Role of the epigenetic regulator Polycomb Repressive Complex 2 in genome regulation and heat stress memory of the unicellular red alga *Cyanidioschyzon merolae*

INAUGURAL-DISSERTATION

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Kowar Teresa
from Miłoćicy/Miltitz, Germany

Berlin, 2022

This doctoral thesis was prepared under the supervision of Prof. Dr. Daniel Schubert in the group of Plant Epigenetics at the Institute of Biology, Freie Universität Berlin, from October 2017 to June 2022.

1st reviewer:	Prof. Dr. Daniel Schubert
2nd reviewer:	Prof. Dr. Reinhard Kunze

Date of defense: 19.10.2022

Statement of Authorship

I hereby certify that I have written my dissertation independently and have not used any sources or aids other than those indicated by me.

Berlin, 26.06.2022

_____

Teresa Kowar

Content

List of Figures

List of Tables

List of Abbreviations

∆E(z)	*C. merolae* strain carrying a full deletion of the *Enhancer of zeste* gene (*CMQ156C*)
µE	micro Einstein, [µmol/m^2/s]
µm	micro meter
A. thaliana /Arabidopsis	*Arabidopsis thaliana*
aa	Amino acid
bp	base pair
C. merolae	*Cyanidioschyzon merolae*
ChIP-Seq	chromatin immuno precipitation followed by sequencing
CLF	Curly leaf
CP	Chloroplast
DAPI	4´,6-diamidino-2-phenylindole
DEG	Differentially expressed genes
DE-RE	Differentially expressed repetitive elements
DNA	desoxyribonucleic acid
E(z)	Enhancer of zeste
ESC	Extra Sex Combs
et al.	*et alumni*
FAIRE	formaldehyde assisted isolation of regulatory elements
FIE-MEA	FERTILIZATION INDEPENDENT ENDOSPERM – MEDEA
FLC	FLOWERING LOCUS C
G3P	glyceraldehyde-3-phosphate
GO	Gene ontology
h	hour
H2AK118/199	Lysine at position 118 or 119 of histone 2A
H3	Histone 3
H3K27me3	trimethylation of lysine at position 27 of histone 3
H3K4me3	trimethylation of lysine at position 4 of histone 3
HMT	histone methyltransferase
HSF	Heat shock factor
HSFA2	HEAT SHOCK FACTOR A2
HSMA	Heat stress memory assay
HSP	Heat shock protein
IGP	indole-3- glycerol phosphate
l	liter
L2FC	log2-fold change
lfc	log fold change
LHC	light harvesting complex
MA2	2x modified Allen´s medium
Mb	Mega base pairs
MEA	Medea
miRNA	micro RNA
N	naive state of cell
n	number of replicates

NAP	Nucleoid associated protein
NL	naive cell state after lag phase
nm	Nano meter
OD	Optical density
P	primed state, after priming
PAM	pulse amplitude modulation
PC	Polycomb
PcG	polycomb group genes
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEP	Plastid endcoded DNA directed RNA polymerase
PH	Polyhomeotic
PL	primed cell after lag phase
PLT	primed cell after triggering
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PSC	Posterior Sex Combs
PSI	photosystem I
PSII	photosystem II
PTM	Post translational modification
RE	Repetitive element
REF6	RELATIVE OF EARLY FLOWERING 6
RNA	ribonucleic acid
RNA-seq	RNA sequenicng followed by sequenicing
s.d.	Standard deviation
SAM	Shoot apical meristem
SCM	Sex Comb on Midleg
Ser5P	Serine-5 phosphorylation
SET domain	Suvar3-9, Enhancer of zester, trithorax
sHsp	Small heat shock protein
SIG	Sigma factor
SU(Z)12	Supressor of zeste 12
SWN	Swinger
T	cell after triggering
T1	Uracil deficient *C. merolae* strain, 10D WT carrying full deletion of *URA3.5*
TE	transposable element, part of repetitive element sequence class in genome
TF	Transcription factor
TOR	Target of rapamycin
trx	Trithorax
URA	Uracil
WT	wild type

Abstract

Polycomb Repressive Complex 2 (PRC2) is a major epigenetic regulator in animals and plants which represses transcription of genomic loci by trimethylation on lysine 27 of histone H3 (H3K27me3), concerning genes and repetitive elements. This molecular activity was linked to several pivotal developmental processes, including maintenance of tissue identities as well as induction of flowering in response to environmental cues. While the function of PRC2 has been extensively studied in higher organisms, its importance in lower eukaryotes, such as unicellular photoautotrophic eukaryotes, remains largely undescribed. To extend this knowledge, the present study aims to determine the role of PRC2 in the single celled, photosynthetic active red alga *Cyanidioschyzon merolae.* For this, the PRC2 loss-of-function mutant Δ*E(z)* was employed, which is deficient in H3K27me3, caused by deletion of *CMQ156C* encoding the PRC2 catalytic subunit ortholog Enhancer of zeste E(z). To evaluate whether H3K27me3 is involved in protecting the genome integrity via repetitive element (RE) silencing, the transcriptome of Δ*E(z)* was compared to 10D wild type (WT) cells. As a result, H3K27me3 absence lead to increased transcription of subtelomeric regions predominantly constituted of REs. Next, I hypothesized, that PRC2 relevance might evolve in stress situations and, moreover, memory thereof. In the context of testing this assumption, a heat stress memory protocol was established, that in fact demonstrated the alga's ability to generate a heat stress memory (termed thermomemory). More specifically, cells pre-exposed to a priming heat stimulus were largely resistant to a subsequent lethal heat exposure. Furthermore, transcriptome analysis revealed a set of genes showing revised transcriptional responses (termed trainability) upon sequential heat exposure. Interestingly, a trainable genomic heat shock locus spanning five genes could be identified. Furthermore, the chloroplast (CP) genome exhibited substantial trainability, which was most pronounced for genes encoding vast parts of the CP genetic system machinery. Finally, a substantial contribution of PRC2 to thermomemory was detected by including Δ*E(z)* in the memory protocol, reflected in aberrant physiological and transcriptional responses.

In summary, PRC2 in *C. merolae* serves the function of maintaining genome integrity via RE silencing and moreover it contributes to heat stress responses, and memory thereof. A putative link between both functions may be found in the lasting, partially H3K27me3-dependent RE repression detected in response to heat.

The presented results contribute to our understanding of how PRC2 function has developed during the course of evolution. As in *C. merolae* PRC2 combines its importance in genome protection, found predominantly in unicellular organisms, with control of memory processes described mainly in multicellular species, this situation may represent a gradual shift in PRC2 function and reflect the phylogenetic position of *C. merolae* situated at the border between plants and lower eukaryotes.

Zusammenfassung

Polycomb Repressive Complex 2 (PRC2) ist ein wichtiger epigenetischer Regulator in Tieren und Pflanzen, der die Transkription genomischer Loci unterdrückt, indem es Lysin 27 des Histons H3 an Genen und repetitiven Elementen methyliert (H3K27me3). Diese molekulare Aktivität ist assoziiert mit mehreren entscheidenden Entwicklungsprozessen, darunter Gewebeidentitäts-Aufrechterhaltung und Blühinduktion. Während die PRC2 Funktion in höheren Organismen beschrieben ist, ist seine Bedeutung in niederen Eukaryonten noch weitgehend unbekannt. Um dieses Wissen zu erweitern, beschreibt die vorliegende Studie die Rolle von PRC2 in der einzelligen Rotalge *Cyanidioschyzon merolae*. Dazu wurde die PRC2-Mutante Δ*E(z)* verwendet, die durch Deletion des Gens C*MQ156C,* welches für das PRC2-Ortholog Enhancer of zeste E(z) kodiert, kein H3K27me3 mehr aufweist. Um festzustellen, ob H3K27me3 am Schutz der Genomintegrität durch Stilllegung repetitiver Elemente (RE) beteiligt ist, wurde das Transkriptom von Δ*E(z)* mit 10D-Wildtypzellen (WT) verglichen. Die Abwesenheit von H3K27me3 führte zu einer verstärkten Transkription von aus RE´s bestehenden Subtelomer – Regionen. Als Nächstes stellte ich die Hypothese auf, PRC2 könnte in Stressgedächtnismechanismen beteiligt sein. Um dies zu prüfen, wurde ein Hitzestressgedächtnis-Protokoll etabliert, das die Fähigkeit der Alge sich an Hitze „zu erinnern“ (Thermomemory) nachwies. So waren Zellen, die Hitze bereits erlebt hatten, weitgehend resistent gegen eine weiteren, tödlichen Hitzereiz. Transkriptom-Analysen offenbarten eine Reihe von Genen, die bei wiederholtem Hitzestress veränderte Transkriptionsreaktionen zeigen (so genannte Trainierbarkeit). Interessanterweise wurde ein trainierbarer Hitzeschock-Lokus identifiziert, der fünf Gene umfasst. Weiterhin zeigte das Chloroplastengenom hohe Trainierbarkeit, was große Teile des genetischen Systems einschloss. Schließlich wurde durch Einbeziehung von Δ*E(z)* in das Memory-Protokoll ein wesentlicher Beitrag von PRC2 zur Stressgedächtnisaufrechterhaltung beschrieben, was sich in abweichenden physiologischen und transkriptionellen Reaktionen zeigte. Zusammenfassend lässt sich sagen, dass PRC2 in *C. merolae* in der Aufrechterhaltung der Genomintegrität durch RE-Repression involviert ist und darüber hinaus zu Hitzestressreaktionen und -gedächtnis beiträgt. Ein mögliches Bindeglied zwischen beiden Funktionen könnte die dauerhafte, teilweise H3K27me3-abhängige RE-Stilllegung als Reaktion auf Hitze darstellen.

Die beschriebenen Ergebnisse tragen dazu bei zu verstehen, wie sich die Funktion von PRC2 im Laufe der Evolution entwickelt hat. So verbindet PRC2 in *C. merolae* seine Bedeutung im Genomschutz, die vorwiegend bei Einzellern zu finden ist, mit seiner Funktion in der Kontrolle von Gedächtnisprozessen, die vornehmlich bei mehrzelligen Arten beschrieben wurde. Dies könnte eine stufenweise Verschiebung der PRC2-Funktion darstellen, und die phylogenetische Position von *C. merolae* an der Grenze zwischen Pflanzen und niederen Eukaryonten widerspiegeln.

1 Introduction

1.1 Chromatin statuses are defined by epigenetic modifications including the repressive histone mark H3K27me3

In multicellular organisms (metazoans), nuclear chromatin modulation via Polycomb group (PcG) proteins is exceptionally relevant, since it ensures cellular identities and its absence in some cases even leads to severe developmental defects (Margueron & Reinberg, 2011). Via deposition and maintenance of a histone mark with repressive nature, namely trimethylation of histone 3 at position lysine 27 (H3K27me3), one of the PcG complexes ensures a working regulation of the packaging of chromatin thereby determining its activity status.

In order to reach a great level of compaction, DNA is found associated with proteins in the nucleus, packaged to a hierarchically organized state called *chromatin*. As chromatin building blocks, *nucleosomes* combine approx. 147 base pairs (bp) of DNA wrapped around a histone protein octamer of four core histones (H2A, H2B, H3, H4) into a nucleoprotein complex (Luger *et al.*, 1997). Arrays of nucleosomes resembling “beads-on-a string” are forming the lowest order of chromatin compaction: a chromatin fiber of 11 nm in diameter (Finch *et al.*, 1977). Following further compaction and integration of a 5th histone variant - the linker histone H1 - a 30 nm fiber is formed (Ramakrishnan, 1997).

Albeit being devoid of true histones, the genomes of the eukaryotic organelles mitochondria and chloroplasts exhibit a degree of DNA compaction via nucleoprotein structures called nucleoids, resembling bacterial DNA structures. 4´,6-diamidino-2-phenylindole (DAPI) staining revealed this structure to be roughly spherical or ovoid and contains proteins associated with a variety in functions ensuring gene realization (i.e. transcription and translation) and genome maintenance (Majeran *et al.*, 2012). Similar to nuclear histones, bacterial nucleoids can be subject to post translational modifications (PTM) but in stark contrast, very little details are known about the how and what. For instance, it was shown that the binding of certain nucleoid associated proteins (NAPs) is prevented by phosphorylation (Y. J. Sun *et al.*, 2004). In addition, other PTMs are likely to be relevant in plastids, too. In fact, the existence of enzymes potentially involved in histone modification regulation, such as homologs of the *Arabidopsis thaliana* (*Arabidopsis*) SET-domain proteins ATXR5, and ATXR6 as well as deacetylases are confirmed in plastid genomes (Powikrowska *et al.*, 2014). However, as research about the chloroplast and mitochondrial nucleoid is very limited, its molecular structure and function largely remains obscure.

Taken together, the multitude of its packaging levels represents a major layer of dynamic chromatin regulation with is realized by the presence or absence of several chromatin

modifications. Hence, chromatin packaging is modulated by DNA methylation, different histone variants, the density of nucleosomes, as well as a whole range of histone modifications (Jenuwein & Allis, 2001). While nucleosomes are globular protein structures, the N-terminal histone ends protrude from the complex and provide an accessible site for a multitude of modifications. Since these modifications are located “on top” of the DNA molecule, they have been termed “epigenetic” (“epi”, Greek for “on”) and include acetylation, methylation, phosphorylation, sumoylation and ubiquitination of single or multiple amino acids (aa) which each itself or in combination lead to specific downstream regulation processes (Kouzarides, 2007). In sum the chromatin environment with its entirety of histone modifications define the activity status of chromatin resulting in an transcriptionally active, accessible euchromatin and repressed, compact heterochromatin in its simplest classification. Accordingly, precise histone marks are associated with one or the other. For instance, while the methylation on lysine 4 (H3K4me3) correlates predominantly with transcriptionally active, gene rich euchromatin, H3K27me3 is found largely in heterochromatic regions associated with transcriptional repression of either inducible silent genes and/or repetitive genomic sequences (Schuettengruber *et al.*, 2007). The active deposition, removal and recognition of each histone mark is provided by regulators of histone marks: the writers, erasers and readers, respectively. The resulting dynamic nature of histone marks provide the cell with great flexibility in the regulation of genomic targets in the response to an alteration of development as consequence of time and in similar extent to changes in the environment the cells is situated in. The fact, that histone marks are heritable through cell divisions (reviewed in Probst *et al.*, 2009) display an epigenetic informational layer in addition to the genetic information coded in the DNA nucleotide sequence. This increase in regulation, flexibility but also information storage provides a tremendous profit during an organism’s life.

1.2 Polycomb group genes and their evolution across eukaryotic kingdoms

PcG genes encode for proteins forming multisubunit complexes (PRCs) and were initially described as repressors of the segment-specific *Hox* genes in the fruit fly *Drosophila melanogaster* (Lewis, 1978). Among repressive complexes formed by *PcG*s, Polycomb repressive complex 1 (PRC1) and 2 (PRC2) are studied best. However, information available for PRC2 exceeds PRC1 knowledge by far. On molecular level, *PcG*s are responsible for gene silencing via histone modifications. They monoubiquitinate H2AK118/119 (PRC1) and trimethylate H3K27 (PRC2) chromatin at target loci, which then leads to appropriate structural changes, and finally, gene repression (Müller & Verrijzer, 2009; Schwartz &

Pirrotta, 2013). Together, since histone modifications set by PRCs are heritable through mitotic divisions, the PcG system provides epigenetic memory (Hugues *et al.*, 2020), which has been shown to be indispensable in a variety of processes in multiple organisms of the plant and animal kingdom.

The Ring Finger E3 ligase complex POLYCOMB REPRESSIVE COMPLEX 1 (PRC1) catalyzes monoubiquitination of histone H2A and is composed of four core proteins. The chromodomain protein Polycomb (PC) recruits the protein complex to chromatin and recognizes H3K27me3. Posterior Sex Combs (PSC) and dRING are involved in the monoubiquitination of H2A while Polyhomeotic (PH) ensures protein-protein interaction. The last subunit, SCM (sex comb on midleg) is important for the spreading of PcG silencing (Schwartz & Pirrotta, 2008). PRC1 subunits were found in mammals, plants, insects, but are incompletely described in primitive organisms (Schuettengruber *et al.*, 2007). In fact, no orthologues of PRC1 subunits could be detected in the genomes of several unicellular organisms, including candidate species of ciliates, ascomycete fungi, diatoms, as well as green and red algae (Berke & Snel, 2015; Mikulski *et al.*, 2017; Shaver *et al.*, 2010). Moreover, in contrast to PRC1 knowledge in animals, information is poor in plant species (Derkacheva & Hennig, 2014). In the model plant *Arabidopsis*, two dRING homologs were described while the Scm subunit could not be identified. Functionally, PRC1 in *Arabidopsis* is linked to vernalization and the control of developmental transitions. Hence, Ring1a/Ring1b double mutants are affected in tissue differentiation resulting in callus like structures (Bratzel *et al.*, 2010).

Contrarily, POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) is highly conserved within the plant and animal kingdom (Bowman *et al.*, 2007; Schwartz & Pirrotta, 2013). The core PRC2 contains four proteins, including the SET (Suvar3-9, Enhancer of zester, trithorax) domain protein Enhancer of zeste (E(z)) with histone methyltransferase (HMT) specificity for H3K27, the WD40 protein EXTRA SEX COMBS (ESC), the histone binding protein p55, and Suppressor of zeste 12 (SU(Z)12). The molecular structure has been determined in mammals (Chammas *et al.*, 2020; Uckelmann & Davidovich, 2021) and several functions could be allotted to it. Exemplarily, in humans, PRC2 malfunction often is found associated with cancer development, hence defining PRC2 as potential drug target in anticancer therapy (Duan *et al.*, 2020). In plants, PRC2 knowledge is in majority retrieved from work in *Arabidopsis* where at least three homologs of the HTM E(z) form at least three different PRC2 complexes with differing composition and function (Derkacheva & Hennig, 2014). Thus, the three *E(z)* homologs *CURLY LEAF (CLF)*, *SWINGER (SWN)* and *MEDEA* (*MEA*) could be linked to differing spatial and temporal distribution. While SWN and CLF are indispensable for vegetative development and phase transitions, MEA is required during gametophyte development and early embryogenesis.

Molecularly, PRC2 complexes act as transcriptional repressors via the histone hallmark H3K27me3, which is set by the *E(z)* homologs CLF, SWN and/or MEA. Biochemical, molecular and genomic studies revealed that H3K27me3 in *Arabidopsis* covers 20-30% of all protein coding and micro RNA (miRNA) genes (X. Zhang *et al.*, 2007) but also transposable elements (TE) (Zervudacki *et al.*, 2018). In summary, H3K27me3 in *Arabidopsis* is required for cell differentiation (Chanvivattana *et al.*, 2004), development (Lafos *et al.*, 2011), regulates phase transitions, and is involved in genome integrity (Antunez-Sanchez *et al.*, 2020). Moreover, the number of studies reporting about H3K27me3 shaping stress responses is gradually increasing during the last decade (J. M. Kim *et al.*, 2015; Lämke & Bäurle, 2017; Shen *et al.*, 2021). The role of H3K27me3 is especially relevant in the vernalization process, as without H3K27me3 regulation plants would refuse to flower in spring (Chouard, 1960). The role of PRC2 and shaping stress responses in plants is addressed in more detail the following Chapter 1.3.

For a long time, studies reporting about PRC2 existence in unicellular species were searched for in vain. This circumstance was especially due to the absence of any PRC2 SET domains in the conventional unicellular models of the *Saccharomycotina*, including *Saccharomyces cerevisiae* (Veerappan *et al.*, 2008). Only in 2007, H3K27me3 was identified in the ciliate *Tetrahymena thermophila* (Garcia *et al.*, 2007; Y. Liu *et al.*, 2007) marking heterochromatin in connection with H3K9me2. Although *T. thermophila E(z)* mutants showed no apparent phenotype, no progeny could be isolated suggesting an essential function of H3K27me3 in the macronuclear development via heterochromatin maintenance (Y. Liu *et al.*, 2007). These results proposed an ancient conservation of PRC2 and its purpose in working developmental processes already in primitive unicellular eukaryotes.

Following this first observation of PRC2 and its mark H3K27me3 in basal single celled organisms, already in the following year, 2008, H3K27me3 was found at subtelomere regions in *Neurospora crassa*, a type of red bread mold of the phylum *Ascomycota* (Smith *et al.*, 2008). Further studies revealed that 6.8 % of the genome was decorated by H3K27me3 leading to transcriptional silencing of its targets. H3K27me3 preferentially localized near to the ends of chromosomes forming broad domains. Consequently, a function of H3K72me3 function in the maintenance of telomere silencing was concluded. At gene rich regions, H3K27me3 domains were found on potentially evolutionary young genes and in mutual exclusion with the H3K9me2. Interestingly, the absence of repressive H3K27me3 levels was not sufficient to completely induce targeted loci (Jamieson *et al.*, 2013).

In the green alga *Chlamydomonas rheinhardtii*, which represents the prevalent model for photosynthetic active unicellulars, only mono- and di-methylation of H3K27 could be detected. Notably, evidence for H3K27me3 existence could not be unambiguously generated. Either, mass spectrometry failed to distinguish the H3K27me3 nominal mass from

acetylation of lysine 27 of the same histone or, as a second possibility, H3K27me3 is entirely absent in *C. reinhardtii* (Shaver *et al.*, 2010). However, PRC2 components are existent in the genome with the exception of *Su(z)12* (Mikulski *et al.*, 2017; Shaver *et al.*, 2010), while no obvious orthologs of PCR1 could be identified. As RNA interference (RNAi)-mediated suppression of the *C. reinhardtii* E(z) homolog *EZH* entailed increased transcription levels of transgenes and retrotransposons, the authors concluded a form of genomic defense response as *EZH* function in *C. reinhardtii*. Yet, due to the lack of H3K27me3 detection, for this photosynthetic active unicellular definite proof for the involvement of H3K27me3 in repetitive sequence silencing could not be provided (Shaver *et al.*, 2010).

A role of PRC2 in the zygotic genome rearrangements in the ciliate *Paramecium tetraurelia* was proposed in 2014 (Lhuillier-Akakpo *et al.*, 2014). Interestingly, in the process of differentiation of the somatic arising from the zygotic nucleus a massive deletion of transposable elements and short, dispersed, single copy sequences is involved which could be linked to the *P. tetraurelia* E(z) homolog Ezl1. Later, it could be shown, that Ezl1 possesses dual substrate specificity and methylates both, H3K9 and H3K27 in this organism in order to repress transposable elements (Frapporti *et al.*, 2019).

In the following year of PRC2 description in *Paramecium*, 2015, Veluchamy and colleagues presented evidence on the existence and activity of PRC2 in the pennate diatom *Phaeodactylum tricornutum* (Veluchamy *et al.*, 2015). Five years later, Zhao *et al.* (2020) showed that H3K27me3, which covers 14% of the genome, was particularly abundant at transposable elements in *P. tricornutum*. Generating PRC2 mutants, they could show, that E(z) is responsible for H3K27me3, and H3K27me2 and that upon the histone marks absence cell morphology is impaired in *P. tricornutum*. This study provided first proof for the relevance of PRC2 in cell differentiation in an unicellular eukaryote only distantly related to animals and plants.

While the role and activity of PRC2 in the green alga *C. reinhardtii* remained enigmatic, H3K27me3 and PRC2 subunits have been identified in the photosynthetic active, red alga *Cyanidioschyzon merolae*. Chromatin immunoprecipitation (ChIP) of H3K27me3 enriched DNA followed by sequencing defined 4% of all protein coding genes and 50% of all repetitive sequences as H3K27me3 targets. Based on single gene analysis, it was deduced that H3K27me3 coverage leads to silencing of corresponding loci (Mikulski *et al.*, 2017). However, the biological function of H3K27me3 in this alga remains undescribed.

Thus, in the year of the commence of the present study, in 2017, more and more evidence arose, that in addition to its known relevance in multicellular organisms, PRC2 plays vital roles in unicellular organisms, too. Hypotheses on PCR2 functional evolution included different H3K27me3 target group specificity as function of the organisms complexity, meaning that H3K27me3 is predominantly associated with genes relevant in development in

multicellular, higher organisms, whereas H3K27me3 silencing might predominate at repetitive elements (RE) in more primitive organisms, including unicellular species. However, as just reviewed, this shift might have happened gradually in the course of evolution since these two genomic silencing functions of PRC2 are separated in some, and unified in other organisms. Only the precise knowledge of PRC2 function across phylogenetic kingdoms can aid in answering the question about PcG evolution as a consequence of multicellularity emergence or whether PcG fill ancient roles in shaping life at all its stages. Thus, many more studies should address the PcG function in unicellular organisms at the evolutionary border between plants and lower eukaryotes.

1.3 Role of chromatin regulation via PcGs in a plant´s life

As PRC2 knowledge in plants is generated in majority in the plant model *Arabidopsis thaliana (Arabidopsis)*, many functions could be attributed to H3K27me3 during the entire life cycle of this plant, ranging from embryogenesis, through flowering and reproduction (Guitton & Berger, 2005), to processing environmental changes in stress responses (Shen *et al.*, 2021). Thus, early in ontogenesis, the *Arabidopsis* endosperm requires tight regulation by FERTILIZATION INDEPENDENT ENDOSPERM – MEDEA (FIE-MEA), which is the endosperm specific PRC2 complex in *Arabidopsis*. By repressing seed-related genes until the time of fertilization, it represents the master regulator of seed development. Following embryo development and maturation, the *Arabidopsis* seed germinates and enters vegetative growth, forming a leaf rosette. Upon environmental and endogenous stimuli, floral transition starts and the shoot apical meristem (SAM) converts from vegetative SAM - forming leaves - to an inflorescence SAM identity. This inflorescence SAM in turn eventually converts to a specialized flower meristem and gives rise to flower organs, the inflorescences. Interestingly, a regulatory function in the control of flowering time and floral transition has been attributed to components of the CLF-PRC2 complex (Guitton & Berger, 2005). Seed germination and the onset of flowering are examples for phase transitions in plant life which are regulated by PRC2.

An fascinating example of PcG involvement in sensing environment and lastingly integrating this information into the life cycle of *Arabidopsis* is given by the vernalization process of this winter annual plant. In 1960, Chouard described vernalization as *“the acquisition or acceleration of the ability to flower by a chilling treatment”* (Chouard, 1960), meaning that a plant flowers only if this experience of cold has been made. This warrants a suitable time for flowering, since the plant ensures that the winter cold period has passed, before it enters the reproductive phase. Interestingly, the involvement of histone modifications including H3K27me3 was reported in this process, enabling a lasting downregulation of a gene which,

as its translated product, prevents flowering: the FLOWERING LOCUS C (FLC) (Bastow *et al.*, 2004). The processes initiated during vernalization, including transcriptional initiations of flowering-required genes, should not be “forgotten” after the cold has passed. This involves memory of the cold, and to some extent relies on lasting methylation levels of H3K9me2 but also H3K27me3 on the *FLC* locus (Bastow *et al.*, 2004).

Not only during vernalization, also when plants experience sudden chilling and/or freezing stress many regulatory cascades could be linked to H3K27me3, which are needed to react properly to the stress and prevent damages following low temperatures (Kwon *et al.*, 2009; Mayer & Charron, 2021). To sum it up, H3K27me3 is involved in chromatin accessibility changes upon cold, and cold responsive genes have to be stably demethylated at H3K27 in order to reach sufficient activity (Shen *et al.*, 2021). A similar situation is encountered in the opposing temperature stress, namely heat. Here, too, active and fast H3K27me3 removal plays a role in guaranteeing rapid activation of heat stress response genes. This is gained by the heat-induced HEAT SHOCK TRANSCRIPTION FACTOR A2 (HSFA2) directly inducing the expression of a H3K27me3 demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) which, in a positive feedback loop, in turn demethylates and activates HSFA2 target loci (J. Liu *et al.*, 2019). This finding nicely depicts, that next to regulating developmental phase transitions, H3K27me3 contributes to stress responses in *Arabidopsis*.

1.4 Chromatin regulation in plant abiotic stress “memories”

In recent decades, there has been an increase in the number of studies reporting the ability of organisms without nervous systems, including plants, to "remember" stressful situations (Hilker *et al.*, 2016). Since many of these organisms do not have a mobility system, they are tied to their surrounding and forced to adapt to any changes in their environment. Hence, it is vitally important to properly react to given cues and appropriately balance their energy investment between ongoing development (and reproduction) and the need of activating defense mechanisms in order to cope with the acute situation. Such stress responses might be directed against the same cue multiple times. Thus, to increase effectivity in these responses, plants (amongst other non-mammal organisms) have an inert ability to “remember” stressful events. The example of the vernalization process in *Arabidopsis* (Chouard, 1960) implies that plants are fairly capable of memorizing a past stress event - a phenomenon which has been termed *priming and memory*. To date, such priming ability is described in a variety of organisms, including archaea bacteria, bacteria, fungi and plants for a broad range of stresses (Hilker *et al.*, 2016). Simply put, a formerly experienced stress acts as priming event and consequently enables an organism to survive when this stress recurs, even if the stress intensity reaches lethal dosages. Experimentally the initial stress is referred

to as *priming* (P), the stress free time following priming is denoted as *lag phase* (L) and the stress reoccurrence is termed *triggering* stress (T). In such experimental stress sequences, the behavior of primed-before (**P**riming-**L**ag phase-**T**riggering, PLT) and non-primed (**T**riggered-only, T) individuals is compared. In case a benefit was generated during priming, PLT individuals react better to the triggering stress than T individuals. Priming and memory mechanisms are described and characterized for various biotic and abiotic stressors (Hilker *et al.*, 2016). Opposing to the situation in higher plants, memory occurrences in non-vascular plant cells are sparsely described. The only report about stress memory in a non-vascular plant cell concerns nutrient starvation (nitrogen depletion) in *Chlamydomonas rheinhardtii* (Mikulski & Santos-Aberturas, 2021). Besides its existence, the mechanisms beneath this metabolic memory in *Chlamydomonas* remains elusive. In *Arabidopsis* abiotic priming phenomena, temperature stress and especially the reaction to heat, is characterized best.

