THE EXPANSION AND DIVERSIFICATION OF THE CCA1-LHY-RVE PROTEIN FAMILY IN MONOCOTS AND DICOTS

Jérôme Gélinas Bélanger

Department of Plant Science

McGill University, Montreal

August 2019

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Master of Science

ABSTRACT

Rice (*Oryza sativa*) is a staple crop for more than 30% of the world population. Maintaining high agricultural yield for this crop is an essential component towards consolidating and achieving food security in numerous developing countries. Currently, climate change is a growing threat for rice culture in tropical areas, and a gradual decrease of 10-15% in the total global rice yield is expected by 2050 because of a higher incidence of abiotic and biotic stresses. The circadian clock is an endogenous timekeeping mechanism found in all prokaryotic and eukaryotic organisms. It coordinates the physiological and molecular responses to abiotic stress, by regulating processes such as stomatal aperture and flowering time in plants. The general goal of this research project aims at characterizing the structural and functional features of the CCA1-LHY-RVE protein family, a MYB transcription factor family involved in the regulation of the circadian rhythm and comprising 11 members in arabidopsis and seven in rice. For the moment, this family has been structurally characterized in arabidopsis, rice and tomato, but no in-depth studies have thoroughly unravelled its role at the molecular level in other model and cultivated monocotyledonous and dicotyledonous species.

This research project comprised three main objectives. The first objective consisted by performing a phylogenetic analysis to structurally characterize the CCA1-LHY-RVE protein family in monocots and dicots. To do so, bioinformatics (phylogenetic analysis, determination of identity ratios between protein sequences, motifs discovery) were performed to evaluate the evolutionary and phylogenetic relationships of the CCA1-LHY-RVE protein family in 10 monocotyledonous species and 21 dicotyledonous species. Results reveal that the RVE8 clade members share a high protein sequence identity, $\geq$ 55%, within the same species and class. Similar observations were made for the LHY clade, whereas the members of the RVE1 clade, the group to which OsEPR1 and AtEPR1 (EARLY PHYTOCHROME RESPONSE 1) both belong, displayed limited inter- and intra-species identity amongst themselves. The second objective was to develop a pipeline for the relative quantification of transcript abundance for six of the seven genes belonging to the CCA1-LHY-RVE family in rice. To achieve this objective,

nine primer pairs (three housekeeping genes and six RVE genes) were designed and validated using standard RT-qPCR validation protocol (bioinformatics validation of primer design, standard curves, melt curve analysis, analysis of housekeeping gene stability, gel electrophoresis). The third and last objective was to quantify the relative transcript abundance using RT-qPCR in six members of the *CCA1-LHY-RVE* family for rice plants that have been subjected to two treatments, a diurnal treatment aiming at observing the daily circadian oscillation of the genes and a heat shock treatment seeking at observing the response of the genes to heat shock. Only two genes, LOC_Os02g45670 and LOC_Os04g49450, were differentially abundant upon a 30 minutes heat shock for the studied timepoints.

The results of this research improve our global understanding of the linkages between the circadian clock and the tolerance to abiotic stresses in plants, especially in monocotyledonous species. Findings of this research project strengthen our understanding of the structural conformation and phylogenetic relationships for the CCA1-LHY-RVE protein family in monocotyledonous species and dicotyledonous species. In a broader perspective, the data generated by this project contributes to the current state of knowledge about the phylogenetic relationships of the plant circadian clock between various cultivated species. Moreover, this research project shines a light on the phylogenetic divergence between monocots and dicots for the RVE1 and LHY clades.

RÉSUMÉ

Le riz (*Oryza sativa*) est un aliment de base pour plus de 30% de la population mondiale et le maintien de rendements élevés pour cette culture s'avère un élément essentiel à la sécurité alimentaire pour de nombreux pays en voie de développement. À l'heure actuelle, les changements climatiques représentent une menace importante pour cette culture en zone tropicale et une réduction graduelle de rendement de 10-15% en lien avec une augmentation des stress hydriques et thermiques est anticipée d'ici 2050. L'horloge circadienne est un mécanisme endogène de gestion du temps impliquée dans la coordination de la réponse aux stress abiotiques et biotiques, telles que la régulation de l'aperture stomatale et l'ouverture des bourgeons floraux, dans les végétaux. L'objectif général de ce projet de recherche visait à la caractérisation des attributs structuraux et fonctionnels de la famille de protéines CCA1-LHY-RVE, une famille de facteurs de transcription MYB impliquée dans la régulation du rythme circadien et comprenant 11 membres dans arabidopsis et sept dans le riz.

Ce projet de recherche était constitué de trois principaux objectifs. Le premier objectif consistait à entreprendre une analyse phylogénétique détaillée afin de caractériser la famille de protéines CCA1-LHY-RVE sur le plan structural pour diverses espèces de monocotylédones et dicotylédones. Afin de réaliser cet objectif, une analyse bio-informatique (analyse phylogénétique, détermination de la similarité entre les séquences protéiques, découvertes de motifs séquentiels récurrents) a été effectuée afin d'évaluer les liens évolutifs et structuraux de la famille de protéines CCA1-LHY-RVE pour 10 espèces de monocotylédones et 21 espèces de dicotylédones. Les résultats générés démontrent que les membres de la clade RVE8 sont caractérisés par une très forte identité séquentielle protéique, ≥55%, à l'intérieur de la même espèce, ainsi que les mêmes classes. Des observations similaires ont été réalisées pour le groupe de protéines LHY, tandis que les membres de la clade RVE1, le groupe auxquels OsEPR1 et AtEPR1 (EARLY PHYTOCHROME RESPONSE 1) appartiennent, démontrent une faible identité inter et intra-espèce parmi eux-mêmes. Le second objectif de ce projet était de développer une procédure pour la quantification relative de l'abondance de transcripts

pour six des sept gènes constituant la famille CCA1-LHY-RVE dans le riz. Pour réaliser cet objectif, neuf paires d'amorces (trois gènes domestiques et six gènes RVE) ont été conçues et validées en utilisant un protocole standard de validation RT-qPCR (évaluation bio-informatique du design des amorces, courbes standards, analyse de la courbe de fusion, détermination de la stabilité des gènes domestiques, électrophorèse sur gel d'agarose). Le troisième et dernier objectif était de quantifier l'expression en mesurant l'abondance de transcripts par technique RT-qPCR des gènes de la famille *CCA1-LHY-RVE* pour des plants de riz ayant été soumis à deux traitements, un traitement diurnal visant à observer l'oscillation circadienne et un traitement de choc thermique ayant pour but l'évaluation de la réponse des gènes face à ce stress. Les données générées démontrent que seulement deux gènes, LOC_Os02g45670 et LOC_Os04g49450, sont différentiellement abondants suivant un choc thermique de 30 minutes.

Les découvertes de ce projet contribuent à une fortification de notre compréhension de la conformation structurelle et des liens phylogénétiques unissant les membres de la famille CCA1-LHY-RVE pour les monocotylédones et les dicotylédones. Dans une perspective plus large, les données générées par ce projet ajoutent une brique à l'état actuel des connaissances à propos des liens phylogénétiques entre les diverses espèces de plantes cultivées, ainsi qu'à la divergence évolutive entre les monocotylédones et dicotylédones pour les clades RVE1 et LHY.

TABLE OF CONTENT

Abstract ii
Résumé iv
Table of contents vi
List of tables and figures ix
List of abbreviations xi
Acknowledgements xv
Preface xvi
CHAPTER I: General introduction 1
1.1 Introduction 1
1.2 Research hypotheses and objectives 3
CHAPTER II: Literature review 4
2.1 Rice: a staple crop for food security and science 4
2.1.1 A pillar for food security and a model plant for monocots in biological studies 4
2.2 Defining features of the circadian clock 5
2.2.1 An introduction to the circadian clock 5
2.2.2 The mechanism of the circadian clock 6
2.2.3 Beyond Arabidopsis: the circadian in other angiosperms 7
2.3 The CCA1-LHY-RVE protein family in arabidopsis: A family of MYB transcription factors 9
2.3.1 The CCA1-LHY-RVE protein family 9
2.3.2 LHY clade 11
2.3.3 RVE8 clade 12
2.3.4 RVE1 clade 13
2.3.4.1 EPR1 14
2.3.4.2 CIR1 15
2.5.4.3 RVE1 16
2.4 The circadian clock and temperature: Temperature compensation, thermal entrainment and thermomorphogenesis 17

2.4.1 The consequences of heat stress on plants 17
2.4.2 Temperature compensation 18
2.4.3 Thermal entrainment and thermomorphogenesis 21
2.5 Connecting statement: upcoming rising challenges are time- and species-dependent 23
2.6 Tables and figures 24
CHAPTER III: Materials and methods 27
3.1 Characterization of the expansion and diversification of the RVE protein family in monocots and dicots 27
3.1.1 Alignment of the CCA1-LHY-RVE family members in sequenced monocot and dicot genomes 27
3.1.2 Identification of conserved motifs in the amino acid sequences of RVE1, RVE8 and LHY clade proteins 28
3.2 Expression of genes in the *CCA1-LHY-RVE* family over a diurnal cycle and in response to heat stress 28
3.2.1 Rice plant material and growth conditions 28
3.2.2 Sampling tissues for transcript measurements 29
3.2.3 RNA extraction and cDNA synthesis of the harvested samples 29
3.2.4 RT-qPCR primer design for members of the CCA1-LHY-RVE family in rice 29
3.2.5 Validation of PCR efficiency 30
3.2.6 Relative quantification of transcript abundance through RT-qPCR assay 30
CHAPTER IV: Characterization of the expansion and diversification of the CCA1-LHY-RVE protein family in monocots and dicots 31
4.1 Introduction 31
4.2 Results 32
4.3 Discussion 36
4.4 Tables and figures 42
CHAPTER V: Development of RT-qPCR assays for members of the *CCA1-LHY-RVE* family and measurement of the gene expression upon heat stress 55
5.1 Introduction 55

5.2 Results 56
5.3 Discussion 59
5.4 Tables and figures 62
CHAPTER VI: General conclusion 77
BIBLIOGRAPHY 82

LIST OF TABLES AND FIGURES

Table 2.1 Protein families forming the circadian clock in arabidopsis.....24

Table 2.2 Members of the CCA1-LHY-RVE protein family in arabidopsis.....25

Table 4.1 RVE gene complements of 10 monocot species.....42

Table 4.2 RVE gene complements of 21 dicot species.....43

Table 4.3 FIMO promoter scanning of HSE and EE in CCA1-LHY-RVE members for arabidopsis and rice 44

Table 5.1 Primer sequences and amplification properties for RVE and housekeeping genes 62

Table 5.2 Similarity between predicted Tm and observed Tm for 4 housekeeping genes and 6 members of the rice CCA1-LHY-RVE protein family.....63

Table 5.3 Validation of primer design and preparation of RT-qPCR samples based on RT-qPCR amplification 64

Table 5.4 Spectrophotometer assessment of the quality of the extracted RNA for samples submitted to diurnal treatment 65

Table 5.5 Spectrophotometer assessment of the quality of the extracted RNA for samples submitted to heat shock treatment.....66

Table 5.6 ANOVA results for p-value Tukey of RT-qPCR assays for diurnal and heat shock treatments.....67

Figure 2.1 Morning and evening loops of the circadian clock gene regulatory network ..26

Figure 4.1 Species tree of the 10 monocot and 21 dicot species included in the analysis of the CCA1-LHY-RVE protein family.....45

Figure 4.2 Sequence logos of the helix-turn-helix MYB DNA binding domain and proline rich domains for the 165 putative members of the CCA1-LHY-RVE protein family in the 31 studied species 46

Figure 4.3 Multiple sequence alignment of the helix-turn-helix MYB DNA binding domain and the proline rich region of the CCA1-LHY-RVE protein members in rice and arabidopsis 47

Figure 4.4 Phylogenetic relationships and subgroup designations of the RVE family proteins in rice and arabidopsis.....48

Figure 4.5 Phylogenetic relationships and subgroup designations of the RVE family proteins in 10 monocot species 49

Figure 4.6 Phylogenetic relationships and subgroup designations of the RVE family proteins in 21 dicot species 50

Figure 4.7 Identity matrix for the putative monocot CCA1-LHY-RVE proteins 51

Figure 4.8 Identity matrix for the putative dicot CCA1-LHY-RVE proteins 52

Figure 4.9 Multiple sequence alignment of the full-length protein sequences for all putative monocot and dicot CCA1-LHY-RVE proteins 53

Figure 4.10 EPR1 and LHY clade proteins are conserved over their full lengths 54

Figure 5.1 Schematic representation of diurnal treatment experimental protocol 68

Figure 5.2 Schematic representation of heat shock treatment experimental protocol 69

Figure 5.3 Primer specificity is confirmed 70

Figure 5.4 Fidelity of PCR assays was confirmed by melt curve analysis 71

Figure 5.5 Amplification plot and standard curve for the *OsEPR1* PCR Assay 72

Figure 5.6 High quality total RNA was extracted from rice leaves 73

Figure 5.7 *OsSUI1* and *OsGAPDH* are the most stable housekeeping genes 74

Figure 5.8 Transcript abundance for six RVE genes in response to diurnal treatment 75

Figure 5.9 Transcript abundance for six RVE genes in response to heat shock treatment 76

LIST OF ABBREVIATIONS

A	Adenine
ABA	Abscisic acid
ASG4	ALTERED SEED GERMINATION 4
At	*Arabidopsis thaliana*
bHLH	Basic helix-loop-helix
bp	Base-pairs
C	Cytosine
CBF	C-REPEAT BINDING FACTOR
CO_2	Carbon dioxide
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
CCT	CONSTANS, CO-like, and TOC1 domain
CDF	CYCLING DOF FACTOR
cDNA	Complementary DNA
CHE	CCA1 HIKING EXPEDITION
CIR1	CIRCADIAN 1
CO	CONSTANS
COL	CONSTANS-LIKE
CRISPR-cas9	Clustered regularly interspaced short palindromic repeats-cas9
CRT	C-repeat
D	Diurnal
DAG	Days after germination

DEAD	Asp-Glu-Ala-Asp
DNA	Deoxyribonucleic acid
DREB	DEHYDRATION RESPONSIVE ELEMENT BINDING
EDS1	ENHANCED DISEASE SUSCEPTIBILITY 1
ELF	EARLY FLOWERING
EPR1	EARLY PHYTOCHROME RESPONSE 1
ERF/AP2	ETHYLENE RESPONSE FACTOR/APETALA 2
FBH1	FLOWERING BHLH1
FT	FLOWERING TIME
G	Guanine
GARP	Glutamic Acid/Alanine-Rich Protein domain
GI	GIGANTEA
HS	Heat shock
HSE	Heat Shock Element
HSF	Heat shock factor
HSFP90/70	HEAT SHOCK PROTEIN90/HEAT SHOCK PROTEIN70
HSFB2B	HEAT SHOCK FACTOR B2b
LBS	LUX binding site
LCL	LHY/CCA1-LIKE
LHCB	LIGHT-HARVESTING CHLOROPHYLL A/B BINDING PROTEIN
LHY	LATE ELONGATED HYPOCOTYL
LNK	NIGHT LIGHT-INDUCIBLE AND CLOCK REGULATED
LUX	LUX ARRHYTHMO

Mb	Megabase
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
mRNA	Messenger RNA
MSA	Multiple sequence alignment
MYB	Myeloblastosis
nt	Nucleotide
Os	*Oryza sativa*
PAS	Per-Arnt-Sim domain
PHY	PHYTOCHROME
PIF	PHYTOCHROME INTERACTING FACTOR
PRR	PSEUDO-RESPONSE REGULATOR
PR	Pseudo Receiver domain
qPCR	Quantitative real-time polymerase chain reaction
RCF1	REGULATOR FOR CBF EXPRESSION 1
ROS	Reactive oxygen species
RVE	REVEILLE
RNA	Ribonucleic acid
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
RT-PCR	Reverse transcription polymerase chain reaction
SIC	SICKLE
SEX4	STARCH-EXCESS 4
T	Thymine
TBS	TCP-Binding Site

TCP	Teosinte branched 1, Cycloidea, and PCF domain
TF	Transcription factor
TOC1	TIMING OF CAB EXPRESSION1
YUC8	YUCCA8
Zt	Zeitgeber
ZTL	ZEITLUPE

ACKNOWLEDGEMENTS

I would like to acknowledge the dedication and help provided by Dr. Olivia Wilkins, my supervisor, throughout this journey. My background in molecular biology was almost non-existent at beginning of my degree, but she still supported and guided me to achieve this research project. Thank you, Olivia, your support and trust mean a lot to me. I also want to highlight the role of my committee meeting formed by Dr. Jessica Head and Dr. Jaswinder Singh. Your advice has been very important in consolidating my knowledge and making the right choices throughout this project. I wish to express my gratitude towards all my colleagues and summer students that helped me to in a way or another achieving these experiments. Ruite Chen for his moral support and humoristic nature. Philip Rosenbaum and Vincent McCarty both for in-depth technical and moral support. Dr. Bulbul Ahmed for great technical support about the cloning experiment, general background in molecular biology and advice on graduate studies. Césarée Morier Gxyoyia for the help concerning the genotyping of the mutants and her natural "joie de vivre". Abraham Azriel Jauregui and Raùl Ernesto for their help as very supportive and motivated summer students. Jonathan Sangiovanni for his help on the phylogenetic analysis. You rock guys.

I would like to give a special thanks to my family, and especially my lovely girlfriend, Michelle Mei Lee Corbu, for her patience, positive attitude and encouragement. I know it has not always been easy for you to see me being stressed all the time and wake up during the night to work. I thank you my dear for your support and your love. I would like to express my sincere gratitude to the funding institutions that supported me through the Alexander Graham Bell Canada Graduate Scholarship (Master program, National Science and Engineering Research Council) and by the B1 Graduate Scholarship (Master program, Fonds de Recherche du Québec – Nature et Technologie). This research was funded by an NSERC Discovery Grant awarded to Dr. Olivia Wilkins on infrastructure purchased through a Canadian Foundation for Innovation John Evans Leaders Fund awarded to Dr. Olivia Wilkins.

PREFACE

All components of this thesis are the author's original work with the exception of Figure 2.1 which is reproduced from the original under the Creative Commons Attribution Non Commercial License (http://creativecommons.org/licenses/by-nc/3.0). This License permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Contribution of co-authors

All the experiments were designed by Dr. Olivia Wilkins and Jérôme Gélinas Bélanger. The experiments and data analyses were carried out by Jérôme Gélinas Bélanger. Jérôme Gélinas Bélanger wrote the thesis.

CHAPTER I: General introduction

1.1 Introduction

Forecasted changes in temperature due to climate change are predicted to be in the range of 2-3°C [1]. Globally, heat stress is a major cause of yield reduction in rice production, and developing countries are the most likely to be suffering from a generalized reduction in rice yields due to their reliance on the crop to fulfill their caloric requirements [2]. Similarly, reductions in yield mediated by climate change and increase in the global temperature are predicted for various other cultivated species such as corn [3,4], soybean [3] and wheat [3]. Heat stress affects grain production by reducing the total amount of grain and the grain-filling period. Heat stress can also affect the plant by causing several kinds of injuries, such as leaf abscission and shoot/root inhibition, and by decreasing the grain quality in rice [5]. Moreover, heat stress can affect insidiously the biochemistry of plants by causing permanent changes in protein conformation and generating reactive oxygen species (ROS) in the plant [6].

Plants have multiple physiological and molecular adaptive strategies to survive stresses caused by heat. Transcriptional regulation through gene regulatory network enables plants, as being sessile organisms, to dynamically counter the damages caused by heat stress by coordinating the timing and pacing of the molecular response at a genome-wide scale [7]. Numerous key transcriptional regulators involved in the heat stress response are also involved in the maintenance and regulation of the circadian clock. The circadian clock is an endogenous timekeeping mechanism and extensive transcriptional network playing the critical role of coordinating the molecular and physiological responses by aligning them with their changing environment [8]. This transcriptional feedback-loop system enables organisms, including rice, to timely control the transcriptional responses according to the relevant times of the day and year in order to react properly to the surrounding stimuli.

The CCA1-LHY-RVE is a family of MYB transcription factors involved in the regulation of the circadian clock [9]. The CCA1-LHY-RVE is a protein family comprising 11 genes in arabidopsis and seven proteins in rice [10]. As a general rule, the

members of this family share two main common structural features. The first one is a single MYB/SANT DNA binding domain of 61 amino acid residues with a conserved SHAQKYF motif [9,11,12]. The second one is a proline-rich domain having a conserved region of 15 characters long, with a conserved PPPRPKRK motif, thought to be the nuclear localization signal of the protein [12,13]. On a phylogenetic level, the protein family is divided in three main clades, the LHY clade, RVE8 clade, and lastly, the RVE1 clade. The RVE1 clade is interesting because it features the protein EPR1 (EARLY PHYTOCHROME RESPONSE 1) in rice and arabidopsis. EPR1 is a promising R1 MYB transcription factor thought to be the entry gate bridging the heat response gene regulatory influence network to the endogenous circadian rhythm in various plant species, including rice and arabidopsis [7]. At the present time, RNA-seq and microarray data have revealed that *EPR1* is upregulated gene upon heat stress in rice and arabidopsis [7,14,15]. This feature might be unique in the CCA1-LHY-RVE family, but more in-depth research is required to confirm this statement due to the sparse state of literature on the question.

Evolutionary speaking, the differences between the genes involved in these respective clades are fascinating on a functional level. *CCA1/LHY* genes are well-known to have a distinct daily circadian oscillation with a peak in the morning [16,17]. Those genes are known to affect the pace and regulation of hundreds of downstream genes, and are often designed as the master oscillators of the morning loop [16,17]. Genes involved in the RVE8 clade are known to act as growth inhibitors at all developmental stages [18,19]. The *RVE8* gene has been shown to be an important player in the regulation of the core oscillator genes *CCA1* and *LHY*, and is also important to maintain rhythmicity in a light- and temperature-dependent manner [11,19]. EPR1/RVE7, RVE1 and CIR1/RVE2, all members of the RVE1 clade, share a limited homology amongst each other [9]. In that sense, EPR1 homology to the two other proteins is mostly confined to the MYB DNA binding and proline-rich domains in arabidopsis, whereas the two other genes are more homologous to each other. To this day, few scientific papers have been published on the structural and functional features of the RVE8 and RVE1 clades.

These differences on the structural and functional levels are not only interesting but are also the pillars that support this thesis. Overall, this thesis revolves around two pivotal questions. The first question is to understand how the members of this family evolved on a structural basis in monocotyledonous and dicotyledonous species in their coding and promoter regions. The second question is to comprehend how the genes member of the *CCA1-LHY-RVE* family are expressed under heat stress and normal growth conditions.

1.2 Research hypotheses and objectives

Hypothesis 1. Monocotyledonous and dicotyledonous species display the same phylogenetic relationships for the three clades of the CCA1-LHY-RVE protein family, the RVE1, RVE8 and LHY clades.

Objective 1. Characterization of the expansion and diversification of the CCA1-LHY-RVE protein family in monocotyledonous and dicotyledonous species.

Hypothesis 2. The members of the *CCA1-LHY-RVE* gene family can be submitted to relative gene quantification using RT-qPCR technology.

Objective 2. Development of RT-qPCR assays for six members of the *CCA1-LHY-RVE* gene family and three housekeeping genes.

Hypothesis 3. *OsEPR1* is the only gene in the *CCA1-LHY-RVE* gene family to be differentially abundant upon heat stress.

Objective 3. Quantification of the gene expression of six members *CCA1-LHY-RVE* gene family and three housekeeping genes.

CHAPTER II: Literature review

2.1 Rice: a staple crop for food security and science

2.1.1 A pillar for food security and a model plant for monocots in biological studies

Asian rice (*Oryza sativa)* is one of the most important staple crops globally, and is the primary source of calories for more than 30% of the world population [20]. In terms of global production, rice is the third most highly-produced crop after sugarcane and maize [21]. However, in regard to human nutrition and caloric intake, rice has the lead over any other plants. Rice cultivation is well-suited to countries with low labor costs and many developing countries are highly dependent on the crop for their food security [22]. It is a versatile and resilient crop that is produced in a wide range of agroecosystems including deep water, upland, rice paddies and lowland rainfed fields in temperate and tropical areas.

Rice is the model cereal. It was chosen because of its relatively small genome size, ease of transformation and its close evolutionary relationship with other widely cultivated cereal grains, such as wheat, barley, and maize [23,24]. Additional physical characteristics, such as a rapid life cycle and a small plant stature, also all contributed to establishing rice as a model for the study of monocots. The rice genome shares a significant synteny with other important cereals, a feature enabling scientists to perform systematic genetic comparisons between species to build on evolutionary conservation analysis and efficient breeding schemes [5,25]. Rice has a larger genome size (Os, 4.3 Mb; At, 1 Mb), number of chromosomes (Os, 2n =24; At, 2n = 10), and generation time (Os, 3-6 months; At, 1-2 months) than the model dicotyledonous plant, *Arabidopsis thaliana* [25]. While it is genetically more complex than arabidopsis, rice has a much more tractable genome than other commercially important members of the *Poaceae* family (grass family), such as bread wheat (*Triticum aestivum*; 2n = 6x = 4, ~17 Mb) and durum wheat (*Triticum turgidum* L. subsp. *durum*; 2n = 4x = 28, ~12 Mb) [26,27].

2.2 Defining features of the circadian clock

2.2.1 An introduction to the circadian clock

The circadian clock is a biological oscillator with a roughly 24 h period that is endogenous to all organisms [28,29]. The clock has been studied in many plants, including poplar, papaya and rice, but arabidopsis remains by far the most studied system for unravelling the clock's structural and functional features [29]. The clock is a transcriptional gene regulatory network which consists of numerous transcription factors that regulate the expression of targets as well as, in some cases, their own expression [18]. The proper functioning of the clock is a *sine qua non* condition to plant fitness as clock-arrhythmic mutants have lower fitness than their wild type counterparts in competition experiments and display irregular phenotypes, such as delayed flowering, decreased chlorophyll synthesis and arrhythmic stomatal opening [8,30,31]. Overall, it is estimated that one third of the arabidopsis transcriptome is directly controlled by the clock [9,18]. Furthermore, in arabidopsis, about 90% of the genes have been demonstrated to be rhythmically expressed, either under circadian regulation (free run) or diel regulation (entrained by cycling environmental cues such as a photocycle or thermocycle) [32]. Overall, the clock comprises more than 20 genes encompassing seven classes of transcription factors [33] (Table 2.1).

The circadian clock has five defining features: 1) it can be entrained to the surrounding rhythmic environmental changes, 2) it is an endogenous mechanism, 3) it has a free-running rhythm (an approximately 24-h rhythm in constant environmental conditions), 4) it can be phase-shifted due to environmental cues, the zeitgebers (German for "time givers"), and 5) it features a temperature compensation mechanism which permits the clock to maintain its period across a range of temperatures [28]. The clock is synchronized to prevailing environmental cues, such as the day-night cycles of light and temperature, through a process called entrainment [34]. Entrainment optimizes the timing of physiological and metabolic behaviours relatively to the environmental conditions [35].

2.2.2 The mechanism of the circadian clock

The clock consists of a central loop and two interlocked transcription-translation feedback loops, the morning loop and the evening loop [36] (Fig 2.1). The morning and evening loops are named for the time of the day when their components are the most expressed, whereas the central loop acts as the main coordinator for all the circadian-regulated genes in a repressilator fashion [37]. The morning loop is comprised of the *AtCCA1*, *AtLHY* and *AtTOC1* genes, three genes expressed from the early morning to the evening [38]. The sequential binding by AtPRR members throughout the day mediates the repression of the complex forming the morning loop [38]. The succession in expression of the *PRR* gene family, from the morning to the night, follows this order: *AtPRR9*, *AtPRR7*, *AtPRR5*, *AtPRR3*, and then *AtTOC1* [39]. In the central loop, also called the second morning loop, AtCCA1 and AtLHY accumulate at dawn to repress *AtTOC1* [40]. The evening loop, or Evening Complex, is a complex formed by AtELF3, AtELF4 and AtLUX which transcriptionally represses itself through a transcriptional feedback loop [41].

