTTTACTAG
CAATACAGAACAGATAATTTGTTTCTACACCAAAACAGCACTAGGAGAAATCCACGCCTCAAGAACCACCAGCACCACCACCTGGGCCCTCAT
TAGCCGGTTTGCGGTTCAAGTATCTTATAACGGCGTCTTCAACCCATGGTTGGTTATAGAAGTCAGCACGTCTTTCCTTCTCAATGTTCCTATTT
GCCTCGCCGGCGACCAGCTTTAAATCTTTGCTCTGGGAAGCTATCAGTGCATTGATGAACTCCACTGGGGACTGGCTGAAACCCAGGAAGAA
TGCCCTCCTCCTGCGGTGCTCATGGATCTTCTTGATCGAAGCAGATATCACCTCGTCGCAGGCCTCAATGTCCTTGTGCTTCTCTGTGTTGGC
AAGGAACGCCATCATCTCTTTCTGCAGCGGGAAGGGAACATCCACCAGCACATCATAGCAAGCACTGCTATGCGCTCCATTCCCTGACAGCTT
AATCTTGTGCTCCAAGTTGATGGGTGGTGGGGCAGACAGATGCTGAGATATCTTCTGTGACAGCATTGCAAACTTGAGCTTGTCCTCCCCAAA
CACCTTCTTCAATTGAGGGTCACACATGAAGAAGGATGGATCATTTGGGTTCTGAAGCTTCTTTGCCTTAATATACTGCCAGAGGGCAGCAATC
ACTCGTGCACGAGTGTCCACCTCCACCCCAAGCACCTCCATAAGCGGCTGTGACAGCTTGAACTTCTCAGGGTTGTAGTTCATCTCCAGCCG
GATGTTTGCAACAAACTCCTTATCCCCTTTCCTCTTTACCTCAAACCCTTCCTGCTGGGCAGCTGTTCGTGCATTCTCCCAGATGATGAGCGGA
TTCTCTGGGTACAGAGACGGGTCCAGCGCTATCGTCACCCTCTTGAAGAATTGGGAGAACTTGGGGTACACTGGGTTGTGCTTGGGCACAAC
ACTGGCAGGATCCAGCTCGGCGCCATCCTCCAGCACGCGACCAATGATCTTGAGCGACCAGGTTGGCGGATCGGCATTCTTGGGTGGTGGG
ATGGTGCGCGGCCCCTGGTTGGCAAAGGTGTTGAAGACGTAGATGCGGAGTGTGCGCTGGAGCGAGGGCGGCGTCTTGAGCGCCTCCTGG
ATGTCTACCTTCTTCCGCGCGAGCGCGGCGTCCACGCGGGCTTCGAACTCGAGCAGCTGCGTGTAGAGCGCGGATTCTGGGAGCAGCGCG
GCGACGCGGTCAGGGAGCTGCTTCTCGGGGAGCTTGCGCTTCTTGTTGCGCCGCGAGGCGGGGGTGAGGTCGACGCCCCCGGAGTTGGC
CACTGCTGCCGCCGCGGC