In general, heat is considered a widespread environmental stress and is becoming increasingly important, especially in the era of climate change. Heat has a major impact on physiology, protein balance, and transcription (J. Zhao *et al.*, 2021). For example, heat-induced oxidative stress generates reactive oxygen species that have high damage potential (Mohammed & Tarpley, 2009). The photosystem machinery is among the first to respond, as high temperatures alter the fluidity of the thylakoid membrane, leading to dissociation of PSII light-harvesting complexes, which in turn leads to decreased PSII integrity (Baker & Rosenqvist, 2004; Janka et al., 2013; Mathur et al., 2011). In addition, transposon activation is a known consequence of heat exposure in Arabidopsis (Cavrak et al., 2014). Many mechanisms that are part of the heat response are refined upon recurrent stress, leading to acquired thermotolerance. i.e. stress memory.

Yet, precise stress memory mechanisms and dynamics are far from being fully understood which is true for all kind of stress memory mechanisms, including heat. Since priming occurs across all three eukaryotic kingdoms, it was suggested that likely the mechanisms underlying the priming phenomenon are ancestral (Hilker *et al.*, 2016). Both, physiological and molecular mechanisms contribute to stress memory formation and generate the rapid acclimation response, which is followed by information storage. Molecularly, numerous mechanisms promote memory formation, including transcription factor activation (and accumulation), RNA polymerase sigma factor (σ^S) relevance, and a variety of epigenetic processes including PRC2 operation (Hilker *et al.*, 2016). Independent of the precise mode of action, all these mechanisms lead to an extensive reprogramming of the transcriptome in case the stressor ever reoccurs. In the event of stress recurrence, such revised responsive genes show specific transcriptional patterns. Accordingly, genes responding similarly to each

stress, no matter if the stressor is faced for the first time or repeatedly, constitute non-memory genes. On the contrary, genes showing altered (higher/lower transcription rates upon stress recurrence) or revised transcriptional responses (newly induced/repressed genes) in a subsequent stress are defined as memory or trainable genes (Ding *et al.*, 2012). Transcriptional memory of genes can refer to two different phenomena, which both describe that a priming event leads to differential expression of the gene, either during the stress-free episode or upon triggering. Thus, the transcriptional response of the memory gene might return to pre-stress levels in between the stress occurrences, but show an altered or revised response during the following stress, showing *modified re-induction*. On the other hand, the transcriptional response upon priming may stay lastingly modified, although the activating cue has passed. These genes show *sustained induction* after the initial stress (Friedrich *et al.*, 2018). Yet, both phenomena entail gene trainability, which describes the differential transcriptional response of a gene, depending on whether the previous priming experience has been made or not. The dissection of the *Arabidopsis* transcriptome facing sequential drought stresses allowed a further characterization of memory genes, defining four distinct types of stress memory genes depending on their transcriptional behavior when a stress recurs: [+/-] and [-/+] memory genes exhibit revised responses upon repetitive stress. Correspondingly, the first stress encounter induces [+/-] genes and this induction reverts back to non-stress levels upon the recurring stress, while [-/+] genes exhibit the opposite behavior: Here, the initial stress (P) leads to downregulation while this repression is alleviated to control levels when the stress recurs (PLT). [+/+] memory genes show induction in both stress encounters, but the response upon the recurring stress (PLT) significantly exceeds the induction upon a single stress (T) only. [-/-] memory genes are repressed by a single stress (T) and their repression is even more pronounced upon stress recurrence (PLT). Both, [+/+[and [-/-] memory genes show an altered transcriptional response (Ding *et al.*, 2013, with [first stress vs non-stressed/ second stress vs first stress]). Memory genes showing sustained induction during lag phase have been described for drought and heat stress memory (Lämke *et al.*, 2016; N. Liu *et al.*, 2014). Furthermore, gene trainability has been identified in systemic acquired resistance and temperature stress of *Arabidopsis* (Jaskiewicz *et al.*, 2011; H.-C. Liu *et al.*, 2018; Mozgová *et al.*, 2015; van Buer *et al.*, 2016) and first attempts are being made to elucidate mechanisms leading to trainability. Thus, how gene trainability is encoded is still matter of debate and depends on stressor kind, the respective gene candidate, and its chromatin environment. In fact, recently chromatin based mechanisms of abiotic memory have been comprehensively discussed in several reviews (Bäurle & Trindade, 2020; Chang *et al.*, 2020; J. H. Kim, 2021; Nishad & Nandi, 2021; Oberkofler *et al.*, 2021; Van Zanten *et al.*, 2022; Xie *et al.*, 2021), and the substantial role of H3K4me3 together with the need in nucleosome abundancy regulation is considered a

general mechanism for plant temperature stress (Friedrich *et al.*, 2018). Likewise, when drought memory genes were identified, lasting association of memory genes with H3K4me3 beyond the duration of acute stress was reported in addition with elevated Serine-5 phosphorylation (Ser5P) of RNA polymerase II levels enabling rapid transcription induction upon repetitive stress (Ding *et al.*, 2012). The role of H3K27me3 in stress memory formation is less clear, since it seems more diverse and might be of indirect regulatory nature. In drought stress, H3K27me3 levels at a subset of dehydration memory genes remained unchanged during stress occurrences, but in a PcG mutant background (*clf*) their transcription was induced. The authors do not consider H3K27me3 a direct memory mark in drought stress, but postulate that it acts in regulating the scope of maximal gene induction and thus provides a fine tuning of the transcriptional answer upon stress (N. Liu *et al.*, 2014). As a contradictory example, on a subset of cold responsive genes, H3K27me3 levels gradually decreased during cold and the levels remained low even after cold has passed. Meanwhile, the transcription of these genes reverted back to control levels. However, the lowered H3K27me3 levels did not lead to a stronger transcriptional response when cold reoccurred uncoupling H3K27me3 from trainability. Yet, this finding shows that, opposing to what was reported for drought stress memory genes, H3K27me3 acts as memory marker for a past experience of cold. Since the cold-induced H3K27me3 decrease was stable for several days, this, for the first time, described H3K27me3 as mediator in translating the experience of a transient stressful cue to long lasting memory (Kwon *et al.*, 2009). A study dissecting the behavior of four histone marks upon hyperosmotic stress reports rapid fractioning of H3K27me3 islands upon priming which occurred within a few hours after the salt stress and lasted for a 10 day growth period (Sani *et al.*, 2013). Alike to what was reported for drought stress, the actual transcription rates of genes could not be predicted from cumulative histone modification levels and *vice versa.* However, the putative contribution of H3K27me3 to the memory process was confirmed (Sani *et al.*, 2013). Stress-induced transcriptional memory was additionally reported for recurring heat stress (Lämke *et al.*, 2016; Lämke & Bäurle, 2017), but the contribution of H3K27me3 to this process remains enigmatic. However, a positive feedback loop of HSFA2 and the H3K27me3 demethylase REF6 was reported to contribute to transgenerational memory of heat stress implying indirect regulating roles of H3K27me3 in *Arabidopsis* thermomemory (J. Liu *et al.*, 2019). In the light of abiotic stress memory in *Arabidopsis*, primed genes seem to associate with lasting changes in chromatin modifications, including H3K4me3 association or loss of H3K27me3 designating these histone marks as relevant in the memory formation of stressful events.

In a nutshell: PcG functions in eukaryotes

In multicellular species, PRC2 knowledge is well established and several important functions could be attributed to it. By ensuring the right timing of the activity of genes which are only needed in a narrow time window while flowering, H3K27me3 plays a role in **regulating developmental phase transitions** (Bastow *et al.*, 2004). Furthermore, decorating many stress responsive genes, **H3K27me3 is involved in regulating the precise and timely appropriate extent of i.e. expression of potentially-inducible genes when stressful situations occur** (Kleinmanns *et al.*, 2017). H3K27me3 contribution to the proper regulation of these genes in the situation of repetitive stress events was reported, a process which is vitally important for sessile organisms like plants (Kwon *et al.*, 2009; J. Liu *et al.*, 2019; Yamaguchi & Ito, 2021). Thus, H3K27me3 **regulates stress memory** in organisms lacking a nervous system. Naturally, these functions might be entangled with each other thus defining H3K27me3 as an actor in balancing plant growth and stress adaptation (Shen *et al.*, 2021). It is more and more accepted that H3K27me3 has differential target preferences, as it is enriched on genes in higher organisms, including plants, and predominantly occupies repetitive sequence elements in lower eukaryotes. Additionally, recently an important role in modulating genome integrity during sexual reproduction was described for *jumonji* demethylases removing H3K27me3. Accordingly, H3K27me3 homeostasis between generations ensures methylation and subsequently silencing of transposable elements (Antunez-Sanchez *et al.*, 2020). The preservation of the transcriptional silence of genome threatening repetitive sequence elements assures the **integrity of the whole genome** and represents yet another function which could be assigned to H3K27me3.

Together, albeit H3K27me3 regulation in plants and mammals is described in some detail already, the knowledge of H3K27me3 is by far not complete; a situation which is even exacerbated for evolutionary simple photoautotrophic eukaryotes. Research of PRC2 evolution including primitive eukaryotes is only in its infancy, and studies in lower, unicellular photosynthetic active species might reveal the purpose of epigenetic regulation in organisms, which, if any, show only very humble developmental life cycles.

1.5 *Cyanidioschyzon merolae* – an arising model organism to dissect chromatin regulation in an unicellular photoautotrophic eukaryote

Cyanidioschyzon merolae belongs to the *Rhodophyta* order *Cyanidiales*, which derived early from other red algal lineages: approximately 1.3 - 1.4 billion years ago (Yoon *et al.*, 2006). The single celled *Cyanidiales* species are considered to belong to the most heat-tolerant photoautotrophic eukaryotes (Ciniglia *et al.*, 2004; Yoon *et al.*, 2006) as they thrive in sulfuric hot springs of geothermal areas worldwide at pH 0.05 - 5 and temperatures between 35°-56°C (Albertano *et al.*, 2000; Merola *et al.*, 1981). This monophyletic group lacks centrosomes and flagellae. Furthermore, like other red algae their chloroplasts contain phycocyanin and chlorophyll a as photosynthetic pigments (Graham & Wilcox, 2000). Yet, due to the absence of phycoerythin, a red protein pigment complex, they appear blue-green (Lin *et al.*, 1990). Similar to *Cyanobacteria*, red algae possess phycobilisomes, huge protein antennae complexes, which are associated with photosystem II (PSII) and serve as light harvesting complexes. The additional existence of light harvesting complexes (LHC) associated with photosystem I (PS I), typically found in *Viridiplantae* (green algae and land plants), positions the red algae phylogenetically between *Cyanobacteria* and *Viridiplantae* (Green & Durnford, 1996; Neilson & Durnford, 2010). As is the case for red algae in general, only asexual reproduction was described for *Cyanidiales* (Kuroiwa *et al.*, 2018).

Three genera compose the *Cyanidiales*, differing in physiology, cell morphology, mode of proliferation, and presence or absence of a cell wall: *Galdieria*, *Cyanidium* and *Cyanidioschyzon*.

Galdieria sulphuraria is the best described *Galdieria* species and is known for its extensive metabolic activities, including photosynthesis, and heterotrophic as well as mixotrophic growth on more than 50 different carbon sources (Gross, 1999; Oesterhelt *et al.*, 1999; Rigano *et al.*, 1977). Morphologically, *Galdieria* are the largest *Cyanidiales* species forming spherical cells surrounded by rigid cell walls. Remarkably, *Galdieria* species proliferate by forming 2 - 32 autospores within mother cells following multiple successive cell divisions before hatching out.

Only three species comprise *Cyanidium*, which are obligate photoautotroph, cell-walled species and can inhabit non-thermal and non-acidic caves across the globe (Kuroiwa *et al.*, 2018). Only recently, a morphological and molecular phylogenetic study described *Cyanidiococcus* as a novel *Cyanidiales* genus comprising a single species: *Cyanidiococcus yangmingshanensis*. Similar to *Galdieria* and *Cyanidium*, *Cyanidiococcus* possesses a cell wall and proliferates by forming 4 daughter cells in a mother cell (S. L. Liu *et al.*, 2020). This

obligate photoautotroph forms small (2 µm diameter), spherical cells with thick walls and chloroplasts occupying most of its cell volume.

Cyanidioschyzon comprises a single species. *Cyanidioschyzon merolae* was isolated from the Campi Flegrei near Naples in Italy from hot, acidic springs (De Luca *et al.*, 1978). In contrast to the remaining *Cyanidiales* members, it is cell wall-less and therefor, the only one not forming autospores but simply dividing by binary fission. Although all relevant meiotic genes are conserved in the genome, only mitotic propagation of the haploid state has been detected (Malik, 2007). In consequence of this trait, and as an additional distinction to its *Cyanidiales* sister genera, *C. merolae* is genetically tractable. Its genome can be manipulated by taking advantage of homologous recombination which occurs when cells propagate. A successful polyethylene glycol (PEG)-mediated transformation protocol was established in 2008 (Ohnuma *et al.*, 2008a) and since then was utilized as basic procedure to generate genetic mutants and investigating e.g. photosynthesis (Zienkiewicz *et al.*, 2018), metabolism (Fujiwara *et al.*, 2015), and cell cycle control (Fujiwara *et al.*, 2020) amongst many more. In addition to genetic tractability, the use of *C. merolae* as model is grounded on its easy lab use. Thus, under optimal (laboratory) conditions, doubling times of 9h can be achieved which for a plant biologist working with *Arabidopsis* incredibly fast. Furthermore, the naturally low pH of culturing media (pH ≈ 3) minimizes the risk of contamination. Taken together, and underlined by the increasing scientific interest in it, *C. merolae* poses a suitable and simple model system for cell and molecular biology (Kuroiwa *et al.*, 2018; Miyagishima & Tanaka, 2021).

The *C. merolae* cell exhibits the simplest structure among eukaryotes possessing an extremely condensed set of organelles (Matsuzaki *et al.*, 2004). Each cell contains a single nucleus, mitochondrion, chloroplast, peroxisome, division cycles of which can be highly synchronized by light/dark cycles (Terui *et al.*, 1995). Further, only a small number of lysosomes is found in the cell and the architecture of its membranous structures, namely endoplasmic reticulum and Golgi body, is reduced, showing a simple shape and only two cisternae, respectively. This simplicity in the cell structure facilitates the investigation of biological processes in each (single) organelle. In consequence, spatial organization as well as morphological information is well described in each cell cycle phase of *C. merolae*. This simplicity described for the cellular architecture and development also applies to the genome structure. All three genomes – nuclear, mitochondrial, chloroplast – have been completely determined (Matsuzaki *et al.*, 2004; Nozaki *et al.*, 2007; Ohta *et al.*, 2003). Interestingly, compared to higher plants, the *C. merolae* chloroplast genome comprises elevated numbers of protein coding genes, namely 200, which include an intact transcriptional machinery. Hence, genes for the plastid DNA-directed RNA polymerase (PEP) are fully covered in the plastid genome which together with one of the four nuclear-encoded sigma factors (SIG)

drive the transcription of specific gene sets (Minoda *et al.*, 2005). For instance, SIG2 activates chloroplast-encoded phycobilisome genes under light stress and SIG4 is responsive to nitrogen depletion with consequent activation of photosystem I genes (Fujii *et al.*, 2013, 2015). While SIG1 and SIG2 seem to activate photosystem genes in nitrogen-repleted conditions, SIG4 drives PSI reaction center genes in nitrogen depleted conditions. SIG3, however, remained unresponsive to changing light and nitrogen conditions (Fujii *et al.*, 2013, 2015). Interestingly, in the *C. merolae* chloroplast genome, bacterial transcriptional activators are kept (*ycf27 – 30*) (Minoda *et al.*, 2010), as well as regulators of translation (*infB*, *infC*, *tufA*, *tsf*). Thus, the *C. merolae* chloroplast is partially self-autonomous in its transcriptional machinery (Kuroiwa *et al.*, 2018). The *C. merolae* nuclear genome is small (16.5 Mb) and was the first complete (sequenced and annotated) algal genome. The 20 chromosomes comprise 5332 genes and remarkably, only 0.7 % are transposable elements. In addition, a novel interspersed repetitive element was found in the *C. merolae* genome and altogether comprise approx. 5% of the genome (Nozaki et al., 2007). Solely 26 introns are detected in the whole genome. Furthermore, only three copies of ribosomal DNA units maintain the nucleolus and interestingly, its histone gene cluster is the smallest known in eukaryotes. Strikingly, despite its compaction, the genome of *C. merolae* contains orthologues of key chromatin regulators of higher eukaryotes such as DNA methyltransferases, thrithorax (Trx) and PRC2 genes, which are lacking in model yeasts. Of those, in particular PRC2 is responsible for developmental phase transitions and cell identity specification in animals and plants, and thus, obvious functions in a unicellular alga remain mysterious. However, recently, PRC2 conservation as well as detectable H3K27me3 has been determined for *C. merolae* in a study aiming at dissecting *PcG* existence in unicellular branches of the green lineage (Mikulski *et al.*, 2017). Following successful identification of PRC2 core genes *in silico*, high sequence conservation of histone H3, and specific H3K27me3 antibody binding on a protein blot, the authors aimed in determining H3K27me3 genome-wide distribution using H3K27me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq). As a result, they defined 242 genes (4% of overall gene number) and 172 repetitive elements (50% of total number) as H3K27me3 targets. Furthermore, H3K27me3 targeting anticorrelated with transcription, which is in accordance to the H3K27me3 silencing function in multicellular organisms. Interestingly, a preferential H3K27me3 binding towards chromosomal ends was described, covering broad regions at (sub)telomeres. Noteworthy is, that several of the H3K27me3 silenced genes of *C. merolae* are potentially stress inducible or are involved in (fast) environmental acclimation. Hence, H3K27me3 targets e.g. a nitrate reductase and -transporter and a respiratory burst oxidase gene. As mentioned earlier, the involvement of PRC2 in stress-related gene expression changes was shown for *Arabidopsis* (Kleinmanns *et al.*, 2017) which could similarly be the

case for *C. merolae*. Although only haploid stages of *C. merolae* cells are described, all necessary core meiotic genes are entailed in its genome (Guo & Yang, 2015). Therefore, similar to the function of regulating phase transition in higher organisms, PRC2 might (indirectly?) prevent the induction of meiosis in *C. merolae*. However, as evidence for potential phase transitions in *C. merolae* is lacking to date, and nor is any of the core meiotic genes targeted by H3K27me3, the relation of PRC2 and meiosis in *C. merolae* remains elusive.

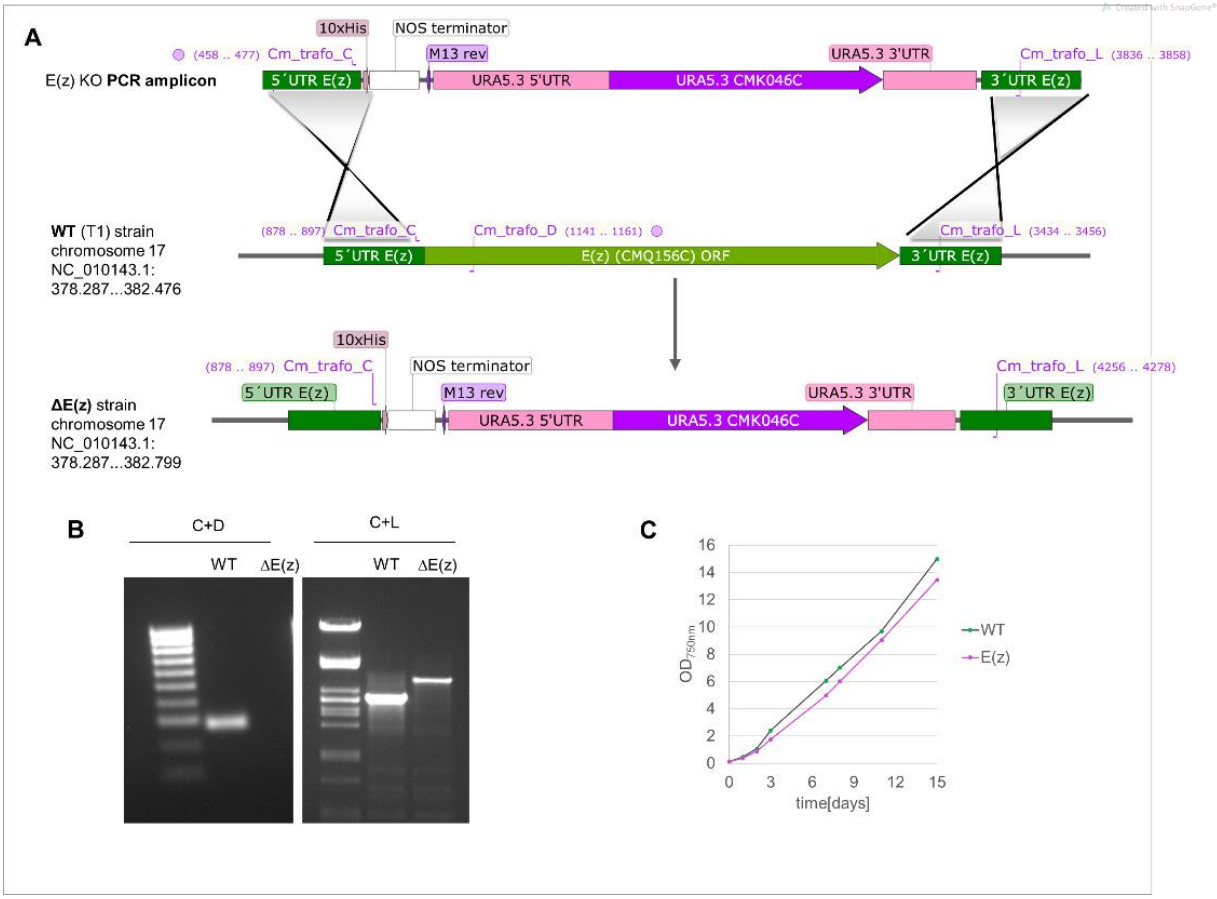


Figure 1.1: Generation, genotyping and phenotyping of *Cyanidioschyzon merolae* strain carrying a deleted *Enhancer of zeste* gene – *ΔE(z)* (Schubert laboratory, unpublished)

A Schematic diagram of the deletion of E(z) via homologous recombination with the URA selection marker using genomic homology arms flanking the E(z) gene in the *C. merolae* genome. The first line represents the introduced PCR amplicon carrying the URA cassette between genomic sequences flanking *E(z)* in the *C. merolae* WT genome. The second line indicates the genomic composition of the uracil-auxotrophic parental strain T1 (derived from 10D WT strain) at the *E(z)* locus *CMQ156C* and the third line shows the expected genomic structure of the generated *ΔE(z)* strain transformant carrying the URA cassette instead the *CMQ156C* ORF. Corresponding primer pair locations used for genotyping are indicated in purple in each line (the sequences are listed in Table S3)

B PCR analysis of the generated *ΔE(z)* strain using primer pairs indicated in A confirmed the homologous recombination event. The WT strain was used as a control. The predicted sizes of the PCR products are: Cm_trafo_D + Cm_trafo_L primer pair: 3401bp for *ΔE(z)* and 2579 bp for 10D WT. Cm_trafo_C + Cm_trafo_D primer pair: 284 bp for 10D WT and no band for *ΔE(z)*.

C Growth curves for the wild-type 10D and *ΔE(z)* strains. 10D and *ΔE(z)* strain were cultured in 2× Allen's medium.

To gain insights into the function of H3K27me3 in *C. merolae*, the laboratories of Prof. Daniel Schubert (Freie Universität Berlin) in collaboration with Prof. Shin-ya Miyagishima´s lab (National Institute of Genetics, Shizouka, Japan) constructed a *C. merolae* mutant strain carrying a full deletion of the methyltransferase subunit of the PRC2, Enhance of zeste (*E(z)*). Such deletion of the histone methyltransferase gene of PRC2 was expected to entail altered, or lost H3K27me3 genomic levels. In detail, the *E(z)* locus (*CMQ156C*) was deleted via homologous recombination using up- and downstream homology arms following the PEG mediated procedure described earlier (Ohnuma *et al.*, 2008b). The used PCR amplicon consisted of an *URA3.5* cassette surrounded by the same up- and downstream sequences flanking *CMQ156C* in the genome (Figure 1.1.A).

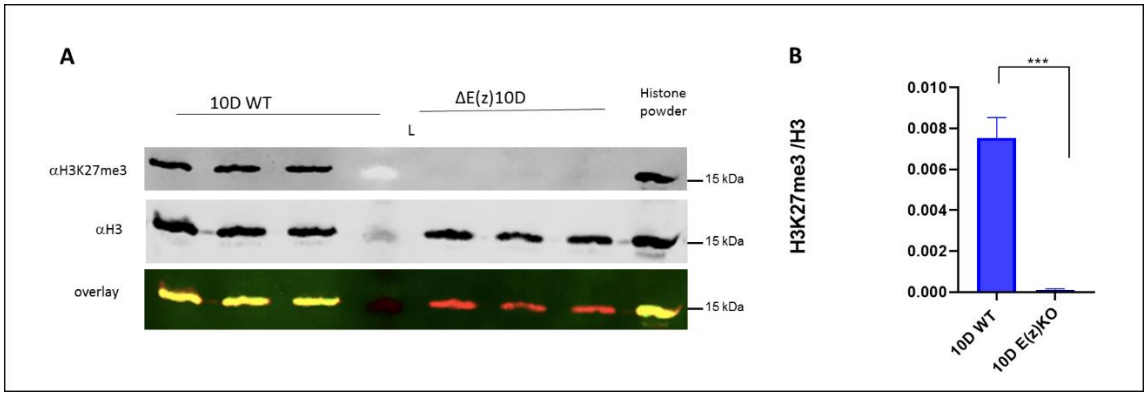


Figure 1.2: Immunoblot showing depletion of H3K27me3 histone modification in *C. merolae* ΔE(z) strain.
A Western blot analysis of WT and ΔE(z) strain using anti-H3K27me3 (Diagenode, #C15410195) and anti-H3 antibody (Diagenode, #C15200011) shows depletion of H3K27me3 levels while H3 levels remain unchanged. Histone powder (HP) was used as internal control. L – ladder (Page Ruler Prestained Protein Ladder (Thermo Scientific).
B Quantification of Western blotting. Band intensities were obtained using Image Studio Software. H3K27me3/H3 protein amount ratio of three independent experiments is shown. p-values were calculated using unpaired t-test, $p<0.001$ (retrieved from Wedler, 2019)

These “sequence arms” were allowed to recombine with homologous sequences of a PCR amplified product in a uracil deficient strain (T1). After ascertaining (a) the absence of any *CMQ156C* sequences in the *ΔE(z)* strain (Figure 1.1.B), (b) no difference in its growth rate (nor in cellular morphology, data not shown) compared to WT (Figure 1.1.C), the resulting ΔE(z) strain was characterized in Daniel Schubert´s lab by BSc Marlene Wedler. Wedler investigated the genomic H3K27me3 levels biochemically via Western blot (Figure 1.2) indicating that H3K27me3 amounts were vanished in the *ΔE(z)* strain. This result suggests, that *E(z)* is the only H3K27me3 methyltransferase in *C. merolae* (Marlene Wedler, thesis 2019).

As discussed by Wedler (2019) the results nicely show, that the entirety of HK27me3 levels depend on the genomic locus *CMQ156C* in *C. merolae*. Albeit PRC2 existence and activity was unequivocally demonstrated, PRC2 targets have been identified, and a PRC2 deletion

strain was generated in *C. merolae*, the precise function of the master epigenetic regulator PRC2 in this unicellular plant organism remains obscure.