AtCCA1 and *AtLHY* transcript abundance peak at dawn [42,43]. Both proteins activate the transcription of the *AtPRR7* and *AtPPR9* genes, two early-late morning genes [36,44]. The AtRVE8 protein, which has peak protein abundance in the mid-afternoon, regulates the transcription of hundreds of target genes, including the core clock components *AtTOC1*, *AtGI*, *AtLUX*, *AtELF4*, *AtPRR9* and *AtPRR5*, through direct binding to the conserved Evening Element in their promoter regions [19]. In a similar fashion, three PRRs, AtPRR5, AtPRR7 and AtPRR9, repress *AtRVE8* [19]. In that sense, *AtRVE8* role is the opposite than *AtCCA1* and *AtLHY* [19]. Moreover, *AtPRR7*, *AtPRR9* and the afternoon-expressed *AtPRR5* are gradually upregulated during the day and gradually repress *AtCCA1* and *AtLHY* [45–47]. This repression is partly mediated by the binding of AtPRR5 to *AtTOC1*, the latter being a transcriptional repressor of AtCCA1 and AtLHY [45–47]. *AtTOC1* is upregulated in the evening [9]. This action of AtPRRs and AtTOC1 constitute the core clock negative transcriptional feedback loop [45–47]. In the evening, AtZTL targets AtTOC1 and AtPRR5 proteins in the cytosol and degrade them via the proteosome pathway [48–50]. AtGI, an evening-expressed gene, interacts

with *AtZTL* upon blue-light stimulation, and provides AtZTL its post-transcriptional oscillation [36,51,52]. *AtGI* oscillation is clock-controlled, and it is this cycling that provides a circadian oscillation to AtZTL as the latter has a constitutive transcription pattern [36,51,52].

AtLUX, *AtELF3* and *AtELF4*, are upregulated during the evening phase and repress *AtPRR9*, a day-phased gene [36,53,54]. The roles of *AtRVE4* and *AtRVE6*, two homologs of *AtRVE8*, are partially redundant to *AtRVE8*, and these genes have been demonstrated to interconnect the morning and evening loops [19]. AtCHE is involved in the central loop and the transcriptional repression of AtCCA1 [36,55]. The transcriptional repression of *AtCCA1* is mediated through the recruitment of AtTOC1 by AtCHE in a transcriptional complex at the *AtCCA1* promoter [36,55].

2.2.3 Beyond arabidopsis: the circadian clock in other angiosperms

First evidence of the circadian rhythm at the molecular level happened concomitantly with the emergence of arabidopsis as a model organism [29,56–58]. Evolutionary conservation is estimated to be strong for the structure, architecture and functions of the circadian clock especially within the components encoding MYB domain transcription factors such as AtCCA1, AtLHY, and AtPRRs [29,59–61]. Contemporary advances in genome sequencing facilitated some breakthroughs with other plant species, thus enabling scientists to demonstrate that AtCCA1, AtLHY and/or AtPRRs homologs are present in *Brachypodium distachyon* [62,63], *Brassica rapa* [64–66], rice (*Oryza sativa*) [63,67–69], grapevine (*Vitis vinifera*) [70], papaya (*Carica papaya*) [71–73], bean (*Phaseolus vulgaris*) [74,75], chestnut (*Castanea sativa*) [76,77], pea (*Pisum sativum*) [78,79], *Lemna gibba* [80,81], soybean (*Glycine max*) [82–84], tomato (*Solanum lycopersicum*) [85,86], and poplar (*Populus spp.)* [63,71,87] [29,88]. A common ancestor in monocots and eudicots had only one copy of the *AtCCA1*/*AtLHY* gene, but poplar and arabidopsis undergone independent duplication events for this gene in their evolution [87,88]. One of the limiting factors hindering the extrapolation of the model plants poplar and arabidopsis data for the specific function of *AtCCA1*/*AtLHY* genes are these independent duplication events. An additional factor is the lack of functional data in non-

model species which still precludes the complete validation and transposition of the arabidopsis model to those plants even for the well-characterized core genes *AtLHY* and *AtCCA1*.

Rice is one exception to this statement as rice clock research is more advanced than in most other plant species, except arabidopsis. Two components mainly differentiate the state of rice research compared to other plant species [29]. The first component is the exhaustive literature providing a functional assessment of the roles of several clock members [67,68,89]. The second component is an extensive literature analyzing the clock's behaviour under field conditions [90–93]. Phylogenetically, rice clock composition is similar to arabidopsis for the clock core as well as peripheral members; RVE family (Os, 7; At, 11), PRR family (Os, 5; At, 5), ELF3 (Os, 2; At, 2), ELF4 (Os, 3; At, 5), LUX (Os, 1; At, 2), GI (Os, 1; At, 1) and ZTL (Os, 3; At, 3) [10]. Measurement of mRNA expression profiles of rice clock members *OsCCA1*/*OsLHY*, *OsZTL*, *OsLUX* and *OsPRRs* revealed that those are similar to their arabidopsis counterparts [67,68,94]. Overexpression studies for *OsPRR1* and *OsZTL* generated similar effects than their overexpressed homologous counterparts in arabidopsis [67,80]. Research on the functional roles of the five *PRR* genes in rice (*OsPPR1*, *OsPRR37*, *OsPRR59*, *OsPRR73*, and *OsPRR95*), performed by respectively transforming *Attoc1* and *Atprr7* arabidopsis loss-of-function mutants with the orthologous genes *OsPRR1* and *OsPRR37*, has demonstrated that *OsPRRs* and *AtPRRs* are, at least partially, genetically interchangeable [67]. Altogether, these evidences demonstrate that specific clock-associated functions are evolutionary conserved between rice and arabidopsis for those genes.

The physiological significance of the circadian clock under field conditions has been investigated in several studies using rice as a model plant, and many of these field experiments have been specifically tackling the global regulatory role of OsGI in the rice transcriptome. AtGI is an interesting protein which has been suggested to bridge the circadian clock to the *AtCO-AtFT* flowering pathway by facilitating the degradation of a family of CYCLING DOF FACTOR (CDF) transcriptional repressors, a family known to activate the expression of *AtCO-AtFT* in relation to ambient temperature and photoperiod

[95,96]. CDFs act downstream of AtGI to repress the *AtCO-AtFT* module and influence the plant response to freezing temperature and growth, but are not required for the maintenance of the clock function [96]. In a time-course transcriptome analysis using a *Osgi* null mutant, it was demonstrated that OsGI controls 75% (false discovery rate of 0.05) of genes amongst the 27,201 studied genes [91]. In this study, OsGI was required to the clock to maintain strong amplitudes in the diurnal rhythm phases of global gene expression in field-grown rice plants [91].

Under laboratory and natural conditions, in *Osgi* mutants, *OsLHY* expression decreases and consequently does not bind to *OsPRR1* [91]. In another time-course global transcriptome analysis using leaves from wild-type and *Osgi* rice mutants, OsGI has been demonstrated to regulate the gene expression of more than half the studied genes [93]. Furthermore, in the same study, *Osgi* mutants have been revealed to exhibit defective transcriptome rhythms upon strong diurnal changes in environmental inputs [93]. Gene expression measured using atypical light/dark conditions demonstrated that the photoperiod of the day influences the rate of *OsGI* transcription [97].

2.3 The CCA1-LHY-RVE protein family in arabidopsis: A family of MYB transcription factors

2.3.1 The CCA1-LHY-RVE protein family

The CCA1-LHY-RVE protein family comprises members of the core clock, namely AtCCA1, AtLHY and AtRVE8, and the peripheral clock, namely AtRVE1-7 and AtRVE7-like [18] (Table 2.2). To this day, the CCA1-LHY-RVE protein family has been structurally characterized in only three plant species; arabidopsis, rice and tomato [18]. In total, there is eleven, nine and six proteins in the CCA1-LHY-RVE protein family encoded in the arabidopsis, rice and tomato genomes, respectively [18]. This family is relevant to research investigating the plant heat stress response as one member, *AtRVE7*/*AtEPR1*, has been revealed to be regulated by heat stress, whereas other members are involved in the cold stress response, as in the case of *AtRVE1* [7,98]. Many

members are also involved in the regulation of downstream targets playing a role in the temperature compensation and thermomorphogenesis mechanisms [99].

All members of this family are single repeat MYB domain-containing transcription factors which share a conserved proline rich-region predicted to act as a nuclear localization signal [11]. Along with *AtCCA1* and *AtLHY*, some members of this family have been demonstrated to oscillate at the mRNA level as well as the protein level [100]. Sub-clustering of the CCA1-LHY-RVE protein family divides it into three distinct groups, the RVE1, RVE8 and LHY clades, based on their respective conserved amino acid sequences homology for the MYB and proline-rich domains [18]. Another sub-clustering system, involving all the members of this family, segregates these genes into two groups, the CCA1 and LCL clades [101]. The LCL subgroup is defined by a conserved region located at the C-terminal end of the proteins [101]. The literature on the LCL clade is not abundant and its members have not been thoroughly characterized as most of the work has been only performed on RVE8, the latter being the model gene for this subgroup [101]. Therefore, characterization of the putative functions of the LCL subgroup are mostly restricted to what has been observed with RVE8. Abundance in the literature is slightly more important for the CCA1-LHY-RVE protein family, although it is not shared uniformly amongst its members. The LHY clade members, CCA1 and LHY, have been functionally and structurally characterized in arabidopsis and in other species, but less is known about the other transcription factors in this group.

Members of the MYB protein family are characterized by a conserved MYB DNA-binding domain [102]. MYB transcription factors are classified into four groups, 1R-MYB, 2R-MYB, 3R-MYB and 4R-MYB, based upon the number and position of MYB repeats in their sequences [102,103]. MYB transcription factors are one of the largest family of genes in plants, and also play important roles in plants such as the regulation of meristem formation, floral and seed development, organ morphogenesis, chloroplast development, cell cycle control, cell fate and identity, primary and secondary metabolism, and light and hormone signalling pathways [103,104]. Moreover, MYB genes also play a crucial role in the regulation of the circadian rhythm and stress responses [103,104]. First found in maize, MYB transcription factors are now well

characterized in various plants, including the model plants arabidopsis and rice, but also grapevine, cotton, poplar and apple [102,103,105]. In rice and arabidopsis, approximately 47.74% and 90.86% of the MYB genes were demonstrated to be diurnal/circadian-regulated [102]. In research tackling rice and arabidopsis promoter regions of MYB genes, no common motifs between the two species have been found which is suggesting an evolutionary divergence in their promoter regions between these two species [102].

2.3.2 LHY clade

The LHY clade is composed of AtCCA1 and AtLHY, two homologous components sharing overlapping core circadian regulatory functions and high rates of sequence homology [106–108]. This homology is related to a gene duplication event in arabidopsis that gave rise to another *AtCCA1*/*AtLHY* copy [106–108]. To repress the expression of evening-phased genes, these proteins need to heterodimerize and bind to the Evening Element, AAAATATCT, an occurring cis-regulatory motif all over the plant kingdom [106,107]. In the clock, AtCCA1 and AtLHY are co-repressors of *AtTOC1* and inhibit its expression until the evening by binding to the Evening Element motif located in its promoter [109,110]. Additionally, again by binding to Evening Element motif, the AtCCA1 and AtLHY transcriptionally repress members of the *PRR* family, such as *AtPRR5* [19]. According to the literature, the main binding target of AtCCA1 and AtLHY seems to be the Evening Element. In arabidopsis, approximately 5056 putative target genes have been identified to have the Evening Element motif in a region 1-kb upstream of their coding regions [111]. Amongst these 5056 putative target genes are many genes involved in the arabidopsis heat stress response, such as *AtDREB2C*, as well as many other types of abiotic stress responses [112].

Differential expression patterns and molecular roles exist between *AtCCA1* and *AtLHY* at different temperatures [108]. AtCCA1 is considered to have a clock buffering function at lower temperature, whereas this role is replaced by AtLHY at higher temperature [108]. At higher temperature, AtLHY and AtGI are thought to maintain a dynamic balance together to keep an accurate rhythmicity of the clock, whereas AtCCA1 replaces AtLHY for this role at lower temperature [113]. AtCCA1 is also involved in the

multiple other pivotal physiological functions, such as the flowering process and the stomatal aperture. Stomatal aperture and flowering are core physiological responses imparting pivotal biochemical and molecular changes in the plant heat stress response. *AtCCA1* and *AtELF3* genes act together through the *AtCO-AtFT* pathway to control the flowering time, and *AtELF3* acts downstream of *AtCCA1* to inhibit the action of *AtPIF4* [40]. Overexpression of *AtCCA1* is known to cause an increase in expression of *AtPIF4* and *AtPIF5*, and recent studies uncovered the involvement of *AtPIF4* in the heat-mediated morphological acclimation and acceleration of flowering [9,114].

2.3.3 RVE8 clade

The RVE8 clade comprises five genetically similar genes in arabidopsis, *AtRVE3*, *4*, *5*, *6* and *8* [11]. AtRVE3 and AtRVE5 have been demonstrated to slightly regulate elements of the core clock, and AtRVE4 and AtRVE6 to shorten the clock pace [18]. Furthermore, the AtRVE4 and AtRVE6 transcription factors have been reported to have redundant roles with AtRVE8 in the clock by interconnecting the morning and evening loops [36]. The AtRVE8, also identified as LHY-CCA1-LIKE5 (AtLCL5), transcription factor is the most studied protein of the clade, and is known to be an important regulator of the core clock oscillator as well as a player in the light signalling mechanism [101,115]. AtRVE8 is required for temperature compensation as high temperatures are known to shorten and lengthen periodicity of overexpressed lines and rve8 mutants, respectively [11]. *AtRVE8* circadian transcription profile is very similar than to *AtCCA1* and *AtLHY* [11]. However, its protein accumulation pattern differs from AtCCA1/AtLHY as the former has its protein accumulation peaking in the subjective afternoon and the latter in the near subjective dawn [11].

AtCCA1 and AtRVE8 regulate the transcription of *AtTOC1* differently despite having similar protein structures and morning phase expression [101,115]. AtCCA1 represses *AtTOC1* transcription by promoting histone deacetylation in the AtTOC1 promoter; in contrast AtRVE8 is associated with a hyper-acetylated state of the AtTOC1 promoter, which may facilitate transcription of *AtTOC1*. Overall, AtCCA1 influence on

AtTOC1 remains greater than the one of AtRVE8, and their combined actions are thought to fine tune the transcription of *AtTOC1* [101].

Again, the modulation in expression of AtRVE8 target genes, such as *AtCCA1*, *AtLHY*, *AtTOC1*, but also *AtPRR5*, *AtGI*, *AtLUX*, and *AtELF4*, is mediated through the binding of AtRVE8 to the Evening Element cis-regulatory motif in their promoters [36]. It has been suggested that AtRVE8 could be a key protein in the central night hub, a module of the night gene regulatory network with a high number of connections [37]. AtRVE8 is thought to play this role because its targets are enriched for the evening phase (AtRVE8-induced ; *AtTOC1*, *AtELF4*) and morning phase (AtRVE8-repressed; *AtCCA1*, *AtLHY*, *AtRVE8*) [37]. A small family of morning-expressed genes, comprising the *NIGHT LIGHT-INDUCIBLE AND CLOCK REGULATED 1*, *2*, *3* and *4* (*AtLNK 1*, *2*, *3* and *4*) genes and playing an important role in the circadian oscillator, has been revealed to interact with the members of the RVE family [116–118]. In that sense, *AtRVE8* is either antagonized or coactivated by AtLNKs [116–118]. The literature is less abundant for the other members of the clade. Nonetheless, it has been demonstrated that triple rve4 rve6 rve8 mutants have an impaired circadian rhythm with a four hours longer delay than that of wild-type plants which suggests that those genes are involved into switching on the afternoon/early evening genes, and that they subsequently activate the expression of morning genes [19].

2.3.4 RVE1 clade

The RVE1 clade is comprised by the *AtRVE1*, *AtCIR1*/*AtRVE2*, *AtEPR1*/*AtRVE7* and *AtRVE7-like* clock-regulated genes, and their peak transcript abundance are near subjective dawn like *AtCCA1* and *AtLHY* [9]. The genes comprising this clade are not core regulators of the clock machinery, but rather have clock-regulated outputs [9,11]. Genes in the RVE1 clade are all either directly or indirectly repressed by AtCCA1 [31]. Amino acid homology amongst AtRVE1, AtCIR1 and AtEPR1 is limited outside the MYB and proline-rich domains as their sequences are only 30% identical and 40% similar [9].

2.3.4.1 EPR1

EPR1 1 is a DNA-binding R1 MYB transcription factor that binds to the canonical Evening Element in rice and arabidopsis [7,63]. *EPR1* has two predicted splice variants in rice and arabidopsis. The variant *OsEPR1.2* lacks the characteristic SHAQKYF motif, but this motif is present in the arabidopsis splice variant *AtEPR.2*. *AtEPR1* and *AtRVE7-like* are two homologous duplicate genes, the latter being specific to *A. thaliana* [9]. In comparison to *AtEPR1*, *AtRVE7-like* is an asymmetrically expressed gene; the copy is expressed at lower levels in all the tissues examined [119]. Accordingly, AtEPR1 shares a 70% sequence homology to the MYB domains of the core clock elements AtCCA1 and AtLHY [42]. Constitutive expression of AtEPR1 represses the *LIGHT HARVESTING COMPLEX B* (*AtLHCB*) gene, a gene coding for a chlorophyll a-b binding protein, without influencing the central oscillators, *AtCCA1* and *AtLHY* [115,120]. AtEPR1 can regulate its own expression, probably by binding to the Evening Element motif located in its promoter [120].

Increased levels of AtEPR1 protein influence downstream phytochrome processes [37]. AtEPR1 is predicted to interact with *AtPIF5* at night and with *AtPIF4* during the day [37]. *AtEPR1* is thought to be the hub connecting the day and night gene modules in arabidopsis, and is regulated by photoreceptors PHYTOCHROME A (AtPHYA) and PHYTOCHROME B (AtPHYB) through far-light perception [37,120]. Studies have demonstrated that, through a low fluence response and using *Atphya* seedlings, AtPHYA modulates the induction of *AtEPR1* expression by far red light [120]. AtEPR1 also interacts with *AtPRR7*, a core-clock component, embedded in a network comprised of AtHSFB2B in which *AtPRR7* is a downstream target [37,44]. AtHSFB2B represses *AtPRR7*, and AtPRR7 plays a role in the thermosensory based regulation of many core and peripheral clock genes, notably *AtCCA1*, *AtLUX*, *AtPIF4*, *AtPIF5*, *AtRVE2*, *AtEPR1* and itself [44]. AtEPR1 is at the centre of a control network including AtCCA1, AtRVE8, STARCH-EXCESS 4 (AtSEX4), and itself [37]. The nexus between AtEPR1 and AtCCA1 interaction is that *AtEPR1* harbours three AtCCA1 recognition motifs in its promoter region [37].

In arabidopsis, overexpression of *AtEPR1* causes a slightly altered photoperiodic

flowering response and induces the opening of cotyledons upon far-red light [120]. *EPR1* mRNA abundance in arabidopsis and in rice is increased by heat stress which suggests that *EPR1* could be an entry gate for the integration of heat stress response with the circadian clock [7,14]. Data available from comparative abiotic stress studies show that *AtEPR1* is only differentially abundant by heat stress and not by any other type of abiotic stresses [14]. This heat stress response mechanism is thought to be related to the presence of a Heat Shock Element upstream of its transcriptional start site [7,14]. Strangely, the average expression of 107 putative targets of OsEPR1, which are all harbouring an Evening Element motif in their promoter regions, is unchanged after heat stress exposure, and thus, maintain their constitutive circadian oscillation patterns.

2.3.4.2 CIR1

CIR1, also known as RVE2, is a transcription factor which regulates itself and as well as the central oscillator [121]. AtCIR1 shares > 70% sequence identity with the MYB domains of AtCCA1, AtLHY and AtEPR1 in arabidopsis [121]. AtCIR1 displays very little homology with AtEPR1 and the central oscillators AtCCA1/AtLHY outside their proline-rich and MYB regions [121]. AtCIR1 regulates the expression of *AtCCA1* and *AtLHY* by diminishing their expression amplitudes, and also shorten *AtTOC1* rhythmicity [121]. It also modifies the rhythmic expression of *AtEPR1* [121]. In arabidopsis, *AtCIR1* overexpression phenotypes are numerous; delayed photoperiodic flowering, increased hypocotyl elongation, and inhibited seed germination [121]. At the molecular level, AtCIR1 reduces the expression of *AtCO* and *AtFT* [121]. AtCIR1 has been suggested to act as a negative regulator of *ENHANCED DISEASE SUSCEPTIBILITY 1* (*AtEDS1*), a gene involved in the salicylic acid-signalling defence network, through post-transcriptional regulation [122]. AtCIR1 has been demonstrated to be embedded into the pathogenic stress resistance network as a negative regulator of disease resistance, and is consequently operating upstream of the salicylic acid, jasmonic acid and ethylene accumulation pathways in *A. thaliana* [123]. In-depth characterization of a cold hypersensitive mutant for the *REGULATOR FOR CBF EXPRESSION 1* (*AtRCF1*) gene, an encoding a cold-inducible DEAD (Asp-Glu-Ala-Asp) RNA helicase

involved the proper splicing of mRNA, revealed that *AtPRR5* and *AtCIR1* are, respectively, cold-inducible negative and cold-inducible positive [124].

2.3.4.3 RVE1

Amino acid sequences of AtRVE1 and AtCCA1 are 80% identical and 94% similar within their MYB-like region [9]. AtRVE1 has been shown to influence the expression of *AtCCA1*, *AtGI* and *AtLUX* [9]. Indeed, in overexpressed *AtRVE1* plants, the expression of these three genes has been shown to be low, and the circadian rhythm to be arrhythmic [9]. *AtRVE1* expression does not affect the pace of the core clock mechanism [9]. *Atrve1*, *Atcir1*, and *Atepr1* single mutants do not display any modifications in the regulation of members of the core clock [9]. Studies demonstrated that AtRVE1 role is mainly confined to the one of a clock output that affects the hypocotyl growth by mediating the circadian regulation of the auxin pathway in arabidopsis [9]. AtRVE1-mediated growth control is performed independently of *AtPIF4* and *AtPIF5*, two genes related to hypocotyl elongation [9]. AtRVE1 plays a significant role in increasing auxin levels and promotes the accumulation of free auxins in the plant [9]. AtRVE1 mediation of the auxin pathway is performed by binding to and subsequently upregulating the downstream target *YUCCA8* (*AtYUC8*) [9]. This action is tissue-specific in the root tip and hydathodes, two areas with active auxin signalling and associated to *AtYUC8* expression [125,126]. AtRVE1 acts as a negative regulator of cold acclimation, and studies have shown that *AtRVE1* circadian oscillation fades when subjected to cold acclimation [98]. When not acclimated, the gene expression of *AtRVE1* was at its highest in the morning hours [98]. Overall, downregulation of *AtRVE1* in all the recombinant inbred lines was associated with high freezing tolerance levels [98]. A significantly higher freezing tolerance was exhibited by *Atrve1* knock-out mutants when subjected to cold temperature, but no visible phenotype was identified when these mutants were grown under typical greenhouse conditions [98].

2.4 The circadian clock and temperature: Temperature compensation, thermal entrainment and thermomorphogenesis

2.4.1 The consequences of heat stress on plants

Climate change is expected to increase temperature by 2-3°C by 2050, and this increase in temperature could have disastrous impacts on food security, including for people living tropical countries, due to the omnipresence of rice in their diet [1]. In nature, drought stress events are habitually related to heat stress, and the specific consequences of each stress on the physiological status of the plant in a time-specific manner are often hard to distinguish. In consequence, drought stress is often tackled in tandem with heat stress in the literature. However, heat stress alone can be a major cause of multifarious alterations leading to quantitative and qualitative yield reduction in rice. Therefore, it is important to address the specificity of heat stress consequences on the plant physiological and molecular statuses in an independent fashion.

The physiological consequences of heat stress alone on cereal plants are multiple and well addressed in the scientific literature. At the physiological level, heat stress alone has been shown to cause a reduction in grain number [127,128], a decrease in the duration of the grain-filling period [127], a reduction of plant growth [6], and an increase in water usage [6]. It also has been shown to inflict tissue senescence [6], pollen sterility [5], ovule abortion [5], and an alteration in the phenological development for the plant [6]. Heat stress can also change the biochemistry by modifying protein conformation, inducing the formation of ROS, perturbing the photosynthetic process, and causing male sterility in crop plants [129]. Furthermore, the effects of heat stress on the molecular architecture of plants are also well covered in the scientific literature concerning the transcriptional landscape of their endogenous circadian rhythm [99,130–132]. Rice is particularly sensitive to heat at the heading and reproductive stages as it can induce spikelet sterility, reduce grain quality (length, width and weight), and reduce the viability of pollen [133,134]. In this aspect, night temperature has a critical importance on total yield production, and data compiled at the International Research Rice Institute from 1979 to 2003 indicated an average reduction of 10% in yield for an increase of 1.13°C during the night and 1°C in the general growing-season temperature [133,135]. Similarly,

high night temperature can result in detrimental changes in the grain biochemistry by increasing the rate of grain filling and decreasing of grain filling matter which is inevitably leading to lower yields [134]. This reduction in yield could have dramatic effects on long-term food security as a 1% annual yield increase will be required to supply the growing world demand [20,136]. Similar reductions are expected for other globally important cereals, such as wheat, corn, and barley.

2.4.2 Temperature compensation

Temperature compensation is defined as the "compensation of the endogeneous rhythm against changes in ambient temperature from one value to another constant value" [137]. Multiple clock components have been shown to be involved the temperature compensation mechanism. Core clock elements, including the genes *AtTOC1*, *AtCCA1*, *AtLHY*, *AtPRR7*, *AtPRR9*, the Evening Complex genes (*AtELF3*, *AtELF4* and *AtLUX*), and the clock peripheral gene *AtGI* have been demonstrated to be involved in this process in arabidopsis [113,138,139]. Heat-induced alternative splicing of clock genes *AtELF3*, *AtTOC1*, *AtPRR7* and *AtCCA1* has also been proposed to be an important factor contributing to this mechanism [38,140]. However, in most cases, the specific role of each splice variant has not been investigated for the temperature compensation response, a factor contributing to our poor understanding of its intricate mechanics.

In the *Atprr7* and *Atprr9* mutants, it has been demonstrated that *AtCCA1* and *AtLHY* expression patterns are hyperactivated, thus triggering temperature overcompensation [138]. Inactivation of *AtCCA1* and *AtLHY* has been shown to fully suppress the temperature overcompensation defects of *Atprr7* and *Atprr9* mutants [138]. *AtTOC1* and *AtGI* RNA expression increase along with temperature from 17 to 27°C in arabidopsis [113]. At 27°C, a dynamic balance between *AtGI* and *AtLHY* is obtained in wild-type arabidopsis plants, and this state of equilibrium enables wild-type arabidopsis to reach optimal temperature compensation [113]. The dynamic balance is generated by AtGI regulating *AtCCA1* and *AtLHY* in a temperature-dependent manner [113]. Accordingly, at 12°C, *Atgi* mutation reduces the expression of *AtCCA1* rather than *AtLHY* [113]. At low temperature, *Atcca1* loss-of-function mutant rhythmic amplitude

diminishes more than the *Atlhy* loss-of-function mutant [113] The reverse, a more diminished *Atlhy* loss-of-function mutant amplitude, has been also demonstrated for a temperature of 27°C, consequently confirming that *AtCCA1* replaces *AtLHY* in the temperature mechanism of the clock at low temperature [113]. *AtRVE8* is a core clock gene involved in the temperature compensation mechanism as *Atrve8* mutants have been demonstrated to have impaired temperature mechanism [11]. *AtRVE8* overexpressed plants displayed the same free-running period than the wild-type at 12°C, but shorter free-running rhythms at 17, 22 and 27°C [11]. *AtRVE8* loss and gain-of-function mutants demonstrated a normal rhythmicity only at 12°C, the lowest temperature assayed [11].