Gene: SUS4
>gg_15361
ACCCTCGCCCTGCAATCCGTGTATAAAAAGGGGCCCGGGGGCTGCCGCTCAACTCACCAATCCATTCCACCTCCGTTCACCCCGTCCATTTG
ATTTGCGTTCGTTCACTGCGTTGCGTTGCGTTTCTTGGGGGGTTGATCTTCTCCTCTCCTCCTTGGATTGGAGTGGAGGTTCCTTCTTCTCCCT
CTCGGAGGAAGGCTTGAGGAGGACCCAACAAGAGGATAGCAATGGGGGAAGCTGCCGGCGACCGTGTCCTGAGCCGCCTCCACAGCGTCA
GGGAGCGCATTGGCGATTCACTCTCTGCCCACCCCAATGAGCTTGTCGCCGTCTTCACCAGGCTGAAAAACCTTGGAAAGGGTATGCTGCAG
CCCCACCAGATCATTGCCGAGTACAACAGTGCAATCCCTGAGGCTGAGCGTGAGAAGTTGAAGGATGGCGCCTTTGAGGATGTCCTGAGGG
CAGCTCAGGAAGCAATTGTTATCCCCCCATGGGTTGCACTTGCCATCCGCCCTAGGCCTGGTGTCTGGGAGTATGTGAGGGTCAACGTCAGC
GAGCTCGCTGTTGAGGAGCTGAGAGTCCCAGAGTACCTGCAGTTCAAGGAACAGCTTGTGGAAGAAGGCCCCAACAACAACTTTGTTCTTGA
GCTGGACTTTGAGCCATTCAATGCATCCTTCCCCCGTCCTTCTCTGTCAAAGTCCATTGGCAATGGTGTGCAGTTCCTCAACAGGCACCTGTC
ATCAAAGCTCTTCCATGACAAGGAGAGCATGTACCCCTTGCTCAACTTCCTTCGCGCCCACAACTACAAGGGGATGACCATGATGTTGAATGA
CAGAATCCGCAGTCTCAGTGCTCTGCAAGGCGCTCTGAGGAAGGCTGAGGAGCACCTGTCCACCCTCCAAGCTGATACCCCATACTCTGAAT
TTCACCACAGGTTCCAGGAACTTGGTCTGGAGAAGGGTTGGGGTGACTGCGCTAAGCGCGCGCAGGAGACTATTCACCTCCTCTTGGACCTT
CTTGAGGCCCCAGATCCGTCCACCCTGGAGAAGTTCCTTGGAACGATCCCCATGGTGTTCAATGTTGTTATCCTCTCCCCTCATGGTTACTTT
GCTCAAGCTAATGTCTTGGGTTACCCTGACACTGGAGGCCAGGTTGTCTACATTTTGGACCAAGTCCGTGCTATGGAGAATGAAATGCTGCTG
AGGATCAAGCAGTGTGGTCTTGACATCACACCAAAGATCCTTATTGTTACCAGGTTGCTCCCTGATGCAACTGGCACCACCTGTGGTCAGCGG
CTTGAGAAGGTCCTTGGCACTGAGCACTGCCATATCCTTCGTGTGCCATTCAGAACAGAAAATGGAATTGTTCGCAAGTGGATCTCACGTTTT
GAAGTCTGGCCGTACCTAGAGACTTACACTGATGATGTGGCACATGAGATTGCTGGAGAGCTTCAGGCCAATCCTGATCTGATCATCGGAAAC
TACAGTGACGGAAACCTTGTTGCGTGTTTGCTTGCGCACAAGATGGGTGTTACTCACTGTACCATTGCCCACGCCCTTGAGAAAACTAAGTAC
CCTAACTCTGACCTCTACTGGAAGAAGTTTGAGGACCACTACCACTTCTCGTGCCAGTTCACCACTGACTTGATTGCTATGAACCATGCTGACT
TCATTATCACAAGTACCTTCCAAGAGATTGCTGGAAACAAGGACACCGTCGGTCAGTACGAGTCACACATGGCATTCACAATGCCTGGTCTGT
ACCGCGTTGTCCATGGTATTGATGTGTTTGACCCCAAGTTCAACATTGTGTCTCCTGGTGCGGACCTGTCTATCTACTTCCCATACACTGAGTC
ACACAAGAGGCTGACCTCCCTCCACCCGGAGATTGAGGAGCTCCTGTACAGCCAAACCGACAACACCGAGCACAAGTTTGTGCTGAACGACA
GGAACAAGCCAATCATCTTCTCCATGGCTCGTCTCGACCGCGTCAAGAACTTGACTGGTCTGGTGGAGCTGTATGGGCGGAACAAGCGCCTG
CAGGAGCTGGTGAACCTCGTGGTTGTCTGCGGTGACCACGGCAACCCTTCCAAGGACAAGGAGGAGCAGGCCGAGTTCAAGAAGATGTTTG
ACCTCATCGAGCAGTACAACCTGAACGGGCACATCCGCTGGATCTCCGCCCAGATGAACCGTGTCCGCAACGGTGAGCTGTACCGCTACATC
TGCGACACCAAGGGCGCCTTCGTGCAGCCTGCTTTCTACGAGGCGTTCGGGCTGACGGTGGTTGAGGCCATGACCTGTGGCCTGCCGACGT
TCGCCACCGCCTACGGTGGTCCGGCTGAGATCATCGTGCACGGCGTGTCTGGCTTCCACATTGACCCGTACCAGGGCGACAAGGCGTCGGC
CCTGCTCGTGGACTTCTTTGAGAAGTGCCAGACAGATTCGAGCCACTGGAACAAGATCTCCCAGGGCGGGCTCCAGCGTATCGAGGAGAAAT
ACACCTGGAAGCTGTACTCGGAGAGGCTGATGACCCTGACCGGCGTGTACGGTTTCTGGAAGTACGTGTCCAACCTGGAGAGGCGCGAGAC
CCGGCGGTACCTGGAGATGCTGTACGCGCTCAAGTACCGCACCATGGCTAGCACCGTGCCGCTGGCCGTGGAGGGAGAGCCCTCCAGCAA
GTGACCCCGCGGCGGCTGGAGACCTGATGAGCGAAAGGGAGCACTTGGAGTCGTGTTTTTTCCTTCCCCTGACCCGGAGGCCAAAAAAGAG
TCTGCTTTTCTTCCTAGGCGGCGGGCGTTCGTTGCTGCTCTTCCCTTCAAGCATTGTTATTACCTTGTCAAGGTCTTGTTTCATCATTGATCCG
GGTGTTGGTTTTAGCAGTCTGATCTGAATTGTTAGTAGTTTGGGTTGAATCGAGCGGTTGAGAGGGATGTTGGGATGGTGGTGGTGTGTGTGC
AGTCGGGTGTGGTGCTCCCTTACTTTCCTGGATGGGATGTTGCTCGTTGAATAATGATCATTGTGGTTGTGGCCTTGGCGCCCTTTTCCCTGA
AATAAGAGTAGCATCCTTGTGGTTCACTTTGCAGCTGGAATCGATGTTTGCCTCAGAAATTTGCTCCCCCCTTTGCCATTTGATTCAGGATCTG
ATGAGTATCGTGA