1.6 Aims of this study

Providing stable and heritable transcriptional silencing, PRC2 is involved in the regulation of phase transitions as well as in the maintenance of cellular identities. In addition to developmental regulation, via its hallmark H3K27me3 PRC2 targets stress responsive genes in plants and ascertain adequate and timed responses in case environmental conditions change (Kleinmanns *et al.*, 2017). In fact, the relevance of regulated H3K27me3 levels during recurring stresses was suggested to ensure the memory thereof, adding a vitally important response to the list of H3K27me3 functions (J. Liu *et al.*, 2019). Moreover, evidence arose, that PRC2 silencing might be involved in preserving genomic integrity by regulating the transcription of genome threatening repetitive elements and/or maintaining telomere silencing (Antunez-Sanchez *et al.*, 2020; Jamieson *et al.*, 2013). Overall, while PRC2 importance is well understood in multicellular organisms, the function of PRC2 in unicellular eukaryotes is less clear. *Cyanidioschyzon merolae*, an unicellular extremophile red alga, is devoid of cellular identities, lacks major developmental transitions due to its asexual reproduction strategy, and exhibits exceptionally low amounts of genomic repetitive sequences. Although PRC2 existence and activity was unambiguously shown in *Cyanidioschyzon merolae* (Mikulski *et al.*, 2017), the precise relevance of this ancestral epigenetic system in a unicellular, development-less plant cell remains obscure.

Therefore, this doctoral thesis aimed to elucidate the role of the PRC2 hallmark H3K27me3 in the red alga *Cyanidioschyzon merolae* by hypothesizing the following functions:

Hypothesis 1: *H3K27me3 silences repetitive elements in C. merolae*.

In order to clarify whether PRC2 affects genome-wide transcription rates, transcriptomic analyses of the *C. merolae* wild type 10D strain in comparison with the PRC2 mutant ***ΔE(z)*** will be reveal whether H3K27me3 absence leads to transcriptional induction of repetitive elements and other genes in *C. merolae*. Conclusions will be drawn, whether transcriptional changes entailed by the loss of PRC2 function regulates specific chromosomal (subtelomeric) regions.

Hypothesis 2: *H3K27me3 represses (stress) inducible genes thereby contributing to proper stress responses in C. merolae.*

To test this hypothesis, reactions of *C. merolae* WT and ∆*E(z)* cells upon stress exposure will be compared. Potential differences in heat responses are monitored in physiologically measuring photosystem II fitness and in stress transcriptomic analyses. Comparison of the entire transcriptome in WT and ***ΔE(z)*** will uncover, whether the histone mark H3K27me3 regulates expression rates of potentially inducible genes under control conditions and stress situations.

Hypothesis 3: *H3K27me3 mediates stress memory in C. merolae*.

As a prerequisite for addressing the question whether H3K27me3 is involved in stress and stress memory responses, the ability to memorize stress as such has to be proven in *C. merolae*. Following published priming and memory schemes, an attempt is made to describe acquired stress tolerance as consequence of sequential stress encounters. Importantly, the existence of stress memory mechanisms in *C. merolae* is undescribed so far. It will be concluded, whether *C. merolae* can “memorize” stress on physiological and transcriptional level to become more resistant to subsequent stresses. Transcriptomic analyses will uncover gene candidates involved this phenomenon. Most importantly, after successful description of *C. merolae* stress memory the PRC2 mutant ***ΔE(z)*** will be included in the stress scheme. Comparing its behavior to WT cells, it will be clarified, whether stress memory mechanisms depend on PRC2 action and first conclusions will be made about potential roles of epigenetics accomplished by PRC2 during stress memory in *C. merolae*.

2 Material and Methods

2.1 Algal culture

Cyanidioschyzon merolae 10D and Δ*E(z)* cells were grown in liquid culture in sterile modified 2x concentrated Allen´s medium (MA2, Table 2.1) at 2.5 < pH < 3.0 under constant white light (80 µmol/m^2/s) at 42°C. Cultures were kept in 50 ml falcon tubes aerated with ambient air supplied through a 1ml serological milk pipett coupled to an aquarium pump. No additional CO_2 was supplied. Cultures were grown to an OD_{750nm} = 1.0 for all experiments to ensure vital, dividing cells in exponential phase. Cells were either grown in the Binder growth chamber KBW 400 (BINDER GmbH) for pre-culturing or the Multi-cultivator MC 1000-OD (Photon System Instruments) when stress assays were performed.

Table 2.1: Composition of MA2 medium
Preparation of stock solutions I – IV used for convenient use in daily laboratory work according to Kobayashi *et al.*, 2010.

Stock solution	***Component***	***Concentration stock***	***Final concentration MA2***
solution I (10x)	$(NH_4)_2SO_4$	400.00 mM	40.00 mM
	$MgSO_4 \cdot 7H_2O$	40.00 mM	4.00 mM
	500x A6 minor salts	20x	2x
solution II (100x)	KH_2PO_4	800.00 mM	8.00 mM
solution III (1000x)	$CaCl_2$	1.00 M	1.00 mM
solution IV[1] (250x)	$FeCl_3$	25.00 mM	0.1 mM
	Na_2EDTA	20.00 mM	0.08 mM
A6 minor salts (500x)	H_3BO_3	46.00 mM	92.00 µM
	$MnCl_2 \cdot 4H_2O$	9.00 mM	18.00 µM
	$ZnCl_2$	0.77 mM	1.54 µM
	$Na_2MoO_4 \cdot 2H_2O$	1.60 mM	3.20 µM
	$CoCl_2 \cdot 6H_2O$	0.17 mM	0.34 µM
	$CuCl_2$	0.30 mM	0.60 µM

For MA2 medium preparation, stock solutions I, II, and III were mixed to 1x concentration and filled ad 1000 ml with desalted water. Then, pH was adjusted with H_2SO_4 to 2.5 – 3.0 and the medium was autoclaved (121°C, 20 min, 101 kPa). After sterilization, solution IV was added to 1x concentration and the MA2 medium stored at room temperature. For MA2 medium supplemented with Uracil (MA2-Ura) needed for uracil-deficient *C. merolae* strains, like the

[1] Filter-sterilized, added to MA2 after autoclavation only.

T1 strain, Uracil solved in water was added to MA2 prior to the autoclavation step to a final concentration of 0.5 mg/ml. In order to dissolve the Uracil in water, pH was increased up to 10 with [4 N] sodium hydroxide (VWR International GmbH).
To estimate growth, optical density at 750 nm (OD_{750nm}) was measured using a Photometer (Amersham Biosciences, Novaspec III, Kluver & Schulz GmbH). When cultures overgrew $OD_{750nm} = 1$, measurements were taken for diluted cultures to ensure reliability of photometric measurements. Subsequently, the values were multiplied back by the dilution factor.

2.2 Heat stress memory assay

2.2.1 Establishment and scheme of the HSMA

Establishing a reproducible priming and memory scheme for heat stress is work as well as time intense and includes continuous trial and error. Additionally, proper read out methods have to be established to determine the actual priming effect. As a result of testing various temperature and duration combinations of priming, lag phase and triggering stress, as well as several methods to determine culture fitness, the priming and memory scheme as shown in Figure 2.1 could be deduced which is referred to as heat stress memory assay (HSMA). This scheme is rated in a chlorophyll fluorescence measurement of photosystem II [F_v/F_m] using a PAM imaging device coupled to survival plate set-up comparing growth of pre-stressed and triggered-only cultures (for methodologies see below).

2.2.2 Heat shock memory assay (HSMA) procedure

In order to exclude undesired circadian variation, the experiment was always conducted in the morning hours between 9:30 and 10:30 AM. To include biological variance which improves reliability of the experiment, four cultures were grown in parallel for three days in advance. On day of experiment, cultures should reach $OD_{750nm} = 1.0$. Pre-grown cultures of each genotype were pooled and subsequently divided to three subcultures of 40 ml in 100 ml glass tubes. These were from hereon treated independently (Figure 2.1): culture N served as control and was kept at control 42°C (black line), culture T was subjected only to the second stress period (triggering, grey dotted line) and culture PLT was experiencing both, priming and triggering stimulus (yellow line), separated by a lag (recovery) phase of different durations: 2h, 24h, 48h. For the 2h and 24h lag phase experiment, culture OD was measured at start of experiment (prior to priming) and after the lag phase (before triggering).

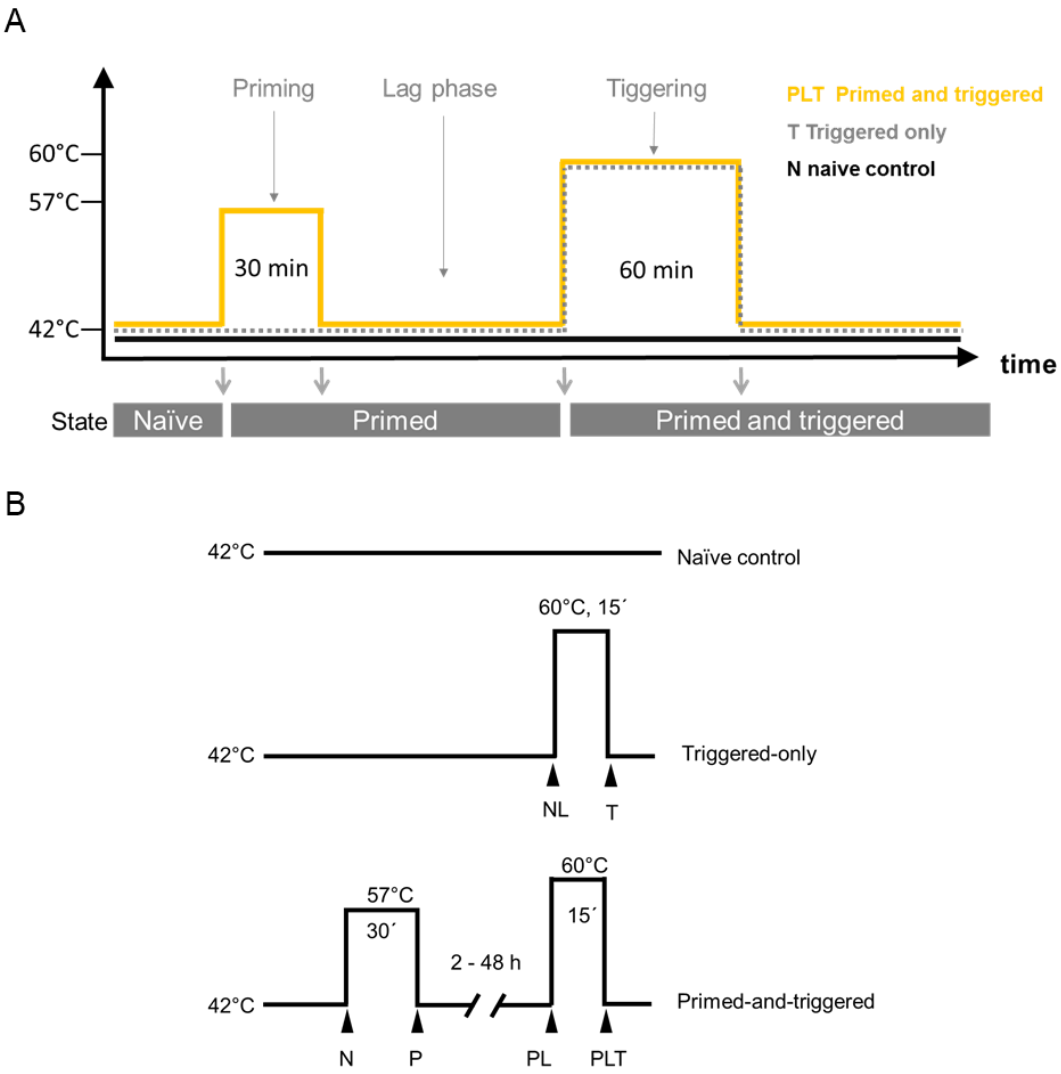


Figure 2.1: Heat stress memory assay (HSMA) scheme.
A Depicted is the duration and strength of priming and triggering stimuli as well as states of cells including naïve (N), primed (P) and primed-and-triggered (PLT) for the survival assay. Application of 30 min at 57°C as priming stimulus, followed by recovery (lag) phase durations between 2h and 24h lead to better performance of these primed cultures upon the triggering stimulus of 60°C for 1h compared to triggered-only (i.e. non-primed) cultures. Black line indicates conditions for control culture N kept at 42°C for the course of experiment. This control cultures served the N samples on survival plates. Grey dotted line indicates conditions for triggered-only (T) culture, which does not encounter the priming stimulus. Orange line indicates conditions for PLT culture, which experiences both, priming and triggering stimulus parted by the lag phase. Arrows indicate sampling time points from cultures resulting in 6 conditions: N, P, NL (after lag phase), PL (after priming and lag phase), T, PLT (see B)
B HSMA scheme used for RNA sequencing. Individual culture conditions are depicted to clarify conditions arising from individual cultures.

During 2h of lag phase, OD did not change substantially, while after 24h culture OD doubled to $1.8 < OD_{750nm} < 2.0$. In the 48h lag phase set-up, cultures were diluted 2h after priming to an $OD_{750nm} = 0.3$ to allow the cells remain in exponential phase for the duration of the lag phase. After 48h, cultures reached OD ranges between 0.8 and 1.4 accounting for a mitotic division rate of 3.5 to 4.5. The amount of cells was counted using a haematocytometer

(Neubauer Zählkammer, Hirschmann Laborgeräte GmbH & co.KG) and ratios calculated in order to estimate growth rates during lag phase (Table 3.2).

2.2.3 Pulse-Amplitude Modulation (PAM) measurements

The maximum quantum yield of photosystem II (PSII) in dark adapted states, $F_v/F_m = (F_m - F_0)/F_m$ was measured using a PAM IMAG-K4B fluorometer (Heinz Walz GmbH) with settings listed in Table 2.2. A saturating light flash of 1000 µmol photons/m^2/s^{-1} was used to measure the F_v/F_m ratio, while F_0 is the minimal (all PSII reaction centers open), F_m is maximal (all PSII reaction centers closed) and F_v is the variable chlorophyll fluorescence level in the dark-adapted state.

A

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2				WT_N_1	WT_N_2	WT_N_3						
3				WT_T_1	WT_T_2	WT_T_3						
4				WT_PLT_1	WT_PLT_2	WT_PLT_3						
5				EZ_N_1	EZ_N_2	EZ_N_3						
6				EZ_T_1	EZ_T_2	EZ_T_3						
7				EZ_PLT_1	EZ_PLT_2	EZ_PLT_3						
8												

B

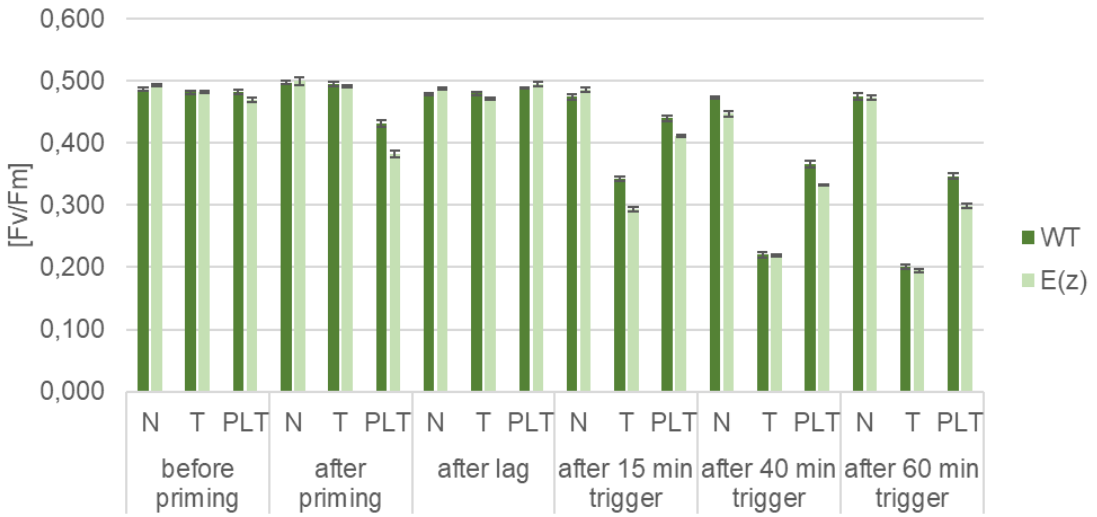


Figure 2.2: Pulse-amplitude-modulation (PAM) measurement
A Loading of 96-well plate for one timepoint measurement. Samples are indicated in the formate *genotype_culture_technical replicate*. Only wells shaded in white were used for measurements. Wells colored in grey are affected by the lense curvature of the camera and resulted in unreliable measurements.
B Representative PAM result of the fulfactorial (i.e. all cultures at all time points) experiment to determine flawless experiment. Note that priming leads to loss in Fv/Fm ratios (only PLT culture was subjected to priming). This loss is reversed after lag (recovery phase). Both genotypes, wild type (WT) and the Δ*E(z)* deletion strain evolved a priming benefit represented by elevated Fv/Fm ratios for PLT cultures upon triggering, independent of triggering length. Each time point, i.e. "before priming", "after priming", after lag", "after 15/40/60 min trigger(ing)" requires one 96 well plate

measurement as shown in A. Note that, N, T, and PLT in this regard are culture names, not conditions. The measurement after priming of PLT culture results in the condition P (=primed).

Each PAM measurement was conducted using a 96 well plate filled with 50 µl of culture in technical triplicates providing three technical measurements per culture and corresponding condition following the plate loading scheme in Figure 2.2A. PAM measurements were taken for each culture (i.e. N, T, PLT) before priming, after priming, after lag phase, and after triggering to ensure flawless execution of experiment. This resulted in a fulfactorial measurement exemplarily shown in Figure 2.2B. The experiment was rated as successful, when N cultures showed Fv/Fm around 0.5 and when primed cultures recovered to control Fv/Fm ratios after lag phase. Prior to PAM measurements, cells in 96-well plates were covered with aluminum foil and dark-adapted in non-transparent boxes for 8 min. Only the Fv/Fm ratio (= YII(0)) was used as a read out for basal fitness of PSII.

Table 2.2: PAM settings

Setting	***Value***
Intensity	3
Frequency	1
Gain	1
Damp	2
Saturation Pulse	5
+ adjust aperture to gain Fm around 0.2	

2.2.4 Survival Plates

Survival growth of treated and non-treated *C. merolae* cells was assessed on 0.75x MA2 agar plates. Plates were prepared according to Fujiwara and Ohnuma (2017). Gellan Gum was replaced by plant agar (Duchefa) and pH adjusted to 2.5 in the 3.3x MA2 solution prior to sterilization. Plates were stored at 4°C and best used after 6 weeks of storage. On day of experiment, 20 µl culture per spot were dropped on plate. After loading all conditions and timepoints according to the scheme in Figure 2.3A, plates and allowed to dry under sterile bench. Then, plates were closed, sealed with Leukopore tape (BSN Medical GmbH) and allowed to grow right side up in the Binder growth chamber under constant light at 42°C. Growth was assessed visually by taking a photograph 7 days after experiment. A representative plate growth of a 2h lag phase experiment is shown in Figure 2.3B.

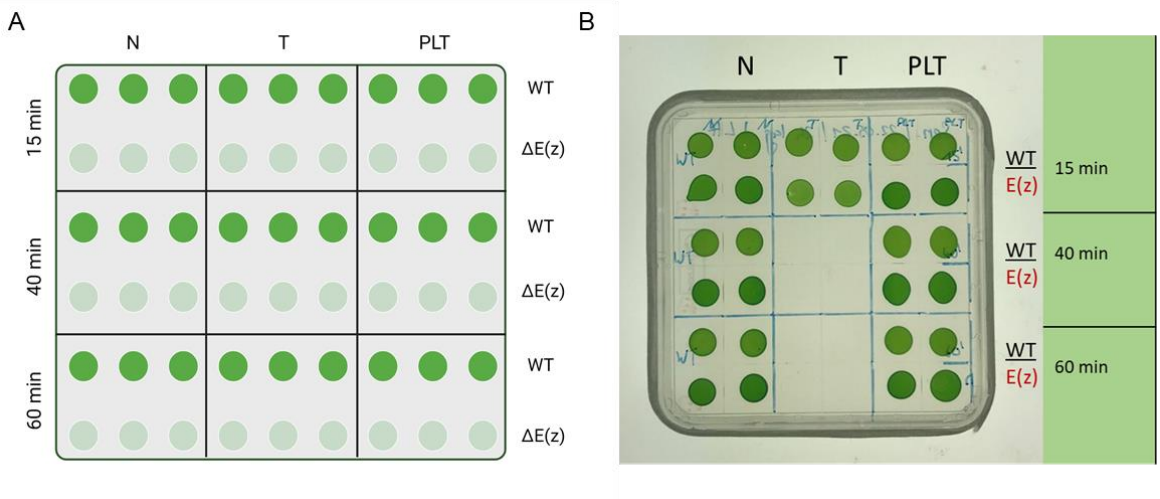


Figure 2.3: Survival plate loading.
A Loading scheme as applied during HSMA. **B** Representative survival plate results showing growth of primed before (PLT) cultures whereas triggered-only cultures refuse to grow. Photograph was takes after 7 days of growth in Binder growth chamber at 42°C under constant light (80 µE) in an 24h lag phase set-up.

2.2.5 Quantitative Real-Time PCR

2.2.5.1 RNA sampling

Importantly, during the HSMA the conditions/samples PLT and T were collected after 15 min in 60°C (= 15 min triggering) to avoid transcriptional responses to apoptosis (reflected exemplarily in no T growth in Figure 2.1B). 2 ml of *C. merolae* cells were harvested via centrifugation 3000 x g for 8 min at 4°C. The supernatant was removed and the pellet-containing reaction tube put into liquid nitrogen for 15 seconds for cell homogenization purposes. RNA was isolated using the pelett by INNUprep RNA Kit (Analytik Jena) according to the manufacturers´ instructions. Here, RL buffer was supplemented with 10µl/ml beta-mercaptoethanol (β-ME) to improve RNA quality. Dissolving the just thawed cell pelett with RL-β-ME buffer was done by pipetting up and down instead of vortexing. RNA was eluted using 40 µl RNAse-free water giving final RNA concentrations of approx. 400 ng/µl.

2.2.5.2 DNAse treatment

Genomic DNA was removed from 1 µg prepared RNA via treatment with DNAse I (Thermo Fisher) according to manufacturers´ instructions. RiboLock RNase inhibitor (Thermo Scientific) was added to a final concentration of 1 u/µl.

2.2.5.3 cDNA synthesis

First strand cDNA synthesis was performed using 0.5 µg DNAseI treated RNA, the RevertAID reverse transcriptase Kit (Thermo Scientific) and oligo(dT) primers according to

manufacturer´s instructions. The obtained cDNA was diluted 1:20 prior to using it in qPCR analyses.

2.2.5.4 Quantitative PCR

qRT-PCR was performed with 5 ng cDNA using the ROX SYBR 2X blue Master Mix (Takyon) and the QuantStudio5 RT-qPCR system (Applied Biosystems). Used primers are listed inTable 2.1. The qPCR program used is given in Table 2.4. Expression levels were measured in technical triplicates. Target gene expression was quantified relative to the mean Ct of three housekeeping (HK) genes (60S, TUB, and TIP41) using ΔΔCT method including primer efficiency (Pfaffl, 2001). For that, a standard curve for each primer was generated to calculate primer efficiency. A pool of of all cDNAs served as template DNA and was used in the dilutions 1:1, 1:10, 1:100. Primer efficiency was calculated as follows: technical replicate Ct values for each cDNA was calculated and plotted against the $\log_{10}$ values of the dilution factors. The slope of the regression between log values and the average Ct values is used to calculate the efficiency for each primer pair in % following the equation: Efficiency (%) = $(10^{-1/slope} - 1) * 100$. Efficiencies between 90 and 110 % in combination with a regression factor $0.95 < R^2$ were considered as suitable for qPCR. Efficiencies were included in the determination of relative expression as follows:

$$relative\ expression = \frac{efficiency\ (target\ gene)^{Ct\ mean\ (target\ gene)}}{efficiency\ (housekeeper)^{\ Ct\ mean\ (housekeeper)}}$$

Table 2.3: qRT-PCR primer

Gene	*Primer name*	*Primer Sequence (5´→3´)*
CMJ099C	*CMJ099C*_for	GAAGCTCCCGGTCTGTATGG
	*CMJ099C*_rev	TCTGGATCTCGCCAACCAAC
CMJ100C	CMJ100C	CGGGGACCTCTTGACCATTC
	CMJ100C	CGACATTCTTCGGCAGCTTG
CMJ101C	*CMJ101C*_for	ATCCCTTCTTCATGTCGCCG
	*CMJ101C*_rev	TCGCAACTAACCACAAACGC
CMJ102C	CMJ102C_for	AGCGCGTCATCTCCTTCATT
	CMJ102C_rev	AGCGAGCCCGAACCAAATAA
CMP040C	*CMP040C*_for	GTATTGGGCTGTGGTTGCAC
	*CMP040C*_rev	ATGATTGAGCGTCGTGGGAG
CMF092C	*CMF092C*_for	TGACGAGAAGAGCGACTGTG
	*CMF092C*_rev	CGTTGCCAAGACGCAAAAGA
60S	CYME_60S_Fw	AAGTTTCGCTGTACGCTTGG

	CYME_60S_Rv	TAACCAGGACCATATCGCCG
CMN304C - TUB	cm_TUB_G_for	TTCACGATCGCTACCCGAAG
	cm_TUB_G_rev	CTGCCGCTGTTCGCATAATC
CMM193C – TIP41	CMM193C_for	TCACGCAAAACAACATCCAT
	CMM193C_rev	TCAACAGCGTTGAATCGAAG

Table 2.4: Cycler program for quantitative real-time PCR

Temperature and Time	***Step***	
95°C for 3 min	Hold stage, initial denaturation	
95°C for 3 sec	Denaturation	40 x
60°C for 30 sec	Annealing & Elongation, data acquisition	
95°C for 15 sec	Melt curve stage , denaturation	
60°C for 1 min	Melt curve stage, complete annealing	
95°C for 15 sec	Melt curve stage, dissociation curve, data acquisition	

2.3 Transcriptomics

2.3.1 Sampling and RNA extraction

WT and Δ*E(z)* cells sampled according to the conditions in Figure 2.1B and sent for sequencing. *C. merolae* cells were sampled as follows: 2 ml of culture were harvested by centrifugation 3000 x g for 3min at 4°C. The supernatant was removed and the pellet-containing reaction tube put into liquid nitrogen for 15 seconds for cell homogenization purposes. RNA was isolated from frozen pelletts using the INNUprep RNA Kit (Analytik Jena) according to the manufacturers´ instructions. In addition, the RL buffer was supplemented with 10µl/ml beta-mercaptoethanol (β-ME) to improve RNA quality. RNA was eluted in RNAse-free water. Genomic DNA was removed from 1 µg prepared RNA via treatment with DNAse I (Thermo Fisher) according to manufacturers´ instructions. RiboLock RNase inhibitor (Thermo Scientific) was added to a final concentration of 1 u/µl.

2.3.2 Sequencing

Library preparation and sequencing was done by Novogene UK (Cambridge). mRNA library was prepared using polyA enrichment and the NEB Ultra RNA Library Prep Kit producing unstranded data. Sequencing was done on an Illumina NovaSeq 6000 platform resulting in 150 bp long paired-end reads. 3G of raw data per sample were delivered. For details on alignment and assignment rates see Table S1.

2.3.3 Bioinformatics

2.3.3.1 Alignment & Assignment

Data computing for read alignment and assignment was done using the high performance computer *Curta* (Bennett *et al.*, 2020) from FU Berlin. *C. merolae* genome reference sequence and annotation files used are listed in Table 2.5 and were retrieved from ENSEMBL (Cunningham *et al.*, 2022).

A. Quality control of reads

Raw read quality was assessed by FastQC (v0.11.9, Andrews, 2010). No read trimming was necessary. All sequences passed quality control (QC) by FasctQC, indicated by Phred scores > 35 along the entire read lengths (Figure S1).

B. Genome Indexing

Prior to aligning the raw reads to it, the *C. merolae* genome was indexed using the GTF file and genome sequence fasta file (Table 2.5) with **STAR aligner** (v2.7.9a, Dobin *et al.*, 2013). The *-- runMode genomendGenerate* was used and options were set to *--sjdbOverhang 149* (read lenghts-1) and *--genomeSAindexNbases 11*.

C. Read mapping/ alignment

Mapping each sample's reads to the genome index was done using the STAR program producing SAM files (*--outReadsUnmapped Fastx --outSAMunmapped Within*) for each sample. More than 92.5 % of the reads were mapped for each sample.

D. Read assignment

For counting reads assigned to genomic features **featureCounts** (Liao *et al.*, 2014) was used (included in the Subread package, v2.0.1) resulting in the count data reporting, for each sample, the number of sequence fragments that have been assigned to each gene and serves as basis for differential gene expression analysis. Only reads with an alignment score higher than 10 were used:

featureCounts --T 8 --s 0 --Q 10 --p --C --a path/to/gtf --o destination/path/to/put/txt_outfile path/to/sam/files/used

Note: When the new_anno gtf file was used for genome indexing and counting, the additional option *--g gene_name* was needed in the above bit of code to refer to GeneIDs within the gtf.