Several other transcription factors, such as FLOWERING BASIC HELIX-LOOP-HELIX 1 (AtFBH1) and HEAT SHOCK FACTOR B2b (AtHSFB2B), are also involved in the temperature compensation mechanism [44,141]. Although these genes are not clock members, their regulatory action is embedded within a network comprising circadian clock genes [44,141]. AtHSFB2B is a repressor binding to the Heat Shock Element cis-regulatory module in the promoter of the morning clock gene *AtPRR7* [44]. *AtHSFB2B* is a circadian gene with a loss-of-function mutant exhibiting impaired temperature compensation upon high temperature [44]. AtHSFB2B mediates temperature-dependent growth, and *AtHSFB2B* overexpressed plants display elongated hypocotyls and late-flowering features [44]. In a similar fashion, it has been demonstrated that AtFBH1 forms a feedback loop with AtCCA1 *in vivo* and thus, plays a clock regulator role [141]. At high temperature and when *AtFBH1* is overexpressed, *AtCCA1* rhythmicity is hindered as plants grown at 28°C display a shorter period length for this gene [141]. This suggests that AtCCA1 ability to compensate temperature effectively is compromised upon those conditions.

Multiple downstream physiological processes are modulated each time the clock's temperature compensation mechanism is required. Amongst all these processes, the regulation of stomatal aperture through ABA-mediated signals in guard cells is one of the most effective ways to limit plant transpiration and, thus, buffering heat stress [142]. Stomatal aperture is mediated through a complex sensing of multiple biotic and abiotic factors including temperature, light quality and intensity, CO_2 level and water availability

[143]. Additionally, stomatal aperture is a physiological process known to be regulated by the circadian clock, notably by the gene *AtCCA1* [30]. The role of *AtCCA1* in stomatal aperture has been recently uncovered by using overexpressed *AtCCA1* arabidopsis plants under the control of the guard cell specific promoter GC1 [144]. *AtCCA1* is involved in the circadian mediated anticipation of stomatal opening as overexpressed *AtCCA1* plants lost their capacity to timely open their stomata based on the light/dark changes [144].

The Evening Complex is involved in the temperature and drought responses in arabidopsis by regulating the *DEHYDRATION-RESPONSIVE ELEMENT-BINDING 2A* and *2B* (*AtDREB2A* and *AtDREB2B*), two *ETHYLENE RESPONSE FACTOR/APETALA2* (*ERF/AP2*), and *PHYTOCHROME INTERACTING FACTOR4* (*AtPIF4*) genes [145,146]. In rice and in arabidopsis, upregulation of *AtDREB2A* is mediated upon heat shock stresses [147]. The Evening Complex regulates the coordination of response to surrounding environmental cues by achieving specific target binding through a combinatorial mechanism in which both LUX Binding Site (LBS) and G-Box motifs are required at the binding sites [146]. In arabidopsis, it has been shown that the expression of the C-REPEAT BINDING FACTOR/DEHYDRATION-RESPONSIVE ELEMENT-BINDING (CBF/DREB) family of transcription factors is gated by the circadian clock [148,149]. The cold-inducible C-repeat (CRT) AtCBF1/AtDREB1b transcription factor is a regulator of *AtLUX* by binding to the CRT motif present in its promoter during freezing stress [150]. CBF/DREB transcription factors are involved in the ABA-independent pathway which mediates stomatal closure by reducing potassium (K+) exchange to the guard cells when the water potential decline [149]. Similar results achieved on soybean, *Glycine max*, suggest that a complex interaction between the circadian clock, light and temperature regulate the expression of CBF/DREB genes to coordinate the general response to the daily water fluctuation through ABA channelling [82].

2.4.3 Thermal entrainment and thermomorphogenesis

While temperature compensation aims to render the clock largely insensitive to temperature shifts, thermal entrainment aims to coordinate the clock to variations in temperature [151,152]. Thermal entrainment can be induced even under small temperature differences (Δ4°C), and is used by plants to sense daily fluctuations in the day-night cycles as well as seasonal changes [153]. It has been proposed that temperature compensation is a requirement to maintain stable phase relationship under different ambient temperatures in temperature entrainment [154]. Consequently, this complexity in segregating these two mechanisms results in the underlying molecular mechanism governing thermal entrainment remaining largely unknown and elusive. An additional hindrance concerning thermal entrainment resides in the inherent experimental challenge that represents identifying an efficiently entrained circadian rhythm with a temperature difference of only 4°C or less [152,155]. In plants, efficient thermal adaptation process as a wide impact on the biochemical and morphological features by virtually modifying all those features, and this factor adds an extra layer of difficulty to identify the core genes triggering this mechanism [152].

Over the years, several thermosensors have been proposed to be involved in the clock's thermal entrainment mechanism although none has been clearly identified for the moment [152]. AtPHYB is a thermosensor which has been proposed to integrate the temperature information to the clock, including the thermal entrainment stimuli [156,157]. Temperature sensing and temperature-induced alterations within the DNA-nucleosome dynamics, two mechanisms conditional to thermal entrainment, are thought to be, at least partially, mediated by AtPHYB [156,157]. Three core clock genes, *AtELF3*, *AtPRR7* and *AtPRR9*, are involved in temperature entrainment [138,158,159]. Both *Atelf3-1* mutant seedlings and *Atprr7-3 Atprr9-1* double mutants are unable to exhibit entrainment by temperature cycles in darkness conditions and upon thermocycles of 22°C/12°C, respectively [99,138,158,159]. Research has also demonstrated that *AtPRR7* is regulated by a novel allele of the *SICKLE* (*AtSIC*) gene [160]. Consistent with an impaired temperature compensation mechanism, *Atsic* mutants exhibit low-amplitude and arrhythmic expression of core clock transcription factors when grown under cool

ambient temperature [160]. Moreover, transcript accumulation of *AtLHY* and *AtCCA1* splice variants was increased in *Atsic* mutant alleles, indicating a potential influence on the core clock [160].

In the case of heat stress response, the function of the clock is to integrate the environmental inputs to shape modulated output pathways that mediate either temperature entrainment, temperature compensation and/or temperature-dependent growth [153,161]. Thermomorphogenesis, also known as temperature-dependent growth, is a feature enabling a plant to reduce the consequences of high temperature on its growth [162,163]. Arabidopsis grows optimally at 22°C, and exhibit stress responses, such as reduced stem and leaf growth, decrease in chlorophyll biosynthesis and impaired seed production, at temperatures lower than 16°C or higher than 22°C [130,152]. When submitted to temperature stresses, arabidopsis display a range of thermomorphogenic adaptation responses, such as hypocotyl elongation, increased leaf hyponasty, and leaf petiole extension, to increase the distance between the warm soil and its heat-labile shoot apical meristem [130,152].

Thermomorphogenesis is mediated by the Basic Helix-Loop-Helix (bHLH) transcription factor AtPIF4 [130]. It has been demonstrated that the interaction between AtTOC1 and AtPIF4 influences the gating of thermoresponsive growth in arabidopsis, and that *AtTOC1* overexpression causes thermoinsensivity [164]. Confirmed by *in vivo* co-immunoprecipitation sequencing, AtTOC1 evening-specific inhibition of thermoresponses is mediated by binding to *AtPIF4* promoter region, thus making *AtPIF4*-mediated growth during the day more responsive to the ambient temperature [164]. Other core and peripheral clock genes belonging to the *AtRVE* group, respectively *AtRVE8* and *AtRVE 3*, *4* and *6*, have been shown to have an influence on leaf size and biomass in arabidopsis due to their action on the *AtPIF4* and *AtPIF5* genes [18]. AtPIF4 is also involved the induction of the flowering process in arabidopsis by activating, through temperature-dependent promoter binding, the floral pathway integrator gene *AtFT* which transcriptionally activates downstream floral identity genes [165–167].

2.5 Connecting statement: upcoming rising challenges are time- and species-dependent

By 2050, worldwide population is expected to reach 9.3 billion, and 40% of the world population will live in countries with insufficient food and water reserves to fulfill the dietary requirements of their own people [168,169]. Climate change is predicted to contribute to a rise in temperature extremes, and is going to reshape global agricultural practices, in particular in terms of crop improvement [129,170]. Studies on the molecular basis of the circadian clock in plants have become, over the years, a focal point of research with numerous contemporary studies now linking the endogenous circadian rhythm to abiotic stress tolerance and heterosis in breeding, two important features in crop improvement [171]. Clock-controlled physiological responses contributing to heat stress tolerance include important features, such as the regulation of stomatal aperture and flower-bud opening [144], as well as the coordination of the general molecular responses at the whole plant level in a timely manner [172]. Efficient breeding strategies to counter these unrelenting problems are to develop crops with a modified flowering process and/or stomatal aperture responses using traditional breeding or by rewiring photoperiodic gene regulatory network using genome editing to manipulate one or multiple clock components and/or downstream targets [173,174].

In the past, unravelling the clock in plants has mostly been achieved using the model plant *A. thaliana*, a dicotyledonous plant. Some basic characteristics differentiate arabidopsis from rice when it comes to deeply understand the clock. For example, one of the most singular problem for research is the difference in flowering pattern between rice, a short day plant, and arabidopsis, a long day plant [69,175,176]. Another concern is that some orthologous genes are not always present in all species, in a 1:1 absolute pairing, which limits the development of a generalized model [175,177]. The focus on arabidopsis has perpetuated a knowledge gap for monocotyledonous plants which is a pressing issue due to the widespread importance of those plants in the human diet. Therefore, there is a case to study the structural and functional commonalities of elements of the clock between species from the same class if we want to solve these global issues.

2.6 Tables and figures

Table 2.1 The arabidopsis circadian clock encompasses many protein families.

Transcription factor families	Family members
Adagio F-box domain	ZEITLUPE (ZTL) [51]
CONSTANS, CO-like, and TOC1 (CCT) domain	CONSTANS (CO), CONSTANS-LIKE (COL), TIMING OF CAB EXPPRESION1 (TOC1) [178]
Glutamic Acid/Alanine-Rich Protein (GARP) domain	LUX ARRHYTHMO (LUX) [179]
Myeloblastosis (MYB) domain	LATE ELONGATED HYPOCOTYL (LHY), CIRCADIAN CLOCK ASSOCIATED1 (CCA1), REVEILLE1-8 (RVE1-8), RVE7-like [112]
Per-Arnt-Sim (PAS) domain	ZTL [180]
Pseudo Receiver (PR) domain	TOC1, PSEUDO-RESPONSE REGULATORS 3, 5, 7 and 9 (PRRs 3, 5, 7 and 9) [68]
Teosinte branched1, Cycloidea, and PCF (TCP) domain	CCA1 HIKING EXPEDITION (CHE) [55]

Table 2.2 Members of the CCA1-LHY-RVE protein family in arabidopsis.

Gene name	Other names	TAIR ID
RVE1	-	AT5G17300
RVE2	CIRCADIAN 1, CIR1	AT5G37260
RVE3	ALTERED SEED GERMINATION 4, ASG4	AT1G01520
RVE4	LHY/CCA1-LIKE 1, LCL1	AT5G02840
RVE5	-	AT4G01280
RVE6	-	AT5G52660
RVE7	EARLY PHYTOCHROME RESPONSIVE 1, EPR1	AT1G18330
RVE7-like	-	AT3G10113
RVE8	LHY/CCA1-LIKE 5, LCL5	AT3G09600
CCA1	-	AT2G46830
LHY	-	AT1G01060

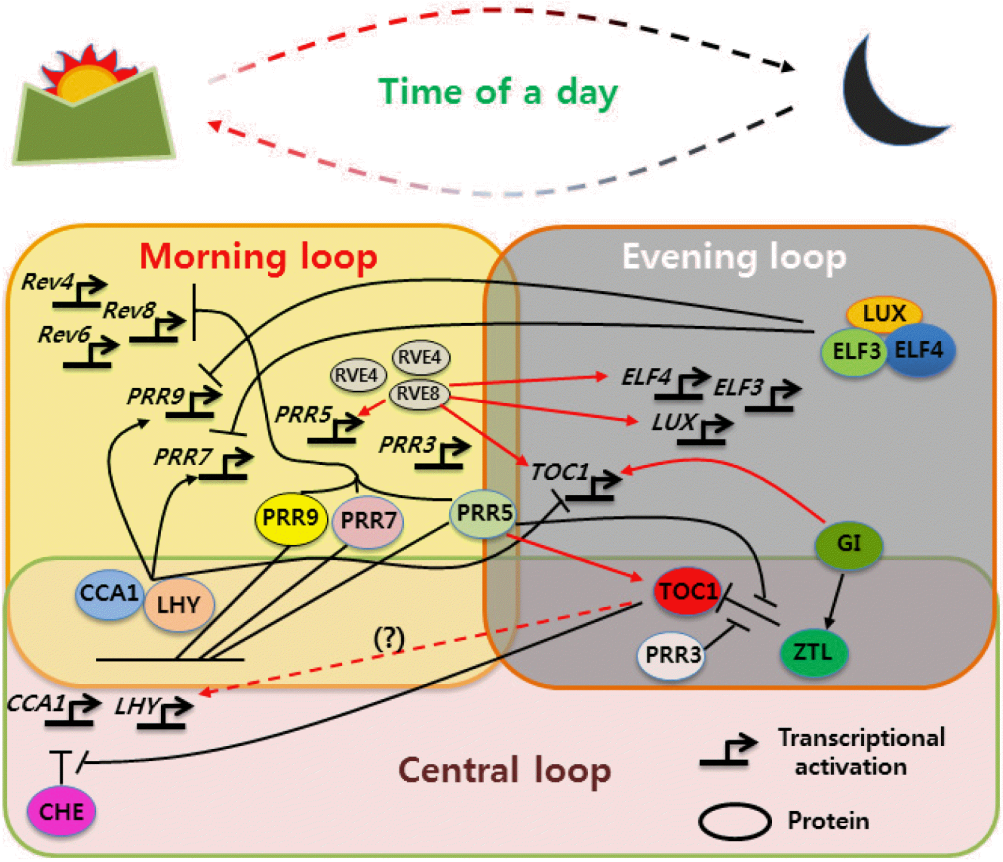


Figure 2.1 Morning and evening loops of the circadian clock gene regulatory network. Red solid lines represent direct transcriptional activation; Red dashed lines indicate indirect activation; Black flathead connectors represent transcriptional repression [36].

CHAPTER III: Materials and methods

3.1 Characterization of the expansion and diversification of the RVE protein family in monocots and dicots

3.1.1 Alignment of the CCA1-LHY-RVE family members in sequenced monocot and dicot genomes

Putative monocot CCA1-LHY-RVE family members were retrieved by querying the whole genomes of the 10 monocot genomes in Phytozome v12.1 [181] with the full-length amino acid sequence of OsEPR1 (LOC_Os06g51260.1) ($E < 1E^{-6}$). The same approach was used to retrieve putative dicot CCA1-LHY-RVE family members in the genomes of the 21 dicot genomes in Phytozome, using the query sequence of AtEPR1 (AT1G18330.1). A multiple sequence alignment of the retrieved sequences was generated with CLUSTALW [182] progressive alignment system in MEGA X using default settings. The putative CCA1-LHY-EPR1 protein family sequences were manually-curated to remove sequences lacking the canonical MYB DNA-binding domain and proline rich domain which define the family [9]. Phylogenetic analysis was performed using the neighbour-joining algorithm in MEGA X using default settings, except for pairwise deletion data treatment for gaps/missing data (500 bootstraps).

Identity matrices were generated by aligning the full-length protein sequences of the CCA1-LHY-RVE protein family members using CLUSTALW alignment system in Sequence Demarcation Tool v1.2 [183]. A species tree to show the relationship between the included taxa was generated with phyloT using NCBI taxonomic ID numbers and imported in Newick format on the Interactive Tree of Life (IToL) [184,185]. Figure built with IToL online system was generated using unrooted mode and all other settings were set as default.

3.1.2 Identification of conserved motifs in the amino acid sequences of RVE1, RVE8 and LHY clade proteins

De novo protein motif discovery was performed using Multiple Em for Motif Elicitation (MEME Suite 5.0.5) [186]. MEME was run in Classic Mode with default settings, except with the maximum motif width set at 80 characters and with the site distribution set to any number of occurrences. The position weight matrix of the AtHSFB2B/AtHSF7 (AT4G11660.1) binding site was selected as the exemplar Heat Shock Element (HSE); the position weight matrix of the AtLHY (AT1G01060.1) binding site was selected as the exemplar Evening Element (EE). The promoters were defined as the region from -1000bp to +200bp relative to the transcriptional start site of each gene. Promoter sequences were downloaded from TAIR [187] for the arabidopsis genes and from the PlantPromoter [188] database for rice genes. Heat shock element motif logo and MEME format were retrieved from the Plant Cistrome database arabidopsisDAPv1 [189] using Tomtom (Database ID: HSF_tnt.HSF7_col_a_m1 (HSF7)) with default settings for the initial TTCnnGAAnnTTC motif found in the literature [7,190]. AtLHY Evening Element logo and MEME format were retrieved from JASPAR database JASPAR2018_CORE_non-redundant [191] using Tomtom (Database ID: MA1185.1) with default settings for the initial AAAATATCT motif found in the literature [192].

3.2 Expression of genes in the *CCA1-LHY-RVE* family over a diurnal cycle and in response to heat stress

3.2.1 Rice plant material and growth conditions

Rice (*Oryza sativa* ssp. *japonica* cv. Nipponbare) seeds (US Department of Agriculture, Stuttgart, AZ, USA) were incubated with abundant moisture in complete darkness for 3 days at 32°C to initiate germination. Germinated seeds were manually transplanted into Agromix G6/PV20 potting mix (Fafard et Frères Ltd., Saint-Remi, QC, Canada). Seedlings were grown in a Percival® E-36VL growth chamber (Percival Scientific inc., Perry, IA, USA) for 7 days after germination (DAG) (12 hours of daylight (28.6°C, 300 µmol/m²/s) and 12 hours of darkness (26.8°C, 0 lumen).

3.2.2 Sampling tissues for transcript measurements

To measure transcript abundance of the *CCA1-RVE-EPR1* genes throughout a diel cycle, the aerial tissues of seven-day old rice plants were harvested at six timepoints (zt 0 h, 4 h, 8 h, 12 h, 14 h, and 22 h). Tissues were flash frozen in liquid nitrogen and stored at -80°C. Each sample included the aerial tissues of four rice plants. The experiment was independently repeated three times. To measure transcript abundance of the *CCA1-RVE-EPR1* genes in response to heat shock, seven-day old rice plants were exposed to 40.0°C for 30 minutes at three timepoints (zt 0 h, 4 h and 8 h). The tissues, replications, and sample handling were as outlined.

3.2.3 RNA extraction and cDNA synthesis of the harvested samples

Total RNA was extracted using the Plant Total RNA Mini Kit (Geneaid, New Taipei, Taiwan). RNA purity was assessed spectrophotometrically using Nanodrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and degradation was assessed by agarose gel electrophoresis (1% agarose, 100 volts, 60-70 min). Reverse transcription was performed using iScript gDNA Clear cDNA Synthesis Kit (Bio rad, Hercules, CA, USA) to generate gDNA free first-strand cDNA. cDNA synthesis was performed with a total of 2800 ng of RNA.

3.2.4 RT-qPCR primer design for members of the CCA1-LHY-RVE family in rice

The coding sequences of the rice *CCA1-LHY-RVE* genes were retrieved from Phytozome as described above. Primers were designed using the NCBI primerBLAST tool [193]. Default settings were used, except for the following: 1) PCR product size was set to between 70-160 bp, and 2) Organism was set to *Oryza sativa* japonica cultivar-group taxid 39947. Primers were selected that were close to 20 bp long and had melting temperature between 58-60°C with a maximum Tm difference between primers in a pair of 2°C, and a GC content close to 50%.

3.2.5 Validation of PCR efficiency

The efficiency of each PCR assay was determined by calculating the slope of a standard curve of known concentrations of genomic DNA. gDNA was extracted with the genomic DNA Mini Kit for plant (Geneaid, New Taipei, Taiwan). Amplifications were performed on a CFX Connect Real-Time PCR Detection System (Bio rad, Hercules, CA, USA) with SsoAdvanced Universal SYBR Green Supermix (Biorad, Hercules, CA, USA). The PCR conditions were as follows: 30 s at 95°C, followed by 40 cycles of 15 s at 95°C and 30 s at 30°C. The melting temperature of the amplicons was determined by a melt curve with 0.5°C increments of 5 sec each from 65°C to 95°C. The assays were evaluated using three criteria: the reaction efficiency between 90-110%, and a $R^2 \geq 0.99$.

PCR specificity was determined through single melt peak analysis, and corresponding temperatures by comparison between predicted Tm and observed Tm. Predicted Tm was determined using uMELT [194] with the following parameters: ([Mono+] 20 mM, free [Mg^{2+}] 3 mM, DMSO 0%, and salt correction as per SL & Hicks (2004) [195]). Tm prediction on uMELT was performed using Blake & Delcourt [196] thermodynamic set. uMELT prediction was made by performing a melt curve analysis of the gDNA amplicon sequence for each studied gene and its associated primer pair into the software. Agarose gel electrophoresis (2%, 60 min, 100 volts) was performed to estimate amplicon length and to confirm amplification of a single product.

3.2.6 Relative quantification of transcript abundance through RT-qPCR assay

Stable housekeeping genes for relative quantification were identified using geNorm software available with CFX Maestro software (Bio rad, Hercules, CA, USA) [197]. Relative quantification was performed using Os*SUI1* and Os*GAPDH* housekeeping genes. Gene expression statistical analysis was performed using an analysis of variance (ANOVA) with CFX Maestro software using a p-value threshold of 0.05.

CHAPTER IV: Characterization of the expansion and diversification of the CCA1-LHY-RVE protein family in monocots and dicots

4.1 Introduction

In this chapter, I investigated the expansion and diversification of the CCA1-LHY-RVE protein family in monocot and dicot lineages. Protein sequences encoded in the genomes of 31 plant species, including 10 monocots and 21 dicots and representing a wide array of cultivated and model plant species, were retrieved (Fig 4.1). No large-scale multi-species phylogenetic studies have been published on the CCA1-LHY-RVE protein family resulting in a knowledge gap detrimental to our understanding of the evolution and phylogenetic relationships of this protein family. Studies of these genes and proteins in rice, tomato and arabidopsis have contributed to the fundamental knowledge of their key features. However, a comparative analysis across a broad range of species has not yet been undertaken. The key steps of my phylogenetic analysis were to identify the putative proteins in the CCA1-LHY-RVE family, perform multiple sequence alignments, characterize their associated identity matrices, and establish the phylogenetic relationships of this protein family across higher plants.

Curation of the retrieved proteins was performed using two conserved structural marker regions, the MYB DNA binding domain, a 61 amino acid motif, and the proline-rich C-terminal region of the MYB domain, a 15 amino acid motif predicted to play a nuclear localization function in these transcription factors (Fig 4.2). MYB transcription factors are involved in a vast array of functions in the plant, including circadian regulation and response to biotic and abiotic stresses, two functions underpinning the subsequent chapter. The generation of the neighbour joining trees was performed to produce the information required for the sub-clustering of the putative proteins into the RVE1, RVE8 and LHY clades for each species using arabidopsis and rice orthologs as markers (Fig 4.3 and 4.4). Based on the multiple sequence alignments, I built identity matrices for the studied monocotyledonous species and dicotyledonous species to perform a refined intra-species sub-classification of the clades and identify elements of conservation amongst the studied proteins.

I also investigated the promoter regions of the *CCA1-LHY-RVE* genes for the occurrence of two cis-regulatory motifs using the region from -1000bp to +200bp relative to the transcriptional start site of each studied gene. The first motif is the Evening Element (AAAATATCT) which is important to the clock machinery. The second motif is the Heat Shock Element (TTCnnGAAnnTTC) which plays a central role in the plant heat stress response. The Evening Element is found all across the plant kingdom where it plays a pivotal role in the circadian clock output. Core and peripheral clock transcription factors bind to the Evening Element in the promoter regions of downstream targets to repress them. Furthermore, the Evening Element is used by clock components to modulate their own expression in a time of the day specific manner through autoregulation. Heat shock factors are helix-turn-helix proteins that bind to the Heat Shock Element upon exposure to heat stress to activate or repress their target genes [198].

Overall, this chapter aimed to fulfill three main objectives: 1) build a phylogenetic pipeline that could be used to understand the phylogenetic relationships between clock components within monocots and dicots, in this case being the CCA1-LHY-RVE protein family, 2) identify putative members of the RVE1, RVE8 and LHY clades, and 3) perform a comparative phylogenetic analysis between monocots and dicots to fill the knowledge gap concerning the phylogenetic relationships for the CCA1-LHY-RVE protein family.

4.2 Results

MYB and proline-rich regions are defining features of the CCA1-LHY-RVE protein family

Initially, 210 protein sequences from monocotyledonous plants were retrieved from 10 species in Phytozome (e-value < $1E^{-6}$). This e-value threshold was chosen since it is the threshold suggested by the National Centre for Biotechnology Information (NCBI) when retrieving sequences for multi-species phylogenetic analysis. This was followed by a two-step curation process to remove sequences that lacked the defining MYB or proline-rich domains, and to remove duplicate sequences that aligned to the

same genetic loci, including predicted splice variants. The first step removed all but the primary amino acid sequence, identified with the notation ".1", for reach genetic locus. The second step involved confirming the presence of the MYB and proline-rich regions in each of the retrieved sequences. These features were identified by performing a multiple sequence alignment of the retrieved sequences and then examining the aligned sequences for the presence of the known motifs from arabidopsis (Fig 4.3). Altogether, this two-step curation reduced the total number of putative sequences belonging to monocotyledonous species from 210 to 74 (Table 4.1 and Fig. 4.5). In monocots, *Panicum virgatum* had the largest number of proteins, 12 in total, whereas *Brachypodium distachyon, B. stacei, Oropetium thomaeum* and *Sorghum bicolor* all displayed six proteins in total for the family, the lowest number of proteins for monocots (Table 4.1 and Fig. 4.5). For the 21 dicotyledonous species, the same steps as with the monocotyledonous species were performed. The initial query recovered 381 putative CCA1-LHY-RVE sequences. After the manual curation, 165 putative proteins were remaining, all of which were displaying conservation in the two regions of interest (Table 4.2 and Fig. 4.6). In dicots, *Glycine max* displayed the highest number of proteins for the family, 21 proteins in total, and the lowest number of proteins was displayed by *Carica papaya*, only three proteins (Table 4.2 and Fig. 4.6).

The CCA1-LHY-RVE protein family is more conserved in monocots than dicots

In arabidopsis, phylogenetic analyses of the RVE family have identified three distinct subgroups or clades [11,18]. My analysis of this family across multiple monocot and dicot species supports the widespread conservation of these subgroups. Of the 74 proteins putatively monocot CCA1-LHY-RVE protein sequences, 10 were most similar to LHY. Each of the monocot genomes encoded exactly one LHY homolog (Table 4.1 and Fig. 4.5). In the case of the 165 putative proteins belonging to dicotyledonous species, 27 were classified as members of the LHY clade (Table 4.2 and Fig. 4.6). Whereas each monocot genome encodes a single LHY type protein, dicot genomes show greater variability in terms of expansion of the family. *C. papaya, Daucus carota, Fragaria vesca, Malus domestica* and *Solanum tuberosum* do not encode any LHY type

proteins, whereas *G. max* encodes four (Table 4.2 and Fig. 4.6). However, most dicot genomes examined encode one or two LHY type proteins. The RVE8 subgroup in monocot comprised between two and six proteins depending on the species homolog (Table 4.1 and Fig. 4.5). In monocots, *S. bicolor* encodes the smallest number of proteins in the RVE8 clade, whereas *P. virgatum* encodes the largest. In dicots, the variation is higher than monocots for the RVE8 clade as the number of proteins is between one and eleven (Table 4.2 and Fig. 4.6). In the case of monocots, the smallest number of proteins for the RVE8 clade is found in *C. papaya* and the highest in *G. max*.