Gene: GRP7, CCR2
>comp165891_c0_seq1
GTAGCCACCGCCGCCGTAGCCGCCGCCGCCACCGCCGTAGCCACCTCCATCACGGCGGCCGTAGCCGCCACCGCCTCCACGGCCGCCGC
CGTAGCCGCCGCCACCCCCGGCGGAGCGCGGGCCGCCGCGGGACTGGGCCTCGTTGACGGTGATGTTGCGGCCGTCGAGCTCCTTGCCG
TTCATGCCCTCGATGGCGCTCCGCATCGCCTCCTCCGTCGAGAAGGTGACGAAGCCGAAGCCGCGGGACCTCTGCGTCTCCCGATCCAGGA
TGATCTTGGACTCGAGGACCTCGCCGTAGGTGCTGAAGGCGGTGTGGAGGGAGTTGTCGTCGGTGGCCCAGGCGAGGCCGCCGACGAAGC
AGCGGTACTCAACGTCCGACGCCGCCATCTCGTTCCGCTCACCCACGAACCCGAACCCGGAGAACCTGACGAACCCGAACCCGAACCCGAA
CCTGAGGGGGAGGAGAAGCCAACCCACCTCTGAGAGAGCACCAACCGTGAGAATA

Gene: LRB2, POB1
>comp198173_c0_seq1
ACGAAACTGCTACAGTAGCGTTTTGCCAAAAATTTTAGCATCCAAACGGTGCCATGGCCATAGCAGGCCAAATTTACCAACGACGAATGTCCG
CCCAAGACCTTTGCATTTGCTGTCAATCGCCGTTACCGAGGAAACAGAACAGTTCGAACCAGAATCTAGCTCACTGGCAGTGAGATACCGGCC
AATCTGGAAACTTGCAGAGAACTCAAGCTTGAGGTTTCCCATCCGACATCTTTCATTTTTTCATCTGGTGTAACAACTTTCACTATACAGAGACA
AAGATCATACAACCTTCATCTAGTGACTAGTGAAACAACAAAAGGCAGCTTTCAGCCTTTTTTTTTCTTTATACTTGTCATCACCAAACTGTCTAT
GAACAAAAGGATTAAGGTGCATGAACAAAAGAAGAAGAAAAACGAAAAGGATATCTGCAGCACACAGCATTTTCCGCAGCATCACCCACGGCC
GGACAAACTCAACCTCACGGCAGCACATCTATGTAGCTGTTCCGTTGCTCTAGGGTTGCTTTATCGTAAGCTCGGCTCTCAGATGCAGCATCC
CGTCGATGAAGAAGAGGCTGTCATCGGCCATGAACGTCTGCCATGGGATTGCAAAGAGATTCCTGTACCCAACTGCCTTTCCGCCGGTGAAC
GTGTAGTTACCCTTGTACTTGCTCACAAACTCGCCAGATGGTCTCGTCCTGGCAGCAAACTCATAGTCCACGGTCACGCTCATTGAGCCCTTC
TCTTGCATCCCCAGGAAGAGGCCAAAGCAGTAGAACGTGCTTTGCTGCTCCATGTTGCAGTGAGCTGAGAGGAAGAAGCCCTGCCCTGCGAG
GTGGAACGCTTGCGAGTATATCCGGCCTGCAGGAAAGAGTCGAGAGCACTCCTCGCGTTTGAGGTCCAAGTATGCTATGCACTGTGGGTAGG
GCCGATCAAACTCAACGACTTTGAGTGGTCTGTACTTGTAAGCCCGCTCAGCGAATTTCCGACAGGTTATTGTGTCCACTGCAAGAGCTCGTT
GCCGGTGTGGTGCATCAGCTTTGTACAGAAGAGCCTCGGTGACACACTTGGTTGCTTGCTCGTGGTCTATATCCGTGCACGTTAGGACCTTC
CGCAGCTTCCTACATGTCATGTGACTGAATCGCACAAGTGGGAGCAGCCGAGAACTAAGGATCTCACGTCTCTCTTCTGATTTCGGGTATTGT
GCACGGGCCCATCTGAGCAAGAAGTCATAGATGGTATCTTCAGATGCCACCTGGAGGTCATTGCTTGACAAGATGGCTTCAATTCCAGCAAGA
GGGATGTTCATCACTTCATCTTGGAACTTAGTCAAATCCTTGTATTTAACAGCAAGGAAATCCTTAGCTGCATCTGTCAGAGGTTGAACTGCAG
CTGCCATTGAAATTGAGCAGGGGAGATCTAGGTAGAGAAGTGCAGATTCTGTGGTCATGGGCAAGCTTGTGAGCAACTGGCTGCAGTACCTC
ATGCACGAAACAACCTCAAATTTGTCGGCAGCCATCAAGATATCCAGCAGAAGAGTGGGCTCAGTAGTTGTCAACTTTCCACTATACATGAAG
CTTAAAAGCTCCATGAGGGCAGTTTCCTCTGTTCCAAGAGAATTAAGAGAAAAAAAAAGGATGACGCACAAAGCAAATATCCACTAGAGAATGA
TTTCTCTCTCTTTTGTGTGTGCGTATGTATGGCAACTGGTTCTCAGAGAACAAATGACGAAAGCTGTACATAAAATGTTCTTTGCATATATAAGC
AGGATGTTCTAGTTTCCTAAAGTTGTCACATTACCTGAATCAGTAATTCTAAGGGTTGCATGTCTCTGATCCGATTCTTTCATGCCGTTTGAGAA
AAGCTTGAAAAAGAAAGGACTCTTCGCAGCAAGAATCGCCGAACTGATATAAATTGTCTTGACTCGTAAAACTGGCGTACCCACCACAGTCCA
AGATGAATCAATACCTTGCCCATCATCACCTGCATCTTGGCCAACATTGGGTGGAGATTCTTCTATCATTGCTACAGGTTCCTCTTGATTTTCTT
CATATGCATCACACTCTTCTGCTTCGACTTCATTGCAACTTGTTTGGTCTGACATGTGAGTTGTAGATTCTTTCTCCTTGAGGAGCTCCTCGCG
GCGGCGCTTGCGGTGGCGCGCCCAGTCGGCGAGGCCACCTCCTCCGGGACCGCCGCCGCTGGAGCCCGGCGCGTCGTCGTCCCCGGCGA
CGACCTCGATCCGCAATTCCCGGTCGGAGAAGTTGACCTCGTTGAAGGCGAACTCAAAGCTGGGCCCCCCGCCCCCGGGCGAGAAGTCCG
GGTCCATCTCCGCCGCCGCGGTCGCGCGCACACCGACCAGCCTCCTCCTCCGCCTACGGCAAGTGGGGGAGGCGTGGAAGGCGGTGGCC
GGAACCCTAGCCGGCGGAGGGCTCGAGGGGGGAGGGCGTGTGGCCGGGAGGGGAAGGAGGAGGAGGAGACAGAAGGGAAGACGGTGGT
GGTTATAATAAGGAGGCAAGGCAGGAGAAAGATTT