E. File conversion for visualization purposes

SAM files were converted to sorted and indexed BAM files using **SAMtools** (v1.11, Li *et al.*, 2009):

samtools view -S -b $FILE_Aligned.out.sam > $FILE_Aligned.out.bam
samtools sort -@ 2 $FILE_Aligned.out.bam -o $FILE_Aligned.sort.out.bam
samtools index $FILE_Aligned.sort.out.bam

To visualize read coverages in the **Integrative Genomics Viewer** (v2.10.3, Robinson *et al.*, 2011) BAM files were converted (compressed) to bigwig file format using **DeepTools** (v3.3.2, Ramírez *et al.*, 2016):

bamCoverage -bs 10 -e -of bigwig -b $FILE_Aligned.sort.out.bam -o $FILE.bw

Table 2.5: Genome reference files

	Format	Download link	Date of download
Genome sequence	FASTA	ftp://ftp.ensemblgenomes.org/pub/plants/release-46/fasta/cyanidioschyzon_merolae/dna/	28.01.2021
Genome annotation file	GTF	ftp://ftp.ensemblgenomes.org/pub/plants/release-46/gtf/cyanidioschyzon_merolae/	28.01.2022

2.3.3.2 Differential expression analysis

Differential expression analysis and visualizations were performed in R Statistical Software (v4.1.0); R Core Team 2021) using the DESeq2 package (v.1.3.20, Love *et al.*, 2014). Prior to DE analysis, proper clustering of samples regarding condition and genotype was analyzed via principle component analysis (PCA) using the *plotPCA()* function from DESeq2. Principal component (PC) 1 accounted for 52% of the variance and separated samples by condition. Priming exhibited highest variability and separated stressed from non-stressed samples as well as triggering from priming. As result, samples separated by condition on PC1 and by genotype PC3/PC4 (Figure 2.4). PCA analysis enabled to detect two outlier samples.

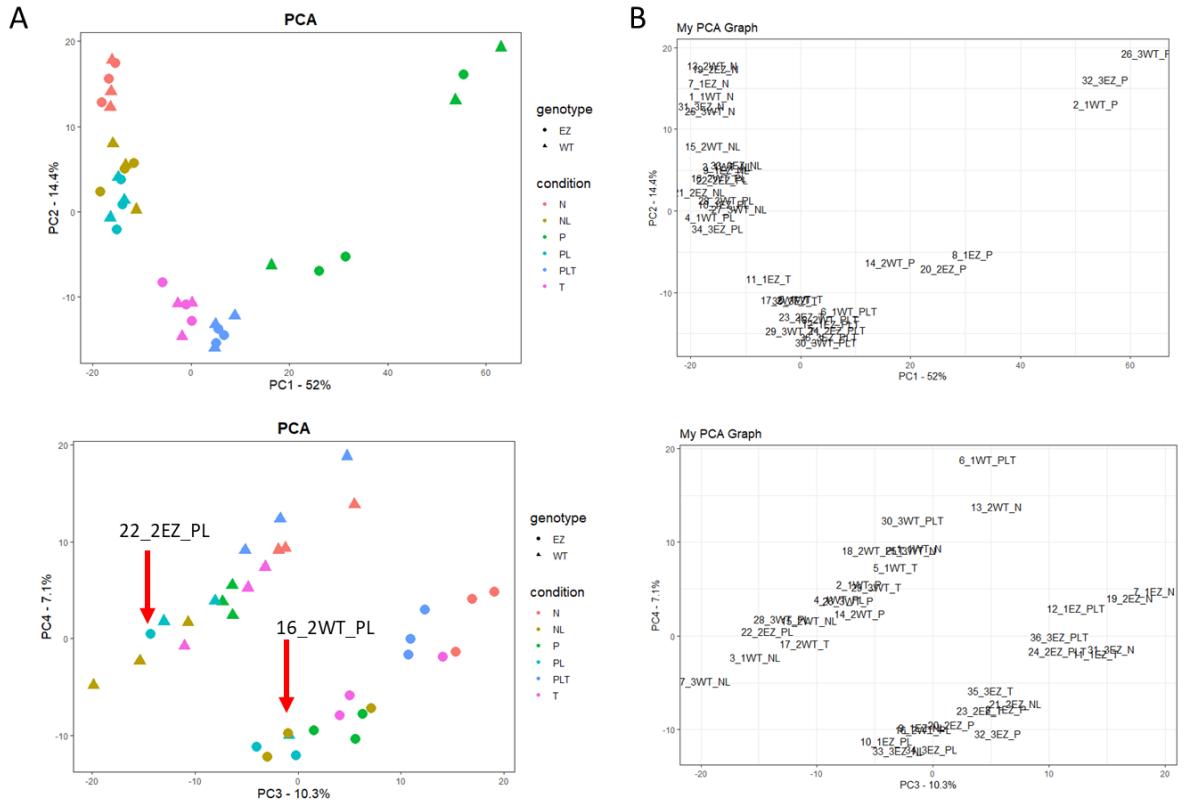


Figure 2.4: Principal component analysis (PCA) of 36 sequencing samples.
PCA was done using the plotPCA() function of DESeq2 (Love *et al.*, 2014). A PCA of PC1/2 (upper panel) and PC3/4 of WT (triangle) and Δ*E(z)* mutant (dot) naïve (N, red), primed (P, green), naïve after lag phase (NL, ochre), primed after lag phase (PL, turquoise), primed and triggered (PLT, blue), and triggered-only (T, pink) samples are shown. Note that priming samples display highest variability.

In detail, sample 22_2EZ_PL clustered with WT on PC3/PC4 and 16_2WT_PL clustered with Δ*E(z)* samples. Read traces of these two samples visualized in IGV for the *E(z)* locus (*CMQ156C*), revealed indeed wrong naming of the samples (Figure 2.5). Since condition and replicate number are identical in those two samples, and read traces unambiguously show that the WT sample shows no reads from gene E(z) (the deleted gene in the Δ*E(z)* mutant) and *vice versa* for the Δ*E(z)* sample it was considered a fact, that samples have been switched during harvesting. As a consequence, *fastq* files delivered from Novogene UK (Cambridge) were renamed. Sample 16 was corrected to 16_2EZ_PL and sample 22 to 22_2WT_PL.

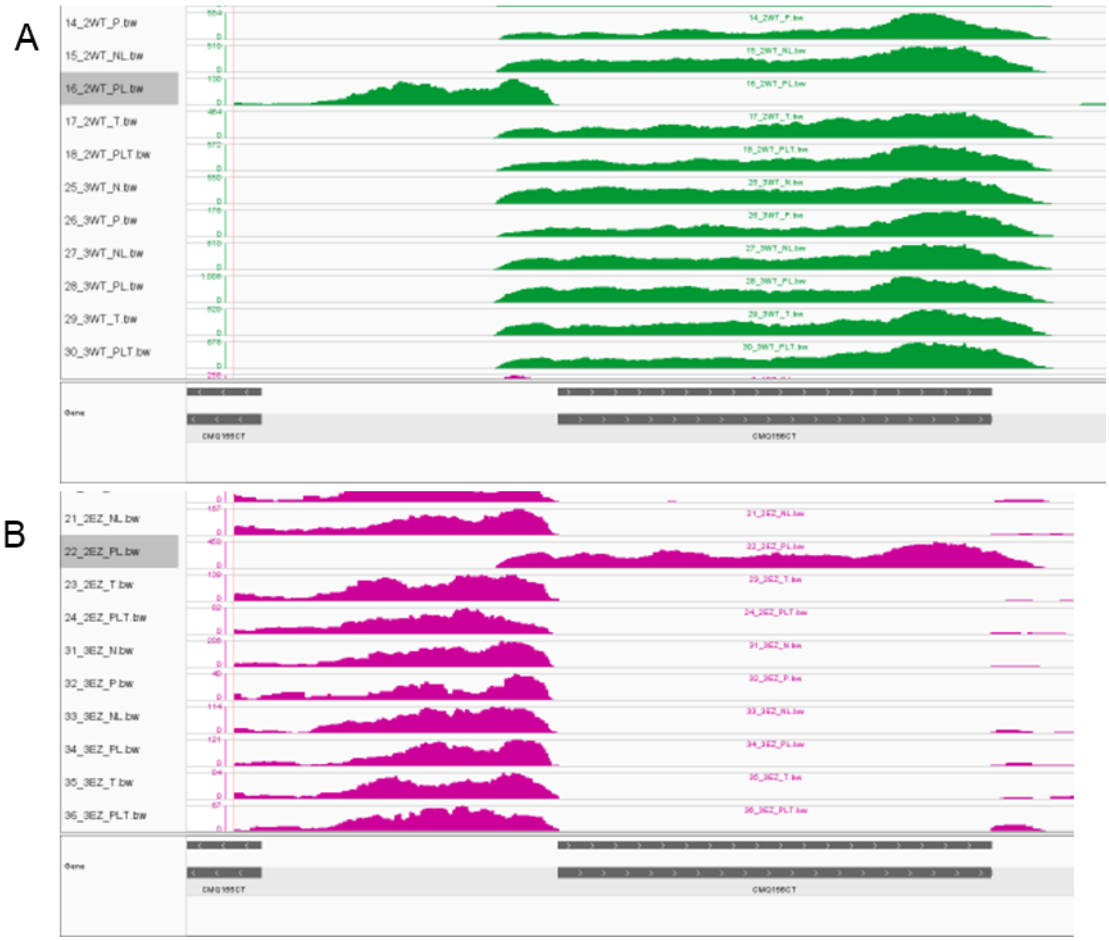


Figure 2.5: Read coverage at the Enhancer of zeste (*E(z), CMQ156C*) gene in the samples detected as outliers in principle component analysis revealed switched samples.
A WT sample read traces of all samples show aberrant behaviour of 16_2WT_PL with no reads of gene *CMQ156C, E(z)*. **B** *ΔE(z)* sample read traces of all samples confirm aberrant behaviour of 22_2EZ_PL reflected in reads for the allegedly deleted gene *CMQ156C, E(z).*

After outlier detection and quality control by visualizations, differential expression analysis was performed using the DESeq2 package. DESeq2 provides methods which base on negative binomial generalized linear models and calculates normalized size factors for each condition followed by median of ratios method. DESeq2datasetObject was created using design formulas according to the DE analysis. To determine differentially expressed genes between WT and *ΔE(z)* under control conditions, only the N samples were processed down to the count data table (including read alignment, assignment, counts table). Accordingly, for this DE analysis in DESeq2 *design = ~ genotype + replicate* was used in DESeq2dataset object generation (*DESeqDataSetFromMatrix()* function). After differential expression analysis the *contrast <- c("genotype", "EZ", "WT")* was used.

When determining memory genes, only WT samples were aligned, assigned and counted and the design formula *design = ~ condition + replicate* was used in DESeq2dataset object generation followed by Wald test using the *contrast <- c("condition", "PLT", "T")*. Results were shrunken via *lfcShrink()* and the false discovery method *ashr* applied

(Stephens, 2017). p-values were adjusted by Benjamini Hochberg procedure. Genes were considered differentially expressed when the adjusted p-value was below 0.05 and the Log2FC exceeded 0.58 which accounts for 1.5 fold change.

2.3.3.3 Visualization

To visualize gene expression, the ggplot2 package was used. For boxplots, normalized counts were retrieved from the DESeq dataset object using the function *counts()* with *normalized=TRUE.* The model based on all 36 samples and the formula *design = ~ genotype + condition + genotype:replicate.* For heatmap generation, matrices from DESeq dataset objects were retrieved using *coef()*. Heatmaps were generated using pheatmap package (Kolde, 2019)

2.4 Data visualization and statistics

The results of chlorophyll fluorescence measurements were analyzed and visualized in GraphPad Prism 9.3.0 for Windows, GraphPad Software (San Diego, California USA, www.graphpad.com). Normal distribution of data was tested using Anderson-Darling and Shapiro-Wilk test. Significances were calculated using unpaired t-test or ANOVA followed by Šidák´s test. qPCR analysis was done in R using the agricolae package (v1.3-5, de Mendiburu & Yaseen, 2020). Venn diagram analysis was done in R using the package *VennDiagram* (Chen & Boutros, 2011) and the significance of overlaps was calculated using a fisher´s exact test.

3 Results

This project aims to uncover the relevance of epigenetic processes implemented by the Polycomb Repressive Complex 2 (PRC2) in a unicellular photoautotrophic eukaryote. *Cyanidioschyzon merolae*, an extremophile red alga with photosynthetic activity, represents a suitable model to address this question, as it possesses PRC2 components and H3K27me3 genomic levels (Mikulski *et al.*, 2017). The absence of both, developmental phase transitions and cellular/organ identities together with the very low number of genomic repetitive element fractions does not imply obvious “conventional” PRC2 functions, such as developmental regulation and maintenance of cellular identities. However, as H3K27me3 decorates half of the repetitive element fraction in the *C. merolae* genome, it likely also silences these genome threatening parasites. Within this project, this hypothesis is tested via an RNA-sequencing experiment comparing the transcriptomes of WT and *ΔE(z)* cells (Chapter 3.1).

Although *C. merolae* is devoid of mitotic memory in relation to cellular identities, PRC2 could contribute to the alga´s (mitotic) memory of environmental cues which generates superior tolerance in case the cells are facing stress repeatedly. This phenomenon, referred to as priming, is well-described for several other non-nervous system organisms and participation of PCR2 within this mechanism was suggested (Hilker *et al.*, 2016). Prior to test the relevance of PRC2 in stress memory responses, it has to be proven, whether *C. merolae* indeed harbors the ability to generate stress memory responses (Chapter 3.2). For that, a working stress memory scheme has to be established allowing the determination of stress memory on physiological level. As *C. merolae´*s most frequent stressor in nature, heat represents a suitable stressor-candidate (Chapter 3.2.1). To test potential epigenetic contribution, inclusion of the Polycomb mutant *ΔE(z)* in this scheme shall answer the question whether and to what extent PRC2 contributes to stress memory formation or maintenance (Chapter 3.2.2). Subsequently, it was of interest to link the physiological memory response to the transcriptional response of specific genes. Therefor, transcriptomic analyses via RNA-sequencing of the stress memory scheme, including WT and *ΔE(z)*, were envisioned (Chapter 3.3). In parallel, transcriptome responses following individual stress stimuli were compared between WT and Δ*E(z)* to determine whether PRC2 regulates the inducibility of stress-responsive genes (independent of memory) (Chapter 3.3.1). Next, physiological memory mechanisms are further characterized by the identification of heat trainable genes, which are specifically regulated by repetitive stress (Chapter 3.3.2). As initial characterization of stress memory mechanisms, peculiar regional and functional regulations are dissected in Chapter 3.3.2 and Chapter 3.4. Comparing the performance and transcriptional responses of *ΔE(z)* to WT behavior might reveal, if *C. merolae* stress

responses and memory thereof depends on PRC2 action thereby describe epigenetic relevance in a unicellular photoautotrophic eukaryote for the first time.

3.1 Polycomb repressive complex 2 contributes to transcriptional silence of both, genes and repetitive elements

In the *Cyanidioschyzon merolae* genome, 50% of all repetitive elements are enriched with the silencing mark H3K27me3 deposited by PRC2. One way to address the question, whether H3K27me3 occupancy also ensures transcriptional repression of these sequences is to compare the transcriptomes of wild type and H3K27me3-lacking, i.e. *ΔE(z) C. merolae* cells via RNA-sequencing. Thus, mRNA of WT and *ΔE(z)* cells growing under control conditions (42°C, pH = 2.5, constant light: 90 µE) was sequenced. Sequencing resulted in at least 10 mio paired-end reads of 150bp in length per sample, in three biological replicates. Quality control was passed for forward and reverse reads of all 36 samples (Figure S1). No read trimming was necessary. Transcriptomic analyses always depend on proper annotated genomes in order to draw conclusions about the constitution of affected genes. For *C. merolae*, the publicly available annotation file does not include any repetitive sequences (RE), such as pseudogenes, rRNA genes and transposable elements. Thus, unfortunately, using the published annotation file no conclusions can be drawn about the potential involvement of *E(z)* in silencing repetitive (potentially regulatory) sequences. Nevertheless, since H3K27me3 was recently described to target repetitive elements in *C. merolae* (Mikulski *et al.*, 2017), it was crucial to test the hypothesis of elevated repetitive element transcription upon the absence of this histone mark. Therefore, and thanks to a tight collaboration with Prof. Dr. Stephen Rader´s lab (University of Northern British Columbia, Prince George, Canada), the investigation of repetitive sequences regulation upon absence of *E(z)* was done using an annotation file under generation. This annotation file, referred to as “new anno” from here on, does not include only genes, but repetitive sequences as well. Due to its unreviewed status still, in this project it served only the purpose to preliminary define transcriptional changes of the repetitive sequence fraction following the absence of H3K27me3 under control conditions. Nevertheless, the published (“old”) annotation file was used in the analysis of (single copy) genes, since usage of the same annotation file ensures comparability to other, published, transcriptomic data and the new anno file is still in revision. To investigate transcriptional changes of repetitive elements upon H3K27me3 loss, the analysis of differential expression in WT and *ΔE(z)* samples was repeated with the new anno

as reference and only results for REs were extracted. Contrarily, the information about DE single copy genes was collected using the published annotation file to ensure comparability with known data sets. Furthermore, due to the unreviewed status of the new anno, using the published annotation file gains more reliable results. Consequently, as indicated in Figure 3.1A, differential expression (DE) analysis results for single copy genes and repetitive elements (RE) were generated using the published, and new anno, respectively. In order to distinguish repetitive element (RE) genes from single copy genes, differentially expressed repetitive elements are referred to as DE-RE from here on.

A

	single copy genes	repetitive sequences
number of DEG´s between WT and ΔE(z)	108	111
total number in genome	5009	393
changed genomic proportion	2,20%	28%
DE H3K27me3 targets	16	50
proportion of H3K27me3 targets within DE elements	15%	45%
annotation file	published	new anno

B

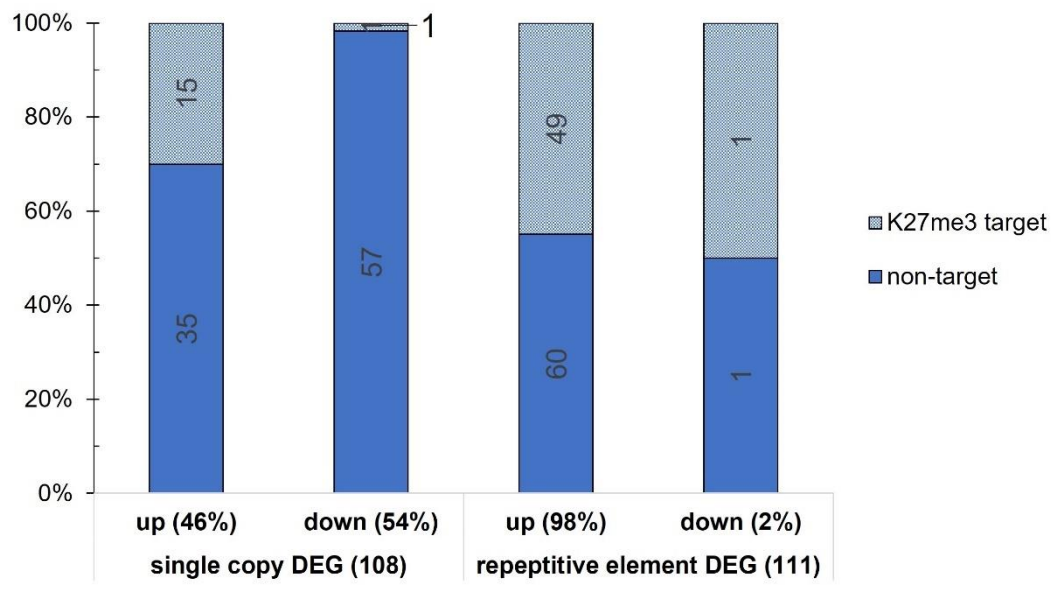


Figure 3.1: Transcriptomic analysis of *C. merolae ΔE(z)* strain.
A Overview of differentially expressed (DE) element constitution including H3K27me3 target identification. DE elements are divided to protein coding genes (genes) and repetitive element coding genes (repetitive elements) and changes are depicted as ratios to total genome numbers. DE repetitive elements were identified using an annotation file under generation (S. Rader, new anno) revealing nearly 1/3 of genomic repetitive elements misexpressed in the *ΔE(z)* strain. DE genes were identified using the publicly available annotation file (Matsuzaki *et al.*, 2004) revealing a proportion of

2.2% being misregulated in *ΔE(z)*. Respective proportions of elements originally targeted by H3K27me3 are noted. **B** H3K27me3 regulated repetitive elements are subject to upregulation in *ΔE(z)*. Total number of DE genes and repetitive elements is shown and divided by their direction of transcriptional changes upon *ΔE(z)*. Proportion and distribution of defined H3K27me3 targets subject to changes in *ΔE(z)* strain are indicated by light blue bits of bars and are predominating in elements showing upregulation. Regulation of genes equally distributes towards up- and downregulation. Numbers in parentheses indicate number of total up- or downregulated elements in genes or repetitive elements. Numbers in bars specify numbers of H3K27me3 targeted (light blue dotted) and non-targeted (darker blue) elements within classes.

3.1.1 Limited H3K27me3 transcriptional regulation of single copy genes

DE analysis revealed that both, genes and repetitive elements repetitive elements are subject to transcriptional changes in the *ΔE(z)* mutant. Consistent with the absence of an obvious phenotype in the *ΔE(z)* mutant, only 2.2 % of the whole gene number are affected by the absence of *E(z)* (Figure 3.1A). Using the published annotation, 114 differentially expressed genes (DEG) were identified, consisting of six retroelements and 108 single copy genes. The six retrotransposons were allocated to the DE-RE fraction form here on. Within the 108 single copy DEGs, directions of changes are distributed in comparable fractions towards upregulation (46%) and downregulation (54%, Figure 3.1B). In order to elucidate, if H3K27me3 enrichment is directly responsible for transcriptional regulation in *C. merolae*, these 108 single copy DEGs were compared to the previously described H3K27me3 target gene list (Mikulski *et al.*, 2017). Consequently, overlap analysis revealed that only 16 out of the 108 single copy DEGs were originally targeted by H3K27me3 and in parallel subject to transcriptional regulation upon the histone mark´s absence. Mikulski *et al.* (2017) defined 246 H3K27me3 gene targets in total. This means, that H3K27me3 transcriptionally regulates only 15% of all H3K27me3 enriched genes under control conditions. Remarkably, all of the 16 regulated H3K27me3 targets showed an upregulation with the exception of one gene. Albeit the very small total number of 16, this fact validates the repressive nature of H3K27me3 (Schuettengruber *et al.*, 2007). Only one gene enriched with H3K27me3 in WT shows the opposing situation and is repressed in the mutant: *CMI166C*. In addition to H3K27me3 targets, non-targets are regulated by the absence of H3K27me3. One explanation for this could be that single transcriptional regulators are subject to H3K27me3 control which entail non-target regulation. Within the upregulated fraction of the DEGs in the mutant, no H3K27me3 targeted transcription factors could be identified. However, four non-H3K27me3 targets encoding for transcriptional regulators found within the list could bear the potential to downstream affect the expression of a multitude of genes leading to the considerable number of regulated non-H3K27me3 targets, including the genes *CMR153*C, *CMN310C*, *CMR155C*, *CMS371C*. As none of them is directly targeted by H3K27me3 the precise mode of H3K27me3 regulating these transcriptional regulators remains elusive.

3.1.2 Repetitive sequences at subtelomeric regions are upregulated upon genomic H3K27me3 loss

Differential analysis using the new annotation file (under generation) as reference identified 105 differentially expressed repetitive elements (DE-RE). Including the six retrotransposons detected in the single copy DE analysis already, the number rises to 111 DE-RE in total. Hence, as the *C. merolae* genome comprises 393 annotated REs to date, more than 1/4 (i.e. 28%) of all *C. merolae* REs lose their transcriptional pattern when genomic H3K27me3 is depleted (Detailed information on REs available on the *C. merolae* genome project website http://czon.jp). Notably, nearly the entire DE-RE set, namely 98%, was upregulated in the *ΔE(z)* mutant, with the exception of two REs only. To elucidate, if H3K27me3 occupancy can be directly linked to transcriptional regulation, the proportion of H3K27me3 targeted REs within the DE-RE set was determined via comparison with the H3K27me3 target gene set (Mikulski *et al.*, 2017). Intriguingly, nearly half of all DE-RE (49 i.e. 44%) are enriched with H3K27me3 and in the same time upregulated in the *ΔE(z)* mutant.

In summary, the relative proportion of regulated loci upon H3K27me3 loss is much higher in RE than in single copy genes (2 % vs 28%). Moreover, the direction of changes is balanced in single copy DEG, whereas nearly all DE-RE are upregulated (46% up in single copy DEG vs. 98% up in DE-RE). This fact suggests a relevant impact of H3K27me3 in repetitive element silencing in *C. merolae* and reminds to findings reported previously for the diatom *Phaeodactylum tricornutum* (Veluchamy *et al.*, 2015). Notably, among the only two repressed repetitive elements is one H3K27me3 target (*CMI001X*). *CMI001X* is located at the subtelomere of chromosome 9 and has been identified as H3K27me3 target by Mikulski *et al.* (2017). Together, when allowed to transpose unrestrained in the course of time, the induction of such high proportions of the genomic repetitive elements might threaten the genome integrity as such.

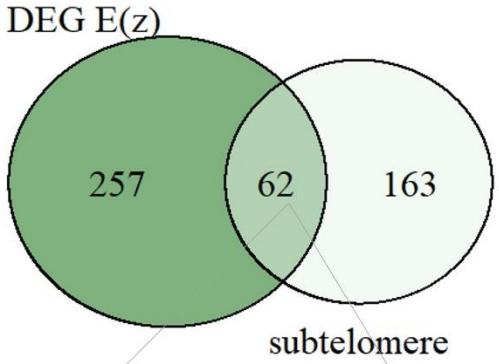
Mikulski *et al.* described a general preferential enrichment of H3K27me3 at chromosomal ends (Mikulski *et al.*, 2017). As just described, *CMI001X* is the first gene on the right arm of chromosome 9 and is repressed upon H3K27me3 absence. To test for a potential transcriptional regulation of sub-/telomeres by H3K27me3, all DEGs identified using the new and published anno file were merged (i.e. 319 DEG) and this dataset was compared to subtelomeric genes described by Nozaki and colleagues (2007). Venn diagram overlap analysis revealed, that the transcription of 27% (62 out of 225 in total) of genes located at subtelomeric regions was affected by the absence of H3K27me3 in *C. merolae* (Figure 3.2A, Table S4). The overlap of those two groups of genes is higher than statistically expected (Figure 3.2A, representation factor = 4.3, $p = 5*10^{-25}$). Further, 35 of these 62 misregulated subtelomeric genes are direct targets of H3K27me3, meaning that 56% of H3K27me3 regulated subtelomeric genes are also enriched with H3K27me3 (Table S4). The direction of

transcriptional change at subtelomeres is biased towards upregulation. Only *CMI001X* and *CMI002C* are downregulated subtelomeric elements, all remaining are (highly) induced in *ΔE(z)*.

These 62 genes are distributed to 26 chromosomal ends of 18 chromosomes. No significant transcriptional changes could be observed on neither end of chromosome 13 and 17 (Figure 3.2B). As 27% of the subtelomeric genes on 26/40 chromosomal ends are losing their natural expression pattern, *E(z)* at least in parts ensures proper subtelomeric regulation. Still, this number is not completely reflecting the formerly described H3K27me3 enrichment at chromosomal ends (Mikulski *et al.*, 2017), Figure 3.2B). There, 36 out of 40 chromosome ends were described to be enriched with H3K27me3. The incomplete (up)regulation upon H3K27me3 loss suggests that this histone mark is necessary, but not sufficient in order to maintain subtelomeric regulation in *C. merolae*. One characteristic of telomeres and subtelomeres is their repetitive sequence nature. As 27 of the 62 regulated subtelomeric genes are indeed repetitive elements, the transcriptional regulation of H3K27me3 at telomeres might be considered as one consequence of the genome wide regulation (silencing) of repetitive elements described above.

Surprisingly, after H3K27 methylation loss, only 18% of genomic *C. merolae* H3K27me3 targets showed changed expression (66 out of 364, Figure 3.1A). Mirroring the behavior of repetitive elements and subtelomeric regions, the vast majority of differentially expressed H3K27me3 targets (64 loci, i.e. 97%) shows significant positive log fold changes, repeatedly confirming the repressive nature of H3K27me3. Out of those 66 H3K27me3 targets, all but nine are either found on subtelomeres or are repetitive elements or both. As the functionality of the telomere as well as proper regulation of repetitive elements together are essential to upkeep the genome stability, these results suggest a decisive role of H3K27me3 in subtelomere regulation via repetitive element silencing. Interestingly, the remaining nine misregulated K27me3 gene targets consist of a nitrate transporter, a purple acid phosphatase, an N6-adenosine methyltransferase, an Fe2OG dioxygenase domain-containing protein and an cathepsin D. Four genes encode for proteins, which are uncharacterized to date.