The RVE1 subgroup comprises between two and five proteins in monocots depending on the species. For this clade and in monocots, three species, *B. distachyon, B. stacei* and *O. thomaeum*, encode for two proteins, the smallest number of proteins, whereas *Zea mays* and *P. virgatum* encode for the highest number of proteins, five. In dicots, the number of proteins found in the RVE1 clade is between zero and six. *Theobroma cacao* does not have any member in this clade, whereas *G. max* encodes for a total of six proteins in the RVE1 clade. Overall, the number of proteins is quite consistent across all monocot species for the three clades, except *P. virgatum*. *P. virgatum* encodes twice as many RVE8 clade proteins than other species. Along with, this number is also close to be doubled in RVE1 which is coherent with the tetraploid nature of the reference *P. virgatum* strain AP13 genome, a high-yielding lowland tetraploid switchgrass clone [199]. In switchgrass, strains are divided between upland (most often octoploid, but sometimes tetraploid) and lowland (pseudotetraploid or allotetraploid) [199]. Dihaploid plants do exist, but are infertile and not used as reference genome [199].

High intra-clade conservation of the CCA1-LHY-RVE family enables the identification of subgroups in monocots, but not in dicots

The extent of sequence conservation varied within each of the three major clades for both monocot and dicot CCA1-LHY-RVE proteins. Within the monocot CCA1-LHY-RVE proteins, high levels of amino acid sequence conservation between species enabled the identification of additional subgroups within the three major clades described above. However, these clear distribution patterns of conservation were not found in the studied

dicots and the high rates of disparity within the clades did not allow the identification of subgroups based on orthologs comparison. In the RVE1 and RVE8 clades of monocots, high sequence identity rates enabled the identification of subgroups forming these clades. In the RVE1 clade of monocots, the subgroups were formed by three subgroups, the LOC_Os02g45670, OsEPR1, and LOC_Os04g49450 subgroups (Fig 4.7). In the OsEPR1 subgroup, identity ratios were all > 60% and the subgroup displayed a higher conservation than the two other members of its clade (LOC_Os02g46030 and LOC_Os04g49450). The LOC_Os02g45670 and LOC_Os04g49450 subgroups were less conserved than the OsEPR1 subgroup, but the distribution pattern and high sequence identity still provided enough information to enable their respective identification. In the RVE8 clade of monocots, the identified subgroups were the LOC_Os06g45840, LOC_Os06g01670, and the LOC_Os02g45670 subgroups (Fig 4.7). These subgroups were all characterized by a high conservation rate amongst their members. LOC_Os02g45670 was the most conserved subgroup amongst the three with identity ratios always > 75%, a rate even higher than the very well-conserved LHY clade. A high intra-subgroup identity rate was also found in the LOC_Os06g45840 subgroup with identity ratios all > 60%.

On the contrary to monocots, the conservation pattern of dicots and the prevalent low identity between putative members of the three clades did not allow the identification and delineation of subgroups (Fig 4.8). In the LHY clade of monocots, pairwise sequence identity was always > 65%, but in dicots the intra-clade conservation within the LHY clade was ranging from 20% to 100%. In the RVE8 clade of dicots, conservation was more prevalent than in the LHY clade, but still lower than in monocots. In the case of the RVE1 clade of dicots, intra-clade identity rates were the lowest amongst the three clades and did not allow the identification of subgroups. Similar conservation patterns between monocots and dicots were also observed using full-length protein multiple sequence alignment and MEME scanning analyses (Fig 4.9 and 4.10). Those permitted the identification of the conserved regions and their associated motifs for each clade both for monocots and dicots. MEME scanning output for dicots was less reliable than the one for monocots based on the respective e-value of the generated motifs. These observations, therefore, shine the light on the fact that that monocot clades are not more only more

conserved than those of dicots, but also that the respective distribution pattern and high identity rates within the monocot clades enable the identification of well-conserved subgroups within these clades.

Heat Shock Element is only enriched in EPR1 in rice and arabidopsis

In arabidopsis, different genes member of the *CCA1-LHY-RVE* family are expressed in response to different environmental and developmental cues [14,15]. I examined the promoters of the *CCA1-LHY-RVE* genes in rice and arabidopsis for the presence of two known *cis*-regulatory motifs, the Evening Element and the Heat Shock Element. High confidence match in my FIMO promoter analysis was associated with motifs meeting three threshold criteria, a 0.05 threshold both for p-value and q-value and a final score >14 (Table 4.3). High confidence match reduced the number of identified motifs from 13 motifs to 6 motifs for the Evening Element. For the Heat Shock Element, 12 motifs were originally retrieved, and 3 motifs were left after the curation process. Generated results indicate that *AtEPR1* in arabidopsis harbours the Evening Element and Heat Shock Element motifs with two motifs for each of the cis-regulatory modules. In rice, the Evening Element is only found in LOC_Os04g49450 and the Heat Shock Element in OsEPR1.

4.3 Discussion

Strong conservation in monocots supports the use of rice as a model plant for circadian studies

For a long time, the circadian clock has been an important research topic to plant scientists as it is central to the plant fitness and response to stress. While it has been widely assumed in the literature that the basic components and functioning of the clock are conserved across the plant kingdom, few studies have taken a systematic approach to annotate circadian clock components in a large number and diversity of plant species. The small number of studies about the CCA1-LHY-RVE protein family has generally

also overlooked this aspect. Currently, research concerning the evolution and phylogenetic relationships of the CCA1-LHY-RVE protein family has been constrained to arabidopsis, *Picea abies*, tomato and rice [10,18,65].

Research on the phylogenetic relationships of CCA1/LHY in plants is an exception to this. Their homologs have been reported in many other plants, such as cucumber (*Cucumis sativa*), pea (*Pisum sativum*), soybean (*Glycine max*), bean (*Phaseolus vulgaris*), *Lemna gibba* and more [200]. Likewise, research tackling the homology of PRRs, LUX, ELF3, ELF4, ZTL and GI in other species is also more advanced than for the proteins belonging to the RVE8 and RVE1 clades [201–203]. Many circadian clock proteins, including those in the CCA1 clade and LCL clade as well as other clock members (PRRs, LUX, ELF3, ELF4, ZTL and GI), have also been studied in several algae species; *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Klebsormidium flaccidum*, *Marchantia polymorpha*, *Anthoceros agrestis*, *Physcomitrella patens*, and *Selaginella moellendorffii* [10]. In *C. reinhardtii* and *O. tauri* only one protein was found and could not be assigned to either the CCA1 clade or the LCL clade [10]. Two of the studied algae species, *M. polymorpha* and *S. moellendorffi,* did not encode proteins for the CCA1 clade and one plant species, *P. abies*, did not have any member in the LCL clade [10]. Overall, the numbers of proteins found in the CCA1 and LCL clades were much lower in the algae species than in the studied reference plant species [10]. For the LCL clade, the number of proteins found in the LCL clade was between one and three, whereas five and six proteins were respectively found in *O. sativa* and *A. thaliana* [10]. For the CCA1 clade, between zero and two proteins were found in the algae species, whereas three, one and six were respectively found in *P. abies*, *O. sativa* and *A. thaliana* [10].

The present analysis used more restrictive criteria for identifying the CCA1-LHY-RVE proteins than some previous analyses. For example, Gray et al. (2017) published a phylogenetic analysis of the CCA1-LHY-RVE protein family in arabidopsis, rice and tomato [18]. In this analysis, nine CCA1-LHY-RVE proteins were identified, whereas in my analysis, I identified only seven. The two proteins that did not meet the criteria I set for being *bona fide* CCA1-LHY-RVE proteins, LOC_Os05g07010 and

LOC_Os01g06320, lack a proline-rich domain. Moreover, LOC_Os05g07010 lack the SHAQKFF motif. Similarly, the CCA1-LHY-RVE protein family in papaya has been described as having 5 members and poplar as having 12 members [65]. Using the criteria I set off for my analysis, papaya has only 3 members and poplar has 11 members. In the case of LHY in papaya, the putative protein was excluded in my study because it lacked a MYB and proline-rich domains, although the rest of its sequence displayed a very high similarity to the LHY gene in arabidopsis. Similar observations have been made with the four other species lacking a protein in the LHY clade, *D. carota, F. vesca, M. domestica* and *S. tuberosum*. In these cases, the core MYB and proline-rich domains were either lacking or missing important motifs, despite the rest of their sequences demonstrating a very high similarity to the reference LHY arabidopsis gene. In that sense, further research using PCR assay could be used to confirm and assess the presence of LHY genes in those species. In the literature, LHY protein is considered pivotal for proper clock regulation, but these four species do not harbour any LHY clade proteins characterized by the two defining features of the CCA1-LHY-RVE protein family, a MYB DNA-binding domain and a proline-rich domain. These evidences suggest that the clock network and its underlying molecular functions might be different in these species than in arabidopsis or that other clock-related genes might have undergone neofunctionalization. Overall, this study consolidated and expanded the current state knowledge on the conserved aspects of the MYB and proline-rich regions with accurate motifs scanning.

The phylogenetic analysis generated in this study suggests that rice is an appropriate model plant to investigate this family in monocotyledonous species. Firstly, the 10 studied monocotyledonous species share a similar number of proteins for the RVE8 clade, approximately three, and only one gene in the LHY clade. Still, for the 10 studied monocotyledonous species, the RVE1 clade is more variable in terms of number of proteins, which is between two and five, but so does the protein sequences of the proteins forming this group on a structural aspect. Secondly, the high structural intra-clade similarity of the protein sequences forming the RVE8 and LHY clades also support the use of rice as a model plant. Along, the same similarity is observed for the protein sequences in the subgroups forming the RVE1 clade (e.g. OsEPR1 subgroup). Globally, this high similarity suggests the conservation of the functional roles of the CCA1-LHY-

RVE protein family in monocots. In dicotyledonous species, phylogenetic divergence is more pronounced which is depicted by the greater disparity in the numbers of proteins forming the three clades and the lower similarity in the identity matrices. On the contrary to monocotyledonous species, the RVE8 and LHY clades are more disparate in terms of the number of species, but still have a higher intra-clade similarity than the RVE1 clade.

Retention of clock proteins is uniform in monocots, but not dicots

Overall, the data generated for monocots and dicots species highlights that the circadian clock is more conserved amongst monocots than amongst dicots. The only species showing a clear variation in the number of proteins is in tetraploid *P. virgatum*. Based on phylogenetic and syntenic analyses, it has been recently demonstrated that before the divergence of Brassicales, the CCA1-LHY-RVE family was comprised of only LHY and five RVE proteins, RVE1, RVE7/EPR1, RVE3, RVE6, and RVE8 [65]. Accordingly, in the arabidopsis genome, CCA1 and RVE2 are respectively descendants of LHY and RVE1 in the β duplication [65]. This observation is consistent with my observations as RVE1 and RVE2 have higher identity rates in the matrix. In the α duplication of the arabidopsis genome, two members of the RVE8 clade, RVE4 and RVE5, emerged from RVE8 and RVE3, respectively [65]. The α whole-genome duplication event is the most recent of the three duplication events (α, β and γ) that happened in arabidopsis, and is thought to be shared by most of the *Brassicaceae* members [204]. The β duplication event happened between the γ event, the oldest, and the α event, and is also widely shared by members of the *Brassicaceae* family [204]. Similarly, this observation is coherent with the generated data in this study.

B. rapa and *A. thaliana* are two species that have a close genetic evolution, but *B. rapa* genome has increased in gene number since its divergence from arabidopsis [65]. Overall, the numbers of genes are evaluated at ~30,000 genes in the arabidopsis genome and ~42,000 genes in *B. rapa* [205]. As demonstrated by these numbers, the number of genes found in *B. rapa* is lower than what would be normally expected in an hexaploid organism and is related to a genic fractionation after triploidization [65]. *V. vinifera*, *P. trichocarpa*, and *C. papaya* did not undergo the α and β duplications, and thus, are good

indicators of the structural conformation of the CCA1-LHY-RVE family before these events. In other studies, common genes that were identified for those species are LHY, RVE1, RVE3, RVE6 and RVE8 [65]. Two members of the RVE8 clade, RVE2 and RVE4, have been suggested to be unique to the Brassicales [65]. RVE7/EPR1 is missing in *V. vinifera* and *C. papaya*, but present in *P. trichocarpa, B. rapa* and *A. thaliana*, suggesting that this gene was lost in *C. papaya* after the divergence of the Brassicales [65]. In total, the number of genes that are members of the CCA1-LHY-RVE family is five in *V. vinifera* and *C. papaya*, but 12 in *P. trichocarpa* [65]. In my study, the number of proteins is significantly lower in *C. papaya* with no proteins figuring in the LHY clade and only one in the RVE8 clade. In the evolution of *P. trichocarpa*, the species has undergone the salicoid duplication, a duplication event independent from the Brassicales events, in which another CCA1/LHY copy arose [65]. Overall, the difference in the gene numbers between my study and Lou et al. (2012) may be partially explained by the use of a different analysis pipeline in which the e-values, Phytozome versions, and curation procedure, differ.

Heat Shock Element and Evening Element are enriched in a limited number of RVE genes in arabidopsis and rice

Promoter scanning reveals that *AtEPR1* in arabidopsis harbours two Evening Element and two Heat Shock Element motifs, whereas OsEPR1 in rice only features one Heat Shock Element. AtEPR1 has been reported binding to the canonical Evening Element *in vitro* [206]. Through an overexpression experiment, it has been demonstrated that AtEPR1 autoregulates itself [120]. In that sense, the Evening Element might be the nexus supporting AtEPR1 autoregulation although it still needs to be proven experimentally. *OsEPR1* in rice is not similar to its arabidopsis counterpart since it only harbours one Evening Element (score of 12.197, a q-value of 0.198 and a p-value $<4.01E^{-05}$). Experiments conducted on EPR1 in arabidopsis and rice has shown that it is upregulated upon heat shock which is coherent with the presence of a Heat Shock Element in their promoter regions [7,15]. In arabidopsis, three other members of the CCA1-LHY-RVE family harbour a Heat Shock Element in their promoters, *AtRVE3*,

AtRVE4, and *AtLHY* (0.05 threshold both for p-value and q-value and a final score >14). Under high temperature stress, *AtLHY* expression is reduced, whereas *AtCCA1* expression is enhanced [207]. The Heat Shock Element is not enriched in *AtCCA1* which suggests a key player for this response mechanism might be the Heat Shock Element.

4.4 Tables and figures

Table 4.1 RVE gene complements of 10 monocot species.

Latin name	Common name	RVE8 clade	LHY clade	RVE1 clade	Total
Brachypodium distachyon	Purple false brome	3	1	2	6
Brachypodium stacei	*Brachypodium stacei*	3	1	2	6
Zea mays	Corn	4	1	5	10
Oryza sativa	Rice	3	1	3	7
Oropetium thomaeum	*Oropetium thomaeum*	3	1	2	6
Panicum hallii	Hall's panicgrass	3	1	3	7
Panicum virgatum	Switchgrass	6	1	5	12
Setaria italica	Foxtail millet	3	1	3	7
Setaria viridis	Wild foxtail millet	3	1	3	7
Sorghum bicolor	Sorghum	2	1	3	6
Total		33	10	31	74

Table 4.2 RVE gene complements of 21 dicot species.

Latin name	Common name	RVE8 clade	LHY clade	RVE1 clade	Total
Arabidopsis thaliana	Arabidopsis	5	2	4	11
Brassica oleracea capitata	Cabbage	2	2	2	6
Brassica rapa	Multiple brassica, including canola and turnip	8	3	5	16
Carica papaya	Papaya	1	0	2	3
Citrus clementina	Clementine	3	1	1	5
Citrus sinensis	Orange	2	1	1	4
Cucumis sativus	Cucumber	4	1	2	7
Daucus carota	Carrot	4	0	3	7
Fragaria vesca	Strawberry	3	0	1	4
Glycine max	Soybean	11	4	6	21
Malus domestica	Apple	7	0	4	11
Manihot esculenta	Cassava	5	2	2	9
Medicago truncatula	Alfalfa	3	1	2	6
Phaseolus vulgaris	Bean	6	2	3	11
Populus trichocarpa	Poplar	5	2	4	11
Prunus persica	Peach	3	2	2	7
Solanum lycopersicum	Tomato	3	1	2	6
Solanum tuberosum	Potato	2	0	2	4
Theobroma cacao	Cacao	3	1	0	4
Trifolium pratense	Clover	3	1	3	7
Vitis vinifera	Grape	3	1	1	5
Total		86	27	52	165

Table 4.3 FIMO promoter scanning of HSE and EE in CCA1-LHY-RVE members for arabidopsis and rice.

Evening Element

Sequence name	Start	Stop	Strand	Score	p-value	q-value	Matched_sequence
AT1G18330	989	998	+	16.3182	2.00E-06	0.0387	AAAAATATCT
AT1G18330	347	356	-	16.3182	2.00E-06	0.0387	AAAAATATCT
AT1G18330	81	90	+	15.6515	5.28E-06	0.0511	CAAAATATCT
AT5G37260	672	681	-	12.5455	3.38E-05	0.198	GATAATATCT
AT5G17300	100	109	+	12.1061	4.21E-05	0.198	TAAAATATCA
AT2G46830	452	461	+	11.9091	4.61E-05	0.198	TATAATATCT
AT5G17300	28	37	-	10.4091	7.09E-05	0.274	AAAAGTATCT
AT5G52660	1022	1031	+	9.92424	7.98E-05	0.281	AGAAATATCC
LOC_Os04g49450	122	131	+	15.9545	3.64E-06	0.047	GAAAATATCT
LOC_Os02g46030	145	154	+	13.3182	1.75E-05	0.136	GGAAATATCT
LOC_Os06g51260	18	27	+	12.197	4.01E-05	0.198	GAAAATATCC

Heat Shock Element

Sequence name	Start	Stop	Strand	Score	p-value	q-value	Matched_sequence
AT1G18330	623	636	-	16.8788	8.20E-07	0.0205	TTCTAGTACCTTCT
AT1G01520	28	41	-	16.6424	9.81E-07	0.0205	GTCTAGAAGGTTCT
AT1G01060	564	577	+	14.4545	5.31E-06	0.0435	TTCTATAAGTTTCT
AT1G18330	527	540	-	14.3394	5.72E-06	0.0435	TTCTCGAAATTTCC
AT5G02840	963	976	-	14.2242	6.24E-06	0.0435	TCCTGGAAACTTCT
AT5G17300	383	396	-	11.4545	3.69E-05	0.189	TTCTCGAGCGTTCC
AT1G01060	514	527	+	11.3818	3.86E-05	0.189	TCCTAGAAACTGCT
AT5G17300	634	647	-	11.2909	4.07E-05	0.189	TTCTCGGAGTTTCT
AT3G09600	1155	1168	+	10.5697	6.26E-05	0.262	TTCAAGAAACTTAT
LOC_Os06g51260	614	627	-	15.4485	2.59E-06	0.0361	TTCTGGAAAATTCC
LOC_Os06g51260	725	738	-	10.1515	7.96E-05	0.303	TTCTAGTTTCCTCT
LOC_Os06g45840	333	346	-	9.87879	9.25E-05	0.323	TTATAGAATCTTAT

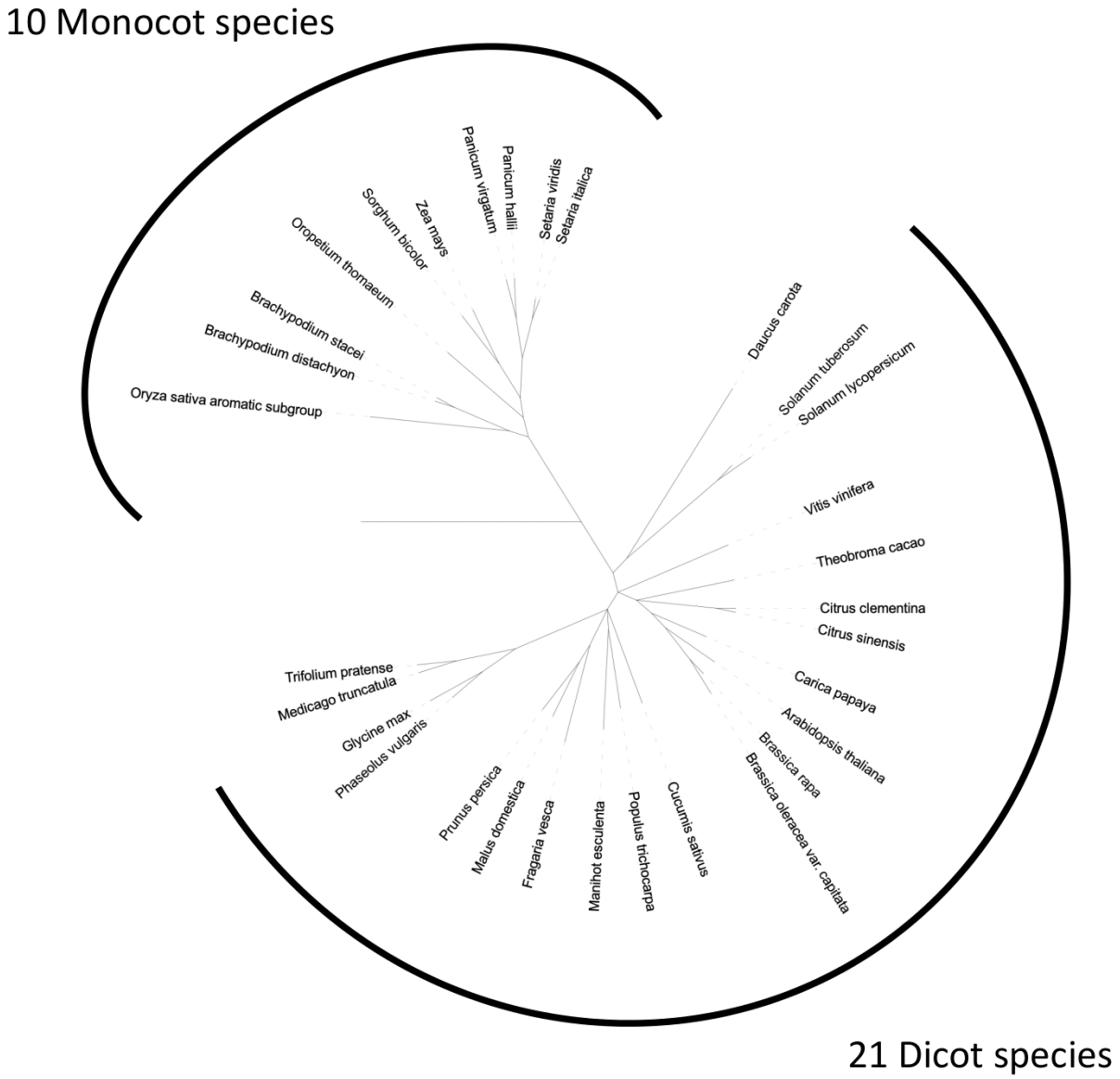
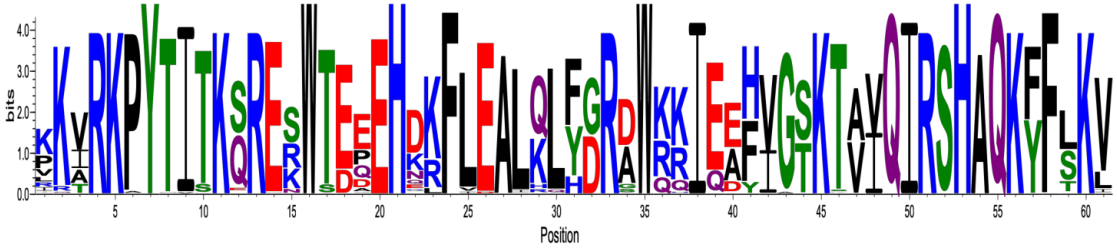


Figure 4.1 Species tree of the 10 monocot and 21 dicot species included in the analysis of the CCA1-LHY-RVE protein family. Displayed order represents NCBI taxonomy database consensus classification. Branch length do not correspond to the evolutionary distance used to infer the phylogenetic tree.

A



B

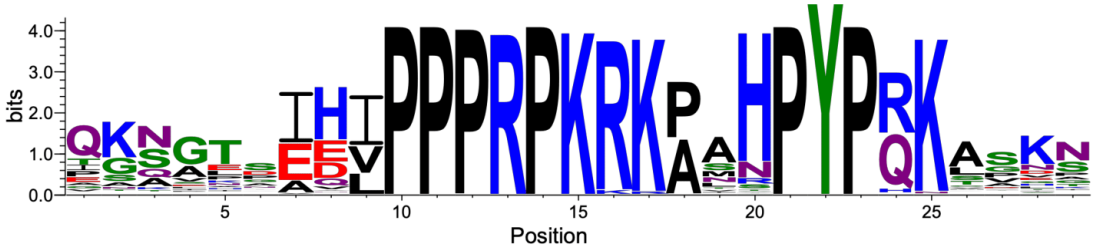


Figure 4.2 Sequence logos of the helix-turn-helix MYB DNA binding domain and proline rich domains for the 165 putative members of the CCA1-LHY-RVE protein family in the 31 studied species. (A) MEME logo for the helix-turn-helix MYB DNA binding domain, and (B) MEME logo for the proline rich domain.

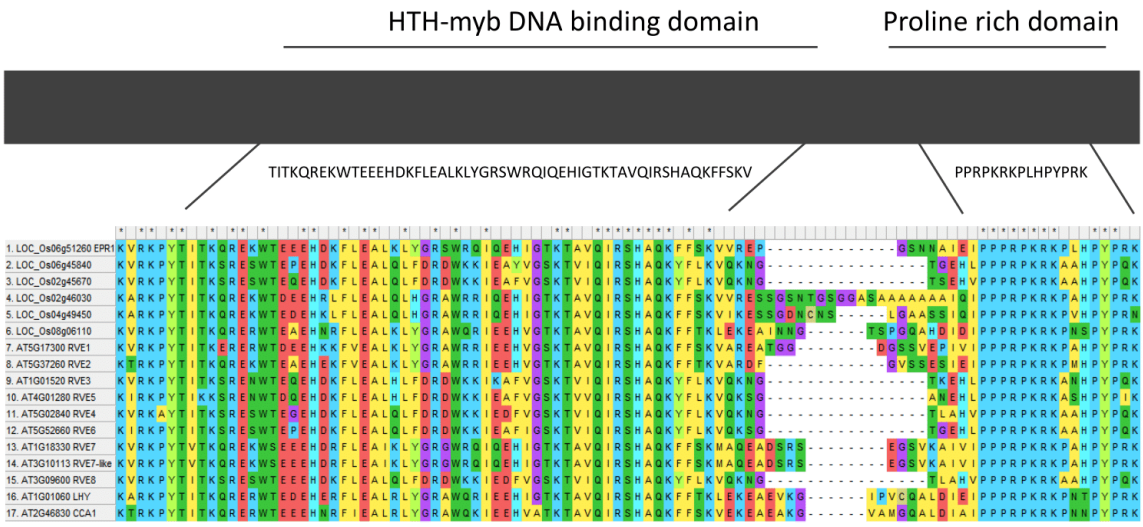


Figure 4.3 Multiple sequence alignment of the helix-turn-helix MYB DNA binding domain and the proline rich region of the CCA1-LHY-RVE protein members in rice and arabidopsis. Asterisk symbol, conserved residues in all input sequences. Colour code representing displayed amino acid. Blue: K, R, P. Red: D, E. Light blue: H. Grey: C. Green: S, Q, T, N. Light green: W, Y. Yellow: V, A, L, I, F, M. Purple: G. White: gap.