Appendix B: Dissociation curve generated via RT-qPCR reactions for endogenous genes (controls) and candidate genes (targets). Single peaks represent high primer specificity



Appendix C: FLOR dataset- Putative orthologs of flowering genes in Sugarcane predicted through BLASTp searches using the COMPGG largest ORFs sequences as queries against a local assembled BLAST database having the *A. thaliana* Flor-ID sequences, with a cutoff of e^{-05}, minimum of 40% of identity, and minimum of 80% of query coverage

	GENE-ID	GENE	PATHWAY	EFFECT ON FLOWERING
1	comp4911_c0_seq1	HXK1, GIN2	SUGAR	POSITIVE
2	comp109063_c0_seq1	GA20OX2	HORMONES	POSITIVE
3	comp121699_c0_seq1	WRKY34, MSP3	VERNALIZATION	NEGATIVE
4	comp126941_c0_seq1	GAI, RGA2	HORMONES	POSITIVE
5	comp144764_c0_seq1	GA3OX1	HORMONES	POSITIVE
6	comp151459_c0_seq1	PGM1	SUGAR	POSITIVE
7	comp165891_c0_seq1	GRP7, CCR2	GENERAL	POSITIVE
8	comp169449_c0_seq1	UGT87A2	GENERAL	POSITIVE
9	comp172176_c0_seq1	ADG1, APS1	SUGAR	POSITIVE
10	comp173852_c0_seq1	AGL16	GENERAL	NEGATIVE
11	comp174508_c0_seq1	MSI1	GENERAL	POSITIVE
12	comp179538_c0_seq1	HDA6, RTS1	GENERAL	POSITIVE
13	comp180010_c0_seq1	CUL3A	VERNALIZATION	POSITIVE
14	comp180194_c0_seq1	PHP, CDC73	VERNALIZATION	NEGATIVE
15	comp181006_c0_seq1	SEF, SWC6	GENERAL	NEGATIVE
16	comp181670_c0_seq1	GA2, ATKS1	HORMONES	POSITIVE
17	comp182838_c0_seq1	AGL16	GENERAL	NEGATIVE
18	comp182845_c0_seq1	AKIN10, SNRK1.1	SUGAR	NEGATIVE
19	comp182845_c2_seq1	AKIN10, SNRK1.1	SUGAR	NEGATIVE
20	comp184054_c0_seq1	ATPRMT10	GENERAL	POSITIVE
21	comp184140_c0_seq1	ESD7, TIL1, ABO4	GENERAL	NEGATIVE
22	comp185692_c1_seq1	SVP	FLOWERING TIME INTEGRATOR	NEGATIVE
23	comp186591_c0_seq1	ATXR7, SDG25	GENERAL	NEGATIVE
24	comp186929_c0_seq1	ASH2R, TRO	GENERAL	NEGATIVE
25	comp187233_c0_seq1	PHYA	PHOTOPERIODISM	POSITIVE
26	comp188103_c0_seq1	LRB1	VERNALIZATION	POSITIVE
27	comp188103_c2_seq1	LRB1	VERNALIZATION	POSITIVE