A



Representation factor = 4.3
$p < 5*10^{-25}$

B

Chromosome		5´telomere	3´telomere	Telomere regulation in ΔE(z) at 5´ telomere / 3´ telomere	Enrichment of K27me3 on chromosomal ends on 5´ end / 3´ end
#		GeneID	GeneID		
1	A	**CMA001C**, **CMA003X**	**CMA142C**,CMA143X, CMA144Z, CMA145C, **CMA146X**	+/+	+/+
2	B		**CMB162C**	-/+	+/+
3	C	CMC002C, **CMC003X**, **CMC004Z**	CMC187X, CMC188C	+/+	-/+
4	D	**CMD002C**, **CMD003Z**, **CMD004X**	**CMD189X**, **CMD191C**	+/+	+/+
5	E	**CME001C, CME002Z, CME003X**		+/-	+/+
6	F	**CMF005X**	**CMF189Z**, **CMF190Z**	+/+	+/+
7	G	CMG221C	CMG219Z	+/+	+/-
8	H	**CMH002Z, CMH003Z, CMH004X**		+/-	+/+
9	I	**CMI001X***, CMI002C *, CMI003X, **CMI005X**	**CMI305X**, CMI306C, CMI307X	+/+	+/+
10	J	**CMJ001X**, CMJ002C, CMJ003X, CMJ005C, **CMJ006Z**		+/-	+/+
11	K		**CMK310X**	-/+	+/+
12	L	CML004C, CML005C	**CML337X**	+/+	+/+
13	M			-/-	+/+
14	N		CMN339Z	-/+	+/-
15	O	**CMO001C**, **CMO002X**		+/-	+/+
16	P		**CMP357X**, CMP358C	-/+	+/-
17	Q			-/-	+/+
18	R	CMR001C, CMR004C	**CMR502C**, **CMR503X**, CMR505Z	+/+	+/+
19	S	**CMS001X,** CMS002C, **CMS003X**		+/-	+/+
20	T	CMT002X, CMT003C, CMT005X, CMT007X		-/+	+/+
		*only genes downregulated			

Figure 3.2: *ΔE(z)* transcriptional changes are dominating at subtelomeres but only partially coincide with H3K27me3 enrichment on chromosomal ends.

A Venn diagram overlap analysis of *ΔE(z)* DEGs (genes and repetitive elements) and subtelomeric genes (Nozaki *et al.*, 2007) indicates similar regulation of both gene sets. *ΔE(z)* DEGs include RNAseq data analysis of WT and *ΔE(z)* under control conditions by DESeq2 (fold change > |1.5|, padj < 0.05) using published annotation file and are merged with results retrieved by same analysis using an annotation file under generation, which includes repetitive genomic sequences. Significance of overlap was calculated by Fisher's exact test.
B Table listing all misregulated subtelomeric genes situated on right (5´ telomere) and left (3´telomere) chromosomal ends in *ΔE(z)* strain on single chromosomes (1-20, A – T). Only two genes downregulated are marked with an asterisk, all remaining are upregulated in the mutant. Repetitive element genes end with the suffix CM…X. H3K27me3 targeted subtelomeric genes are indicated in bold-faced letters. Telomere regulation is listed in 4th column and indicates presence (+) or absence (-) of gene regulation at 5´ and 3´ telomere. Enrichment of H3K27me3 on chromosomal ends is listed in last column. Enrichment data are retrieved from Mikulski *et al.*, 2017 describing presence (+) and absence (-) of H3K27me3 enrichment on corresponding chromosome ends.

3.1.3 *E(z)* might regulate the response to changing environment

Initially it was hypothesized, that H3K27me3 might control genes which are inducible under certain conditions, including stress situations. 108 DEGs were identified in the *ΔE(z)* mutant. Usually, to describe the function of a given gene set, gene ontology (GO) analysis would provide insights into the affected gene functions. Unfortunately, the small number of single copy DEGs in the *ΔE(z)* mutant accompanied by incomplete annotation/characterization do not allow for statistically convincing GO term/pathway/KEGG analyses. Exemplarily, only 73/108 genes had a GO term for biological function allotted, which did not allow to conduct statistical analyses. As a consequence, to confer potential functions of *E(z)* regulation the DEG list could to be analysed manually in order to find obvious categories, potential regulators. Alternatively, a comparison of the regulated gene sets to other published transcriptomic data could reveal similar regulatory effects.

To address the question, whether H3K27me3 regulates (stress) inducible genes in *C. merolae,* as it was hypothesized in the beginning, the *ΔE(z)* regulated set of genes was compared to various other stress transcriptomes available for *C. merolae*. Comparing regulated gene sets of heat shock (Kobayashi *et al.*, 2014) and differing light qualities (Tardu *et al.*, 2016) to *ΔE(z)* DEGs did not reveal any similarities in regulation (Table 3.1). No significant overlap was detected for the heat shock response; neither for the heat shock 28°C to 50°C nor for the 42°C to 62°C cells. Likewise, the effects of different light qualities on growing *C. merolae* cells showed independently regulated genes sets.

However, limiting nitrogen (LN stress) applied for a duration of 24h (S. Rader, unpublished) revealed a slight overrepresentation of *ΔE(z)* regulated genes (Table 3.1 and Figure 3.3A). Due to the discrepancy in the dimensions of the datasets (nearly 3000 genes vs. 114) the overlap loses statistical strength, however, 75% of the *ΔE(z)* is affected by LN stress, too.

Table 3.1: Overlap analysis results
Comparison of differentially expressed genes (DEGs) in the *ΔE(z)* mutant with several transcriptomic studies describing stress responses in *C. merolae.* Representation factor < 1 indicates bigger gene overlap than expected by chance. (n.s.: not significant)

dataset	DEGs in dataset	genes in overlap	Representation factor	p-value*	Reference
ΔE(z)	**114**				this project
62°C heat shock	1466	29	0,9	n.s.	(Kobayashi *et al.*, 2014)
50°C heat shock	1387	32	1,0	n.s.	
Blue light	1204	35	1,3	n.s.	Tardu *et al.*, 2016
Red light	1116	27	1,1	n.s.	
Rapamycin	144	12	3,7	****	Imamura *et al.*, 2018
Limiting nitrogen	2982	86	1,3	****	S. Rader, pers. communication

*significance was assessed by Fisher´s exact test

A central regulator of nitrogen pathways in eukaryotes is the serine-threonine kinase Target of Rapamycin (TOR). Comparing the effects of an inhibited TOR action (rapamycin dataset, Imamura *et al.*, 2018), a significant overlap could be observed (Figure 3.3B). 8% of genes are commonly regulated, most of it towards upregulation in both data sets. The fact that both datasets consist of small DE gene numbers (rapamycin = 144, DEG *E(z)* =114), results in statistical significance in the overlap, although only 12 genes are commonly regulated. However, it cannot be excluded, that TOR and E(z) act in a similar pathway which might cross in the response to nitrogen regulation. In addition to the regulated H3K27me3 target genes mentioned earlier, these observation hints towards a potential function of E(z) in regulating genes, which are subject to changes due to an adaptation effort to demanding environmental changes of *C. merolae*, in this case nutrient deprivation.

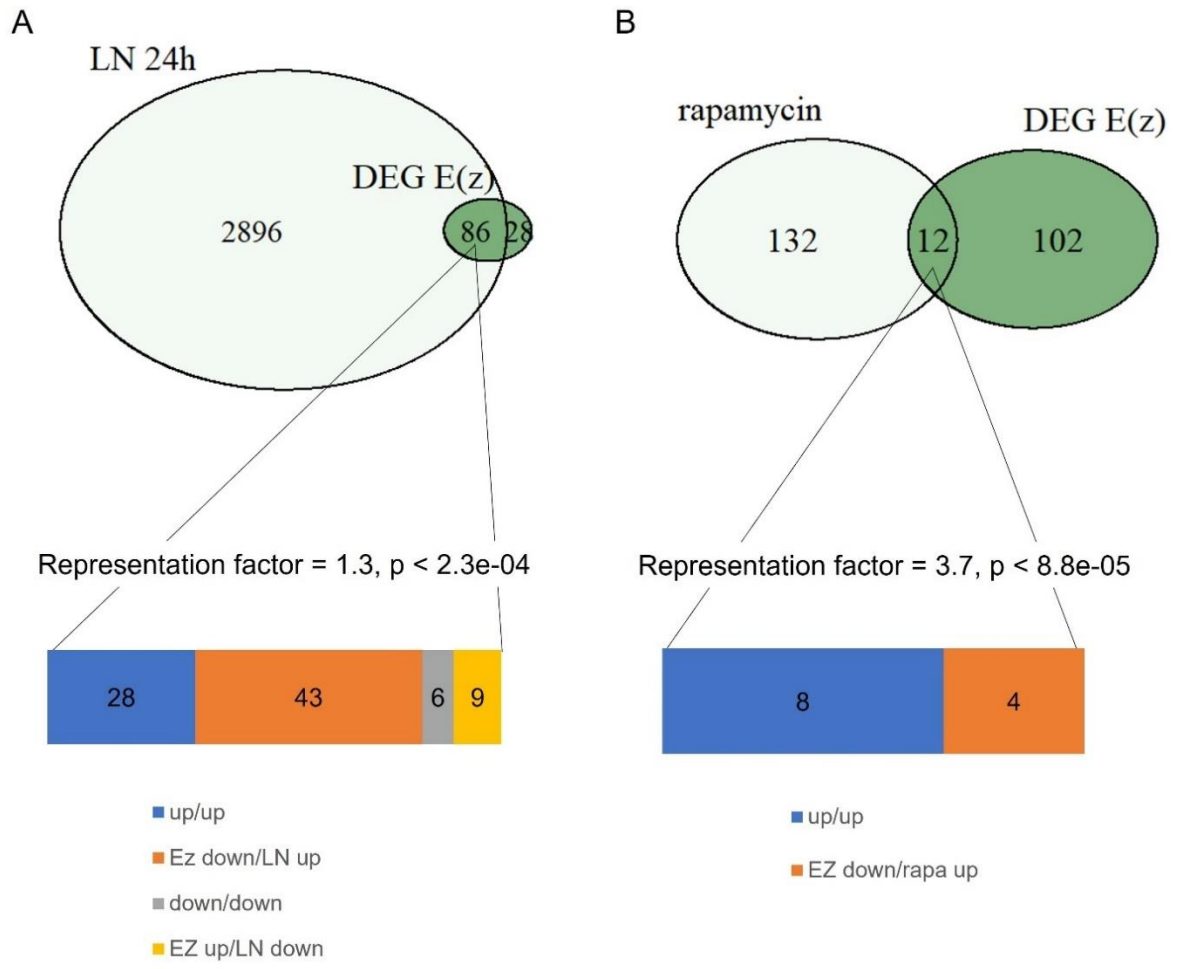


Figure 3.3: *ΔE(z)* regulated gene set overlaps with limiting nitrogen (LN) conditions and Target of rapamycin (TOR) regulation in *Cyanidioschyzon merolae.*
A Venn diagram overlap analysis of differentially expressed genes (DEGs) defined in *ΔE(z)* mutant under control conditions compared to gene expression changes observed upon a 24 hour nitrogen depletion (LN) stress (S. Rader, unpublished) shows similarly regulated gene sets. Significance of overlap was calculated by Fisher's exact test. Bar under Venn diagram depicts similarities of directions of gene expression changes of commonly regulated genes between *ΔE(z)* and LN. Majority of 86 genes regulated in both conditions are subject to downregulation in *ΔE(z)* and upregulation in LN. Nearly all remaining genes are upregulated in both, LN and *ΔE(z)*.
B Venn diagram analysis of DEGs in *ΔE(z)* compared to gene expression changes under rapamycin treatment representing loss of function TOR performance (data retrieved from Imamura *et al.*, 2018 applying significance thresholds same to *ΔE(z)* data set) identifies significant overlap between both gene sets. Significance of overlap was calculated by Fisher's exact test. Majority of commonly regulated genes are upregulated under both conditions, rapamycin treatment and *ΔE(z)* mutant. Bar under Venn diagram depicts similarities of directions of gene expression changes of commonly regulated genes between TOR and E(z) regulation.

It can be concluded, that H3K27me3 acts as a repressive mark of genes and predominantly repetitive elements in *C. merolae*. By silencing repetitive elements, *E(z)* ensures the transcriptional control of (sub)telomeres. Furthermore, H3K27me3 enriched genes, which are involved in stress responses are only partly regulated by *E(z).* Yet, the overlap of *ΔE(z)* regulated genes with transcriptomic responses to limited nitrogen availability provides an exemplary link between H3K27me3 and the response to nutrient starvation stress. Indeed,

the participation of H3K27me3 in stress answers, and moreover, stress memory mechanisms, is a known phenomenon in plants (Kleinmanns *et al.*, 2017; Kwon *et al.*, 2009; Yamaguchi & Ito, 2021). Consequently, it cannot be excluded that the relevance of *E(z)* regulation in *C. merolae* is hidden under control conditions, explaining the absence of any significance comparing the *ΔE(z)* DEG set with e.g. transcriptomic heat stress gene sets. Instead, the importance of *E(z)* in *C. merolae* could arise only when a stress situation is encountered. Hence, in order to further determine the function of *E(z)* in *C. merolae* it seemed essential to expose the *ΔE(z)* mutant line to stress and compare its response to WT behaviour. In doing so, the potential contribution of PRC2 mediated transcriptional regulation during stress responses can be tested.

3.2 Mitotically stable heat stress memory in *C. merolae* and the impact of *E(z)* on memory formation

The ability to memorize stress exposures was described for many organisms lacking a nervous system. Such memory can entail a substantial benefit of the respective individuum in case the stress ever returns compared to naïve organisms - a process which was termed stress *priming* (reviewed in Hilker *et al.*, 2016). These priming and memory processes include cellular, biochemical and/or transcriptional answers. Furthermore, epigenetic regulators substantially contribute to these mechanisms. In fact, heat stress priming is discussed as epigenetic phenomenon (Friedrich *et al.*, 2018) and the contribution of H3K27me3 in regulating the responses to and memory of stress occurrences is a known process in plants (reviewed in Kleinmanns & Schubert, 2014; Shen *et al.*, 2021). In order to test for a similar role of epigenetics in *C. merolae* stress responses it was a prerequisite to prove whether it is able to memorize stress at all. For that, a priming and memory stress scheme had to be established. Only then it can be deciphered, whether epigenetic regulators, such as E(z), contribute to this mechanism. As part of the *Cyanidiaceae*, *C. merolae* is considered to belong to the most heat-tolerant photosynthetic eukaryotes (Ciniglia *et al.*, 2004; Yoon *et al.*, 2006). Thriving in hot springs, temperature variations might represent the most prominent stressor this red alga has to face in its natural habitat. Furthermore, the heat shock response in *C. merolae* is well described (Kobayashi *et al.*, 2014). Based on this combination of (a) knowledge about heat memory as epigenetically driven phenomenon, (b) available data about heat shock responses in *C. merolae* as well as (c) temperature fluctuations as relevant natural challenge, heat stress was chosen to dissect whether epigenetics (via PRC2) are shaping stress memory mechanisms in *C. merolae*.

3.2.1 *C. merolae* can be primed for enhanced heat stress tolerance

To elucidate, whether epigenetic processes shape heat stress memory in *C. merolae* similar to what was shown in plant thermomemory (Friedrich *et al.*, 2018), it was necessary to establish a protocol of sequential, repetitive stress which proves that *C. merolae* indeed memorizes stress. The relevance of epigenetic regulators, such as E(z), can only be determined after this phenomenon could be described for *C. merolae*.

The establishment of a priming and memory scheme is comprehensive and time intense. In order to determine conditions which lead to heat stress memory in *C. merolae*, various temperature and duration combinations were tested for priming and triggering, in multiple culturing vessels. Overall, this lead to establishment of a highly reproducible heat stress memory scheme following the *priming – lag phase – triggering* sequence described in the introduction (Chapter 1.3).

In detail, three cultures are treated in parallel emanating from the 42°C control culture: Culture N (**N**aïve) is used as internal control in each experiment and is kept under control conditions (42°C) for the entire assay duration. Culture T (**T**riggered) is subjected to a severe heat shock of 60°C for 1h, termed *triggering stimulus*. The 3rd culture, PLT (**P**rimed-**L**ag-**T**riggered) faces - prior to the triggering - a *priming stimulus* of 57°C for 30 min. Priming and triggering stimulus are separated by a recovery phase of 2h, designated *lag phase* (Figure 3.4A).

To determine a benefit generated by the priming experience, a survival assay was established. Here, samples of each culture after the triggering timepoint (i.e. N(*)/T/PLT according to Figure 3.4A) were allowed to grow under control conditions on solidified culture medium (MA2) for one week. In detail, *C. merolae* WT cells subjected to 60°C for 1h are not able to grow on solidified MA2 medium containing survival plate (Figure 3.4B, T culture). However, if cells are facing a preparing priming stimulus preceding the severe triggering stress, they prosper on the survival plate (Figure 3.4B, PLT culture) seen in comparable growth of PLT and N cultures. This result indeed reports the ability of *C. merolae* to convert the priming stress experience into a benefit which is substantial for survival. Therewith, this result describes the ability of *C. merolae* to memorize heat stress for the first time.

Next, to quantitatively measure the priming benefit, cell fitness on physiological level was assessed using a pulse-amplitude modulation (PAM) fluorometer. Here, chlorophyll fluorescence measurement of photosystem II (PSII) in dark adapted states was determined during the whole time course of the heat stress memory assay (Figure 3.4C). PAM measurements are reflected in Fv/Fm ratios and mean values of n = 4 independent replicates ± SD are reported. As control (naïve) value Fv/Fm_N = 0.48 ± 0.03 was defined. Upon the

priming stress of 57°C for 30 min, Fv/Fm decreases insignificantly, whereas the triggering stimulus of 60°C for 60min leads to dramatic loss of PSII fitness (Figure 3.4C). After 2h of lag phase no substantial difference between the primed-before and non-primed culture could be reported suggesting similar PSII fitness states as starting point prior to triggering stress. This full recovery after priming is substantial in order to ascribe the differential triggering responses as consequence of (priming-) information storage and not as consequence of sustained cellular damage as result of priming. The dramatic decrease in the T culture Fv/Fm was alleviated, when the culture was primed before. Here, significantly higher Fv/Fm were measured in PLT cultures, suggesting an elevated PSII fitness upon priming (unpaired t-test, p = 0.01). Thus, the *priming benefit* observed on survival plates could be confirmed on physiological level in PAM measurements (Figure 3.4C). The sequence of priming – lag phase – triggering leading to successful memory of heat in *C. merolae* will be referred to as heat stress memory assay (HSMA) from hereon. The HSMA includes both, the survival plate assay and PAM measurement.

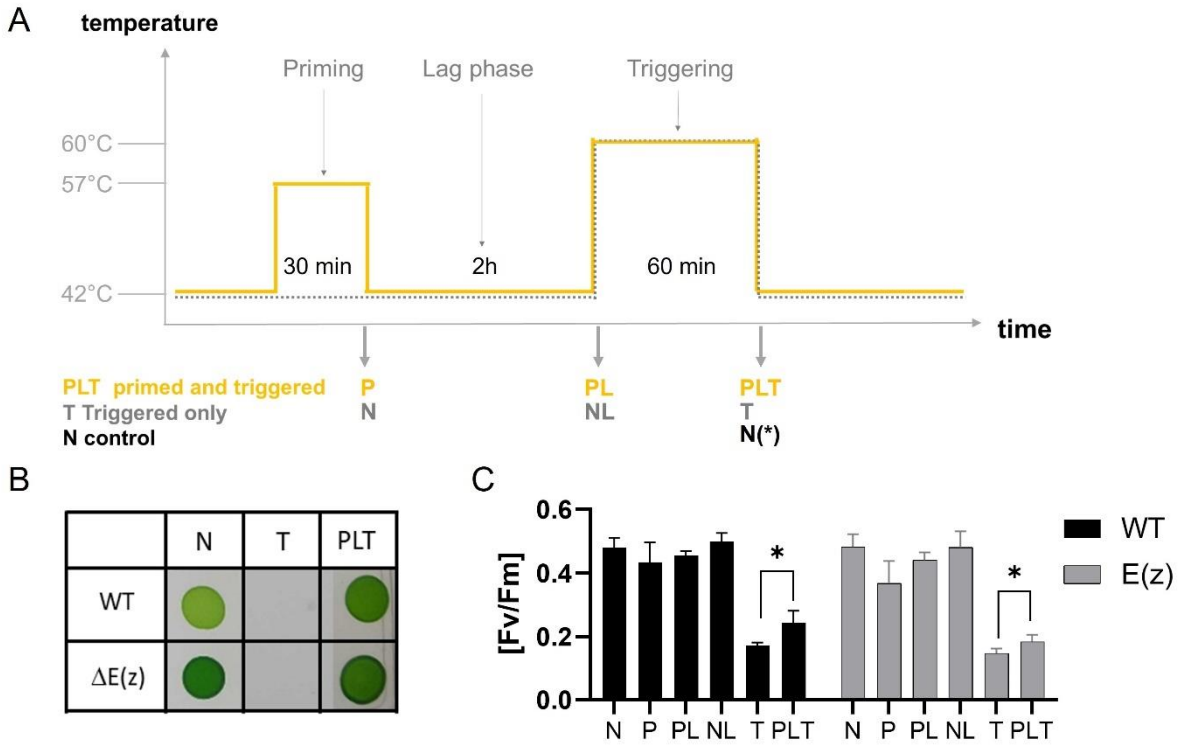


Figure 3.4: *Cyanidioschyzon merolae* is able to memorize heat stress.
A Heat stress memory assay (HSMA) scheme depicting duration and strength of priming and triggering stimuli as well as length of lag phase leading to memory formation in primed and triggered (PLT) cultures. Black line indicates conditions for naïve (N) control culture kept at 42°C for the course of experiment. Grey dotted line indicates conditions for triggered-only (T) culture, which does not encounter the priming stimulus. Orange line indicates conditions for the primed-and-triggered (PLT) culture, which experiences both, priming and triggering stimulus parted by the recovery (lag) phase. Arrows and colours indicate sampling time points and origin (culture) resulting in 6 conditions: N, P (primed), NL (after lag phase), PL (after priming and lag phase), T, PLT. (*) extra sampling point of naïve culture for Survival growth plate.

B Survival growth of wild type (WT) and Enhancer of zeste deletion *ΔE(z)* strain in naïve N, T and PLT conditions. Sampling was done after 60 min of 60°C triggering for all three cultures. 20 µl of corresponding culture were applied on solidified 2x Modified Allen´s medium (MA2) and growth was documented 7 days after experiment. Picture represents result of n = 10 biological replicates.
C Maximum photochemical efficiency of photosystem II depicted by variable over maximum fluorescence [Fv/Fm] measured using a Pulse-amplitude modulation (PAM) fluorometer in each condition: N, P, NL, PL, T, PLT in WT and *ΔE(z)* cells. Fv/Fm values were higher in PLT compared to T in both genotypes, indicating higher photosynthetic activity upon triggering when cultures were primed before. Each value is an average ± SD of n = 4 biological replicates. Asterisk indicate significant difference between PLT and T assessed by unpaired t-test, $p<0.05$. (SD: standard deviation)

In a following step, the persistence of memory was of interest to see whether *C. merolae* can maintain the information about priming heat for longer than 2h. Furthermore, by stretching the lag phase the maximally possible memory preservation should be determined.

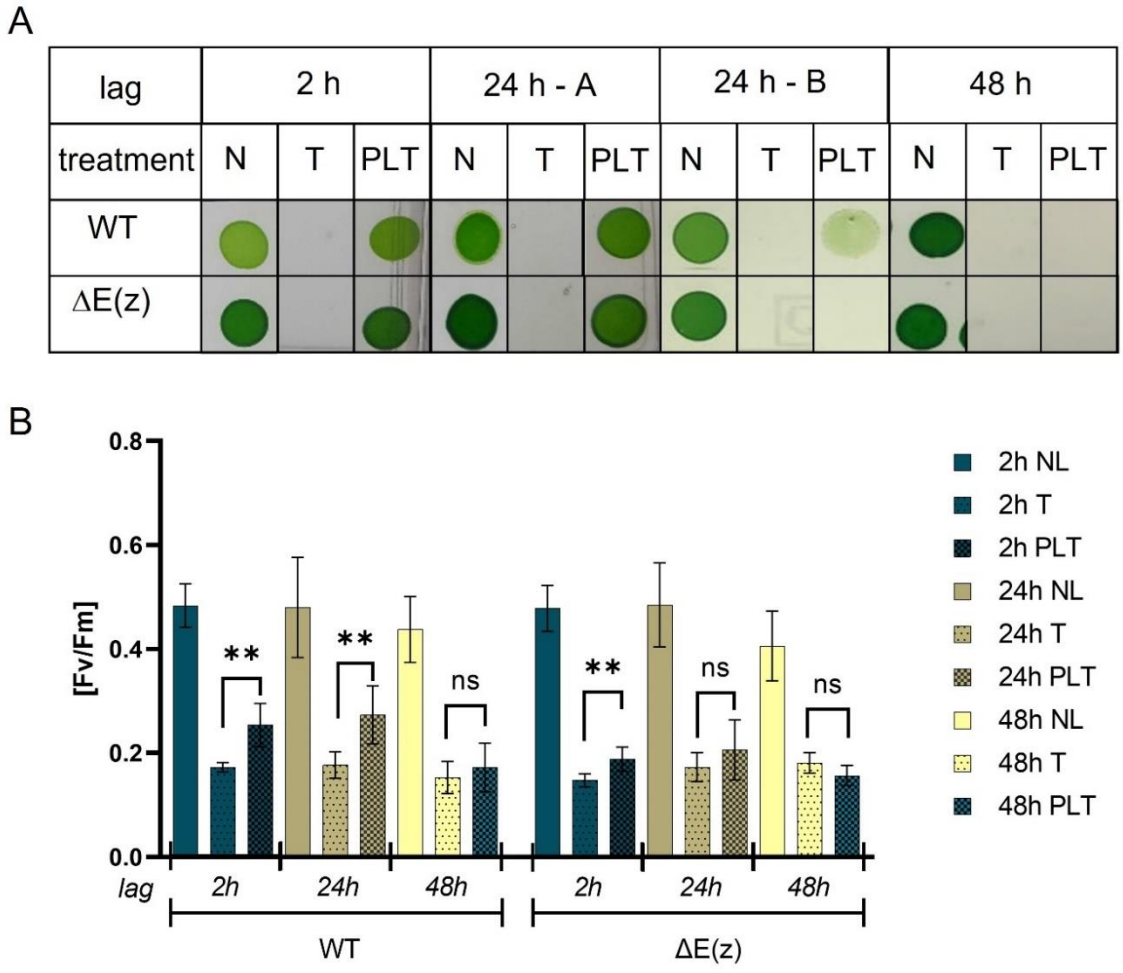


Figure 3.5: Prolongation of lag phase in heat stress memory assay results in memory decay after 48h in *Cyanidioschyzon merolae* and reveals attenuated memory in *ΔE(z)* cells.
A Survival growth assay upon application of different lag phase durations. 2h, 24h and 48h lag phase were tested using the scheme depicted in Figure 6A. 20 µl of naïve (NL), triggered-only (T) and primed-and-triggered (PLT) cells of wild type (WT) and *ΔE(z)* were allowed to grow at 42°C under constant light on solidified 2x Modified Allen´s medium (MA2) and growth was documented 7 days after experiment. Representative growth documentation. 2h: n=5, 24h - A: n= 5, 24h – B: n = 1, 48h: n=4

B Maximum quantum yield of photosystem II represented by Fv/Fm values acquired using a pulse-amplitude modulation (PAM) fluorometer of NL, T, PLT in WT and *ΔE(z)* cells upon application of different lag phases. Values are means of Fv/Fm ± SD of n = 5 (2h lag), n = 6 (24h lag) and n = 4 (48 h lag) biological replicates. Blue, brown and yellow bars represent values obtained for 2h, 24h, and 48h, respectively. Unpaired t-test (p = 0.05) assessed the differences between T and PLT pairs in different lag phase and genotype.

Functions attributed to *E(z)* in various organisms usually include cellular division, meaning that H3K27me3 is mitotically stable (Hugues *et al.*, 2020) and can confer information to daughter cells or even next generations. Thus, it seemed important to stretch the lag phase up to a length which includes cellular divisions. As the doubling time of *C. merolae* accounts for 18h in our lab, a lag phase of 24 h was tested next. Besides, application of a 24h memory phase allowed to administer the triggering stimulus at the same time of day like the priming stimulus.