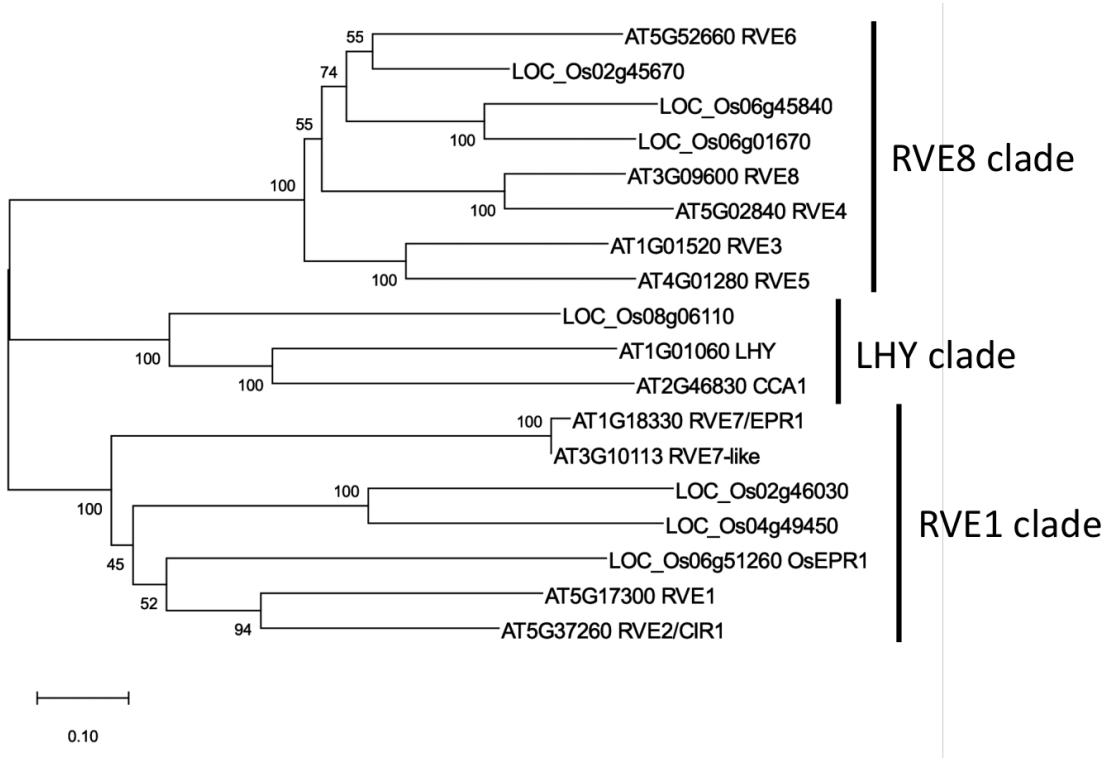


Figure 4.4 Phylogenetic relationships and subgroup designations of the RVE family proteins in rice and arabidopsis. The neighbour-joining tree includes seven RVE proteins from rice and nine from arabidopsis. The proteins are clustered into three subgroups, which have been highlighted and named for their most well-studied member. Five hundred bootstrap replicates analyses were conducted and the support for each branch is indicated on the tree. Evolutionary distances were calculated using the Poisson method. Branch length corresponds to the evolutionary distance used to infer the phylogenetic tree.

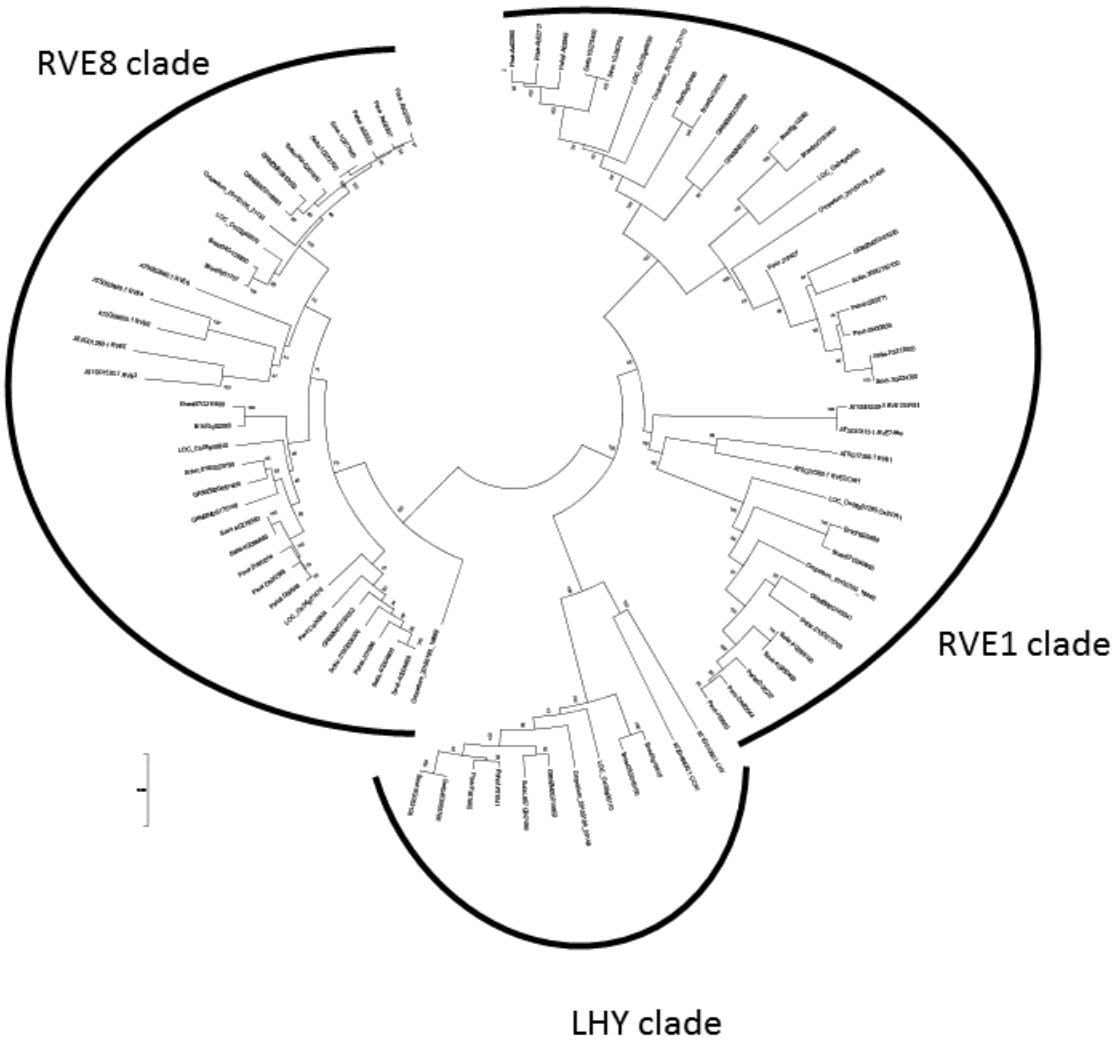


Figure 4.5 Phylogenetic relationships and subgroup designations of the RVE family proteins in 10 monocot species. The neighbour-joining tree includes the 74 RVE proteins identified in Table 4.1 The proteins are clustered into three subgroups, which have been highlighted and named for their most well-studied member. Five hundred bootstrap replicates analyses were conducted and the support for each branch is indicated on the tree. Evolutionary distances were calculated using the Poisson method. Branch length corresponds to the evolutionary distance used to infer the phylogenetic tree.

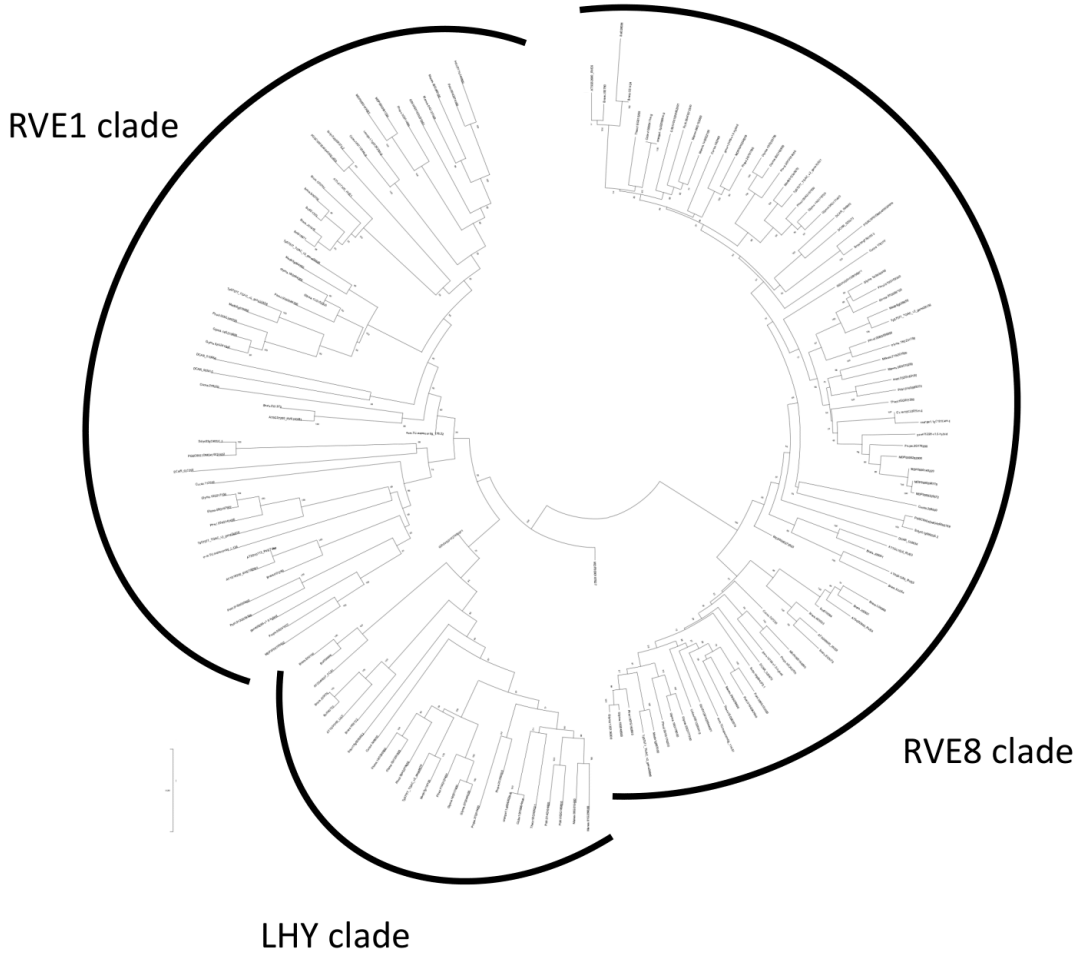


Figure 4.6 Phylogenetic relationships and subgroup designations of the RVE family proteins in 21 dicot species. The neighbour-joining tree includes the 165 RVE proteins identified in Table 4.2 The proteins are clustered into three subgroups, which have been highlighted and named for their most well-studied member. Five hundred bootstrap replicates analyses were conducted and the support for each branch is indicated on the tree. Evolutionary distances were calculated using the Poisson method. Branch length corresponds to the evolutionary distance used to infer the phylogenetic tree.

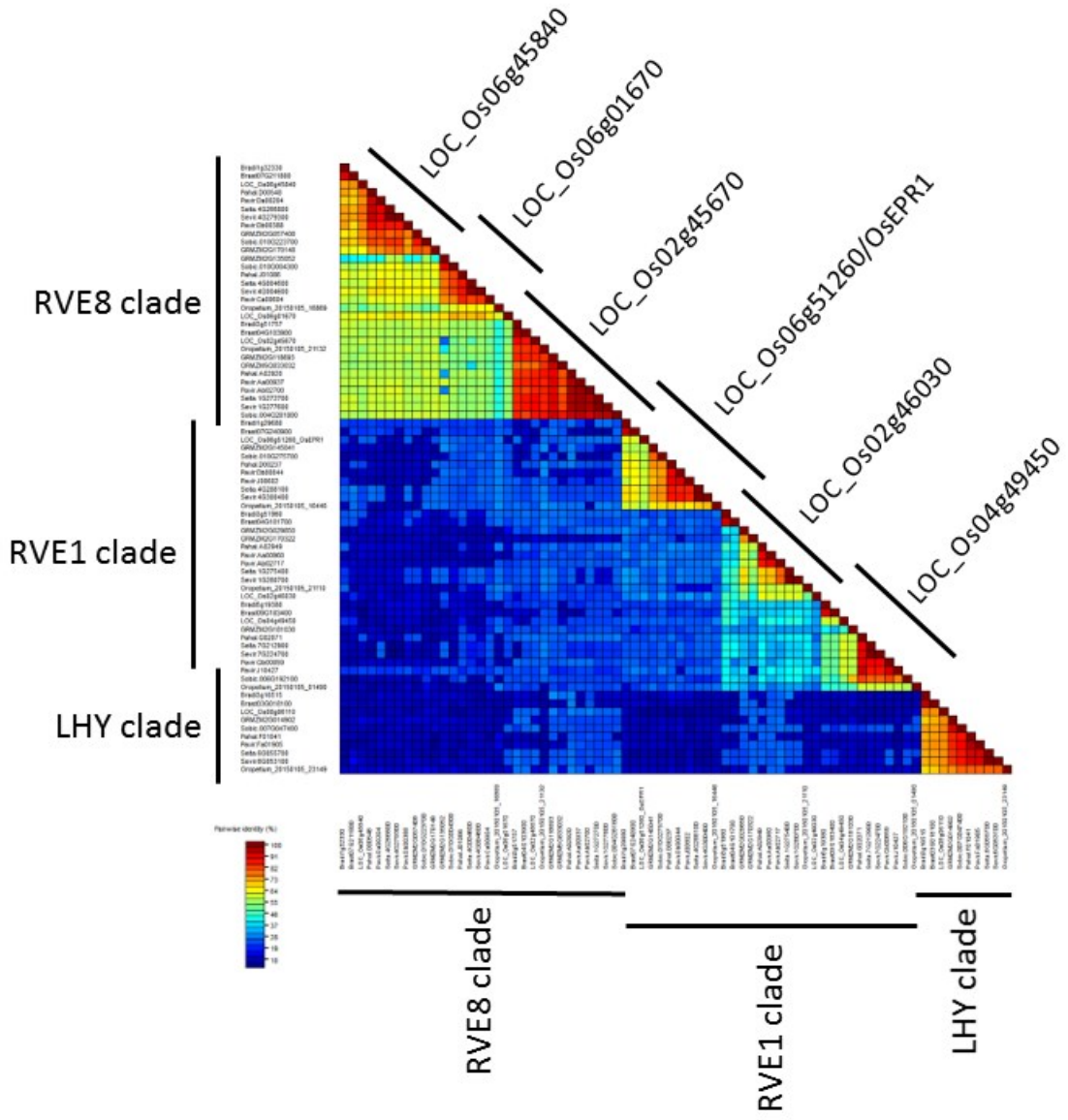


Figure 4.7 Identity matrix for the putative monocot CCA1-LHY-RVE proteins. Identity matrix of the 74 proteins identified in Table 4.1. MSA of full-length proteins was performed using CLUSTALW in SDTv1.2. Colour coded matrix highlights pairwise identity score for each MSA generated. Identity scores calculated as 1-(M/N) where M is the number of mismatching nucleotides and N the total number of positions along the alignment at which neither sequence has a gap character. Major clades have been identified and highlighted. Subclades have been identified and named based on model organism *Oryza sativa* subsp. *japonica* corresponding loci.

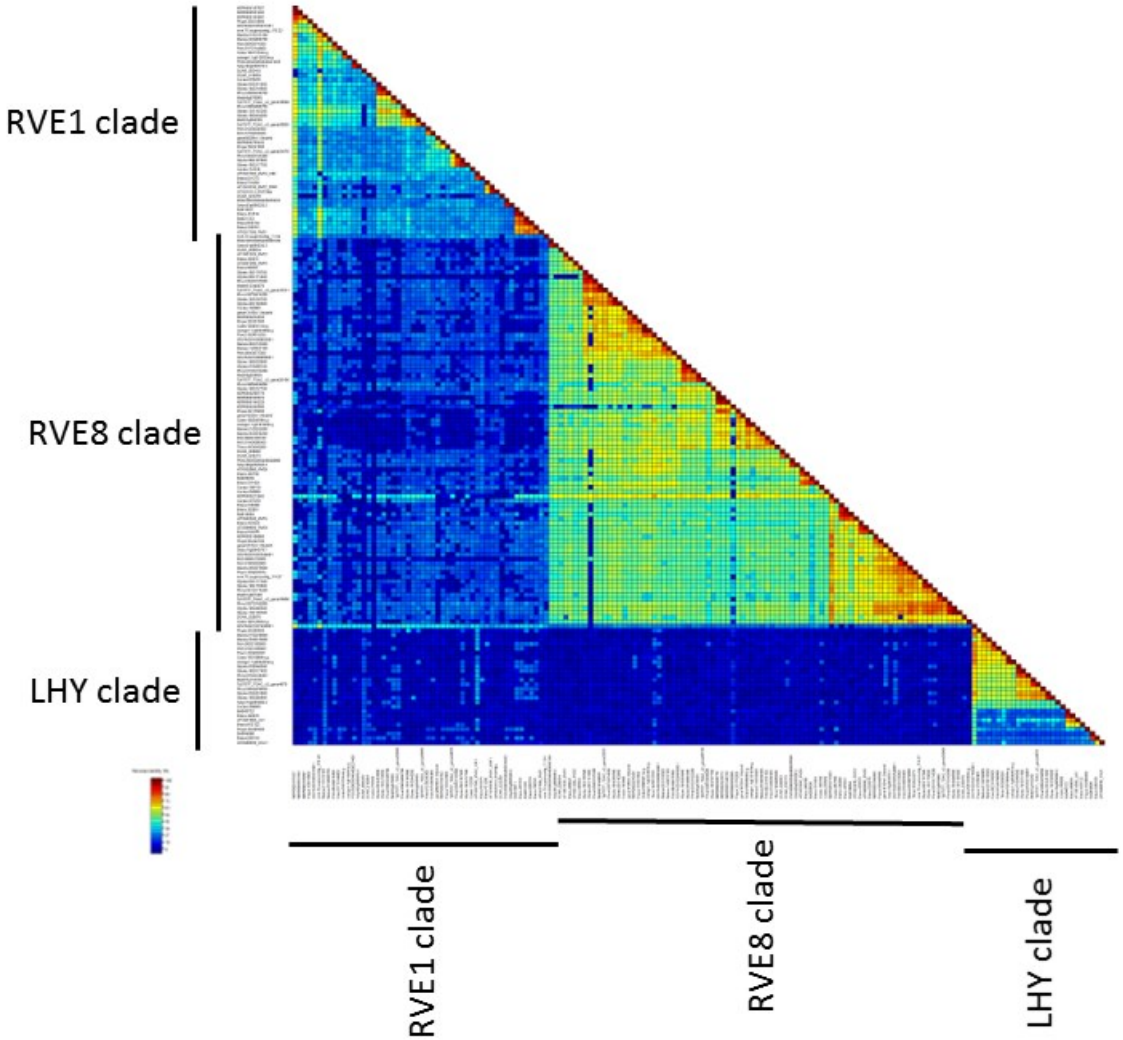


Figure 4.8 Identity matrix for the putative dicot CCA1-LHY-RVE proteins. Identity matrix of the 165 proteins identified in Table 4.2. MSA of full-length proteins was performed in using CLUSTALW in SDTv1.2. Colour coded matrix highlights pairwise identity score for each MSA generated. Identity scores calculated as 1-(M/N) where M is the number of mismatching nucleotides and N the total number of positions along the alignment at which neither sequence has a gap character. Major clades have been identified and highlighted.

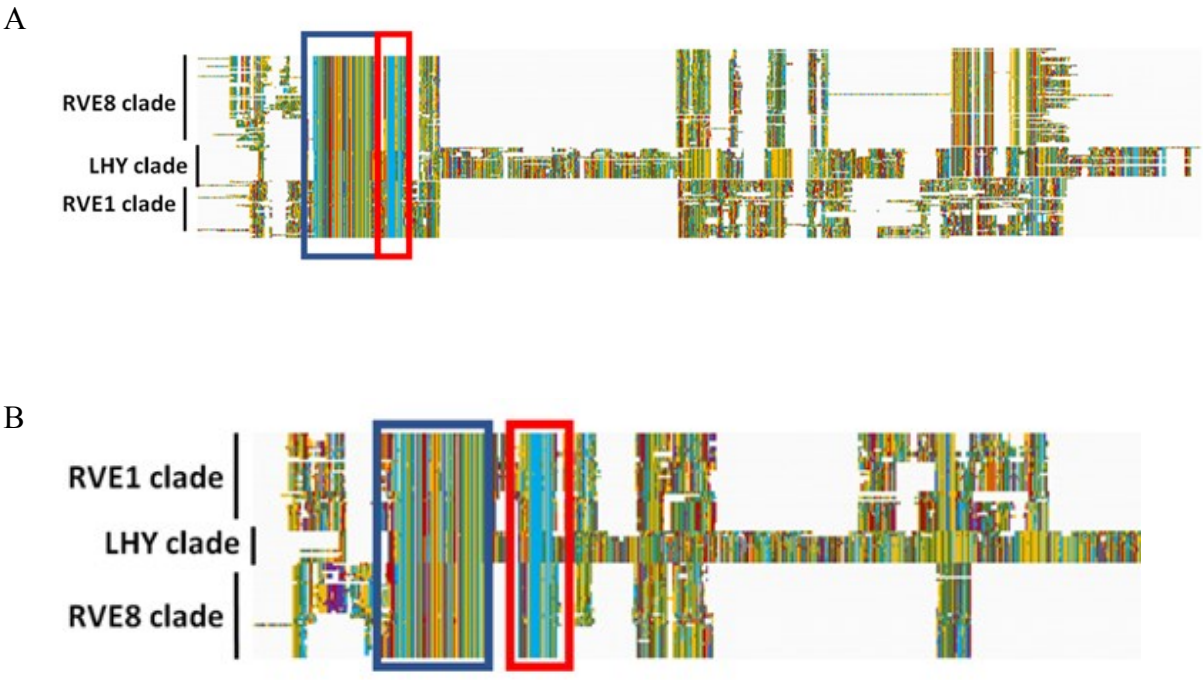
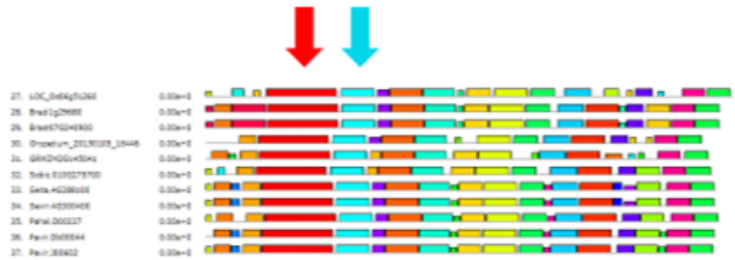


Figure 4.9 Multiple sequence alignment of the full-length protein sequences for all putative monocot and dicot CCA1-LHY-RVE proteins. (A) The MSA of 165 dicot RVE protein sequences, and (B) 74 monocot RVE sequences. In each panel, the proteins are ordered based on the phylogenetic relationships defined above. The blue box denotes the conserved MYB DNA-binding domain; the red box denotes the conserved proline rich domain. Colour code represents amino acid residues-based side chain properties. Blue: K, R, P. Red: D, E. Light blue: H. Grey: C. Green: S, Q, T, N. Light green: W, Y. Yellow: V, A, L, I, F, M. Purple: G. White: gap.

A



B

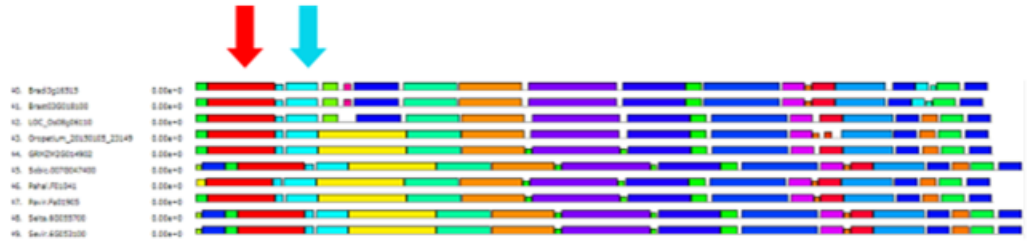


Figure 4.10 EPR1 and LHY clade proteins are conserved over their full lengths. (A) EPR1 orthologs in 10 monocot species, and (B) LHY orthologs in 10 monocot species Colour blocks represent conserved motifs. Red arrow indicates MYB DNA binding domain, which is shown as a red box, and pale blue arrow indicates proline rich domain, shown as a pale blue box.

CHAPTER V: Development of RT-qPCR assays for members of the *CCA1-LHY-RVE* family and measurement of the gene expression upon heat stress

5.1 Introduction

In this chapter, I developed an assay to measure the transcript abundance of six members of the *CCA1-LHY-RVE* family using a quantitative reverse transcription polymerase chain reaction (qRT-PCR) assay. After developing the assay, I measured the abundance of the *CCA1-LHY-RVE* transcripts at six times of day to determine which of them had diel patterns of transcript accumulation and at what times peak abundance was detected. I also measured transcript abundance in rice leaves that were exposed to brief periods of high temperature. Because of the predicted role of the *CCA1-LHY-RVE* genes at the interface of circadian rhythm and stress response, I measured their transcript abundance in rice plants exposed to high temperature stress at one of three times of day, at either 0zt, 4zt, or 8zt.

The first objective was to develop qRT-qPCR assays to measure the abundance of three housekeeping (*OsEF1*, *OsSUI1* and *OsGAPDH*) genes and six genes members of the *CCA1-LHY-RVE* family. To achieve this objective, I followed a four-step plan comprising the 1) design of qRT-PCR expression primers that were specific to each of the target genes, 2) validation of the specificity and fidelity of the primers by melt curve and gel electrophoresis analysis, 3) assessment of the stability of the transcript abundance of the housekeeping genes across the experimental conditions, and 4) evaluation of the efficiency of PCR assays through standard curve quantification. The qRT-PCR assays were developed in agreement with the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines to provide the minimal information necessary to reach publication standards for RT-qPCR data [208]. The second objective was to use the validated qRT-PCR assay to quantify the *CCA1-LHY-RVE* transcripts in order to detect their circadian oscillations and monitor their responses to heat stress. To achieve this objective, transcript abundance of the *CCA1-LHY-RVE* transcripts was relatively quantified using the stable housekeeping genes identified above

upon diurnal and heat shock conditions. Ultimately, the aim of this objective was to determine if the response to high temperature of these genes is influenced by their endogenous circadian oscillation pattern.

5.2 Results

Development of specific and efficient qRT-PCR assays

Primer pairs were designed for each locus using primerBLAST (Table 5.1). Primer were designed to meet four criteria seeking to maximize their efficiency and reliability. These criteria were 1) amplicon length between 70-160 bp, 2) an approximate primer length of 20 bp, 3) a melting temperature between 58-60°C with a maximum Tm difference between primers in a pair of 2°C, and 4) a GC content close to 50%. Primer specificity was confirmed using gel electrophoresis, and PCR assays were required to produce a single band of size that matched the predicted amplicon lengths (Fig 5.3). Results show that all of the primer pairs, except *OspUBQ*, amplified a single product of the predicted size. *OspUBQ* primer pair displayed multiple bands with the wrong amplicon length and was, therefore, excluded for further relative quantification of the *CCA1-LHY-RVE* target genes.

A second independent assay of primer specificity was performed by measuring the melting temperature of the amplicons (Table 5.2). From this assay, it was possible to confirm if a single amplicon was produced based on the shape of the melt curve, and to determine if the amplicon melted at the theoretically predicted temperature. The predicted amplicon melting temperatures obtained from uMELT were then compared to the observed melting temperatures which were obtained by RT-qPCR amplification of the serial diluted cDNA samples (Table 5.2). All the tested primer pairs displayed only one amplification peak, thus confirming the specificity of the reaction (Fig 5.4). Furthermore, the analysis confirmed that for all, but two of the PCR assays, the predicted melt temperatures were ≤ 1°C of the observed melting temperatures (Table 5.2). This provided confirmation that the amplified products were not only of the expected size, but also of the expected base pair composition. Two primer pairs, LOC_Os04g49450 and *OspUBQ*,

were not in the ≤ 1°C range between their observed melting temperatures and predicted melting temperatures as both displayed a 2.5°C Tm difference.