28	comp188565_c0_seq1	YAF9A, TAF14B	GENERAL	NEGATIVE
29	comp188881_c0_seq1	PGM1	SUGAR	POSITIVE
30	comp189287_c1_seq1	VOZ1	PHOTOPERIODISM	POSITIVE
31	comp189499_c0_seq1	ICE1, SCRM	VERNALIZATION	NEGATIVE
32	comp189698_c0_seq1	WDR5A	VERNALIZATION	NEGATIVE
33	comp190463_c0_seq1	UGT87A2	GENERAL	POSITIVE
34	comp190976_c0_seq1	ADG1, APS1	SUGAR	POSITIVE
35	comp191126_c0_seq1	ESD4	GENERAL	NEGATIVE
36	comp191239_c0_seq1	JMJ32, JMJD5	GENERAL	NEGATIVE
37	comp191266_c0_seq1	ATJ3	GENERAL	POSITIVE
38	comp191301_c0_seq1	HXK1, GIN2	SUGAR	POSITIVE
39	comp191649_c0_seq1	HAM1	GENERAL	NEGATIVE
40	comp192064_c0_seq1	REF6, JMJ12	GENERAL	POSITIVE
41	comp192150_c0_seq1	BAF60, CHC1	GENERAL	POSITIVE
42	comp192296_c0_seq1	PUB13	GENERAL	POSITIVE
43	comp192856_c0_seq1	FVE, MSI4	GENERAL	POSITIVE
44	comp193070_c0_seq1	HXK1, GIN2	SUGAR	POSITIVE
45	comp193361_c0_seq1	ELF8, VIP6	GENERAL	NEGATIVE
46	comp193746_c0_seq1	GA2OX8	HORMONES	NEGATIVE
47	comp193806_c1_seq1	ICE1, SCRM	VERNALIZATION	NEGATIVE
48	comp194394_c0_seq1	COL9	PHOTOPERIODISM	NEGATIVE
49	comp194712_c0_seq1	ARP6, SUF3, ESD1	GENERAL	NEGATIVE
50	comp195240_c0_seq1	CKB4	CIRCADIAN CLOCK	POSITIVE
51	comp195337_c1_seq1	AKIN10, SNRK1.1	SUGAR	NEGATIVE
52	comp195715_c1_seq1	ARP4	GENERAL	NEGATIVE
53	comp195751_c0_seq1	VIP5	GENERAL	NEGATIVE
54	comp195801_c1_seq1	ATJ3	GENERAL	POSITIVE
55	comp195844_c0_seq1	ABH1, CBP80	GENERAL	NEGATIVE
56	comp196172_c0_seq1	SSRP1, HMG, NFD	GENERAL	NEGATIVE
57	comp196353_c0_seq1	CUL4	GENERAL	NEGATIVE
58	comp196438_c1_seq1	EBS	PHOTOPERIODISM	NEGATIVE
59	comp196444_c0_seq1	HDA5	GENERAL	POSITIVE

60	comp196577_c0_seq1	CRY1	PHOTOPERIODISM	POSITIVE
61	comp196789_c0_seq1	SVP	FLOWERING TIME INTEGRATOR	NEGATIVE
62	comp197050_c1_seq1	UGT87A2	GENERAL	POSITIVE
63	comp197114_c0_seq1	LWD1, ATAN11	CIRCADIAN CLOCK	NEGATIVE
64	comp197385_c1_seq1	MSI1	GENERAL	POSITIVE
65	comp197388_c0_seq1	VIL1, VRN5	VERNALIZATION	POSITIVE
66	comp197504_c0_seq1	VIP3, SKI8	GENERAL	NEGATIVE
67	comp197644_c0_seq1	DCL3	GENERAL	POSITIVE
68	comp197644_c2_seq1	DCL3	GENERAL	POSITIVE
69	comp197651_c0_seq1	PUB13	GENERAL	POSITIVE
70	comp197749_c1_seq1	GA20OX1	HORMONES	POSITIVE
71	comp197815_c0_seq1	HDA5	GENERAL	POSITIVE
72	comp197939_c2_seq1	BBX19	PHOTOPERIODISM	NEGATIVE
73	comp197947_c0_seq1	VIP4	GENERAL	NEGATIVE
74	comp198087_c0_seq1	GAI, RGA2	HORMONES	POSITIVE
75	comp198109_c0_seq1	TPS1	SUGAR	POSITIVE
76	comp198173_c0_seq1	LRB2, POB1	VERNALIZATION	POSITIVE
77	comp198370_c2_seq1	SUS4	SUGAR	POSITIVE
78	comp198410_c0_seq1	TOE1, RAP2.7	AGING	NEGATIVE
79	comp198591_c2_seq1	VOZ1	PHOTOPERIODISM	POSITIVE
80	comp198712_c0_seq1	MED16, SFR6	GENERAL	POSITIVE
81	comp199135_c0_seq1	FT	FLOWERING TIME INTEGRATOR	POSITIVE
82	comp199200_c0_seq1	GA1	HORMONES	POSITIVE
83	comp199200_c1_seq1	GA1	HORMONES	POSITIVE
84	comp199267_c0_seq1	SDG26, ASHH1	GENERAL	POSITIVE
85	comp199378_c0_seq1	SIZ1	GENERAL	NEGATIVE
86	comp199380_c1_seq1	ELF6	GENERAL	NEGATIVE
87	comp199615_c0_seq1	UBP12	GENERAL	NEGATIVE
88	comp199615_c1_seq1	UBP13	GENERAL	NEGATIVE
89	comp199615_c3_seq1	UBP13	GENERAL	NEGATIVE
90	comp199692_c1_seq1	DET1, FUS2	PHOTOPERIODISM	NEGATIVE
91	comp199837_c0_seq1	LDL1	GENERAL	POSITIVE