When *C. merolae* cells were triggered 24h after the priming stimulus, a stable priming benefit was detected on survival plate (Figure 3.5A, 24h - A) as well as in the PAM measurement (Figure 3.5B, 24h). After further extending the lag phase to 48h, the priming benefit was abolished. Similar to T, no growth was detected on the survival plate for the PLT culture. Similarly, the significant difference in Fv/Fm ratios detected between PLT and T for the 2h and 24h lag phase was not measurable anymore. It is noteworthy, that although not significant, for the 48h lag phase there still is a tendency in higher PSII fitness in the primed-before culture accounting for Fv/Fm_{PLT} = 0.17 ± 0.05 compared to the non-primed culture Fv/Fm_{T} = 0.15 ± 0.03. After lag phase, cell numbers were determined by counting and OD measurements revealing that on average cells underwent approx. 3.5 mitotic divisions during 48h of lag phase (Table 3.2). Thus, these results suggest, that the thermomemory in *C. merolae* fades after 48h including approx. 3.5 mitotic divisions.

Table 3.2: Growth during 48h lag phase
Estimates based on optical density (OD) measurements at 750nm. 2h after triggering cells were diluted to an OD_{750nm} = 0.3. Estimation of cells newly developed during lag phase was done based on the OD:cell ratio counted for OD = 0.3. (WT: wild type, EZ: Δ*E(z)* strain, PLT; primed-and-triggered. N°: number, before: before lag phase)

replicate	*sample*	*OD_{750nm}*	*N° of cells*	*ratio to "before"*	*N° of new cells*	*% new cells*
1	before	0,3	1,77E+07			
	WT PLT	0,918	5,41E+07	3,1	3,64E+07	67%
	EZ PLT	0,746	4,40E+07	2,5	2,63E+07	60%
2	before	0,3	1,77E+07			
	WT PLT	1,034	6,09E+07	3,4	4,33E+07	71%
	EZ PLT	1,07	6,31E+07	3,6	4,54E+07	72%
3	before	0,3	1,77E+07			
	WT PLT	1,15	6,78E+07	3,8	5,01E+07	74%
	EZ PLT	1,082	6,38E+07	3,6	4,61E+07	72%
4	before	0,3	1,77E+07			
	WT PLT	1,37	8,07E+07	4,6	6,31E+07	78%
	EZ PLT	1,37	8,07E+07	4,6	6,31E+07	78%

3.2.2 Attenuated mitotic heat stress memory in *ΔE(z)* mutants

Initially, it was unclear whether *C. merolae* can memorize stress events at all. After determination of heat memory as shown in the previous Chapter for WT cells, now the contribution of epigenetic mechanisms to this process was analyzed. To that end, the *ΔE(z)* mutant was tested in the HSMA.

When a lag phase of 2h was administered, *ΔE(z)* cultures showed WT-like ability to profit from the formerly experienced heat shock. Accordingly, the PLT culture showed survival growth, whereas no growth could be documented for the un-primed T culture (Figure 3.4 B). On physiological level, however, some differences were noted for the PSII fitness (Figure 3.4C). As control (naïve) value Fv/Fm_N = 0.48 ± 0.04 was determined for Δ*E(z)* which resembles the WT control value. Interestingly, upon the priming stress of 57°C for 30 min Fv/Fm decreased to Fv/Fm_P = 0.36 ± 0.07 in the mutant, representing a stronger reaction of *ΔE(z)* PSII upon priming compared to what was seen for WT. To test, whether E(z) indeed shapes the priming response on the level of chlorophyll fluorescence measurement, Fv/Fm ratios of N and P samples in WT and mutant were compared (Figure 3.6A). Accordingly, whereas there was a visible, yet no statistical difference between N and P in WT, in *ΔE(z)* Fv/Fm_P was significantly lower compared to the control Fv/Fm_N. Furthermore, while Fv/Fm_N values did not vary between WT and mutant, Fv/Fm_P values differed significantly (Figure 3.6) which suggests that *ΔE(z)* PSII reacts more sensitive towards heat. Contrary to what was

seen for the 57°C heat stress, WT and *ΔE(z)* cells did not respond differently to the 60°C stress and exhibited comparable Fv/Fm values (Figure 3.6B). As the triggering stimulus is quite harsh and leads to no growth on survival plates in either genotype (Figure 3.4B, Figure 3.5A), the potential difference in the response to 60°C might only be detected when WT and *ΔE(z)* cells are subjected to shorter triggering stimuli. Hence, PAM measurements after 15 min of 60°C in both, WT and *ΔE(z)*, revealed a detectable, yet insignificant difference in Fv/Fm values of T between both genotypes (Figure 3.6C). The fact that *ΔE(z)* PSII seemed to react stronger to heat on chlorophyll fluorescence level compared to WT hints towards a lowered thermotolerance of *ΔE(z)* photosystem II. The significant loss of PSII fitness after priming was recovered after lag phase in *ΔE(z)*. No substantial difference between the primed-before culture (PL) and non-primed culture (NL) could be reported for *ΔE(z)* akin to WT behavior. The dramatic decrease seen for the T culture was attenuated, when the culture was primed before (PLT). This results in a significantly higher Fv/Fm in PLT compared to T cultures (unpaired t-test, p = 0.01, Figure 3.4C) and represents a priming benefit in *ΔE(z)*, too. Consequently, the memory formation described for WT is not altered in the *ΔE(z)* strain when a 2h lag phase was administered. Hence, albeit it seems to ensure a stable heat response on PSII fitness level, the involvement of *E(z)* in the formation of a heat stress memory could not be detected when a lag phase of 2h was applied.

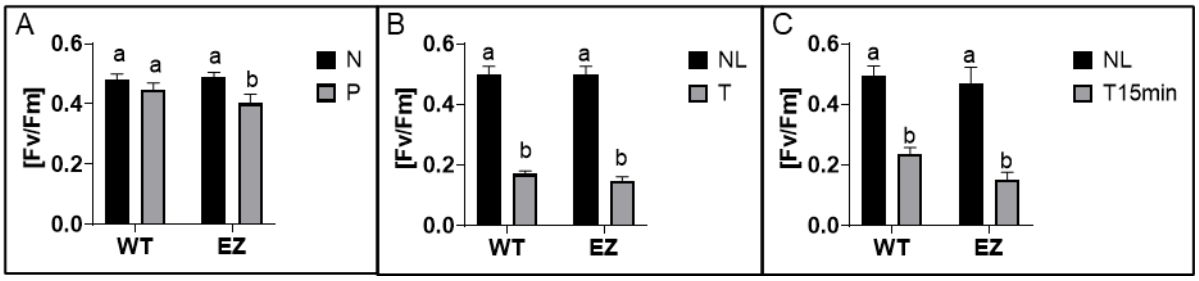


Figure 3.6: *ΔE(z)* cells exhibit elevated photosystem II sensitivity to 57°C heat stress.
Maximum quantum yield of photosystem II (PSII fitness) represented by Fv/Fm values acquired using a pulse-amplitude modulation (PAM) fluorometer of primed (P) and triggered-only (T) *C. merolae* wild type (WT) and PRC2 loss of function mutant *ΔE(z)* strain.
A PSII fitness of naïve (N) and cells subjected to 57°C 30 min heat (P) indicates increased heat sensitivity in *ΔE(z)* primed cells.
B Fv/Fm ratios of naïve after lag phase (NL) and cells subjected to 60°C 60 min (T) indicate similar responses of WT and *ΔE(z)* cells.
C PSII fitness of NL cells and cells subjected to 60°C for 15 min (T15min) in WT and *ΔE(z)*. Mutant cells show insignificantly lower Fv/Fm values when heat shocked in 60°C for 15 min. Differences between the given conditions and genotypes are assessed individually for each graph by 2way ANOVA followed by Šidák´s test of n=4 (A), n= 4 (B), n= 3 (C) replicates. (SD: standard deviation)

The storage of priming information for 2h leading to enhanced performance upon triggering could be described as independent of PRC2 action. However, the possibility remained, that E(z) contributes to longer-term storage of the priming information. To that end, *ΔE(z)* cells

were subjected to the HSMA with a 24h lag phase separating priming from triggering stimulus. As a result, by stretching the lag phase to 24h, the *ΔE(z)* mutant line differed in the HSMA answer. Although the majority of survival plate replicates resulted in growth for PLT whereas there was none for T, in one case (1/6 replicates) the *ΔE(z)* PLT culture would not grow suggesting a loss of priming ability in *ΔE(z)* cells when a 24h lag phase was administered (Figure 3.5A, 24 h - B). It has to be noted, that in that case the WT PLT growth is weakened, too, which assumes decreased cell numbers loaded onto the survival plate. As the control cultures (N) are comparable to other replicates for both genotypes, a possible experimental explanation might be elevated water bath temperature during triggering. Consequently, higher triggering temperature would increase the selective pressure for primed cells and lead to lower cell concentrations in the samples. As it turned out in the course of establishing the survival assay growth, the applied samples must exceed certain cell concentrations (determined by OD_{750nm}) in order to achieve detectable growth on the plates after 7 days (data not shown). It is tempting to speculate, that the majority of *ΔE(z)* cells were primed resulting in growth of the survival plate (Figure 7 24h – A) whereas a minor fraction which is better represented when cell number is lowered (potentially daughter cells of the primed parents grown during lag phase) is not primed. Subseqeuntly, this is reflected in the absence of *ΔE(z)*_PLT growth (Figure 7, 24h-B). The notion of memory attenuation in *ΔE(z)* upon a lag phase of 24h detected on survival plate is undermined in PAM measurements, where the significant difference in Fv/Fm between PLT and T observed for WT is vanished in the *ΔE(z)* mutant (Figure 3.5B, 24h lag). Whereas the WT still benefits from the formerly experienced priming event on the level of PSII fitness, the *ΔE(z)* PLT culture behaves more alike to non-primed T cultures suggesting an inhibited memory maintenance in *ΔE(z)*. Once the lag phase is stretched to 48h, however, no memory effect can be observed anymore in *ΔE(z)*, neither on survival plate (Figure 3.5A), nor in PAM measurements (Figure 3.5B). Actually, the response in *ΔE(z)* is even converted in tendency exhibiting higher T than PLT values (Figure 3.5B).

In summary, the HSMA of WT and *ΔE(z)* suggest that *C. merolae* is able to memorize heat stress for 2h independent of E(z) activity. This memory is vanished, once the cells divided approx. 3.5 times in both, WT and *ΔE(z)* cells. However, upon *E(z)* mutation heat priming memory storage for 24h is attenuated which is seen in impeded survival and no observable difference in PSII fitness between primed and non-primed cultures during the application of a 24h lag phase. Together, these results imply that *E(z)* plays a role in the maintenance of memory (24h lag), rather than establishing it (2h lag). In sum the idea of *E(z)* as contributing factor to thermomemory regulation in *C. merolae* is strengthened.

3.3 Heat stress memory transcriptome and its reliance on PRC2 action

After showing that *C. merolae* harbors the ability to memorize heat stress and that this answer is partially dependent on PRC2 activity (Chapter 3.2) it was of interest, which transcriptional changes accompany this process. It is established, that in addition to physiological memory mechanisms, information about a given stress encounter can be stored via customization of chromatin modifications which eventually can lead to differential transcriptional responses of single genes, termed transcriptional trainability or memory. To uncover potential memory mechanisms on transcriptional level (i.e. trainable genes) during heat stress memory in *C. merolae* and in the same time determine the impact of the epigenetic repressor H3K27me3 on this process, transcriptomic analyses of WT and ∆*E(z)* were conducted.

Thus, in order to dissect the physiological heat stress memory of *C. merolae* described in Chapter 3.2 and in the same time elucidate the impact of epigenetic control via H3K27me3 in this process, a comprehensive whole transcriptome analysis was performed using RNA sequencing (RNAseq). Since the relevance of the H3K27me3 histone methyltransferase E(z) was strongest when a 24h lag phase was applied, mRNA of N, P, NL, PL, T, PLT (according to the experimental scheme in Figure 3.7) for WT and *ΔE(z)* cells in three biological replicates collected during the 24h lag phase HSMA was sent for sequencing, 36 samples in total. It has to be pointed out, that differently to the HSMA scheme used in Chapter 3.2, the triggering sampling time point (concerning T and PLT condition) was preponed to 15 min at 60°C, rather than 60 min. This decision was considered crucial to focus on memory responses and in the same time minimize transcriptomic reactions initiated by cell death. Accordingly, as a reminder, no growing cells could be detected for the T culture on survival plate after 60 min at 60°C. Sequencing resulted in 9 - 20 mio paired-end reads of 150 bp length per sample (Table S1). To ensure reproducibility of results as well as comparability to published data, only the fully reviewed published annotation file (old anno) served as reference for differential expression analysis in the heat (memory) response of *C. merolae*.

3.3.1 Transcriptional answer to heat depends on stimulus strength

Differential expression analysis was performed for the following pairwise comparisons in order to describe the amount of differentially expressed genes (DEG) as a result of advanced age, heat stress and recurring heat (Figure 3.7). NLvsN describes transcriptional changes due to age or, more precisely, OD_{750nm} increase. To get insights into heat shock responses

the comparisons PvsN and TvsNL were conducted, defining DEGs due to a heat shock of 57°C for 30 min and 60°C for 15 min, respectively. In PLTvsPL all DEGs are comprised which are responding to a 60°C triggering stimulus in a formerly primed culture. Finally, the comparison PLTvsT yields genes reacting differently to the triggering stimulus solely due to the differential past i.e. the presence (PLT) or absence (T) of the priming event. All comparisons were made in WT as well as in *ΔE(z)* cells.

The impact of increasing cell density is moderate, accounting for 274 DEGs in the comparison NvsNL (Figure 3.7, WT NvsNL). Progressing age in *C. merolae* seems to correlate with the upregulation of genes, as 78% i.e. 207 genes are (hyper)induced. Transcriptomic changes during the diurnal cell cycle are well addressed in Fujiwara *et al.*, 2020 already and not further scrutinized. In this project the increasing age control (NL) was crucial to discriminate between gene expression changes due to age and (recurring) heat responses.

The experience of the priming stimulus, i.e. 57°C for 30 min lead to massive reprogramming of the transcriptome involving expression changes of 1973 DEGs (Figure 3.7, WT PvsN). This number accounts for 40% of the whole genome (5009 protein coding genes) and reminds on the reaction to limited nitrogen condition and light stress in *C. merolae* (described in Chapter 3.1.3). The majority of DEGs are subject to upregulation (1173, i.e. 59%) suggesting a prevalent need of elevated transcription upon a 57°C heat stress. The same is true for the triggering stimulus, i.e. 60°C for 15 min, where 58% of the DEGs are upregulated (Figure 3.7, WT TvsNL). Although the proportion of gene expression changes are equally distributed upon priming and triggering stress, the total DEG number is dramatically lower upon triggering comprising only 312 DEGs (Figure 3.7, WT TvsNL). This accounts for only 1/5 of the DEG number upon priming and in terms of total DEG numbers is comparable to the age affect (NLvsN). This different impact on DEG number found for the two stresses might result from the different stress durations (priming 30min, triggering 15 min). However, it is known, that *C. merolae* is able to sense heat in a way, that the transcriptional onset of heat responsive genes is controlled by the absolute temperature in °C. The impact of temperature difference plays a minor role (Kobayashi *et al.*, 2014). Thus, it is unclear if the difference in DEG number found in PvsN compared to TvsNL is due to the total temperature (57°C vs 60°C) or duration of stress encounter (30 min vs 15 min). In sum, the priming stimulus entails 5-fold higher DEG numbers than the sole triggering stress, while the distribution of changes (up- or downregulation) remain similar in both stresses.

Next, the impact of the triggering stimulus was examined on a culture which was subjected to a priming stimulus 24h before (PLT). The degree of changes in PLT cultures is slightly higher than the DEG number detected for T (Figure 3.7, WT PLTvsPL) meaning that a triggering stress leads to an enhanced transcriptional answer in primed-before cultures compared to

triggered-only cultures. Thus, the priming event impacts the later triggering response on gene transcription level. It entails higher numbers of genes reacting to the stress and might constitute one reason for the priming benefit observed in physiological memory in *C. merolae* (Chapter 3.2). Further, the direction of gene expression changes slightly shifted towards upregulation. Here, 62% (222) of the DEGs in PLTvsPL are upregulated. Therefore, the priming benefit of formerly primed cultures (PLT) can be linked to elevated gene expression.
Subsequently, the relevance of PRC2 transcriptional regulation in single and repetitive heat encounter was investigated. All comparisons were made for the *ΔE(z)* mutant, too. Upon H3K27me3 absence in *C. merolae,* the number of genes changing their expression due to age (Figure 3.7, EZ NLvsN) was slightly increased compared to WT accounting for 330 and 274 DEGs for *ΔE(z)* and WT, respectively. When heat stress occurred, however, this situation reverted: *ΔE(z)* was less responsive and did not reach WT DEG numbers after priming as well as after triggering (Figure 3.7, EZ PvsN and TvsNL). Especially the triggering reaction differed. Accordingly, only 50% of the WT changes were detected for the mutant (EZ:WT = 157/312 DEG).

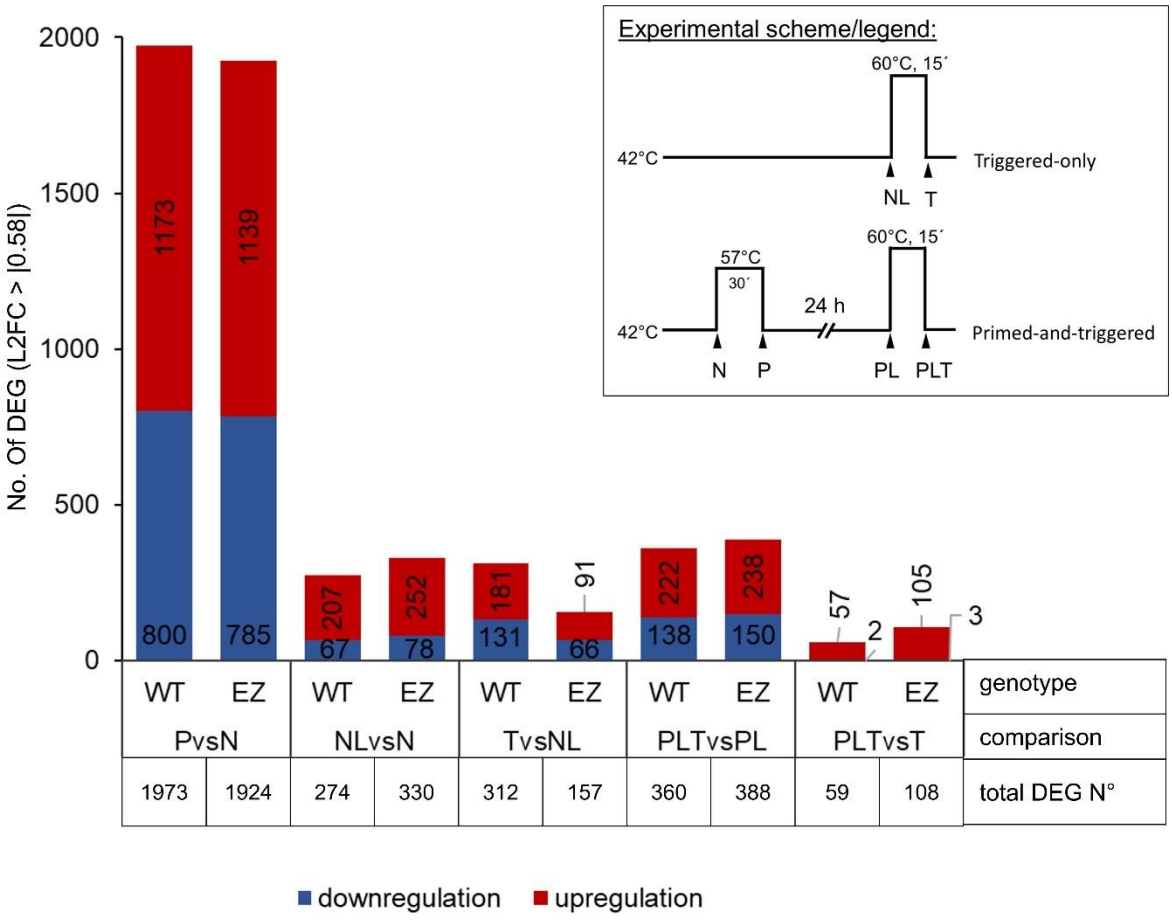


Figure 3.7: Global transcriptional response of *Cyanidioschyzon merolae* in heat stress memory assay conditions

Total number of differentially expressed genes (DEGs) upon priming stimulus (PvsN), after lag phase (NLvsN), upon triggering stimulus in triggered-only (TvsNL), upon triggering stimulus in primed-and-triggered (PLTvsPL) as well as DEGs between triggered-only and primed-and-triggered cells upon triggering (PLTvsT) are shown, in wild type (WT) and *ΔE(z)* (EZ) cells. Note that DEG number is higher in *ΔE(z)* mutant upon increasing age (NvsNL), but decreased upon heat shock (PvsN and TvsNL). DEGs were identified by DESeq2 using the parameters log2-fold change (L2FC) > |0.58| and padj < 0.05. Numbers of upregulated (red) and downregulated (blue) gene fractions are depicted in each bar, vertically and horizontally oriented numbers are numbers of downregulated and upregulated DEGs, respectively. (No.: number, padj: adjusted p-value)

This bears the possibility that genes important in the response to heat might lack (sufficient) induction in *ΔE(z)*. The observed incomplete transcriptomic reaction of *ΔE(z)* to heat reminds of the observation in Chapter 3.2.1, where *ΔE(z)* cells responded more sensitively to the priming stimulus in PAM measurements (Figure 3.6). Necessary genes in the heat response may be prevented from expression which could lead to the worse thermotolerance in the mutant. Traditionally, heat shock factors mediate the expression of heat responsive genes. The genome of *C. merolae* possesses three annotated heat shock factors: *CML277C*, *CMO187C*, and *CMT597C*. In addition, the protein encoded by *CMR316C* has a heat shock factor domain. To decipher whether the attenuated heat response of ∆*E(z)* might be due to insufficient HSF regulation, the expression of all four HSF genes during the HSMA was scrutinized and compared with H3K27me3 occupancy around the four loci (Figure S2). For *CML277C* and *CMT597C* no altered expressions were observed in the mutant compared to the WT (Figure S2 A,C) and accordingly, these regions are barely decorated with H3K27me3. Contrarily, *CMO187C* and *CMR316C* exhibited aberrant expression patterns in ∆*E(z)* compared to WT (Figure S2 B,D). *CMO187C*, in WT downregulated by priming, is upregulated in the mutant when priming occurs reflecting a misexpression of a putative HSF when heat occurs. Interestingly, the neighbouring gene *CMO188X* encodes an repetitive element which exhibits H3K27me3 levels. The lowered H3K27me3 levels in the ∆*E(z)* mutant might entail upregulation of this repetitive element which might affect the transcription of the neighbouring gene *CMO187C*, too. Interestingly, *CMR316C* was the HSF showing the strongest aberration in ∆*E(z)* as its expression rates decreased upon H3K27me3 absence. Comparing *CMR316C* expression upon priming and upon triggering, respectively, the mutant expression rates were 1.8-fold (padj = $1.5 * 10^{-5}$) and 1.6-fold (padj = $2 * 10^{-13}$) lowered compared to WT. It is noteworthy, that there is in fact a small number (n = 27) of genes, which require the heat stimulus in order to be misexpressed in the mutant, meaning they show differential expression between WT and ∆*E(z),* only in the context of heat, i.e. the direct comparison PvsP and TvsT, respectively (data not shown). In fact, *CMR316C* was among those. Thus, it can be concluded that a potentially relevant HSF in the heat shock response is not sufficiently expressed in the ∆*E(z)* mutant which coincides with its worse general heat reaction.

The intergenic region directly downstream of *CMR316C* is enriched with a pronounced H3K27me3 peak under WT control conditions (Figure S2 D). Albeit the detected downregulation of *CMR316C* in the mutant is counterintuitive - as this H3K27me3 peak is vanished in *ΔE(z)* - this result indeed suggests H3K27me3 regulation of *CMR316C* expression. The precise regulatory mechanism, however, remains obscure. The conjecture, that this gene might contribute to the failure of gene expression upon stress in the *ΔE(z)* mutant has to be tested in future studies. Together, *CMR316C* is defined as promising candidate in regulating thermotolerance of *C. merolae*.

When *ΔE(z)* cells were primed before triggering they reach WT DEG numbers upon triggering. In fact, their reaction is slightly stronger on gene expression level (Figure 3.7, EZ:WT = 388/360 DEG in PLTvsPL) which did not reflect the memory attenuation of *ΔE(z)* in the HSMA (Chapter 3.2).

In conclusion, sole priming and triggering reactions on transcriptomic level strongly varied in *C. merolae* exhibiting 5-fold higher DEG numbers in priming compared to triggering stimulus. Since both, duration as well as strength of stress were different between priming and triggering, it is unclear to date what exactly caused this difference. Further, albeit the scope of changes was higher in the mutant upon age (NLvsN), *ΔE(z)* had difficulties to reach WT responses when stress occurred which coincided with decreased expression levels of an HSF domain containing protein which might support expression upon heat stress. Notably, the worse reaction to the sole triggering stress was annihilated when *ΔE(z)* cultures were primed before, i.e. PLT cultures. As this observed WT like behaviour of *ΔE(z)* PLTvsPL does not reflect the attenuated memory formation of *ΔE(z)* in the HSMA, a more detailed analysis of T versus PLT condition was performed to identify so called trainable genes – i.e. genes, showing altered transcriptional responses to the 60°C triggering stress only because they were primed before. Parallel analysis of trainable gene behaviour in *ΔE(z)* might gain first insights into epigenetic regulation of those genes. Subsequently, the comparison of PLTvsT transcriptome was analysed next.

3.3.2 Identification, characterization, and application of heat stress memory genes in *C. merolae*

Genes reacting differently in primed (PLT) and non-primed (T) cultures when they are subjected to triggering stress may reveal insights into mechanisms underneath the observed heat stress memory in *C. merolae*. These genes can be identified comparing the PLT and T transcriptomes. The PLTvsT comparison includes genes which react differently to the 60°C stress only because they were facing a 30 min 57°C stress 24h before the triggering stress.

As a result, DESeq2 analysis using the contrast PLTvsT in *C. merolae* WT cells yielded 59 genes which differ at least 1.5 x in expression strength (abs L2FC > |0.58|, padj < 0.05, Figure 3.8A, Table S2). When genes are showing a revised transcriptional response during a second stress episode they are referred to as trainable or memory genes. This kind of genes are known for several abiotic stress responses in plants (Ding *et al.*, 2012; reviewed in Xie *et al.*, 2021). Accordingly, the 59 DEGs identified to respond differently when facing stress for the first time (T) compared to a second stress encounter response (PLT) are designated *C. merolae* heat trainable or memory genes within this project. In the following, the dissection of this memory gene set, including its constitution, dynamics during the HSMA, and genomic origin served the intention to draw conclusions about molecular mechanisms underlying thermomemory in *C. merolae*.

3.3.2.1 Memory genes follow distinct expression patterns during the HSMA

By differential expression analysis of PLT compared to T conditions of the HSMA, 59 genes were identified as memory genes in *C. merolae* WT cells. Examining the results of differential expression analysis revealed nearly exclusively positive log2 fold-changes (L2FC) thus describing the vast majority of memory genes as higher expressed in PLT than in T. Yet, two genes form an exception from that behaviour and are downregulated in PLT (Figure 3.8A). These two are both encoding for retroelements and scrutinized in Chapter 3.3.2.3.

Analysing the summarized expression behaviour of all 59 memory genes during the entire HSMA revealed a distinct memory expression pattern which the memory genes have in common (Figure 3.8B). In detail, the expression of all 59 memory genes in summary was highly elevated upon priming (P). Indeed, comparing this observation with the list of DEGs upon priming, 56 of the 59 memory genes show significant positive fold changes in PvsN (Table S5). After lag phase, the strong upregulation seen in P in sum recovered to control amounts reflected in similar expression levels in PL and NL (Figure 3.8B). Likewise, no gene was detected differentially expressed between PL and NL (PLvsNL, Table S5)). In response to the triggering stimulus memory genes in both, primed-before (PLT) and non-primed (T) cultures showed a general upregulation (Figure 3.8B). While the majority of memory genes was significantly induced in PLTvsPL, only 8 memory genes are likewise significantly upregulated in TvsNL (Figure 3.8C significance indicated by asterisk). Together, the upregulation in PLT exceeded the T regulation across memory genes (Figure 3.8B) and therewith is indicative for a need of gene hyperinduction in order to generate a benefit from priming.

After a common expression pattern of memory genes could be shown which hinted towards a general gene hyperinduction in primed and triggered cultures, it was asked whether within

those 59 memory genes expression patterns differ and if these patterns can be used to draw conclusions about the dynamics of memory gene expression. As described in the introduction (Chapter 1.3), memory mechanisms of several stresses demand higher induction rates of [+/+] trainable genes upon a second stress and/or a sustained induction during lag phase, both leading to considerable higher transcription rates of corresponding genes when they have been primed before. Accordingly, it was questioned whether the 59 *C. merolae* thermomemory trainable genes follow similar patterns. Thus, individual memory gene expression was examined and visualized in a heatmap depicting L2FC values for the T(vsNL) and PLT(vsPL) conditions. k-means clustering allowed the identification of five main gene clusters (Figure 3.8C). Following, the summarized expression profiles of each individual cluster was examined (Figure 3.9A) showing slight but important differences between clusters in the memory gene induction throughout the full factorial experiment. As a result, cluster 1 comprised eight memory genes, reflecting the common memory gene expression pattern during the HSMA: These 8 genes were strongly upregulated during priming, returned to basal levels after lag phase, and showed significantly higher levels in PLT cultures compared with T cultures (Figure 3.9A).