Afterward, a standard curve experiment based on a five-fold serial dilution was performed using cDNA synthesized from high quality RNA (Table 5.4 and 5.5, and Fig. 5.7) to evaluate the efficiency of the primers (Table 5.3 and Fig. 5.6). As per the MIQE guidelines, efficiency threshold was fixed between 90-110% which consequently indicates that slope ratio is between -3.1 and -3.6 and that the R^2 score is ≥0.99. All the studied primer pairs met this efficiency threshold of 90-110%, except *OspUBQ* which had an efficiency of 59.3% (Table 5.3). To find the most stable housekeeping genes, an analysis of housekeeping gene stability was performed on *OsEF1*, *OsSUI1*, *OsGAPDH* and *OspUBQ* using geNorm (Fig 5.8). In this analysis, both *OsSUI1* and *OsGAPDH* had the highest stability with both having an average M value of 0.496 and a stability of 0.700 Ln(1/Avg M). Overall, *pUBQ* had the lowest stability (0.184 Ln(1/Avg M) and 0.832 average M value). Based on these results and on the housekeeping gene stability analysis, *OspUBQ* was deliberately excluded for future relative quantification procedures. *OsEF1* stability was higher than *pUBQ* and considered "acceptable" per geNorm standards. However, only *OsSUI1* and *OsGAPDH* were found suitable for the relative quantification of the transcript abundance as their stability values were higher than OsEF1 and considered "idea" per geNorm standards.

Daily circadian oscillation pattern is not uniform amongst the members of the CCA1-LHY-RVE family

The daily circadian oscillation pattern was measured (Fig. 5.9) and statistically analyzed for six members of the *CCA1-LHY-RVE* family (Table 5.6; D vs D, 4 h gap). In total, four genes were differentially abundant at least once amongst the five studied timelapses (LOC_Os02g45670, *OsLHY*, LOC_Os04g49450, and *OsEPR1*). Analysis of diurnal vs diurnal treatments reveals that for all the conditions studied, only two genes, LOC_Os06g45840 and LOC_Os06g01670, were not differentially abundant. On the four differentially abundant genes, only *OsEPR1* was differentially abundant more than once. Overall, LOC_Os02g45670, OsEPR1 and OsLHY were all differentially abundant at

timelapse 14zt vs 22zt. Conversely, LOC_Os04g49450 was differentially abundant at early in the morning at timelapse 0zt vs 4zt. Finally, *OsEPR1* was also differentially abundant in the middle of the day at timelapse 8zt vs 12zt.

Daily circadian oscillation pattern is not modified amongst the members of the CCA1-LHY-RVE family upon heat shock

The daily circadian oscillation pattern was measured (Fig. 5.10) and statistically analyzed for the six members of the *CCA1-LHY-RVE* family upon a 30 min heat shock treatment (Table 5.6; HS vs HS, 4 h gap). This pattern was studied for two timelapses, 0zt vs 4zt and 4zt vs 8zt. Upon heat shock, none of the studied genes were differentially abundant for any of the studied timelapses. LOC_Os04g49450 was differentially abundant upon normal conditions at 0zt vs 4zt, but not under heat stress. This suggests a modification in the expression pattern for this gene upon heat shock for this timelapse.

LOC_Os02g45670 and LOC_Os04g49450 are differentially abundant upon heat stress.

Differential gene expression between diurnal and heat shock conditions (Fig. 5.9 and 5.10) was measured by comparing diurnal and heat shock treatments for the same timepoint (Table 5.6; HS vs D, 0 h gap). A comparative analysis of the heat shock and diurnal treatments reveals that only two genes were differentially abundant upon those conditions, LOC_Os02g45670 both at timepoints 0zt and 8zt, and LOC_Os04g49450 at 0zt. These results suggest that these genes respond rapidly to heat shock at these timepoints, whereas the others are either non-heat responsive at these timepoints and/or for the specific conditions of the heat shock treatment.

5.3 Discussion

Time of the day contributes to differential transcriptional response of OsEPR1 and LOC_Os06g45840 to heat stress

Overall, most of the genes studied in this experiment did not show differential level of abundance upon a 30 min heat stress in this experiment. Only LOC_Os06g45840 and LOC_Os06g01670 were downregulated upon this treatment. In a microarray analysis, *OsEPR1* and LOC_Os06g45840 were upregulated in rice flag leaf upon a 20 min, a 60 min and a 2 h heat shock at 40°C [209]. Similarly, in field studies, *OsEPR1* has been shown to be upregulated upon heat stress [7]. In arabidopsis, *AtCCA1* is differentially upregulated at 6zt, whereas *AtLHY* is differentially downregulated, upon a 1 h heat stress at 37°C at 1zt [207]. Upregulation of *AtCCA1* and downregulation of *AtLHY* have been also observed with a 30 min heat stress treatment [210]. *AtRVE1-4* are differentially upregulated upon moderate cold stress at 10°C at 1zt, whereas *AtRVE2*, *AtRVE7*/*AtEPR1* and *AtRVE7-like* are differentially upregulated upon heat stress at 37°C [207]. At 6zt, all members of the *CCA1-LHY-RVE* family, except *AtRVE4*, *RVE7* and *RVE7-like*, are differentially upregulated upon moderate cold stress [207]. Similarly, still at 6zt, all the *CCA1-LHY-RVE* family members are also differentially upregulated upon heat stress, except *AtRVE4*, *AtRVE6*, *AtRVE8* and *AtLHY* [207]. Overall, the RT-qPCR data generated in this study are consistent with the literature for LOC_Os06g45840 being upregulated upon heat stress, but not for *OsEPR1* upon the same conditions.

Differences in expression results can be observed between the Blair et al. (2019) dataset and Albihlal et al. (2018) datasets as *AtCCA1* is not significantly upregulated in the former, and significantly upregulated in the latter. Explanations supporting these differences, as provided by Blair et al. (2019), are the growth conditions, duration of 37 °C treatment, and difference in the analysis pipeline. Along with these explanations, the choice of housekeeping genes (*OsSUI1* and *OsGAPDH*) may also have influenced the data generated in this experiment. Altogether, these factors may explain why *OsEPR1* is not differentially upregulated in my experiment. Additional trials, through a multi-

temporal experimental analysis and with other housekeeping genes, might be necessary for a more in-depth analysis of the expression pattern of the *CCA1-LHY-RVE* family.

Time of the day contributes to differential expression in all RVE genes, except LOC_Os06g45840 and LOC_Os06g01670, upon diurnal conditions

In arabidopsis, some members of the *CCA1-LHY-RVE* family, such as *AtCCA1* and *AtLHY*, are known to oscillate at the mRNA level [100]. At least two previous published microarray circadian datasets for arabidopsis demonstrate the specific oscillation patterns by the members of this family [11,211,212]. In these published datasets, *AtCCA1* and *AtLHY* have been demonstrated to be upregulated at the beginning of the day. Subsequently, the transcript abundance reduces over the day until it reaches his minimal expression around 12zt, the middle of the day. In my experiment, *OsLHY* follows the same circadian oscillation with a peak in mRNA abundance at the beginning of the day and a basal expression between 12zt and 14zt. In the case of *OsLHY*, only two timepoints were differentially abundant (8zt vs 12zt and 14zt vs 22zt) which respectively correspond to the moment when it reaches its lowest and highest transcript abundance.

Two members of the RVE8 clade, LOC_Os06g45840 and LOC_Os06g01670, did not show any differential expression for all the timepoints analyzed (D vs D). Overall, production of transcripts for those genes was generally low and stable for all the timepoints. In arabidopsis, three genes member of the RVE8 clade (*AtRVE4*, *AtRVE5* and *AtRVE6*) do not feature a pronounced oscillation pattern when compared to the rest of the genes forming the *CCA1-LHY-RVE* family. In a similar fashion to LOC_Os06g45840 and LOC_Os06g01670, these three genes demonstrate a low circadian expression over the course of the day suggesting that the low constant circadian oscillation pattern of the RVE8 clade has been conserved in rice and arabidopsis. In the published dataset for arabidopsis, *AtEPR1* displays an increase in expression from 0zt to 8zt [11,211,212]. This trend has not been observed for *OsEPR1* as revealed by the data from my study. Overall, in my dataset, *OsEPR1* shows a plateau in gene expression from 0zt to 12zt which follows a decrease from 12zt to 14zt. Subsequently, *OsEPR1* demonstrates a swift increase in expression from 14zt to 22zt. Altogether, *OsEPR1* is the only gene showing a

differential expression for two timelapses (8zt vs 12zt and 14zt vs 22zt). Overall, based on the differences between my dataset for rice and the published datasets for arabidopsis, it is possible to state that *OsEPR1* and *AtEPR1* do not follow the same circadian oscillation pattern. This novel oscillation pattern in rice suggests a modification of the molecular function for *OsEPR1*, but further research is required to assess its role and extent on the prevailing downstream molecular processes.

5.4 Tables and figures

Table 5.1 Primer sequences and amplification properties for RVE and housekeeping genes.

Target mRNA		Primer sequence	Amplicon length (bp)	Tm primer (°C)	GC (%)
OsEF1, LOC_Os07g43730	F	AAGAGGAAGTCAGCGGCTAAG	77	59.79	52.38
	R	CAGAATGGGCAGGAAAATACA		56.20	42.86
OspUBQ, LOC_Os06g46770	F	CAGCAGCGCCTCATCTTC	67	58.59	61.11
	R	GGATGTTGTAGTCAGCCAAGG		58.64	52.38
OsGAPDH, LOC_Os04g40950	F	AAGCCAGCATCCTATGATCAGATT	79	59.96	41.67
	R	CGTAACCCAGAATACCCTTGAGTTT		60.80	44.00
OsSUI1, LOC_Os07g34589	F	GCTGCAATGGTACTGTTGTCC	100	59.80	52.38
	R	CCGGCCTGAACAAGAAAAT		56.10	47.37
LOC_Os02g45670	F	TGGTGAAAGAAGATTTAGGTGCT	97	58.01	39.13
	R	GTTTCACCAGGTTGCCATGC		60.32	55.00
LOC_Os06g45840	F	AATGCTTGTGCAGTTTCGCC	131	60.32	50.00
	R	TCCCAGGAGGGTACTACAGC		60.03	60.00
LOC_Os06g01670	F	TCCTGGGATAGTGCTCTTGC	93	59.17	55.00
	R	TCCAGATTGGCTCTCTACGC		59.25	55.00
OsLHY, LOC_Os08g06110	F	AGCTTTTCTCCTCCGCAAGT	130	59.60	50.00
	R	CATCCGCTGTGTCGAGTTCT		60.11	55.00
LOC_Os04g49450	F	AGGGAAGGATCTTGGACGGA	159	59.96	55.00
	R	GGAACCTGGGCTCACACATT		60.25	55.00
OsEPR1, LOC_Os06g51260	F	GCTTCTTCATCGGTTGGGGA	132	60.04	55.00
	R	CCACACGTGTAAAGGGGGAA		59.89	55.00

Table 5.2 Theoretical and measured amplicon melting temperatures

Target genes	Predicted Tm (°C)	Observed Tm (°C)
OsEF1	80.5	80.5
OspUBQ	86	88.5
OsGAPDH	80.5	81
OsSUI1	81.5	82.5
LOC_Os02g45670	84.5	84.5
LOC_Os06g45840	83	83
LOC_Os06g01670	79	79.5
OsLHY	81.5	81.5
LOC_Os04g49450	80.5	83
OsEPR1	81.5	81.5

Table 5.3 Validation of primer design and preparation of RT-qPCR samples based on RT-qPCR amplification.

Target genes	Efficiency (%)	Slope	R^2
OsEF1	93.1	-3.50	0.991
OspUBQ	59.3	-4.930	0.992
OsGAPDH	91.9	-3.532	0.990
OsSUI1	93.5	-3.488	0.990
LOC_Os02g45670	106.7	-3.171	0.992
LOC_Os06g45840	95.8	-3.427	0.990
LOC_Os06g01670	94.6	-3.460	0.999
OsLHY	91.2	-3.553	0.993
LOC_Os04g49450	97.8	-3.375	0.997
OsEPR1	101.0	-3.298	0.998

Table 5.4 Spectrophotometer assessment of the quality of the extracted RNA for samples submitted to diurnal treatment.

Samples	Concentration (volume/ul)	A260/A280	A260/A230
0zt diurnal rep 1	502.60	2.02	2.49
0zt diurnal rep 2	1061.80	2.07	2.28
0zt diurnal rep 3	504.30	2.05	2.16
4zt diurnal rep 1	580.90	2.05	2.18
4zt diurnal rep 2	638.50	2.03	2.21
4zt diurnal rep 3	466.20	2.09	2.93
8zt diurnal rep 1	515.10	2.06	2.23
8zt diurnal rep 2	605.40	2.03	2.23
8zt diurnal rep 3	446.90	2.10	2.97
12zt diurnal rep 1	352.60	2.13	2.31
12zt diurnal rep 2	415.60	2.14	2.33
12zt diurnal rep 3	630.60	2.04	2.21
14zt diurnal rep 1	414.80	2.13	2.32
14zt diurnal rep 2	436.00	2.12	2.24
14zt diurnal rep 3	651.10	2.08	2.05
22zt diurnal rep 1	159.70	2.07	2.20
22zt diurnal rep 2	492.30	2.13	2.29
22zt diurnal rep 3	616.10	2.08	2.15

Table 5.5 Spectrophotometer assessment of the quality of the extracted RNA for samples submitted to heat shock treatment.

Samples	Concentration (volume/ul)	A260/A280	A260/A230
0zt heat shock rep 1	284.50	2.07	2.15
0zt heat shock rep 2	372.20	2.10	2.31
0zt heat shock rep 3	386.90	2.11	2.20
4zt heat shock rep 1	269.40	2.09	2.04
4zt heat shock rep 2	316.00	2.07	2.19
4zt heat shock rep 3	307.20	2.11	2.17
8zt heat shock rep 1	238.40	2.10	2.19
8zt heat shock rep 2	212.50	2.07	2.17
8zt heat shock rep 3	218.30	2.10	2.14

Table 5.6 ANOVA results for p-value Tukey of RT-qPCR assays for diurnal and heat shock treatments.

Target mRNA	Conditions									
	HS vs HS		HS vs D			D vs D				
	0zt vs 4zt	4zt vs 8zt	0zt vs 0zt	4zt vs 4zt	8zt vs 8zt	0zt vs 4zt	4zt vs 8zt	8zt vs 12zt	12zt vs 14zt	14zt vs 22zt
LOC_Os02g45670	1.0000	0.2493	0.0103 *	0.0547	0.0251 *	0.9883	0.8076	0.1385	0.1385	0.0008 *
LOC_Os06g45840	0.9999	0.5776	0.9631	0.9771	0.9965	0.2560	0.9999	0.9998	0.9416	0.9829
LOC_Os06g01670	0.9020	0.9993	0.8112	0.9714	0.3345	0.6561	0.9973	0.9726	1.0000	0.9494
OsLHY	0.3891	0.3959	0.9749	0.8249	0.5861	0.6733	0.6424	0.0785	0.9999	0.0398 *
LOC_Os04g49450	0.4915	0.6499	0.0313 *	0.9999	0.9989	0.0002 *	0.9987	0.9983	0.3036	0.0525
OsEPR1	0.7452	0.9991	0.9999	0.9956	0.9991	0.4863	0.9959	0.0073 *	0.2169	0.0004 *

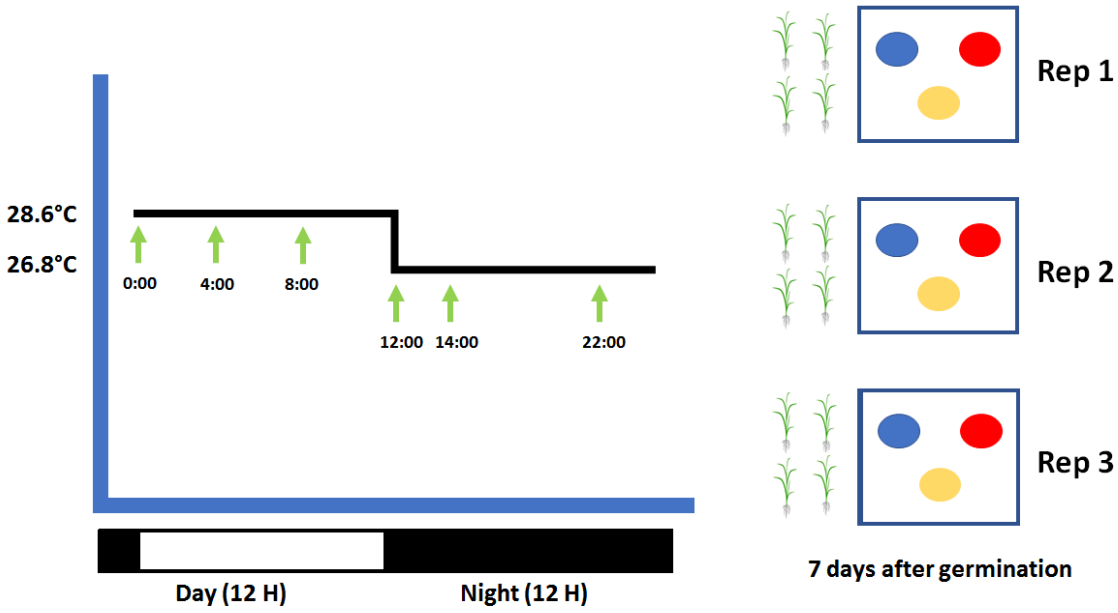


Figure 5.1 Schematic representation of diurnal treatment experimental protocol. Green arrows represent collection timepoints. Day temperature was set at 28.6°C and night temperature at 26.8°C. Three biological replicates were harvested for each timepoint. Each biological replicate comprised three technical replicates (yellow, blue and red dots) of four rice plants each. Rice plants were harvested seven days after germination.

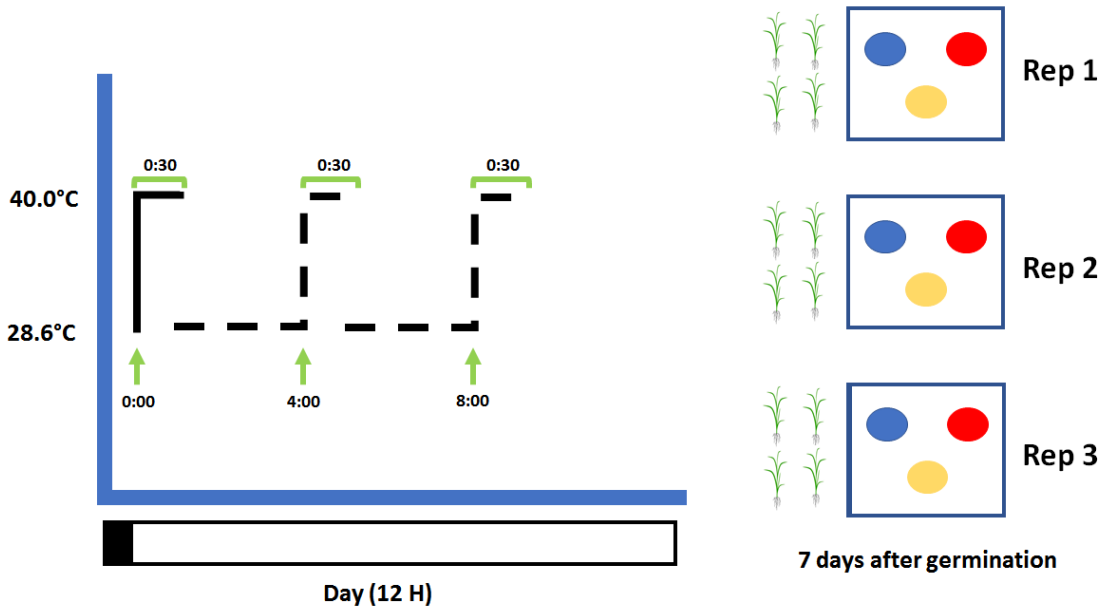


Figure 5.2 Schematic representation of heat shock treatment experimental protocol. Green arrows represent collection timepoints. Day temperature was set at 28.6°C and increased at 40°C during 30 minutes after which the plants were harvested by freeze drying. Three biological replicates were harvested for each timepoint (0zt, 4zt and 8zt). Each biological replicate comprised three technical replicates (yellow, blue and red dots) of four rice plants each. Rice plants were harvested seven days after germination.

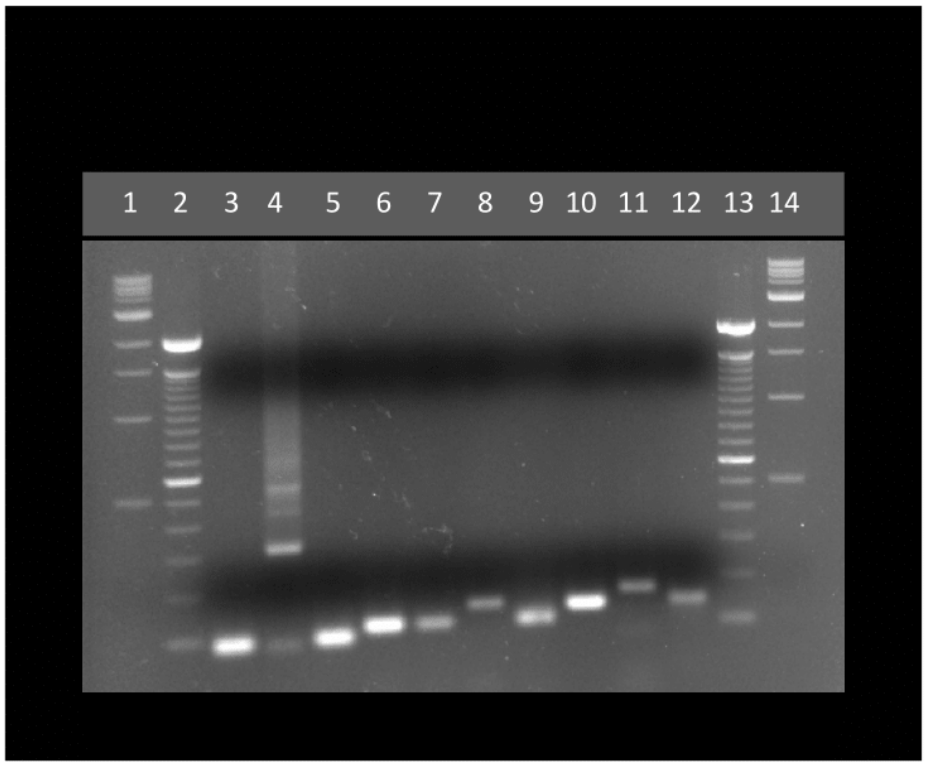


Figure 5.3 Primer specificity is confirmed. RT-PCR amplification performed on pooled cDNA rice samples (36 ng total RNA) subjected to heat stress treatment, timepoint 0zt. Electrophoresis was performed on a 2 % agarose gel electrophoresis performed with 1X TAE buffer. Amplicon size as indicated in Table 5.1. Lane 1, 1kb ladder; Lane 2, 100 bp ladder; Lane 3, *OsEF1*; Lane 4, *OspUBQ*; Lane 5, *OsGAPDH*; Lane 6, *OsSUI1*; Lane 7, LOC_Os02g45670; Lane 8, LOC_Os06g45840; Lane 9, LOC_Os06g01670; Lane 10, *OsLHY*; Lane 11, LOC_Os04g49450; Lane 12, *OsEPR1*; Lane 13, 100 bp ladder; Lane 14, 1kb ladder.

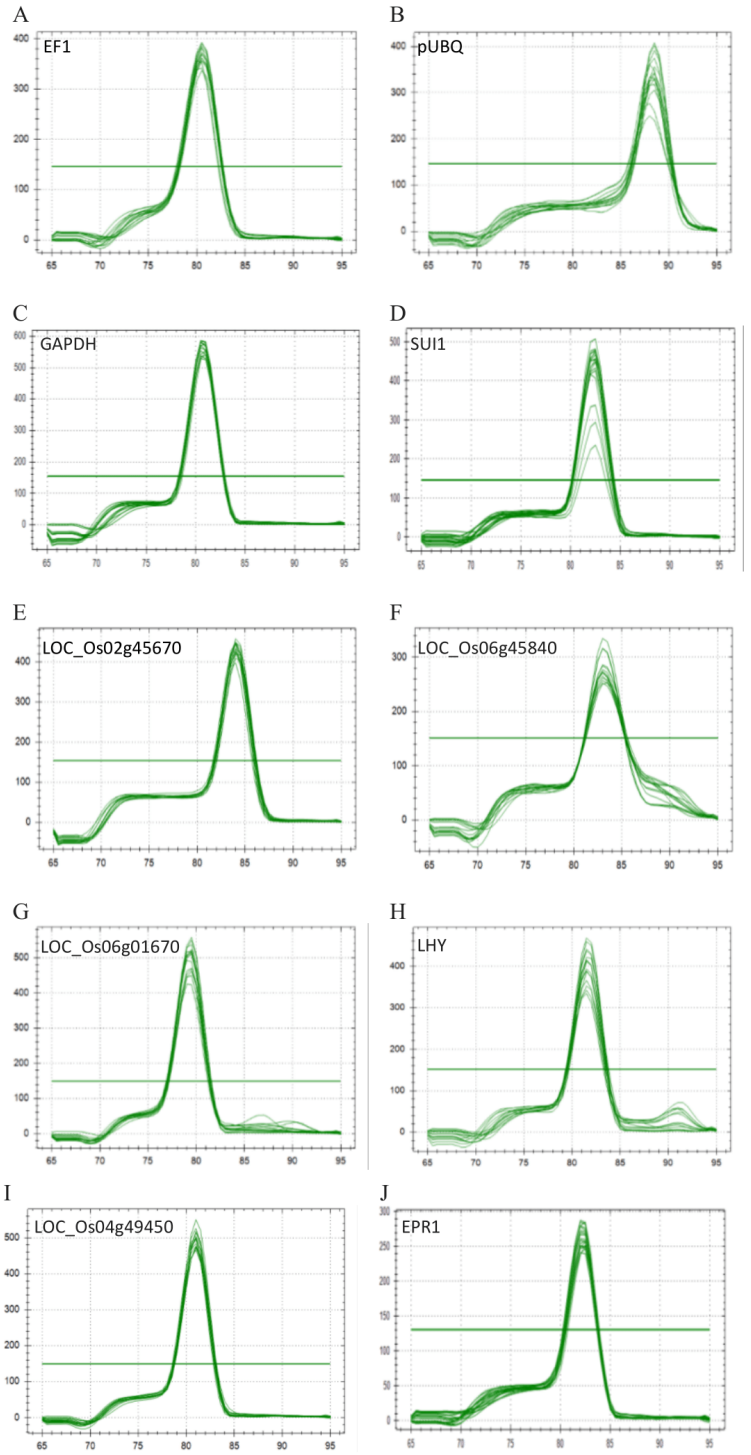
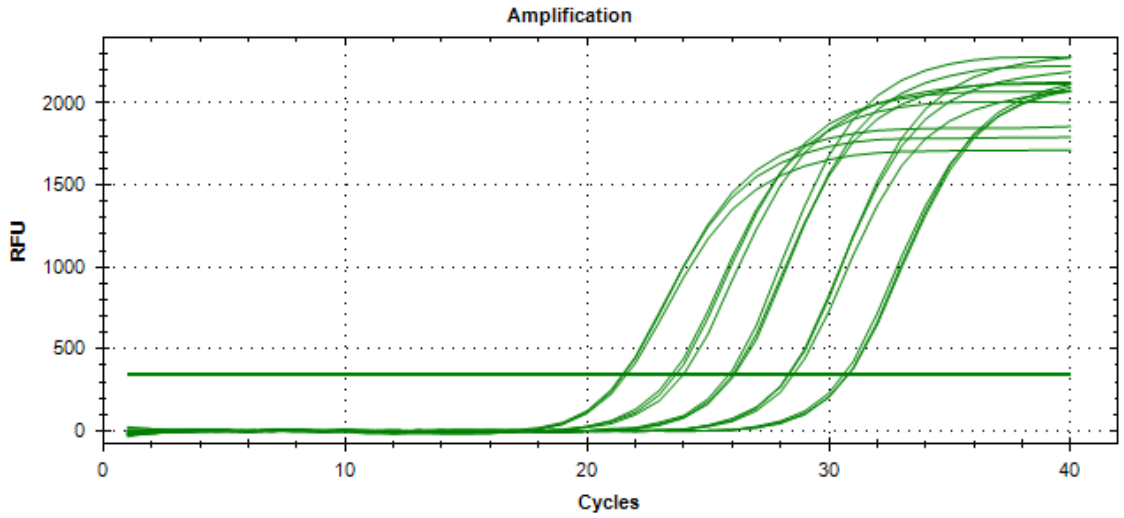


Figure 5.4 Fidelity of PCR assays was confirmed by melt curve analysis. The fidelity of each PCR assay was determined by monitoring the fluorescence of SYBR green as temperature was increased from 65°C to 95°C. The transformed fluorescence intensity is plotted as a function of temperature. (n=15 wells). A single peak indicates that a simple product was amplified. X axis, temperature in degree Celsius; Y axis, fluorescence in -d(RFU)/dT.