92	comp199837_c1_seq1	LDL1	GENERAL	POSITIVE
93	comp199879_c0_seq1	BRM, CHR2	GENERAL	NEGATIVE
94	comp199984_c0_seq1	SUS4	SUGAR	POSITIVE
95	comp200190_c1_seq1	CLF, SDG1	VERNALIZATION	NEGATIVE
96	comp200275_c0_seq1	CUL3A	VERNALIZATION	POSITIVE
97	comp200408_c0_seq1	CSTF77	GENERAL	POSITIVE
98	comp200516_c0_seq1	SUF4	VERNALIZATION	NEGATIVE
99	comp200591_c0_seq1	CRY1	PHOTOPERIODISM	POSITIVE
100	comp200591_c1_seq1	CRY1	PHOTOPERIODISM	POSITIVE
101	comp200785_c0_seq1	TOE1, RAP2.7	AGING	NEGATIVE
102	comp201070_c0_seq1	CBP20	GENERAL	NEGATIVE
103	comp201152_c0_seq1	UBC2	GENERAL	NEGATIVE
104	comp201398_c0_seq1	RRP6L2	GENERAL	POSITIVE
105	comp201653_c0_seq1	ATX2	GENERAL	NEGATIVE
106	comp201728_c1_seq1	NF-YB3, HAP3C	PHOTOPERIODISM	POSITIVE
107	comp201729_c0_seq1	EMF2, CYR1	GENERAL	NEGATIVE
108	comp201813_c1_seq1	MRG1	GENERAL	POSITIVE
109	comp201878_c0_seq1	CSTF64	GENERAL	POSITIVE
110	comp201934_c0_seq1	ZTL	CIRCADIAN CLOCK	NEGATIVE
111	comp202087_c0_seq1	PRR7	CIRCADIAN CLOCK	POSITIVE
112	comp202116_c2_seq1	COP1	PHOTOPERIODISM	NEGATIVE
113	comp202161_c2_seq1	EMF2, CYR1	GENERAL	NEGATIVE
114	comp202287_c0_seq1	GA2OX2	HORMONES	NEGATIVE
115	comp202441_c2_seq1	GA2, ATKS1	HORMONES	POSITIVE
116	comp202464_c0_seq1	PRP8, EMB33	GENERAL	POSITIVE
117	comp202522_c1_seq1	SKB1, PRMT5	GENERAL	POSITIVE
118	comp202645_c0_seq1	LIF2	GENERAL	NEGATIVE
119	comp202805_c0_seq1	SDG7, ASSH3	VERNALIZATION	NEGATIVE
120	comp202963_c1_seq1	UGT87A2	GENERAL	POSITIVE
121	comp203015_c0_seq1	DCL1	GENERAL	POSITIVE
122	comp203057_c1_seq1	FIO1	CIRCADIAN CLOCK	NEGATIVE
123	comp203344_c0_seq1	CRY2	PHOTOPERIODISM	POSITIVE

124	comp203449_c0_seq1	MYR1	PHOTOPERIODISM	NEGATIVE
125	comp203529_c0_seq1	FLD, RSI1	GENERAL	POSITIVE
126	comp203529_c1_seq1	FLD, RSI1	GENERAL	POSITIVE
127	comp204063_c1_seq1	HXK1, GIN2	SUGAR	POSITIVE
128	comp204072_c0_seq1	PHYB	PHOTOPERIODISM	NEGATIVE
129	comp204236_c0_seq1	GAI, RGA2	HORMONES	POSITIVE
130	comp204294_c1_seq1	SPA1	PHOTOPERIODISM	NEGATIVE
131	comp204402_c0_seq1	PHYA	PHOTOPERIODISM	POSITIVE
132	comp204503_c1_seq1	FLK	GENERAL	POSITIVE
133	comp204664_c0_seq1	HUB2	GENERAL	NEGATIVE
134	comp204750_c2_seq1	UBP12	GENERAL	NEGATIVE
135	comp204774_c0_seq1	EBS	PHOTOPERIODISM	NEGATIVE
136	comp204886_c0_seq1	PHYC	PHOTOPERIODISM	NEGATIVE
137	comp205075_c0_seq1	DCL3	GENERAL	POSITIVE
138	comp205284_c2_seq1	UBP26	GENERAL	NEGATIVE
139	comp205398_c0_seq1	SIZ1	GENERAL	NEGATIVE
140	comp205442_c0_seq1	INO80	GENERAL	POSITIVE
141	comp205467_c1_seq1	JMJ30	GENERAL	NEGATIVE
142	comp205469_c2_seq1	PFT1, MED25	PHOTOPERIODISM	POSITIVE
143	comp205700_c0_seq1	PIE1	GENERAL	NEGATIVE
144	comp205770_c0_seq1	LDL2	GENERAL	POSITIVE
145	comp205837_c0_seq1	UBP12	GENERAL	NEGATIVE
146	comp206078_c0_seq1	GIS5, EMB2780	GENERAL	NEGATIVE
147	comp206124_c0_seq1	PRR7	CIRCADIAN CLOCK	POSITIVE
148	comp206240_c3_seq1	CPK6	PHOTOPERIODISM	POSITIVE
149	comp206323_c0_seq1	FKF1, ADO3	PHOTOPERIODISM	POSITIVE
150	comp206536_c0_seq1	TPL	AGING	NEGATIVE
151	comp206637_c0_seq1	GI	CIRCADIAN CLOCK	POSITIVE
152	comp206769_c0_seq1	DCL4	GENERAL	POSITIVE
153	comp206868_c0_seq1	GLK1, GPRI1	GENERAL	NEGATIVE
154	comp206981_c0_seq1	SPT16	GENERAL	NEGATIVE
155	comp206981_c1_seq1	SPT16	GENERAL	NEGATIVE