Overall these genes showed basal transcription rates of around 100 reads only. Moreover they - with the exception of *trxA* - showed a significant hyperinduction in T (Figure 3.8C). Since upregulated in all heat stress contexts (P, T, PLT), cluster 1 genes probably comprise [+/+] trainable genes essential in the heat stress response, which fully recover after lag phase. Accordingly, *CMJ101C* is found among cluster 1 genes which encodes a small heat shock protein and is well described in the *C. merolae* heat shock response (Kobayashi *et al.*, 2014).

Cluster 2 genes follow an expression pattern which is similar to cluster 1 genes. The difference, however, was found in weaker induction in T cultures. Accordingly, although significantly induced by priming (PvsN, Table S5) and recovered after lag phase, all cluster 2 genes missed significant upregulation in TvsNL (Table S5) with the exception of *infB*. Basal transcription rates are similar to cluster 1 genes situated around 100 – 1000 reads. As described by Hilker *et al.* 2016, via the organism´s experience of a previous priming stimulus, the velocity, onset, sensitivity and strength of the organismic response might be shaped. As a consequence, for trainable gene expression, this might entail a faster, earlier, more sensitive or stronger induction in PLT when compared to T. Genes in cluster 2 are not significantly upregulated in T. While cluster 1 genes show stronger responses in PLT, cluster 2 genes might comprise genes which react earlier to stress reoccurrence when primed. To test this hypothesis, a time series during triggering should reveal whether cluster 2 genes reach the induction level of PLT in cultures which were subject to extended triggering-only, e.g. 60°C for 30 min.

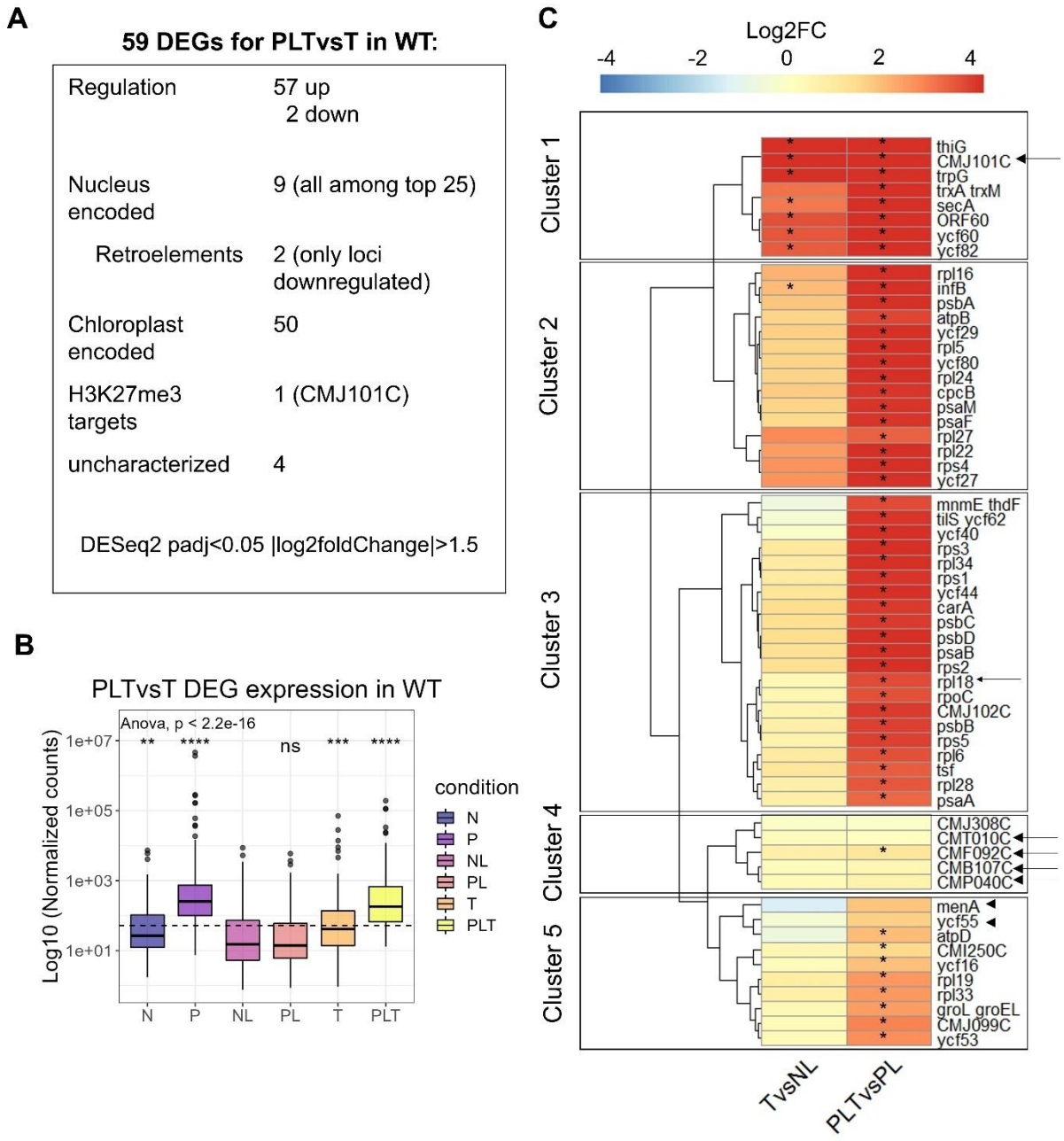


Figure 3.8: Identification and initial characterization of heat stress memory genes in *Cyanidioschyzon merolae*

59 genes were identified as differentially expressed in primed-and-triggered (PLT) compared to triggered-only (T) wild type (WT) cells

A Overview of 59 memory gene constitution. Differential expression was assessed using DESeq2 with parameter indicated.

B Box plots reveal general expression pattern of memory genes across conditions in WT and ΔE(z) mutant strain. Y-axis depicts relative expression of 59 PLTvsT DEGs given as log10 transformed normalized counts. Note that the average expression is highly induced in priming (P), reverts back to control levels (PL compared to NL) and is hyperinduced in PLT compared to T. Significance was tested using 2-way ANOVA followed by comparison of all conditions to the NL condition as reference. Asterisks denote significant difference between conditions of summarized expression (59 memory genes). Median of normalized counts across memory genes accounts for approx. 200 counts.

C Heatmap depicting expression changes of 59 memory genes in triggered-only (TvsNL) and primed-and-triggered (PLTvsPL) WT cells upon triggering. k-means clustering defined 5 major clusters in expression behaviour. Heatmap color coding shows log2 fold changes (Log2FC) of single

genes in each condition (red = high, blue = low). Asterisks indicate significant regulation in corresponding comparison (padj < 0.05, Log2FC > |1.5|). Genes marked by black arrows were not identified as DEGs in the *ΔE(z)* mutant strain in the comparison PLTvsT and subsequently are designated as genes losing trainability in the mutant. (n.s.: not significant)

Cluster 3 represents the largest cluster, comprising 21 genes. Interestingly, all of them originated from the chloroplast genome. These genes were all upregulated by priming, but barely to not induced in T. In PLT, however, they showed all significant and strong upregulation. Here, too, gene transcription rates returned to control levels after lag phase. In summary, while cluster 3 gene behaviour is comparable to the pattern found in cluster 1 and 2 genes, the important difference here was found in the lack of T induction. If those genes, which represent [-/+] trainable genes, are induced in later triggering time points remains subject to future investigations.

Within the 4th cluster were genes, which hardly exhibited gene expression changes in response to heat stress. The sole triggering stress in either culture (TvsNL or PLTvsPL) did not lead to significant changes in expression. Only one gene, *CMF092C*, surpassed the significance thresholds of upregulation in the PLT culture (Figure 3.8C, PLTvsPL specified by asterisk). Albeit the summarized induction upon priming was lower than in clusters 1-3 (PvsN, Table S5) cluster 4 genes in majority show significant induction in PvsN (Table S5), PvsN_WT). Only the gene *CMB107C* constitutes an exception and is not induced by priming, nor by triggering (Table S5). Still, all five genes situated in this cluster are differentially expressed between PLT and T directly. It is worth noting, that the priming induced upregulation of these genes did not completely revert to control levels after lag phase showing somewhat higher expression levels in PL compared to NL. Taken together, cluster 4 genes could represent a separate group of heat trainable genes in *C. merolae* which are not necessarily responsive to high temperature (i.e. 60°C) stress.

Finally, the 10 genes in cluster 5 were upregulated upon priming, returned to control levels after lag phase, and were in majority significantly induced in PLTvsPL (Figure 3.8C, Figure 3.9A). However, cluster 5 genes remained unresponsive to triggering, i.e. in T. Similar to cluster 3 genes, those genes exhibit transcriptional rates between 100 and 1000 reads and additionally, the [-/+] trainable gene behaviour. The strength of induction in PLT, still, is considerable lower, undermined by the fact that two genes in this cluster did not react significantly to the triggering stimulus in PLT. It is tempting to speculate, that cluster 5 genes differ to cluster 2 and 3 memory genes in their extent of memory specificity.

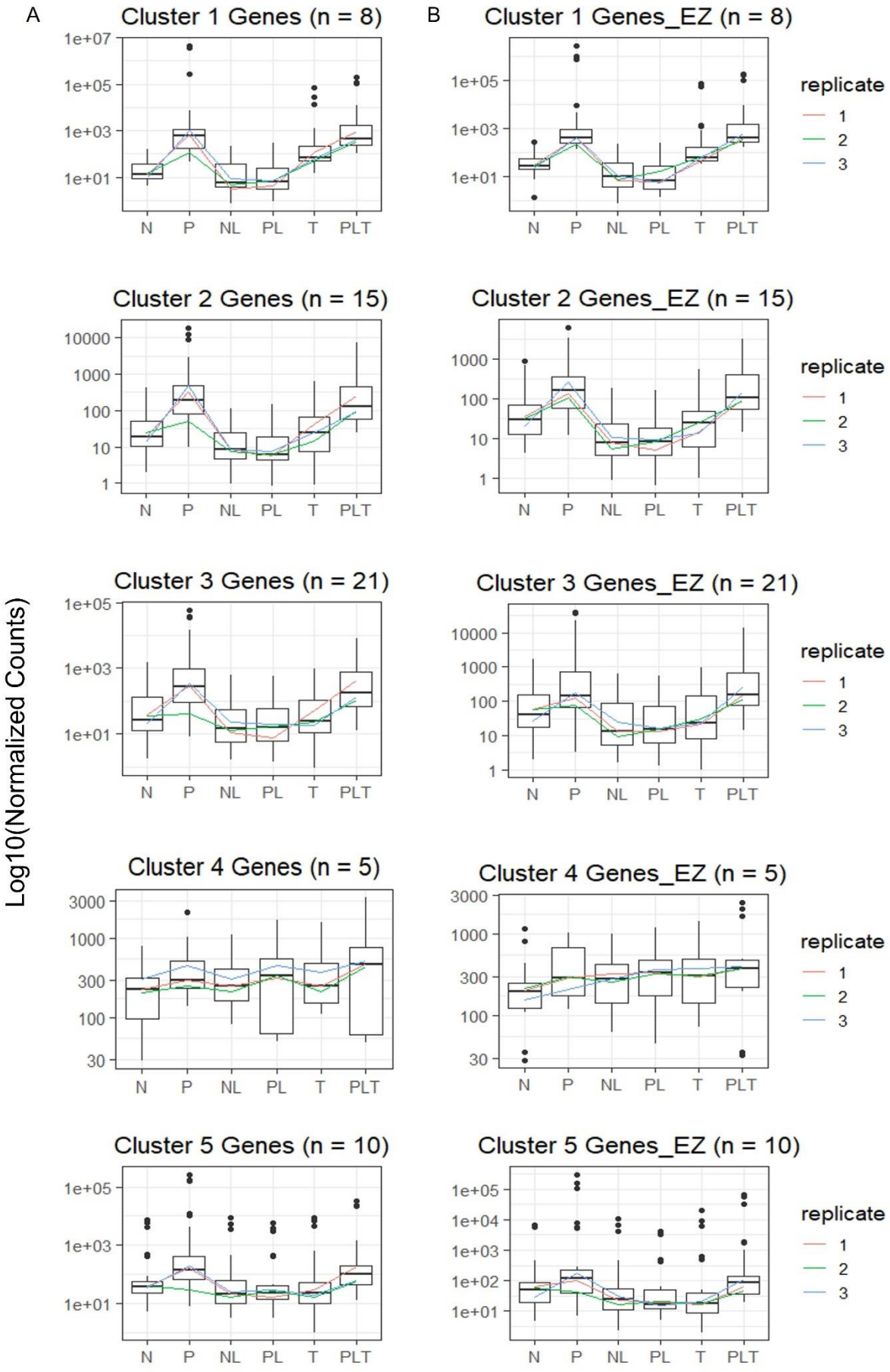


Figure 3.9: Expression profiles of memory genes in clusters 1 – 5

Memory genes were clustered in R by k-means clustering in Figure 9 for wild type (**A**) and *ΔE(z)* (**B**). The y-axis represents the log10 transformed standardized normalized counts of genes, the x-axis represents the different conditions: naïve (N), primed (P), naïve after lag phase (NL), primed after lag phas (PL), triggered-only (T), and primed-and-triggered (PLT). Plotted is the average temporal expression profile in each sequenced replicate (1-3) and boxplots describe the median, upper and lower quartiles. Outliers are marked with black dots.

Overall, a set of 59 genes was identified in transcriptomic analyses of the HSMA describing the event of heat stress memory formation in *C. merolae*. Within memory genes, two distinct groups could be identified. While cluster 4 genes seem to be unresponsive to heat but still lastingly regulated by the priming event, genes in clusters 1, 2, 3, and 5 show a stronger reinduction upon triggering when primed before, which might be the result of a faster response upon an already known, recurring stress stimulus. As these memory genes can be used as read out of successful memory formation, this set will aid in the identification of major regulators of *C. merolae* thermomemory.

3.3.2.2 E(z) acts on cluster 4 heat stress memory genes which are only moderately heat responsive

Dissection of the transcriptome underlying heat memory in *C. merolae* revealed a set of 59 memory genes. The majority of those showed stronger reinduction upon repetitive heat in comparison to single heat encounters. Yet, a smaller group of these memory genes (situated in cluster 4) showed only moderate responses to heat *per se*, however their expression was altered during recurring heat, when they have experienced the priming stimulus before.

Next is was asked, whether PRC2 mediated transcriptional regulation contributes to proper expression of this set of WT trainable genes. Therefore, the transcriptional behaviour of the WT memory genes in the *ΔE(z)* mutant was explored. Indeed, examination of L2FC and padj values for the 59 memory genes in *ΔE(z)* PLTvsT, revealed that 51 out of 59 WT memory genes were differentially expressed in PLTvsT in the mutant, too, leaving 8 WT memory genes which exhibited a lack of PLTvsT differential expression in *ΔE(z)* (Table S2, indicated in red). Consequently, these 8 genes were designated as genes loosing trainability in the *ΔE(z)*. By further characterizing their behaviour in WT and mutant, these genes could reveal the importance of H3K27me3 in *C. merolae* thermomemory. First, these 8 genes were localized in the heatmap showing all WT memory genes to allocate them to either of the clusters 1-5 (Figure 3.8C, indicated by arrows). Interestingly, 4 of these 8 were situated in cluster 4 memory genes. Indeed, when comparing expression profiles of genes in clusters 1 - 5 in *ΔE(z)* to those in WT, no considerable difference could be detected for clusters 1 – 3 (Figure 3.9 B). Basal transcriptional rates and/or expression dynamics throughout the HSMA remained same. Solely the magnitude of induction in PvsN, TvsNL and PLTvsPL sometimes differed between WT and mutant. In summary, the mode of action of cluster 1 – 3 genes was

maintained in *ΔE(z)*. Further, cluster 5 genes in *ΔE(z)* seem to exhibit slightly decreased induction in P, but are able to retain the expression profile. In fact, comparing memory gene behaviour in *ΔE(z)* to WT, the most pronounced difference was found in cluster 4. As stated earlier, 4 genes displaying loss of significance between PLTvsT in *ΔE(z)* are situated in this cluster, immediately drawing attention to this group of genes.

In detail, in WT cluster 4 genes followed the general memory expression pattern of upregulation in P, return of transcriptional rates to control levels in PL followed by stronger induction in PLT compared to T (Figure 3.9A). In *ΔE(z)*, this pattern seemed disturbed. No difference between the medians of P and NL could be detected as well as induction levels in T were comparable to PLT (Figure 3.9B) nicely depicting the fact, that 4 out of 5 genes in this cluster do not show differential expression in *ΔE(z)* PLTvsT (Figure 3.8C). In fact, examination of the expression profile of cluster 4 genes revealed that in N and PLT outliers arose in the mutant, which could lead to the summarized aberration in expression pattern. These results suggest, that *E(z)* might act predominantly on memory genes which are only moderately induced by heat.

Efforts were then narrowed to further describe the mode of *E(z)* action in transcriptional regulation of the 8 genes. Consequently, their annotation and individual gene expression during the HSMA was scrutinized. The corresponding 8 genes included the nuclear encoded genes *CMB107C*, *CMF092C*, *CMJ101C*, *CMP040C*, *CMT010C,* as well as *ycf55* (*CMV038C*), *menA* (*CMV116C*) and *rpl18* (*CMV179C*) the latter three originating from the chloroplast genome (Figure 3.10).

It is important to note, that the memory gene *CMT010C*, encoding for a retrotransposon, is one out of two memory genes showing repression in PLTvsT. Representing this exceptional reaction among trainable genes, those two repressed memory genes are separately examined in Chapter 3.3.2.3. Thus, to avoid redundancy, *CMT010C* was excluded from Figure 3.10. The first memory gene, which shows trainability-loss in *ΔE(z)*, *CMB107C*, was moderately upregulated in WT PLT compared to T (Figure 3.10 and Table S2). This gene is not characterized yet, but is predicted to possess a thioredoxin domain (https://www.uniprot.org/uniprot/) and consequently was annotated as “thioredoxin-like” within this project. Belonging to cluster 4 memory genes, *CMB107C* showed only moderate changes upon heat. Interestingly, it showed a slight aberration in the expression pattern of memory genes in WT: After priming, normalized counts seemed not to recover entirely after lag phase (PL levels higher than NL) and while in PLT the induction further increases, the T condition barely reacts at all showing control (NL) transcriptional levels (Figure 3.10A).

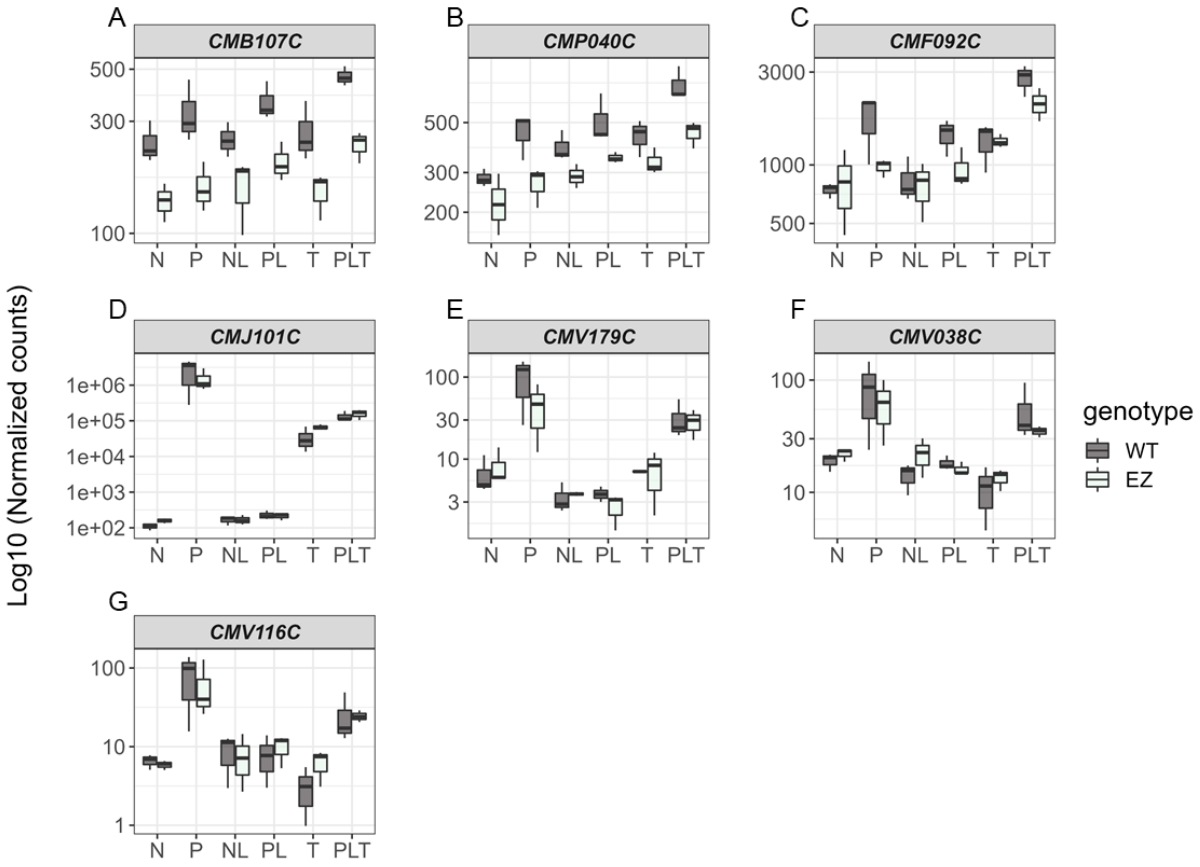


Figure 3.10: Genes exhibiting aberrant expression (patterns) in *ΔE(z) C. merolae* cells
RNAseq expression data in boxplots for **A** *CMB107C*, **B** *CMF092C*, **C** *CMP040C*, **D** *CMJ101C*, **E** *CMV179C*, **F** *CMV038C*,**G** *CMV116C*, identified as memory genes in *C. merolae* heat stress. Y-axis depicts relative expression given as log10 transformed normalized counts in different conditions collected during the heat stress memory assay (HSMA): naïve (N), primed (P), naïve – after lag phase (NL), primed – after lag phase (PL), triggered-only (T) and primed-and-triggered (PLT). Wild type (WT) and *ΔE(z)* counts are depicted in black and light green, respectively. All genes show aberrant trainability in *ΔE(z)* cells compared to WT leading to loss of significantly different expression in PLT compared to T. Note that *CMB107C* and *CMP040C* show divergent expression patterns during the HSMA.

In *ΔE(z)* the transcription rate was slightly lowered under control conditions (EZvsWT, L2FC = -0,43, padj = 0.02). In fact, *CMB107C* transcription rates in *ΔE(z)* did not reach WT levels in any condition. Furthermore, the induction in P seen for WT is close to vanished in the mutant (L2FC = 0.4, padj = 0.03, PvsN) and equally the induction in WT_PLTvsPL accounts for 0.3 padj=0.007 while in EZ_PLTvsPL is lowered to L2FC = 0.14 padj =0.5. The incomplete induction of the gene *CMB107C* in the mutant suggests that *CMB107C* reaction to heat depends on H3K27me3. In summary, it is tempting to speculate, that the absence of *CMB107C* activation in *ΔE(z)* PLT might be due to the missing activation upon priming itself. When the priming stimulus lacked, there was barely activation upon triggering, which was reflected in no induction in T(vsNL). However, when the experience of a priming stimulus was made it lead to a hyperinduction in PLT(vsPL), which underlines the need of a prior heat experience in order to guarantee a proper PLT induction.

A slightly similar situation was detected for two other genes exhibiting memory loss in *ΔE(z)*: *CMP040C* and *CMF092C*. The former encodes for a protein which is not characterized in *C. merolae* to date. No similarity to any known protein domain was detected. Likewise, *CMF092C* is an protein of unknown function possessing two disordered domains and is assigned to the chaperonin_RbcX_superfamily by the database InterPro. The chaparonin_RbcX_sf is a protein family which assists in the correct assembly of RbcL and RbcS subunits during RuBisCO biogenesis and supports the maximal activity of this carbon fixating enzyme. Notably, this protein shows no similarities to higher plants but is unique to red, brown and yellow-green algae (blastp). Albeit uncharacterized, both genes were designated trainable (Table S2) and belong to cluster 4 of memory genes (Figure 3.8C). Similar to *CMB107C*, *CMF092C* and *CMP040C* were both strongly activated upon priming in the WT and this activation was not entirely back to control levels after lag phase (Figure 3.10 B,C). In the mutant, however, the induction in *ΔE(z)*_P did not reach WT_P levels and both genes were only insignificantly induced in *ΔE(z)* PLT suggesting that the lack of activation upon priming in *ΔE(z)* might cause the paltry induction in *ΔE(z)* PLT. It can be concluded, that the necessity of a priming stimulus in order to react to the triggering stimulus accounts for those two genes, too. Furthermore, via preventing transcriptional levels to return to control levels after the priming induction, *E(z)* seems to contribute to elevated transcription of those genes upon stress reoccurrence in PLT.

In contrast to *CMB107C*, *CMP040C* and *CMF092C* the small heat shock protein of the HSP20 family *CMJ101C* is an example of a gene, which loses significance in *ΔE(z)* PLTvsT due to an overreaction in *ΔE(z)*_T compared to WT_T. *CMJ101C* exhibits 1.8-fold higher induction in PLT compared to T (padj = 0.02, Table S2) in WT. *CMJ101C* belongs to cluster 1 memory genes (Figure 3.8C), and in WT follows the pattern of a [+/+] trainable gene, with full recovery after lag phase. In the mutant, expression of this gene differs in the extent of induction upon heat (Figure 3.10D). While the *CMJ101C* induction in *ΔE(z)* upon priming is reflecting WT_P levels, E(z)_T exceeds WT_T levels representing a hyperinduction of this gene upon triggering in the mutant. No further significant increase can be detected in *ΔE(z)*_PLT suggesting a near-to-saturated level in T already. Thus, *CMJ101C* reflects a gene which loses trainability in *ΔE(z)* due to an hyperinduction resulting in a vanished difference in PLTvsT for the mutant. *CMJ101C* encodes for the small heat shock protein CmsHSP2 (Kobayashi *et al.*, 2014). Interestingly, *CMJ101C* is the only memory gene which is enriched with H3K27me3 in the wild type (Mikulski *et al.*, 2017). The observed hyperinduction in the *ΔE(z)* mutant upon triggering ascribes once more a repressive function to H3K27me3 and suggests the possibility, that *E(z)* might ensure a sensitive, regulated answer to heat stress (buffering function). Intriguingly, CmsHSP2 was predicted to be targeted to the chloroplast or mitochondrium (Kobayashi *et al.*, 2014) designating this gene

as potential candidate for a crosstalk between nucleus and chloroplast.
Since chloroplast genomes are devoid of histones (Salganik *et al.*, 1991), the histone mark H3K27me3 is absent from the chloroplast genome. Still, three chloroplast genes lost their WT transcriptional memory pattern upon the mutation of *E(z)* suggesting nuclear-chloroplast crosstalk of certain upstream, H3K27me3 dependent regulator(s). These genes, namely *ycf55*, *menA* and *rpl18,* followed the general memory transcriptional pattern: After strong induction upon priming they recovered to control levels after lag phase and subsequently during triggering were induced to higher extent in PLT than in T. For those three genes, the loss of significance in the mutant was not found in an overall deviation of gene expression in some conditions. Rather, the only slight differences in expression strengths did not allow to pass significance thresholds. Thus, either the strength of induction was not entirely reached in *ΔE(z)* PLT (*ycf55*) or transcriptional levels were slightly elevated in *ΔE(z)* T leading to reduced difference in expression strength between PLT and T (*menA* and *rpl18*). All three genes exhibit very low transcription rates. In detail, the detected read number did not exceed 100 reads. Differential analysis reaches its limits in reliable detection once read numbers fall below 10 reads, which is the case for these three genes under control and T conditions. Thus, further investigation is needed to confirm the actual biological meaning of these three memory genes.

Taken together, the influence of H3K27me3 on memory gene transcription is limited. Only 8 out of 59 identified memory genes show aberrant memory behavior including one retroelement which is addressed in the following section. Expression data revealed that the reason for missing the significance threshold between T and PLT for 7 out of those 8 genes is either an upregulation in E(z)_T compared to WT_T (*CMJ101C*) or the entire HSMA expression pattern (i.e. upregulation in priming, recovery after lag phase and higher re-induction in PLT than in T) is disturbed. For the chloroplast memory genes *ycf55*, *menA* and *rpl18* the detected changes are too subtle to gain significance. Remarkably, E(z) seems to act predominantly on memory genes of cluster 4 comprising genes which are not substantially activated by heat but still are affected by repetitive heat. Whether the memory process is accompanied by chromatin modification remodelling including H3K27me3 remains to be solved in next experiments.