A



B

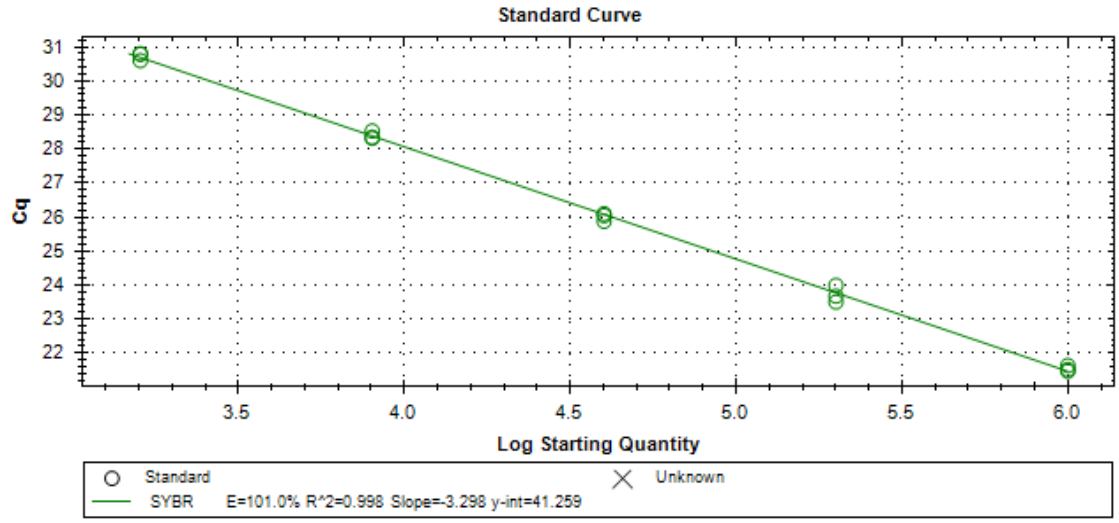
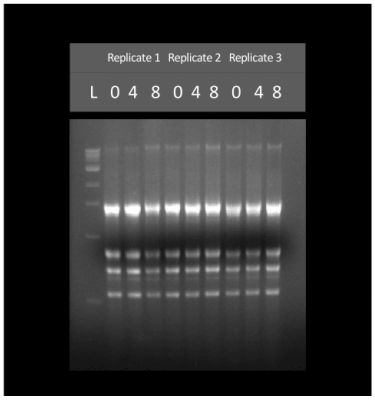
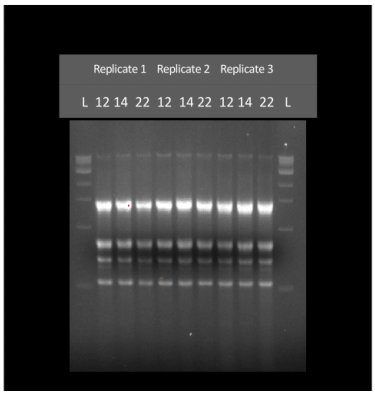


Figure 5.5 Amplification plot and standard curve for the *OsEPR1* PCR Assay. (A) Examplar amplification plot of a one-in-five dilution series of genomic DNA for *OsEPR1*. (B) Exemplar standard curve for one-in-five dilution series for *OsEPR1*. Each dilution series has been performed in triplicates.

A



B



C

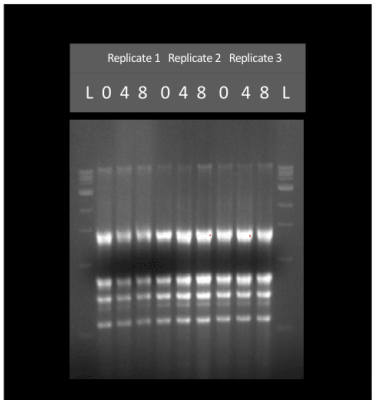


Figure 5.6 High quality total RNA was extracted from rice leaves. (A) Diurnal treatment, 0zt, 4zt and 8zt, (B) Diurnal treatment, 12zt, 14zt and 22zt, and (C) Heat shock treatment, 0zt, 4zt and 8zt. Timepoints are indicated by the following numbers: 0, 0zt; 4, 4zt; 8, 8 zt; 12, 12zt; 14, 14zt; 22, 22zt. Ladder is indicated as L, 1 kb ladder.1 ug total RNA was incubated prior to gel electrophoresis at 60°C during 5 min to inhibit the formation of secondary structures

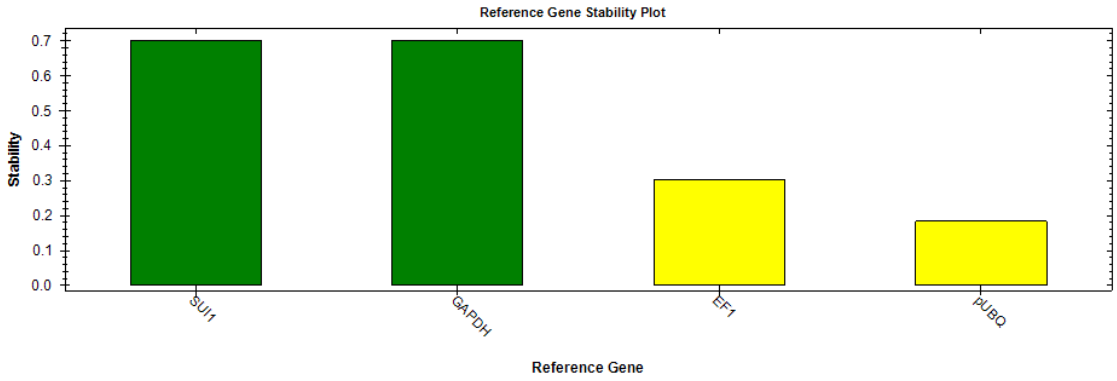


Figure 5.7 *OsSUI1* and *OsGAPDH* are the most stable housekeeping genes. Green, ideal stability. Yellow, acceptable stability. Stability unit is Ln(1/AvgM). Stability was calculated based on a geNorm analysis of heat shock and diurnal treatments for timepoints 0zt, 4zt and 8zt. geNorm analysis was performed on three replicates for each timepoint.

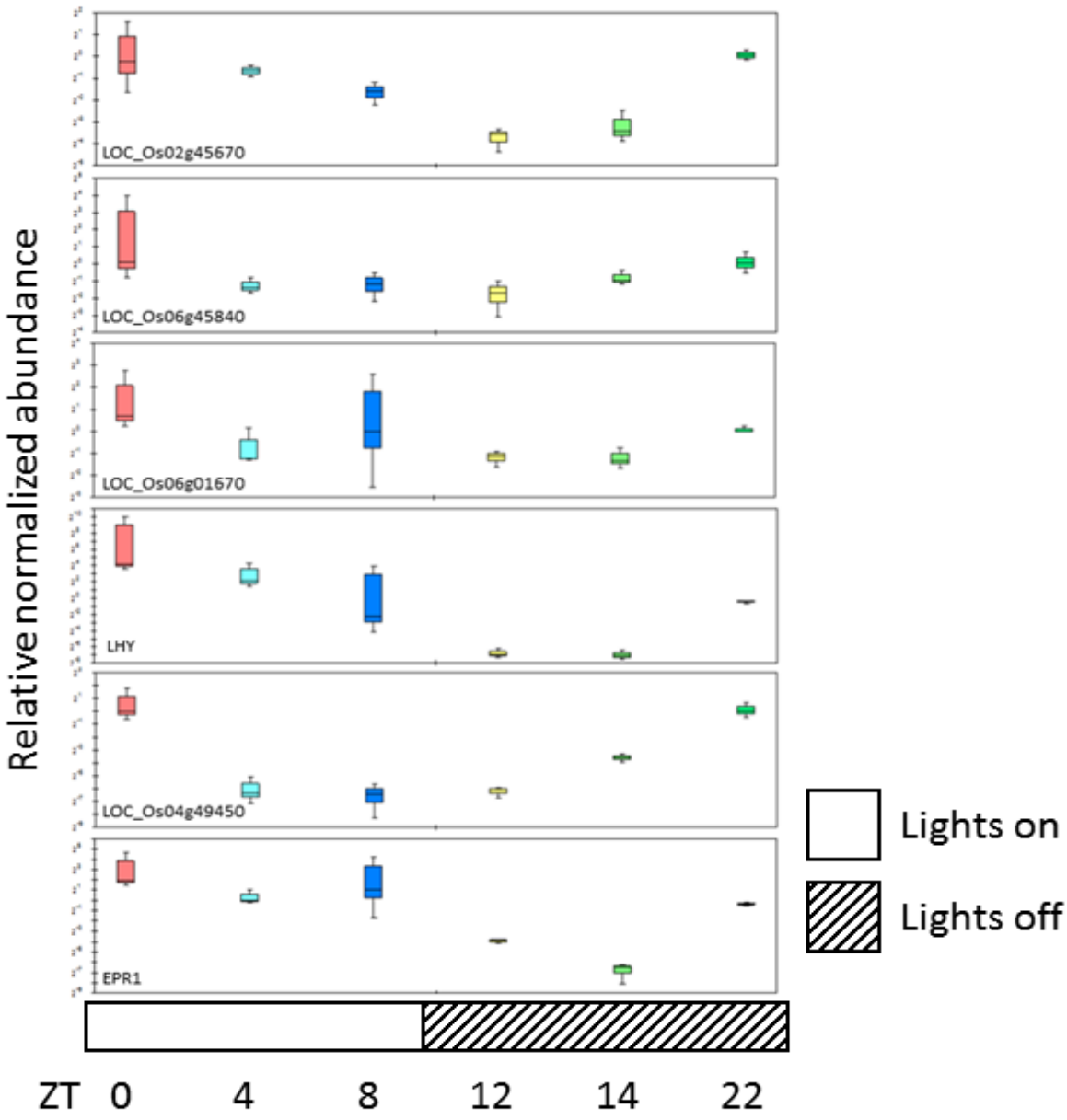


Figure 5.8 Transcript abundance for six RVE genes in response to diurnal treatment. RT-qPCR amplification results for all diurnal treatments on seven-day old rice plants. Error bars show 95% confidence interval of a minimum of three biological replicates. Results shown for all six timepoints and normalized with two housekeeping genes, *OsSUI1* and *OsGAPDH*.

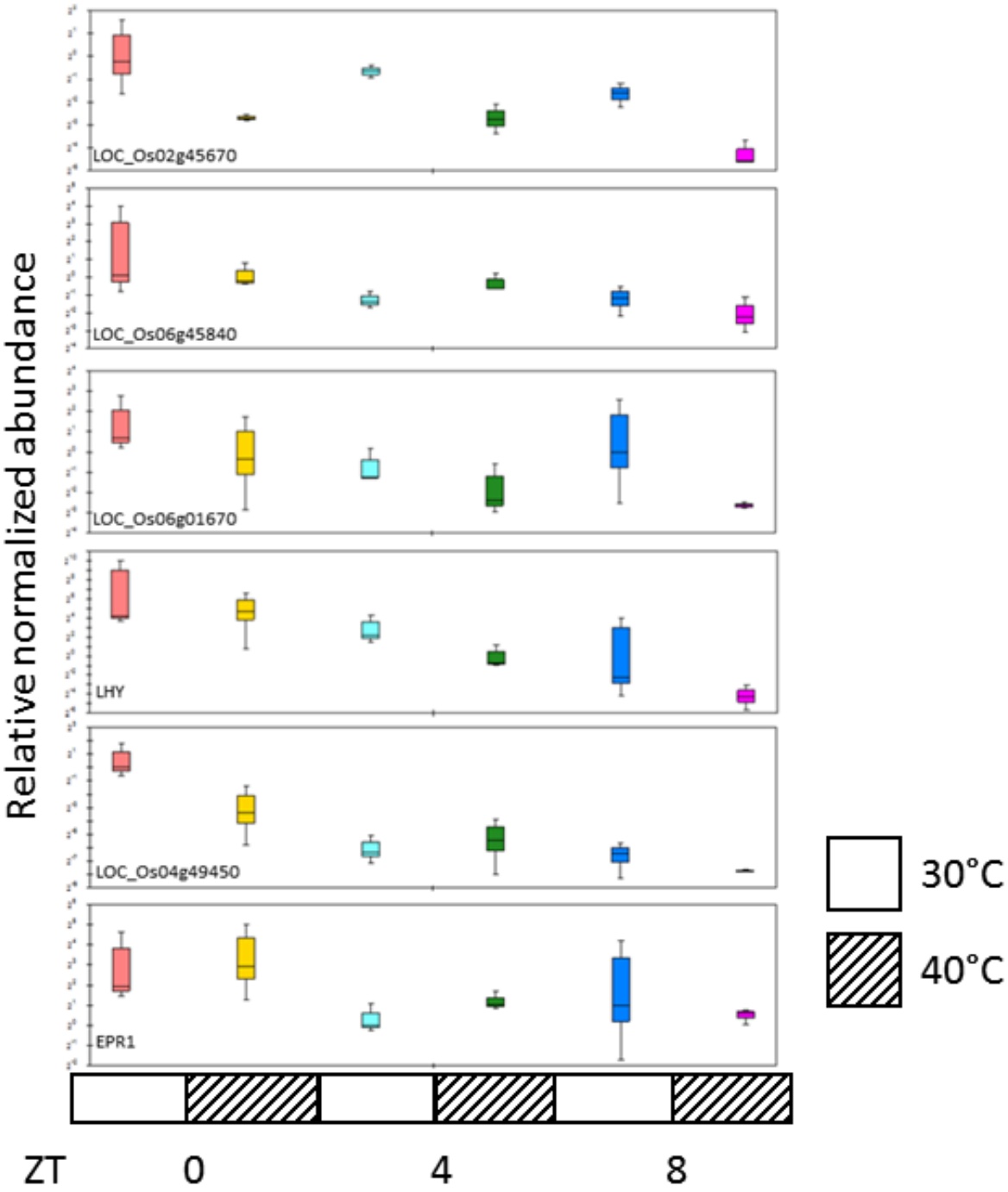


Figure 5.9 Transcript abundance for six RVE genes in response to heat shock treatment. RT-qPCR amplification results for all heat shock treatments on seven-day old rice plants. Results shown for all six timepoints and normalized with two housekeeping genes, Os*SUI1* and Os*GAPDH*.

CHAPTER VI: General conclusion

In the last decades, scientists have been able, pieces by pieces, to decipher parts of the intricate circadian clock machinery. Many years have gone by since the first observations on the daily leaf movements of the tamarind tree, *Tamarindus indicus*, made by Androsthenes in the fourth century, and those experimentally proven by French astronomer de Mairan on the daily leaf movements of *Mimosa pudica* in 1729 [213–215]. Nowadays, the story is very different as progress is much more sustained with scientific papers being published almost everyday. Contemporary advances in the field are mainly achieved using molecular techniques and observations are made using not *T. indicus* nor *M. pudica*, but *A. thaliana* as a model plant because it harbours multiple advantages (small genome, diploid, fast grower, small size) for scientists [177]. The current state of research on the circadian clock in arabidopsis offers promising avenues in terms of targets for breeding, whether it be for the core clock or peripheral clock genes, and the functional characterization of some clock elements for their role in the plant heat stress response in relation to the circadian rhythm of the plant is progressing. However, the pace of progress is still very slow due to the daunting densities of these networks.

Throughout the years of research, over 20 key clock players have been identified and characterized in arabidopsis as well as several other promising peripheral genes. Past research on the clock has also shown that core and peripheral clock members could be involved in the abiotic stress tolerance in plants. Starting from those observations, the CCA1-LHY-RVE protein family, a group of 11 members in arabidopsis and seven in rice, has become a novel player bridging the abiotic stress tolerance and the circadian clock concepts. The general goal of this research was to expand our understanding of the structural, compositional and functional features of this protein family in other plant species, mostly cultivated ones. This study is the first to provide a detailed phylogenetic analysis of the protein family on a large scale, 10 monocot and 21 dicot species in total, as well as a thorough gene expression analysis of six genes member of this family depicting the circadian oscillation pattern upon heat shock and control conditions for several timepoints.

The first objective of this study was to investigate the phylogenetic relationships of the CCA1-LHY-RVE protein family in 31 plant species. In the literature, the circadian clock is most generally regarded as a conserved system. Factual data concerning the evolutionary schemes and phylogenetic relationships for a broad range of species and clock components are often missing to support this claim. The CCA1-LHY-RVE protein family is an interesting case study since the family encompasses the core clock components CCA1 and LHY, often described as the master regulators of the clock, and lesser-known genes from the RVE8 and RVE1 clades, thought to be peripheral in their actions on the clock. Generated data demonstrate that identity rates are high for the RVE8 and LHY clades in monocots and dicots, but not for the RVE1 clade. Overall, monocotyledonous plants displayed higher rates of conservation for the three protein clades in terms of the number of proteins and amino acid sequence identity in comparison to dicotyledonous plants. Indeed, structural composition of the protein family in the studied dicots is very disparate with a highly varying number of proteins in all clades, from three for *C. papaya* to 21 for *G. max*. Along with these observations, the CCA1-LHY-RVE protein family for these five species (*C. papaya*, *D. carota*, *F. vesca*, *S. tuberosum* and *M. domestica*) is characterized by a small number of proteins. These differences may be linked to the prevailing differences between the pipeline used this study and the pipelines used in the other experiments. Further analyses are required to assess what are the effects of a minimal number of genes on the clock general functions. Promoter scanning in rice and arabidopsis reveal that a minimal number of promoters in arabidopsis and rice harbour a Heat Shock Element (At, 1; Os, 1) and an Evening Element (At, 4; Os, 1) for the family. Altogether, these observations suggest that arabidopsis is not an optimal model plant to study the circadian clock for monocots. Furthermore, the high rates of disparity and phylogenetic divergence in dicots for the CCA1-LHY-RVE protein family suggests that other model plant species might be necessary to perform thorough analyses of the structural and functional features of circadian clock components.

In the second objective, a RT-qPCR assay was developed to measure the transcript abundance of six genes belonging to the *CCA1-LHY-RVE* family. Overall, gene specific primer pairs were successfully designed and validated for three housekeeping

genes and six RVE target genes using standard validation protocols for RT-qCPR (standard curves, gel electrophoresis, geNorm analysis and Tm comparison between predicted and observed). These primers pairs were subsequently used in the third objective to quantify the relative gene expression of the six genes member of the CCA1-LHY-RVE family. For each of the gene, the daily circadian oscillation upon diurnal and heat shock conditions was measured for respectively six and three timepoints. Comparative analysis between heat shock and diurnal treatments for the same timepoints demonstrate that two genes are differentially abundant upon those conditions, LOC_Os02g45670 both at 0zt and 8zt and LOC_Os04g49450 at 0zt. Additionally, analysis of the diurnal treatments between themselves with a 4 h gap reveal that only two genes, LOC_Os06g45840 and LOC_Os06g01670, are not differentially abundant upon those conditions. Under these conditions, only *OsEPR1* was differentially abundant for more than two timelapses (8zt vs 12zt and 14zt vs 22zt). The results generated in this study differ slightly from the published dataset available in the literature, but these differences might be related to the differences in the experimental pipeline as well as the housekeeping genes used. Finally, I demonstrated that *EPR1* has a different circadian oscillation pattern in rice than arabidopsis. This difference may suggest a different functional role, but more in-depth research remains necessary to confirm that statement.

As a whole, this research project shines the light on the CCA1-LHY-RVE protein family structural, compositional and functional characteristics in multiple species. The underlying objective of this project was to expand the current research on the circadian in other plant species that are more commonly used as crops than arabidopsis. Thought this study is not the first to tackle the evolutionary conservation and phylogenetic relationships of the CCA1-LHY-RVE protein family, it is the first to investigate this question using a large number of cultivated species. By doing so, I have been able to determine the pairwise sequence identity ratios for each putative sequence in 31 plant species. These ratios prove that the protein family is much more conserved in monocots for each clade and which suggests that, at the functional level, the specific roles of each gene may be more conserved within monocots than dicots. Similarly, this experiment is the first to tackle specifically the gene abundance of *CCA1-LHY-RVE* genes in rice using

RT-qPCR. As a result, an analysis and quantification pipeline has been built and is now available for six RVE genes and three housekeeping genes.

In terms of future directions, I have also laid the groundwork and generated insights for the expansion of this research project for other important parts of the circadian clock, such as the PSEUDO-RESPONSE REGULATOR group and the ELF3, ELF4, LUX, GI and ZTL proteins. Recent work concerning these clock components between various plants and algae species has demonstrated that a high degree of variation exists at the structural and compositional levels between each species [10]. Still, as in the case of the CCA1-LHY-RVE family, a knowledge gap needs to be filled concerning cultivated plant species [10]. Another avenue would be to quantify the expression pattern for each of these clock components in rice upon various abiotic stresses including heat shock. Along with that, I think that it may be necessary to further expand the available research protocol and perform a multi-temporal analysis for the heat shock assay (i.e. 30 min, 1 h, 2 h and 3 h) instead of a single time frame that was used in this experiment, 30 min. Relative quantification using other housekeeping genes may also be an appropriate additional step to validate the data generated in this study.

Novel molecular techniques set the scene to deciphering the intricate mechanisms controlled by the clock. In that regard, new technological advances in the field of genome editing offers new possibilities of research. The traditional method of functional characterization, as depicted by numerous cited studies in this thesis, is to overexpress and/or knocking out components of the clock and analyze the downstream phenotypes generated. Often, the phenotype generated is arrhythmic plants which prohibits the use of these plants for agricultural uses [216]. Genome editing with innovative systems, such as CRISPR-cas9 and TALENs, has been proposed as an alternative to circumvent these issues. Recent studies have use the CRISPR-cas9 system to target the cis-regulatory element in the promoter to induce slight quantitative variation in the expression of a transcription factor's target, as demonstrated with the CLAVATA-WUSCHEL stem cell circuit controlling meristem size in tomato which resulted in larger fruit size [217]. To my knowledge, no such experiment has been conducted on quantitative trait loci involved in the clock regulation, but this strategy could be efficiently used to overcome the

associated issues related to targeting the coding sequence of the core regulators involved in the circadian cascade.

As a final word, the world is changing rapidly. At the core of these changes are the issues concerning global warming, fast pace in technological development and science, rise in world population and the development of efficient communication systems. These concomitant changes represent challenges and opportunities for the current and future scientists. In a timely manner, the global network that is now the world scientific community now must find answers to solve these challenges using the new opportunities that represent novel technologies. In that sense, deciphering the intricate role of the circadian clock in molecular processes in model and cultivated plants embodies a promising avenue to solve these rising challenges.

BIBLIOGRAPHY

1 Griggs, D.J. and Noguer, M. (2002) *Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, 57

2 Leone, A. *et al.* (2012) Plant Tolerance to Heat Stress: Current Strategies and New Emergent Insights. *Abiotic Stress. Plants* DOI: 10.1007/978-94-017-0255-3_1

3 Zhao, C. *et al.* (2017) Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.* 114, 9326–9331

4 Leng, G. and Huang, M. (2017) Crop yield response to climate change varies with crop spatial distribution pattern. *Sci. Rep.* 7, 1–10

5 Barnabás, B. *et al.* (2008) The effect of drought and heat stress on reproductive processes in cereals. *Plant, Cell Environ.* 31, 11–38

6 Hasanuzzaman, M. *et al.* (2013) Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684

7 Wilkins, O. *et al.* (2016) EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments. *Plant Cell* 28, 2365–2384

8 Schöning, J. *et al.* (2006) Clockwork green - The circadian oscillator in Arabidopsis. *Biol. Rhythm Res.* 37, 335–352

9 Rawat, R. *et al.* (2009) REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proc. Natl. Acad. Sci.* 106, 16883–16888

10 Linde, A.-M. *et al.* (2017) Early evolution of the land plant circadian clock. *New Phytol.* DOI: 10.1111/nph.14487

11 R., R. *et al.* (2011) REVEILLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the arabidopsis circadian clock. *PLoS Genet.* 7

12 Bateman, A. *et al.* (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169

13 Filichkin, S.A. and Mockler, T.C. (2012) Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes. *Biol. Direct* 7, 1–15

14 Kilian, J. *et al.* (2007) The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50, 347–363

15 Winter, D. *et al.* (2007) An "electronic fluorescent pictograph" Browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2, 1–12

16 Kamioka, M. *et al.* (2016) Direct Repression of Evening Genes by CIRCADIAN

CLOCK-ASSOCIATED1 in the Arabidopsis Circadian Clock. *Plant Cell* 28, 696–711

17 Mizoguchi, T. *et al.* (2002) LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. *Dev. Cell* 2, 629–641

18 Gray, J.A. *et al.* (2017) The REVEILLE Clock Genes Inhibit Growth of Juvenile and Adult Plants by Control of Cell Size. *Plant Physiol.* 173, 2308–2322

19 Hsu, P.Y. *et al.* (2013) Accurate timekeeping is controlled by a cycling activator in Arabidopsis. *Elife* 2, 1–20

20 FAO/OECD (2012) *Building resilience for adaptation to climate change in the agriculture sector: Proceedings of a Joint FAO/OECD Workshop*

21 UN Food and Agriculture Organization, C.S.D. (FAOSTAT) (2017) , Crops/Regions/World list/Production Quantity (pick lists), Rice (paddy), 2016. . [Online]. Available: http://www.fao.org/faostat/en/#data/QC. [Accessed: 02-Aug-2019]

22 UN Food and Agriculture Organization, C. (1990) , Rice in human nutrition - Rice consumption and nutrition problems in rice consuming countries. . [Online]. Available: http://www.fao.org/3/t0567e/T0567E04.htm. [Accessed: 10-Aug-2019]

23 Somers, D.J. *et al.* (2009) *Plant genomics : methods and protocols*, NV-1 o.Humana Press.

24 Kersey, P.J. *et al.* (2018) Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* 46, D802–D808

25 Izawa, T. and Shimamoto, K. (1996) Becoming a model plant: The importance of rice to plant science. *Trends Plant Sci.* 1, 95–99

26 Ozdemir, B.S. *et al.* (2008) Brachypodium genomics. *Int. J. Plant Genomics* 2008, 1–7

27 Vogel, J.P. *et al.* (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463, 763–768

28 Jolma, I.W. *et al.* (2010) Circadian oscillators in eukaryotes. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2, 533–549

29 McClung, C.R. (2013) Beyond Arabidopsis: The circadian clock in non-model plant species. *Semin. Cell Dev. Biol.* 24, 430–436

30 Dodd, A.N. *et al.* (2005) Cell biology: Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science (80-.).* 309, 630–633

31 Green, R.M. *et al.* (2002) Circadian rhythms confer a higher level of fitness to Arabidopsis plants. *Plant Physiol.* 129, 576–584

32 Gehan, M.A. *et al.* (2015) Transcriptional networks-crops, clocks, and abiotic

stress. *Curr. Opin. Plant Biol.* 24, 39–46

33 Finn, R.D. *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199

34 Seo, P.J. and Mas, P. (2015) STRESSing the role of the plant circadian clock. *Trends Plant Sci.* 20, 230–237

35 Rubin, M.J. *et al.* (2017) Circadian rhythms vary over the growing season and correlate with fitness components. *Mol. Ecol.* 26, 5528–5540

36 Kim, T. *et al.* (2014) The Importance of the Plant Circadian Clock to Confer Heat Tolerance. 2014, 313–321

37 Smieszek, S.P. (2015) Uncovering the dynamic architecture of circadian gene expression in plants. *PhD thesis R. Hollow. Univ. London*

38 Kwon, Y.J. *et al.* (2014) Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC Plant Biol.* 14, 136

39 Gautier, L. *et al.* (2004) Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307–315

40 Lu, S.X. *et al.* (2012) CCA1 and ELF3 Interact in the control of hypocotyl length and flowering time in Arabidopsis. *Plant Physiol.* 158, 1079–1088