156	comp207438_c0_seq1	MED13, GCT, MAB2	GENERAL	POSITIVE
157	comp207537_c0_seq1	MED12, CCT, CRP	GENERAL	POSITIVE
158	comp207551_c1_seq1	AP1	FLOWER DEVELOPMENT	POSITIVE
159	comp207801_c0_seq1	ESD7, TIL1, ABO4	GENERAL	NEGATIVE
160	gg_00125	TOE1, RAP2.7	AGING	NEGATIVE
161	gg_00570	FLD, RSI1	GENERAL	POSITIVE
162	gg_00621	GRP7, CCR2	GENERAL	POSITIVE
163	gg_00669	CKB4	CIRCADIAN CLOCK	POSITIVE
164	gg_00950	MYR2	PHOTOPERIODISM	NEGATIVE
165	gg_01014	GA1	HORMONES	POSITIVE
166	gg_01112	COP1	PHOTOPERIODISM	NEGATIVE
167	gg_01706	ESD6, HOS1	PHOTOPERIODISM	POSITIVE
168	gg_02384	ADG1, APS1	SUGAR	POSITIVE
169	gg_02647	JMJ32, JMJD5	GENERAL	NEGATIVE
170	gg_02840	JMJ30	GENERAL	NEGATIVE
171	gg_02988	LRB1	VERNALIZATION	POSITIVE
172	gg_03035	MED18	GENERAL	POSITIVE
173	gg_03050	GI	CIRCADIAN CLOCK	POSITIVE
174	gg_03849	LRB1	VERNALIZATION	POSITIVE
175	gg_04124	HDA9	GENERAL	NEGATIVE
176	gg_04311	AKIN10, SNRK1.1	SUGAR	NEGATIVE
177	gg_04347	GA2, ATKS1	HORMONES	POSITIVE
178	gg_04388	MRG1	GENERAL	POSITIVE
179	gg_04469	BAF60, CHC1	GENERAL	POSITIVE
180	gg_04634	SEF, SWC6	GENERAL	NEGATIVE
181	gg_04771	UBP12	GENERAL	NEGATIVE
182	gg_04836	CUL3A	VERNALIZATION	POSITIVE
183	gg_04891	LIF2	GENERAL	NEGATIVE
184	gg_05025	PGM1	SUGAR	POSITIVE
185	gg_05144	ZTL	CIRCADIAN CLOCK	NEGATIVE
186	gg_05438	GLK2	GENERAL	NEGATIVE
187	gg_05493	PUB13	GENERAL	POSITIVE

188	gg_05531	HUB2	GENERAL	NEGATIVE
189	gg_05696	AP1	FLOWER DEVELOPMENT	POSITIVE
190	gg_05775	GA2, ATKS1	HORMONES	POSITIVE
191	gg_05858	NF-YB1, HAP3A	PHOTOPERIODISM	NEGATIVE
192	gg_05947	MYR2	PHOTOPERIODISM	NEGATIVE
193	gg_06280	SVP	FLOWERING TIME INTEGRATOR	NEGATIVE
194	gg_06367	FLK	GENERAL	POSITIVE
195	gg_06574	CSTF77	GENERAL	POSITIVE
196	gg_06658	TEM2, RAV2	AMBIENT TEMPERATURE	NEGATIVE
197	gg_06688	HTA11	GENERAL	NEGATIVE
198	gg_06966	MOS1	GENERAL	POSITIVE
199	gg_07023	AKIN10, SNRK1.1	SUGAR	NEGATIVE
200	gg_07145	MSI1	GENERAL	POSITIVE
201	gg_07147	WDR5A	VERNALIZATION	NEGATIVE
202	gg_07255	FIE1, FIS3	GENERAL	NEGATIVE
203	gg_07330	AFR2	PHOTOPERIODISM	NEGATIVE
204	gg_07428	PRR7	CIRCADIAN CLOCK	POSITIVE
205	gg_07446	TOE1, RAP2.7	AGING	NEGATIVE
206	gg_07497	HXK1, GIN2	SUGAR	POSITIVE
207	gg_07545	FY	GENERAL	POSITIVE
208	gg_07573	CDF2	PHOTOPERIODISM	NEGATIVE
209	gg_07611	VIP5	GENERAL	NEGATIVE
210	gg_07924	FTIP1	PHOTOPERIODISM	POSITIVE
211	gg_08235	ATC	PHOTOPERIODISM	NEGATIVE
212	gg_08285	FUL, AGL8	FLOWERING TIME INTEGRATOR	POSITIVE
213	gg_08436	MED16, SFR6	GENERAL	POSITIVE
214	gg_08550	UBC2	GENERAL	NEGATIVE
215	gg_08777	SUF4	VERNALIZATION	NEGATIVE
216	gg_08993	FVE, MSI4	GENERAL	POSITIVE
217	gg_09130	ATJ3	GENERAL	POSITIVE
218	gg_09147	DCL4	GENERAL	POSITIVE
219	gg_09307	FT	FLOWERING TIME INTEGRATOR	POSITIVE