41 Berns, M.C. *et al.* (2014) Evening Expression of Arabidopsis GIGANTEA Is Controlled by Combinatorial Interactions among Evolutionarily Conserved Regulatory Motifs. *Plant Cell* 26, 3999–4018

42 Wang, Z.Y. and Tobin, E.M. (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* 93, 1207–1217

43 Schaffer, R. *et al.* (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* 93, 1219–1229

44 Kolmos, E. *et al.* (2014) HsfB2b-mediated repression of PRR7 directs abiotic stress responses of the circadian clock . *Proc. Natl. Acad. Sci.* 111, 16172–16177

45 Liu, T. *et al.* (2013) Direct regulation of abiotic responses by the Arabidopsis circadian clock component PRR7. *Plant J.* 76, 101–114

46 Nakamichi, N. *et al.* (2010) PSEUDO-RESPONSE REGULATORS 9, 7, and 5 Are Transcriptional Repressors in the Arabidopsis Circadian Clock. *Plant Cell* 22, 594–605

47 Gendron, J.M. *et al.* (2012) Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci.* 109, 3167–3172

48 Fujiwara, S. *et al.* (2008) Post-translational regulation of the Arabidopsis circadian

clock through selective proteolysis and phosphorylation of pseudo-response regulator proteins. *J. Biol. Chem.* 283, 23073–23083

49 Más, P. degradation of T. by Z. *et al.* (2003) Targeted ZTL modulates circadian function in Arabidopsis thaliana. *Nature* 426, 567–70

50 Kiba, T. *et al.* (2007) Targeted Degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL Complex Regulates Clock Function and Photomorphogenesis in Arabidopsis thaliana. *Plant Cell Online* 19, 2516–2530

51 Kim, W.Y. *et al.* (2007) ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* 449, 356–360

52 Somers, D.E. *et al.* (2000) ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis. *Cell* 101, 319–329

53 Huang, W. *et al.* (2012) Mapping the Core of the Arabidopsis Circadian Clock Defines the Network Structure of the Oscillator. *Science (80-.).* 336, 75–78

54 Wang, L. *et al.* (2012) Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. *Proc. Natl. Acad. Sci.* 110, 761–766

55 Jose L. Pruneda-Paz, Ghislain Breton, Alessia Para, S.A.K. (2009) A Functional Genomics Approach Reveals CHE as a Component of the Arabidopsis Circadian Clock. *Science (80-.).* 323, 1481–1485

56 Pruitt, R.E. and Meyerowitz, E.M. (1985) Arabidopsis thaliana and plant molecular genetics. *Science (80-.).* 229, 1214–1218

57 Estelle, M.A. and Somerville, C.R. (1986) The mutants of Arabidopsis. *Trends Genet.* 2, 89–93

58 Redei, G.P. (1975) Arabidopsis as a Genetic Tool. *Annu. Rev. Genet.* 9, 111–127

59 Nagel, D.H. and Kay, S.A. (2012) Complexity in the wiring and regulation of plant circadian networks. *Curr. Biol.* 22, 648–657

60 Carré, I. and Veflingstad, S.R. (2013) Emerging design principles in the Arabidopsis circadian clock. *Semin. Cell Dev. Biol.* 24, 393–398

61 McClung, C.R. (2011) The Genetics of Plant Clocks. *Genet. Circadian Rhythm. TA - TT* - 74, 105–139

62 Huang, H. *et al.* (2017) Cross-species complementation reveals conserved functions for EARLY FLOWERING 3 between monocots and dicots. *Plant Direct* 1, 1–14

63 Filichkin, S.A. *et al.* (2011) Global profiling of rice and poplar transcriptomes highlights key conserved Circadian-controlled pathways and cis-regulatory modules. *PLoS One* 6

64 Yi, H. *et al.* (2016) Natural variation in CIRCADIAN CLOCK ASSOCIATED 1

is associated with flowering time in Brassica rapa . *Genome* 60, 402–413

65 Lou, P. *et al.* (2012) Preferential Retention of Circadian Clock Genes during Diploidization following Whole Genome Triplication in Brassica rapa . *Plant Cell* 24, 2415–2426

66 Kim, J.A. *et al.* (2012) Comparative mapping, genomic structure, and expression analysis of eight pseudo-response regulator genes in Brassica rapa. *Mol. Genet. Genomics* 287, 377–388

67 Murakami, M. *et al.* (2007) Characterization of the Rice Circadian Clock-Associated Pseudo-Response Regulators in Arabidopsis thaliana . *Biosci. Biotechnol. Biochem.* 71, 1107–1110

68 Murakami, M. *et al.* (2003) The Evolutionarily Conserved OsPRR Quintet: Rice Pseudo-Response Regulators Implicated in Circadian Rhythm. *Plant Cell Physiol.* 44, 1229–1236

69 Matsuzaki, J. *et al.* (2015) Punctual Transcriptional Regulation by the Rice Circadian Clock under Fluctuating Field Conditions. *Plant Cell* 27, 633–648

70 Carbonell-Bejerano, P. *et al.* (2014) Circadian oscillatory transcriptional programs in grapevine ripening fruits. *BMC Plant Biol.* 14, 1–15

71 Takata, N. *et al.* (2010) Phylogenetic footprint of the plant clock system in angiosperms: Evolutionary processes of Pseudo-Response Regulators. *BMC Evol. Biol.* 10, 1–14

72 Ming, R. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). *Nature* 452, 991–996

73 Zdepski, A. *et al.* (2008) Conserved Daily Transcriptional Programs in Carica papaya. *Trop. Plant Biol.* 1, 236–245

74 Kaldis, A.D. *et al.* (2003) Light and circadian regulation in the expression of LHY and Lhcb genes in Phaseolus vulgaris. *Plant Mol. Biol.* 52, 981–997

75 Galeou, A. *et al.* (2018) Investigation of the Phaseolus vulgaris circadian clock and the repressive role of the PvTOC1 factor by a newly established in vitro system. *J. Plant Physiol.* 222, 79–85

76 Ibañez, C. *et al.* (2008) Overall alteration of circadian clock gene expression in the chestnut cold response. *PLoS One* 3

77 Iba, C. *et al.* (2005) Ramos A, Perez-Solis E, Ibanez C, Casado R, Collada C, Gomez L, et al. Winter disruption of the circadian clock in chestnut. 102, 7037–7042

78 Liew, L.C. *et al.* (2009) DIE NEUTRALIS and LATE BLOOMER 1 Contribute to Regulation of the Pea Circadian Clock. *Plant Cell* 21, 3198–3211

79 Hecht, V. (2005) Conservation of Arabidopsis Flowering Genes in Model Legumes. *Plant Physiol.* 137, 1420–1434

80 Serikawa, M. *et al.* (2008) Functional Conservation of Clock-Related Genes in Flowering Plants: Overexpression and RNA Interference Analyses of the Circadian Rhythm in the Monocotyledon Lemna gibba. *Plant Physiol.* 146, 1952–1963

81 Okada, M. *et al.* (2017) Synchrony of plant cellular circadian clocks with heterogeneous properties under light/dark cycles. *Sci. Rep.* 7, 1–10

82 Marcolino-Gomes, J. *et al.* (2014) Diurnal oscillations of soybean circadian clock and drought responsive genes. *PLoS One* 9

83 Liew, L.C. *et al.* (2017) A novel role of the soybean clock gene LUX ARRHYTHMO in male reproductive development. *Sci. Rep.* 7, 1–16

84 Liu, H. *et al.* (2009) Analysis of clock gene homologs using unifoliolates as target organs in soybean (Glycine max). *J. Plant Physiol.* 166, 278–289

85 Facella, P. *et al.* (2008) Diurnal and Circadian Rhythms in the Tomato Transcriptome and Their Modulation by Cryptochrome Photoreceptors. *PLoS One* 3

86 Müller, N.A. *et al.* (2015) Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.* 48, 89–93

87 Takata, N. *et al.* (2009) Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2. *New Phytol.* 181, 808–819

88 McClung, C.R. (2012) A modern circadian clock in the common angiosperm ancestor of monocots and eudicots. *J. Biol.* 9, 23

89 Murakami, M. *et al.* (2007) Comparative overviews of clock-associated genes of Arabidopsis thaliana and Oryza sativa. *Plant Cell Physiol.* 48, 110–121

90 Izawa, T. (2012) Physiological significance of the plant circadian clock in natural field conditions. *Plant, Cell Environ.* 35, 1729–1741

91 Izawa, T. *et al.* (2017) Os-GIGANTEA Confers Robust Diurnal Rhythms on the Global Transcriptome of Rice in the Field Published by : American Society of Plant Biologists (ASPB) Linked references are available on JSTOR for this article : Os-GIGANTEA Confers Robust Diurnal Rhythms.

92 Izawa, T. (2015) Deciphering and prediction of plant dynamics under field conditions. *Curr. Opin. Plant Biol.* 24, 87–92

93 Itoh, H. and Izawa, T. (2011) A study of phytohormone biosynthetic gene expression using a circadian clock-related mutant in rice. *Plant Signal. Behav.* 6, 1932–1936

94 Izawa, T. *et al.* (2002) Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes Dev.* 16, 2006–2020

95 Shrestha, R. *et al.* (2014) Molecular control of seasonal flowering in rice, Arabidopsis and temperate cereals. *Ann. Bot.* 114, 1445–1458

96 Fornara, F. *et al.* (2015) The GI-CDF module of Arabidopsis affects freezing tolerance and growth as well as flowering. *Plant J.* 81, 695–706

97 Itoh, H. *et al.* (2010) A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. *Nat. Genet.* 42, 635–638

98 Meissner, M. *et al.* (2013) Mapping quantitative trait loci for freezing tolerance in a recombinant inbred line population of Arabidopsis thaliana accessions Tenela and C24 reveals REVEILLE1 as negative regulator of cold acclimation. 1, 1256–1267

99 Liu, J. *et al.* (2015) Genetic and epigenetic control of plant heat responses. *Front. Plant Sci.* 06, 1–21

100 McClung, C.R. *et al.* (2004) The Arabidopsis Circadian System. *Arab. B.* 1, e0044

101 Farinas, B. and Mas, P. (2011) Functional implication of the MYB transcription factor RVE8/LCL5 in the circadian control of histone acetylation. *Plant J.* 66, 318–329

102 Chinnusamy, V. *et al.* (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* DOI: 10.1186/1471-2164-13-544

103 Ambawat, S. *et al.* (2013) MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* 19, 307–321

104 Du, H. *et al.* (2009) Biochemical and molecular characterization of plant MYB transcription factor family. *Biochem.* 74, 1–11

105 Dubos, C. *et al.* (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci.* 15, 573–581

106 Green, R.M. and Tobin, E.M. (2002) Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression. *Proc. Natl. Acad. Sci.* 96, 4176–4179

107 Calixto, C.P.G. *et al.* (2015) Evolutionary Relationships Among Barley and Arabidopsis Core Circadian Clock and Clock-Associated Genes. *J. Mol. Evol.* 80, 108–119

108 Kusakina, J. *et al.* (2014) A fast circadian clock at high temperatures is a conserved feature across Arabidopsis accessions and likely to be important for vegetative yield. *Plant, Cell Environ.* 37, 327–340

109 Alabadí, D. *et al.* (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science (80-.).* 293, 880–883

110 Michael, T.P. (2002) Phase-Specific Circadian Clock Regulatory Elements in Arabidopsis. *Plant Physiol.* 130, 627–638

111 Muthusamy, S.K. *et al.* (2017) Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in

bread wheat. *J. Plant Physiol.* 211, 100–113

112 Nagel, D.H. *et al.* (2015) Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4802-10

113 Gould, P.D. *et al.* (2006) The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell* 18, 1177–1187

114 Proveniers, M.C.G. and Van Zanten, M. (2013) High temperature acclimation through PIF4 signaling. *Trends Plant Sci.* 18, 59–64

115 Farinas, B. and Mas, P. (2011) Histone acetylation and the circadian clock: A role for the MYB transcription factor RVE8/LCL5. *Plant Signal. Behav.* 6, 541–543

116 Sanchez, S.E. and Kay, S.A. (2016) The Plant Circadian Clock: From a Simple Timekeeper to a Complex Developmental Manager. *Cold Spring Harb. Perspect. Biol.* 8, a027748

117 Rugnone, M.L. *et al.* (2013) LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proc. Natl. Acad. Sci.* 110, 12120–12125

118 Xie, Q. *et al.* (2014) LNK1 and LNK2 Are Transcriptional Coactivators in the *Arabidopsis* Circadian Oscillator. *Plant Cell* 26, 2843–2857

119 Owens, S.M. *et al.* (2013) Asymmetric functional divergence of young, dispersed gene duplicates in Arabidopsis thaliana. *J. Mol. Evol.* 76, 13–27

120 Kuno, N. (2003) The Novel MYB Protein EARLY-PHYTOCHROME-RESPONSIVE1 Is a Component of a Slave Circadian Oscillator in Arabidopsis. *Plant Cell Online* 15, 2476–2488

121 Zhang, X. *et al.* (2007) Constitutive expression of CIR1 (RVE2) affects several circadian-regulated processes and seed germination in Arabidopsis. *Plant J.* 51, 512–525

122 Carstens, M. *et al.* (2014) Increased resistance to biotrophic pathogens in the Arabidopsis constitutive induced resistance 1 mutant is EDS1 and PAD4-dependent and modulated by environmental temperature. *PLoS One* 9, 4–11

123 Murray, S.L. *et al.* (2007) Characterization of a Novel, Defense-Related Arabidopsis Mutant, cir1 , Isolated By Luciferase Imaging . *Mol. Plant-Microbe Interact.* 15, 557–566

124 Guan, Q. *et al.* (2013) A DEAD Box RNA Helicase Is Critical for Pre-mRNA Splicing, Cold-Responsive Gene Regulation, and Cold Tolerance in *Arabidopsis*. *Plant Cell* 25, 342–356

125 Ljung, K. *et al.* (2001) Sites and homeostatic control of auxin biosynthesis in Arabidopsis during vegetative growth. *Plant J.* 28, 465–474

126 Ljung, K. (2005) Sites and Regulation of Auxin Biosynthesis in Arabidopsis

Roots. *Plant Cell Online* 17, 1090–1104

127 Hatfield, J.L. and Prueger, J.H. (2015) Temperature extremes: Effect on plant growth and development. *Weather Clim. Extrem.* 10, 4–10

128 Fu, G. *et al.* (2016) Heat Stress Is More Damaging to Superior Spikelets than Inferiors of Rice (Oryza sativa L.) due to Their Different Organ Temperatures. *Front. Plant Sci.* 7, 1–16

129 Bita, C.E. and Gerats, T. (2013) Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* 4, 1–18

130 Quint, M. *et al.* (2016) Molecular and genetic control of plant thermomorphogenesis. *Nat. Plants* 2, 15190

131 Haak, D.C. *et al.* (2017) Multilevel Regulation of Abiotic Stress Responses in Plants. *Front. Plant Sci.* 8, 1–24

132 Roy, S. (2016) Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signal. Behav.* 11, e1117723

133 Mohammed, A.R. and Tarpley, L. (2011) High night temperature and plant growth regulator effects on spikelet sterility, grain characteristics and yield of rice (Oryza sativa L.) plants . *Can. J. Plant Sci.* 91, 283–291

134 Song, X. *et al.* (2015) Effects of high night temperature during grain filling on formation of physicochemical properties for japonica rice. *J. Cereal Sci.* 66, 74–80

135 Peng, S. *et al.* (2004) Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci.* 101, 9971–9975

136 Redfern N//Binamira, J S, S.K. (2012) Rice in Southeast Asia: facing risks and vulnerabilities to respond to climate change. *Build. Resil. Adapt. to Clim. Chang. Agric. Sect.* at <http://www.fao.org/fileadmin/templates/agphome/documents/faooecd/oecd_proceedings.pdf#page=302>

137 Bodenstein, C. *et al.* (2012) Temperature compensation and entrainment in circadian rhythms. *Phys. Biol.* 9, 36011

138 Salome, P.A. *et al.* (2010) The Role of the Arabidopsis Morning Loop Components CCA1, LHY, PRR7, and PRR9 in Temperature Compensation. *Plant Cell Online* 22, 3650–3661

139 Ford, B. *et al.* (2016) Barley (Hordeum vulgare) circadian clock genes can respond rapidly to temperature in an EARLY FLOWERING 3-dependent manner. *J. Exp. Bot.* 67, 5517–5528

140 Filichkin, S.A. *et al.* (2015) Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant* 8,

207–227

141 Nagel, D.H. *et al.* (2014) FBH1 affects warm temperature responses in the *Arabidopsis* circadian clock. *Proc. Natl. Acad. Sci.* 111, 14595–14600

142 Cai, S. *et al.* (2017) Evolutionary Conservation of ABA Signaling for Stomatal Closure. *Plant Physiol.* 174, 732–747

143 Araújo, W.L. *et al.* (2011) Control of stomatal aperture: A renaissance of the old guard. *Plant Signal. Behav.* 6, 1305–1311

144 Hassidim, M. *et al.* (2017) CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and the circadian control of stomatal aperture. *Plant Physiol.* 4, pp.01214.2017

145 Riechmann, J.L. *et al.* (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science (80-.).* 290, 2105–2110

146 Ezer, D. *et al.* (2017) The evening complex coordinates environmental and endogenous signals in Arabidopsis. *Nat. Plants* 3, 17087

147 Gu, L. *et al.* (2016) ZmGOLS2, a target of transcription factor ZmDREB2A, offers similar protection against abiotic stress as ZmDREB2A. *Plant Mol. Biol.* 90, 157–170

148 Fowler, S.G. (2005) Low Temperature Induction of Arabidopsis CBF1, 2, and 3 Is Gated by the Circadian Clock. *Plant Physiol.* 137, 961–968

149 Du, S. (2016) Functional Characterizations of the Reciprocal Interaction of the Circadian Clock gene Time for coffee (TIC) with Stress and Energy in Arabidopsis. *PhD thesis Rheinischen Friedrich-Wilhelms-Universität Bonn*

150 Chow, B.Y. *et al.* (2014) Transcriptional regulation of LUX by CBF1 mediates cold input to the circadian clock in Arabidopsis. *Curr. Biol.* 24, 1518–1524

151 Boikoglou, E. *et al.* (2011) Environmental memory from a circadian oscillator: The Arabidopsis thaliana clock differentially integrates perception of photic vs. Thermal entrainment. *Genetics* 189, 655–664

152 Gil, K.E. and Park, C.M. (2019) Thermal adaptation and plasticity of the plant circadian clock. *New Phytol.* 221, 1215–1229

153 Mizuno, T. *et al.* (2014) The EC night-time repressor plays a crucial role in modulating circadian clock transcriptional circuitry by conservatively double-checking both warm-night and night-time-light signals in a synergistic manner in Arabidopsis thaliana. *Plant Cell Physiol.* 55, 2139–2151

154 Rand, D.A. *et al.* (2006) Uncovering the design principles of circadian clocks: Mathematical analysis of flexibility and evolutionary goals. *J. Theor. Biol.* 238, 616–635

155 Thines, B. and Harmon, F.G. (2011) Four easy pieces: Mechanisms underlying circadian regulation of growth and development. *Curr. Opin. Plant Biol.* 14, 31–37

156 Jung, J.-H. *et al.* (2016) Phytochromes function as thermosensors in Arabidopsis. *Science (80-.).* 354, 886–889

157 Casal, J.J. *et al.* (2016) Phytochrome B integrates light and temperature signals in Arabidopsis . *Science (80-.).* 354, 897–900

158 Thines, B. and Harmon, F.G. (2010) Ambient temperature response establishes ELF3 as a required component of the core Arabidopsis circadian clock. *Proc. Natl. Acad. Sci.* 107, 3257–3262

159 Mcclung, C.R. and Salome, P.A. (2005) PSEUDO-RESPONSE REGULATOR 7 and 9 Are Partially Redundant Genes Essential for the Temperature Responsiveness of the Arabidopsis Circadian Clock. *Plant Cell* 17, 791–803

160 Marshall, C.M. *et al.* (2016) The Arabidopsis sickle Mutant Exhibits Altered Circadian Clock Responses to Cool Temperatures and Temperature-Dependent Alternative Splicing . *Plant Cell* 28, 2560–2575

161 Mizuno, T. *et al.* (2014) Ambient temperature signal feeds into the circadian clock transcriptional circuitry through the EC night-time repressor in Arabidopsis thaliana. *Plant Cell Physiol.* 55, 958–976

162 Crawford, A.J. *et al.* (2012) High temperature exposure increases plant cooling capacity. *Curr. Biol.* 22, R396–R397

163 Gray, W.M. *et al.* (2002) High temperature promotes auxin-mediated hypocotyl elongation in Arabidopsis. *Proc. Natl. Acad. Sci.* 95, 7197–7202

164 Zhu, J.Y. *et al.* (2016) TOC1-PIF4 interaction mediates the circadian gating of thermoresponsive growth in Arabidopsis. *Nat. Commun.* 7, 1–10

165 Kumar, S.V. *et al.* (2012) Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature* 484, 242–245

166 Thines, B.C. *et al.* (2014) The time of day effects of warm temperature on flowering time involve PIF4 and PIF5. *J. Exp. Bot.* 65, 1141–1151

167 Fernández, V. *et al.* (2016) Photoperiodic and thermosensory pathways interact through CONSTANS to promote flowering at high temperature under short days. *Plant J.* 86, 426–440

168 Ibarrola-Rivas, M.J. *et al.* (2017) Is the available cropland and water enough for food demand? A global perspective of the Land-Water-Food nexus. *Adv. Water Resour.* 110, 476–483

169 Fischer, T. *et al.* (2014) Crop yields and global food security. ACIAR Monograph No. 158. DOI: ISBN 978 1 925133 06 6 (PDF)

170 Hatfield, J.L. (2013) Climate Change: Challenges for Future Crop Adjustments. *Clim. Chang. Plant Abiotic Stress Toler.* DOI: doi:10.1002/9783527675265.ch01

171 Shen, G. *et al.* (2015) The regulatory network mediated by circadian clock genes is related to heterosis in rice. *J. Integr. Plant Biol.* 57, 300–312

172 Wang, W. *et al.* (2011) Timing of plant immune responses by a central circadian regulator. *Nature* 470, 110–115

173 Nakamichi, N. (2015) Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant Cell Physiol.* 56, 594–604

174 Brambilla, V. *et al.* (2017) The Importance of Being on Time: Regulatory Networks Controlling Photoperiodic Flowering in Cereals. *Front. Plant Sci.* 8, 1–8

175 Galbiati, F. *et al.* (2016) Hd3a, RFT1 and Ehd1 integrate photoperiodic and drought stress signals to delay the floral transition in rice. *Plant Cell Environ.* 39, 1982–1993

176 Plant, S. *et al.* (2001) Genetic Control of Flowering Time in Rice , a NATURAL VARIATIONS : A NEW RESOURCE FOR. DOI: 10.1104/pp.010710.1

177 Rensink, W.A. (2004) Arabidopsis to Rice. Applying Knowledge from a Weed to Enhance Our Understanding of a Crop Species. *Plant Physiol.* 135, 622–629

178 Brambilla, V. and Fornara, F. (2017) Y flowering? Regulation and activity of CONSTANS and CCT-domain proteins in Arabidopsis and crop species. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1860, 655–660

179 Onai, K. and Ishiura, M. (2005) PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock. *Genes to Cells* 10, 963–972

180 Vogt, J.H.M. and Schippers, J.H.M. (2015) Setting the PAS, the role of circadian PAS domain proteins during environmental adaptation in plants. *Front. Plant Sci.* 6, 1–10

181 Goodstein, D.M. *et al.* (2012) Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* 40, 1178–1186

182 Larkin, M.A. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948

183 Muhire, B.M. *et al.* (2014) SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS One* 9, e108277

184 Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128

185 Ciccarelli, D. et al. (2006) of a Highly Resolved Tree of Life. *Science (80-.).* 311, 1283–1288

186 Bailey, T. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Second Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, 28–36

187 Berardini, T.Z. *et al.* (2015) The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485

188 Hieno, A. *et al.* (2014) Ppdb: Plant promoter database version 3.0. *Nucleic Acids Res.* 42, 1188–1192

189 O'Malley, R.C. *et al.* (2016) Erratum: Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape (Cell (2016) 165(5) (1280–1292)). *Cell* 166, 1598

190 Gupta, S. *et al.* (2007) Quantifying similarity between motifs. *Genome Biol.* 8, R24

191 Khan, A. *et al.* (2018) JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 46, D260–D266

192 Mikkelsen, M.D. and Thomashow, M.F. (2009) A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. *Plant J.* 60, 328–339

193 Ye, J. *et al.* (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. DOI: 10.1186/1471-2105-13-134

194 Dwight, Z. *et al.* (2011) uMELT: Prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics* 27, 1019–1020

195 SantaLucia, J. and Hicks, D. (2004) The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33, 415–440

196 Blake, R.D. and Delcourt, S.G. (1998) Thermal stability of DNA. *Nucleic Acids Res.* 26, 3323–3332

197 Tanaka, E. *et al.* (2015) Cost-eff ectiveness of tocilizumab, a humanized anti-interleukin-6 receptor monoclonal antibody, versus methotrexate in patients with rheumatoid arthritis using real-world data from the iorra observational cohort study. *Mod. Rheumatol.* 25, 503–513

198 Guo, M. *et al.* (2016) The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Front. Plant Sci.* 7, 114

199 Casler, M.D. *et al.* (2011) The Switchgrass Genome: Tools and Strategies. *Plant Genome J.* 4, 273

200 Blodgett, T.M. *et al.* (2011) Similarities in the circadian clock and photoperiodism in plants. *Clin. Imaging* 27, 320–331

201 Toda, Y. *et al.* (2019) Evolutionary Insight into the Clock-Associated PRR5 Transcriptional Network of Flowering Plants. *Sci. Rep.* 9, 1–14

202 Cockram, J. *et al.* (2012) Genome Dynamics Explain the Evolution of Flowering Time CCT Domain Gene Families in the Poaceae. *PLoS One* 7, e45307

203 Song, Y.H. *et al.* (2010) Similarities in the circadian clock and photoperiodism in plants. *Curr. Opin. Plant Biol.* 13, 594–603

204 Tiley, G.P. *et al.* (2016) Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol. Evol.* 8, 1023–1037

205 Wang, X. *et al.* (2011) The genome of the mesopolyploid crop species Brassica rapa. *Nat. Genet.* 43, 1035–1040

206 Weirauch, M.T. *et al.* (2015) Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* 158, 1431–1443

207 Blair, E.J. *et al.* (2019) Contribution of time of day and the circadian clock to the heat stress responsive transcriptome in Arabidopsis. *Sci. Rep.* 9, 36–40

208 Bustin, S.A. *et al.* (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622

209 Zhang, X. *et al.* (2013) Transcriptome profile reveals heat response mechanism at molecular and metabolic levels in rice flag leaf. *Gene* 530, 185–192

210 Albihlal, W.S. *et al.* (2018) Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b regulates multiple developmental genes under benign and stress conditions. *J. Exp. Bot.* 69, 2847–2862

211 Covington, M.F. and Harmer, S.L. (2007) The circadian clock regulates auxin signaling and responses in Arabidopsis. *PLoS Biol.* 5, 1773–1784

212 McEntee, C. *et al.* (2008) The Diurnal Project: Diurnal and Circadian Expression Profiling, Model-based Pattern Matching, and Promoter Analysis. *Cold Spring Harb. Symp. Quant. Biol.* 72, 353–363

213 McClung, C.R. (2006) Plant Circadian Rhythms. *Plant Cell Online* 18, 792–803

214 Monceau, D. du La physique des arbres. 2 vols. . (1758) , 307, 432

215 (2008) Botanische Forschungen des Alexanderzuges. *Nature* 68, 292–293

216 Nakamichi, N. *et al.* (2009) Transcript profiling of an Arabidopsis PSEUDO RESPONSE REGULATOR arrhythmic triple mutant reveals a role for the circadian clock in cold ctress response. *Plant Cell Physiol.* 50, 447–462

217 Rodríguez-Leal, D. *et al.* (2017) Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* 171, 470–480