220	gg_09326	JMJ30	GENERAL	NEGATIVE
221	gg_09327	CRY1	PHOTOPERIODISM	POSITIVE
222	gg_09418	UGT87A2	GENERAL	POSITIVE
223	gg_09633	ABH1, CBP80	GENERAL	NEGATIVE
224	gg_09684	DET1, FUS2	PHOTOPERIODISM	NEGATIVE
225	gg_09702	TPL	AGING	NEGATIVE
226	gg_09821	UBP26	GENERAL	NEGATIVE
227	gg_10072	PRP8, EMB33	GENERAL	POSITIVE
228	gg_10086	LIF2	GENERAL	NEGATIVE
229	gg_10091	HDA6, RTS1	GENERAL	POSITIVE
230	gg_10332	SKB1, PRMT5	GENERAL	POSITIVE
231	gg_10422	APL, WDY, FE	PHOTOPERIODISM	POSITIVE
232	gg_10726	SPT16	GENERAL	NEGATIVE
233	gg_10979	UBC2	GENERAL	NEGATIVE
234	gg_11337	ARP6, SUF3, ESD1	GENERAL	NEGATIVE
235	gg_11383	CUL4	GENERAL	NEGATIVE
236	gg_11571	CPK6	PHOTOPERIODISM	POSITIVE
237	gg_11586	LDL1	GENERAL	POSITIVE
238	gg_11892	EBS	PHOTOPERIODISM	NEGATIVE
239	gg_12051	DCL1	GENERAL	POSITIVE
240	gg_12215	GAI, RGA2	HORMONES	POSITIVE
241	gg_12309	EMF2, CYR1	GENERAL	NEGATIVE
242	gg_12336	TPS1	SUGAR	POSITIVE
243	gg_12377	SSRP1, HMG, NFD	GENERAL	NEGATIVE
244	gg_12460	LATE	PHOTOPERIODISM	NEGATIVE
245	gg_12463	HXK1, GIN2	SUGAR	POSITIVE
246	gg_12613	GI	CIRCADIAN CLOCK	POSITIVE
247	gg_12814	VIP3, SKI8	GENERAL	NEGATIVE
248	gg_12882	ESD4	GENERAL	NEGATIVE
249	gg_13242	VOZ1	PHOTOPERIODISM	POSITIVE
250	gg_13427	PGI1	SUGAR	POSITIVE
251	gg_13443	COL9	PHOTOPERIODISM	NEGATIVE

252	gg_13446	SOC1, AGL20	FLOWERING TIME INTEGRATOR	POSITIVE
253	gg_13984	HULK3	GENERAL	POSITIVE
254	gg_14121	GIS5, EMB2780	GENERAL	NEGATIVE
255	gg_14131	HDA5	GENERAL	POSITIVE
256	gg_14198	ELF3	CIRCADIAN CLOCK	NEGATIVE
257	gg_14212	PFT1, MED25	PHOTOPERIODISM	POSITIVE
258	gg_14467	HDA5	GENERAL	POSITIVE
259	gg_14523	PHYB	PHOTOPERIODISM	NEGATIVE
260	gg_14709	ELF7, VIP2	GENERAL	NEGATIVE
261	gg_14784	BBX19	PHOTOPERIODISM	NEGATIVE
262	gg_14818	UBP12	GENERAL	NEGATIVE
263	gg_14859	CRY2	PHOTOPERIODISM	POSITIVE
264	gg_14938	HAM2, HAG5, HAC11	GENERAL	NEGATIVE
265	gg_14947	UBC2	GENERAL	NEGATIVE
266	gg_15000	YAF9A, TAF14B	GENERAL	NEGATIVE
267	gg_15070	DCL4	GENERAL	POSITIVE
268	gg_15108	ARP4	GENERAL	NEGATIVE
269	gg_15162	FT	FLOWERING TIME INTEGRATOR	POSITIVE
270	gg_15361	SUS4	SUGAR	POSITIVE
271	gg_15435	GI	CIRCADIAN CLOCK	POSITIVE
272	gg_15551	ADG1, APS1	SUGAR	POSITIVE
273	gg_15632	FBH4, AKS3	PHOTOPERIODISM	POSITIVE
274	gg_15635	SPA3	PHOTOPERIODISM	NEGATIVE
275	gg_15862	ESD7, TIL1, ABO4	GENERAL	NEGATIVE
276	gg_15932	EMF2, CYR1	GENERAL	NEGATIVE
277	gg_16034	ATX2	GENERAL	NEGATIVE

Appendix D: Dissociation curves generated via RT-qPCR reactions for endogenous genes (controls) and candidate genes (targets) in ethephon treated samples. Single peaks represent high primer specificity

SRO1-IAC-SP

eEF1a-IAC-SP

GAPDH-IAC-SP

ATPRMT10-IAC-SP

BRM-IAC-SP

ESD7-IAC-SP

SYD-IAC-SP

TPS1-IAC-SP