3.3.2.3 Two retroelements are persistently suppressed upon priming event

When identifying the 59 memory genes and determining their direction of change, two genes attracted special attention, since they were – opposing to the remaining 57 – suppressed by priming and exhibited negative fold changes in PLTvsT (Table S2). Remarkably, those two genes, namely *CMT010C* and *CMJ308C*, both encode for retrotransposons. According to

their expression pattern in primed-before and triggered-only cultures, they were allocated to cluster 4 genes (Figure 3.8 C). During the HSMA they show remarkable behavior (Figure 3.11). Thus, in WT, both are significantly upregulated via priming. Similar to what was noted for other cluster 4 genes, such as *CMB107C*, after lag phase (PL) this induction did not revert entirely to control (NL) levels for *CMJ308C* and *CMT010C*. Indeed, PL levels undercut NL levels showing -1.2 fold (padj = 0.01) and -1.4 fold (padj = 0.0007) reduced levels for *CMT010C* and *CMJ308C*, respectively, suggesting an repressive effect of priming on the transcription of both genes which evolves during lag phase. As transcription levels were lower in PLT than in T, it seems that the occurrence of priming leads to lasting suppression of those two retrotransposons, especially in the case of stress reoccurrence.

Albeit the common regulation during the HSMA for both genes in WT, they show fundamental differences in *ΔE(z)*. While *CMJ308C* retains its transcriptional memory pattern in the *ΔE(z)* mutant, *CMT010C* is unable to react to heat stress exhibiting no difference in transcriptional rates across conditions for *ΔE(z)* (Figure 3.11). This might be explained by the basal induction of *CMT010C* upon the *E(z)* deletion. Accordingly, *CMT010C* was identified as misregulated gene in *ΔE(z)* under control conditions in Chapter 3.1 which is reflected in significantly higher control (N) levels in the mutant (Figure 3.11, Table S4). However, *CMT010C* was not identified as H3K27me3 occupied gene in WT (Mikulski *et al.*, 2017). The WT expression pattern of *CMT010C* during the full factorial experiment was not detectable anymore in the mutant. Moreover, *CMT010C* expression seemed to be induced by age (NL) and to similar extents by priming in the mutant. No further expression changes could be detected upon triggering, neither in T nor PLT. In conclusion, the repressive (memory) effect gained by priming in WT_PLT was probably lost in *ΔE(z)* because the overall transcription of *CMT010C* was hyperinduced in the mutant.

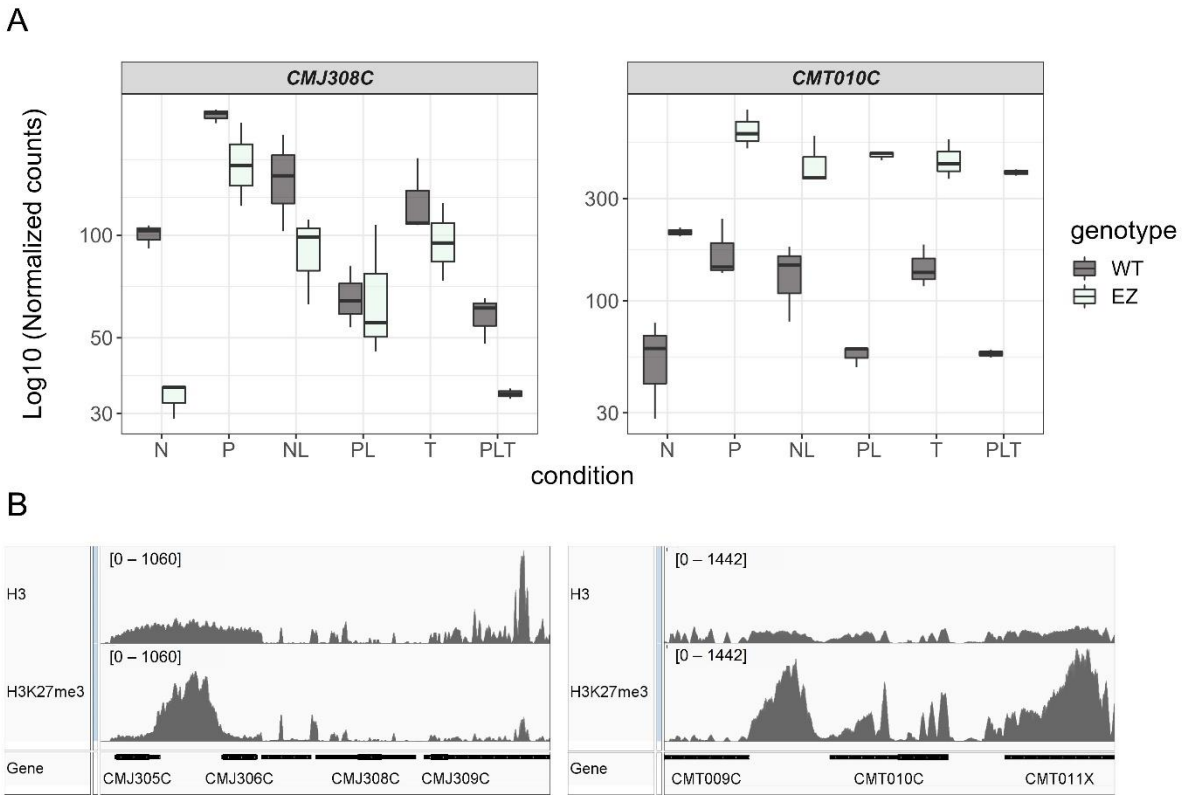


Figure 3.11: Two retrotransposons are lastingly repressed by priming heat in primed-and-triggered *Cyanidioschyzon merolae* cells

RNAseq expression in boxplots (**A**) and H3 as well as H3K27me3 occupancy data (**B**) for retrotransposon encoding genes *CMJ308C* and *CMT010C*. Both genes were identified as memory genes, however, those are the only two exhibiting lasting repression when primed.

A Y-axis depicts relative expression given as log10 transformed normalized counts in different conditions collected during the heat stress memory assay (HSMA): naïve (N), primed (P), naïve – after lag phase (NL), primed – after lag phase (PL), triggered-only (T) and primed-and-triggered (PLT). Wild type (WT) and *ΔE(z)* counts are depicted in black and mint green, respectively. *CMT010C* shows aberrant trainability in *ΔE(z)* cells compared to WT leading to loss of significantly different expression in PLT compared to T. *CMJ308C* shows E(z)-independent memory.

B Genome browser view of H3 and H3K27me3 read coverage at the loci *CMJ308C* (right panel) and *CMT010C* (left panel). Depicted histone H3 and H3K27me3 ChIP-seq signals in naïve WT cells (Mikulski *et al.*, 2017). Both tracks exhibit three merged biological replicates. Numbers in brackets above each track denote the scale read count. ChIP-seq data are retrieved form (Mikulski *et al.*, 2017). The “Gene” track indicates gene position as given by the “new” reference annotation file (S. Rader, unpublished).

As *CMT010C* clearly was regulated by the H3K27me3 absence albeit not identified as direct H3K27me3 target, and furthermore, as repetitive sequences in *Arabidopsis* are discussed to be partially regulated by epigenetic regulations in order to be resilenced after heat stress activation (Pecinka *et al.*, 2010), in Figure 3.11B the H3 and H3K27me3 occupancy at the direct genomic vicinity of both, *CMJ308C* and CMT01C was visualized. The neighboring regions to *CMT010C* show elevated H3K27me3 levels to both sides. Interestingly, the 3´peak covers another repetitive element (*CMT011X*), and the H3K27me3 peak 5´ to *CMT010C* concerns an intergenic region. Similarly, an H3K27me3 peak adjacent to *CMJ308C* could be observed. However, as *CMJ308C* is located on the reverse strand, the H3K27me3 peak is

located downstream to the gene, opposing to the intergenic H3K27me3 peak upstream to *CMT010C*.

In summary, in WT the expression pattern across conditions is similar for both, *CMJ308C* and *CMT010C*: they are upregulated by priming heat, but not by triggering heat. However, when priming occurred earlier, triggering leads to downregulation suggesting a repressive effect of the priming event on their transcription upon a triggering stimulus (Figure 3.11). However, while H3K27me3 seems to regulate repression and thus memory of *CMT010C* (without being a direct target of this mark), *CMJ308C* shows PRC2-independent transcriptional memory. Since repetitive elements in majority are excluded from the current annotation, it will be interesting whether the repression via a priming stimulus is true for more than these two repetitive elements.

3.3.2.4 Validation of identified memory genes in extended lag phase HSMA

The validity of the identified genes in the Chapters 3.3.2.1 and 3.3.2.2 as memory markers was tested in primed and triggered samples collected when different lag phases were applied in order to determine the maximal memory preservation (Chapter 3.2.1). Several candidates were tested in RT-qPCR to see whether their expression associates with the detected memory behaviour on survival plate and PAM measurements (Figure 3.5). In the same time, some differentially expressed genes identified by the RNAseq were validated for the 24h lag phase.

To do so, primers were designed for the following genes: *CMJ101C* and *CMJ099C*, which according to the analysis showed memory responses in addition to high transcription rates. Furthermore, *CMJ101C* represents the only H3K27me3 target and is a cluster 1 memory gene, which is highly heat responsive and shows [+/+] trainability. *CMJ099C* in turn is a cluster 2 gene and represents a [=/+] trainable gene, which is not induced upon a triggering stimulus unless the prior exposure to priming was made. Further, two cluster 4 genes were included in the validation, which were described to incompletely return to control levels after lag phase and additionally loosing trainability in *ΔE(z)* mutant: *CMP040C* and *CMF092C*. Since in this validation, the trainability, i.e. the difference between PLT and T, was of interest, only the conditions NL, T, PLT were looked at, in WT and *ΔE(z)* samples. Similar to the mRNA sent for RNAsequencing, T and PLT samples were collected after 15 min in 60°C to prevent transcriptional responses connected to cell death.

As a result, two groups of genes could be verified. The first group includes *CMJ101C* and *CMJ099C* and mirrors the memory emergence seen on physiological level. Accordingly, after 2h lag phase *CMJ099C* and *CMJ101C* were induced in the T culture, but the expression in PLT was elevated (Figure 3.12 A). Although the difference in expression between PLT and T is clearly visible for both genes, it did not reach significance. This can be possibly explained

by high variations between replicates suggesting some variability in the response to triggering after 2h lag phase. Consistent with the RNAseq results, the expression of *CMJ099C* and *CMJ101C* was higher in PLT compared to T cultures after 24h lag phase (Figure 3.12 B). Albeit expression of *CMJ101C* was elevated more than 2000-fold in PLT compared to T, significance could not be reached. Notably, for both genes the expression in the mutant was elevated in PLT which lead to significant differences between PLT and T for *CMJ099C* in the mutant. In the RNA-sequencing analysis *CMJ101C* was identified as a gene losing trainability due to hyperinduction in T already in the mutant. This observation is partially mirrored in qPCR experiments, since the expression in *ΔE(z)*_T likewise is slightly elevated compared to WT_T.

After 48h lag phase, the transcriptional difference between PLT and T expression for those two genes vanished when the lag phase lasted 48h (Figure 3.12 C). These results were consistent with what was detected on survival plate and in PAM measurements (Figure 3.5): memory emergence in 2h lag phase as well as 24h lag phase followed by memory loss when the lag phase lasted for 48h. *CMJ099C* and *CMJ101C* showed slightly higher, yet comparable expression patterns in *ΔE(z)* which equally reflected the emergence and cease of memory formation relative to the length of lag phase. In the case of 24h lag phase the expression behaviour of *CMJ099C* and *CMJ101C* can be linked to the priming benefit in 24h – A (Figure 3.5 A).

Contrarily, the two genes *CMP040C* and *CMF092C* exhibited memory loss in the mutant strain *ΔE(z)* displaying the second group of memory genes. In detail, in WT cells *CMF092C* and *CMP040C* expression was not significantly different between PLT and T in 2h (Figure 3.12A), but in the 24h lag phase (Figure 3.12B). Upon 48h lag phase extension this significant difference is not detectable anymore but the trend in expression hinted towards an additional increase for both genes (Figure 3.12C). However, in *ΔE(z)* the expression for both genes remained unresponsive over the whole time course showing barely T or PLT induction in none of the lag phases. Accordingly, the absence of WT trainability of those genes in *ΔE(z)* is consistent with the result of the survival plate in 24h B of Figure 3.5A as well as the loss of significance in PAM measurements in Figure 3.5B.

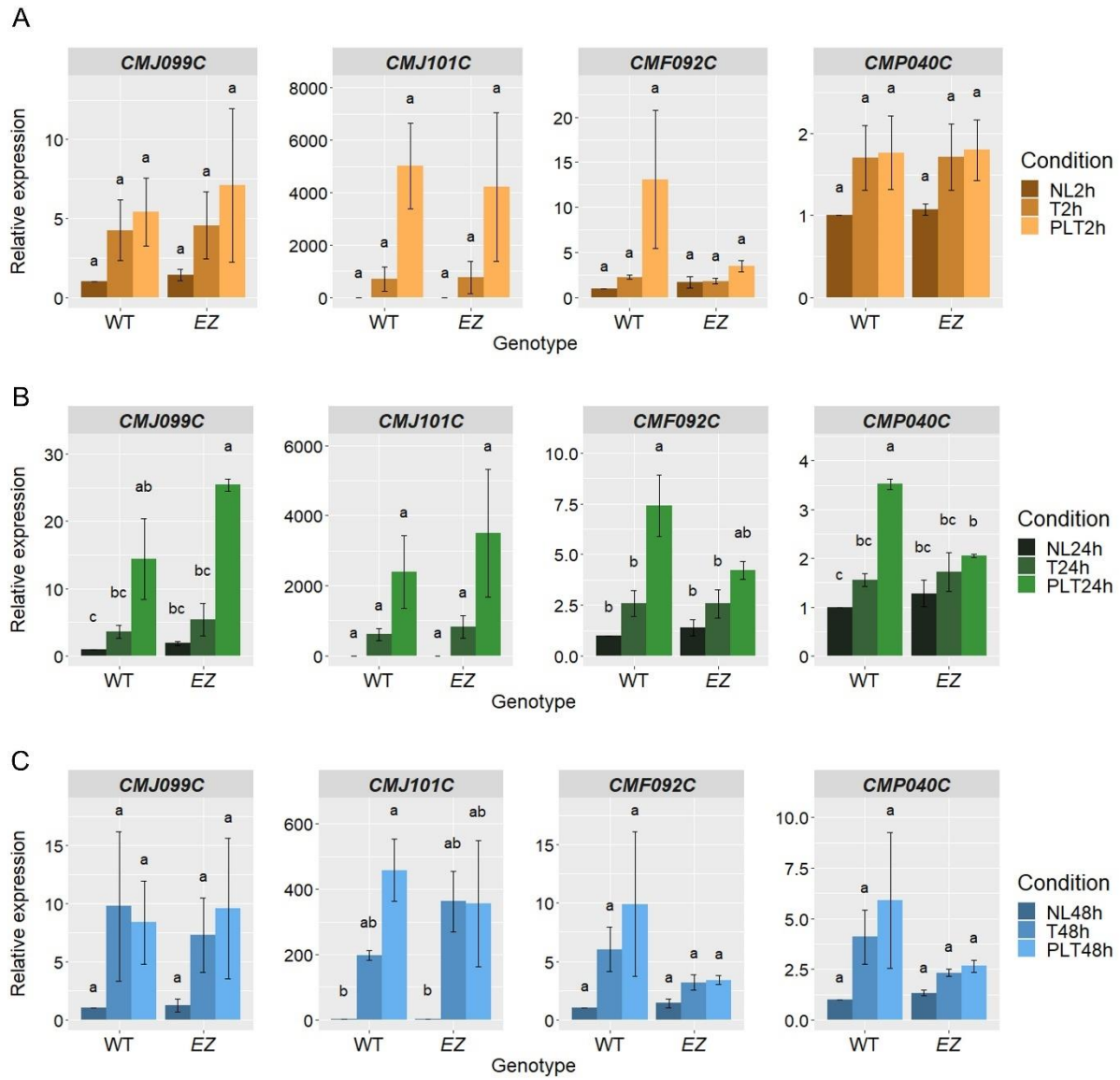


Figure 3.12: Memory gene expression reflects physiological memory in *Cyanidioschyzon merolae*

Gene expression given as normalized counts of memory genes *CMJ099C*, *CMJ101C*, *CMF092C* and *CMP040C* in WT and *ΔE(z)* (*EZ*) cells exposed to lag phase durations of 2h (**A**), 24h (**B**), and 48h (**C**) in the heat stress memory assay is shown. For the memory genes *CMJ099C* and *CMJ101C* the shown expression associates with PAM measurements. Significance arises when lag phase is stretched to 24h and fades completely after 48h lag. *CMF092C* and *CMP040C* memory genes show loss of trainability in *ΔE(z)* with strongest effect in the 24h lag phase. Samples were taken after 15 min triggering in 60°C waterbath of naïve (N), triggered-only (T) and primed-and-triggered (PLT) cells in wild type (WT) and *ΔE(z)* cells. RNA was isolated from 2 ml cultures at an OD750nm reached after lag phase (see Table 3.2). Transcript levels were measured by reverse transcription and qPCR. Mean of the reference genes TIP41, 60S, and TUB was used as internal control. Data was normalized to WT NL condition in corresponding lag phase. Error bars represent the SD of three biological replicates. Statistical significance was tested by 2-way ANOVA followed by Tukey´s test ($p < 0.05$). Identical letters indicate no significant difference. (SD: standard deviation, OD750nm: optical density at 750 nm)

As a result, the two genes *CMF092C* and *CMP040C* could be identified and verified losing trainability in *ΔE(z)*. Understanding the mechanism which leads to the inability of those genes

to respond in a memory-like manner might be an important step towards unravelling the function *E(z)* has in the memory formation or maintenance process.

In sum, the memory dynamic over time, i.e. upon lag phase extension, seen on physiological level in Chapter 3.2.3 could be reproduced on transcriptional level. Memory emerged in 2h lag phase, was stably established in 24h lag phase and undetectable when the lag phase was 48h in length. The connection between physiological and transcriptional memory using the memory genes undermines the potential of the memory gene set to aid in the dissection of the precise mechanism underneath thermometry in *C. merolae*.

In the same time, results from the RNAseq analysis for the 24h lag phase could be successfully verified using RT-qPCR. Consistent with RNAseq data, *CMF092C* and *CMP040C* were identified as *E(z)*-dependent memory genes in *C. merolae.*

3.3.2.5 PRC2 independent memory regulation of small heat shock protein (sHSP) genic region on chromosome 10

As presented in Chapter 3.3.2.1, 59 genes were identified as heat memory genes in *C. merolae*. According to their origin they can be classified to 9 nuclear and 50 chloroplast encoded genes. When searching for co-regulation of genes at a common genomic position, three nuclear encoded gene candidates attracted special attention, since they were situated in direct vicinity to each other in the genome: *CMJ099C*, *CMJ101C* and CMJ102C - located on chromosome 10 (Figure 3.8C, Heatmap). The trainable behaviour of the former two was analysed for different lag phases in the previous Chapter. Opposing to the RNA-sequencing differential expression analysis, which identified *CMJ101C* as gene losing trainability in the *ΔE(z)* mutant, gene expression measurements across lag phases defined both, *CMJ101C* as well as *CMJ099C*, as E(z) independent memory genes. Following, it was of interest to determine, whether more than these three genes are affected around those genes.

For that purpose the examination of transcriptional behaviour was widened around that locus spanning 11 genes from *CMJ096C* to *CMJ107C* and the regulation across all HSMA conditions was surveyed. This allowed the broadening of common regulation at that genomic region to 5 genes, namely *CMJ099C* to CMJ103C (including *CMJ100C*, which is excluded from the current annotation file, Figure 3.13A). The two bordering genes *CMJ098C* and *CMX002X* (the suffix X designates a repetitive element) are not regulated by heat anymore (Figure 3.13A, C). In sum, the genes *CMJ099C* – *CMJ103C* seem to be similarly affected upon the heat memory scheme being highly upregulated in PLT and somewhat in T.

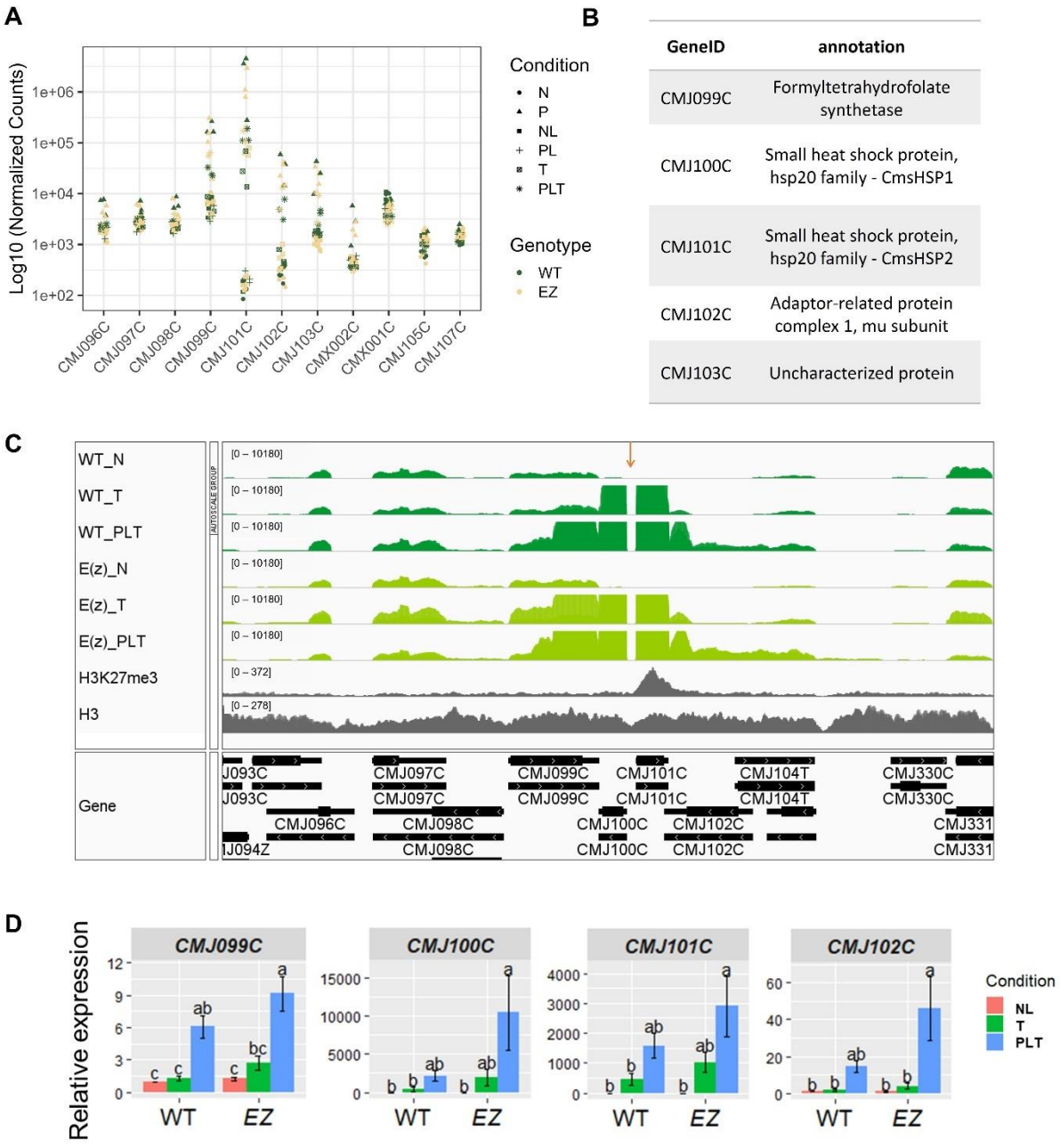


Figure 3.13: Five neighboring genes on chromosome 10 exhibit common regulation in heat stress memory of *Cyanidioschyzon merolae*

A Expression data on chromosome 10 stretching from gene *CMJ096C* to *CMJ107C* restrict similar regulation to 5 genes: *CMJ099C – CMJ103C*. Depicted is the relative expression given as log10 transformed normalized counts for wild type (WT, dark green) and *ΔE(z)* (yellow) in following conditions: naïve (N), primed (P), naïve after lag phase (NL), primed after lag phase (PL), triggered only (T) and primed and triggered (PLT) in three replicates each. Note that *CMJ100C* is not included in the current annotation file and no sequencing data are available subsequently.

B Annotation of five similarly regulated genes on chromosome 10 in heat stress memory including *CMJ100C* (annotation retrieved from czon.jp)

C Genome browser view of memory region defined in A and B. Depicted are RNAseq normalized counts in N, T, and PLT cells for WT and *ΔE(z)* (track 1-6) compared to histone H3 and H3K27me3 ChIP-seq signals in naïve WT cells (Mikulski *et al.*, 2017, track 7 - 8). Only *CMJ101C* is enriched with H3K27me3. Note, that region between *CMJ100C* and *CMJ101C* exhibits slightly decreased H3 signals. All tracks exhibit three merged biological replicates. Numbers in brackets above each track denote the scale depicted in normalized counts, RNAseq data are set to same scale. The lower track indicates gene position as given by the reference annotation file.

D Gene expression changes of the genes *CMJ099C*, *CMJ100C*, *CMJ101C*, *CMJ102C* in the center of the commonly regulated region on chromosome 10, including *CMJ100C* which is not included in the published annotation file. Samples were taken from untreated naïve (NL) cells and after 15 min triggering in 60°C waterbath of triggered-only (T) and primed-and-triggered (PLT) cells in wild type (WT) and *ΔE(z)* strain. RNA was isolated from 2 ml cultures. Transcript levels were measured by reverse transcription and qPCR. TIP41 was used as internal control and data was normalized to WT_NL. Error bars represent the SD of three biological replicates. Statistical significance was tested by 2-way ANOVA followed by Tukey´s test ($p < 0.05$). Identical letters indicate no significant difference. (SD: standard deviation, ChIP: chromatin immunoprecipitation)

After the commonly regulated region could be extended to 5 neighbouring genes, next their constitution was scrutinized. *CMJ099C* is situated 5´ of this region. It is annotated as formyltetrahydrofolate synthetase (FTHS). FTHS encodes for an enzyme, which catalyzes the formylation of tetrahydrofolate - an initial step in the reduction of carbon dioxide and other one-carbon precursors to acetate (InterPro data base, Blum *et al.*, 2021). Notably, the two genes forming the centre of this region, *CMJ100C* and *CMJ101C*, encode for small heat shock proteins and are annotated as CmsHSP1 and CmsHSP2, respectively (Kobayashi *et al.*, 2014). Formerly it was proposed that *CmsHSP1* and *CmsHSP2* - located on opposite strands - are driven by a common promoter and are transcribed as well as translated rapidly upon heat stress. Furthermore, prediction studies suggest that those two sHSPs are targeted to different cellular compartments: *CMJ100C* to the cytosol and *CMJ101C* to the mitochondria or chloroplast (Kobayashi *et al.*, 2014) designating these two heat shock proteins as potential heat information spreaders across the cell. The following gene, *CMJ102C*, encodes for the mu-subunit of adaptor-related protein complex 1 which is needed for the coupling of clathrin lattices with particular membrane proteins by self-phosphorylation via a mechanism that is still unclear (Nakayama *et al.*, 1991). Lastly, *CMJ103C* encodes for an uncharacterized protein of 442 aa in length and is unique to *C. merolae*. No predicted domains have been found and sequence alignments using BLAST (Altschul *et al.*, 1990, 1997) resulted in no hits in any other species.

Examination of transcriptomic data for these genes revealed that *CMJ101C* is heat responsive, and generally activated in all heat contexts (PvsN, TvsNL, PLTvsPL, Table S5). This fact could not be reproduced in a significant manner in qPCR experiments (Figure 3.13D), however *CMJ101C* expression was induced approx. 500-fold in T compared to the NL condition and significance might be hidden under the tremendously high induction of this gene in other conditions. *CMJ100C* is likewise induced by heat, but not to similar high extends which is in contrast to the fact, that *CMJ101C* is targeted by the allegedly repressive mark H3K27me3 (Mikulski *et al.*, 2017, Figure 10A) whereas *CMJ100C* is not. The other three genes (*CMJ099C*, *CMJ102C, CMJ103C*) are only induced, when heat recurs, i.e. in the PLT condition (Figure 3.13) which presumes that the heat shock proteins *CMJ100C* and *CMJ101C* act as perceivers of heat information and only subsequently the genes *CMJ099C*, *CMJ102C* and *CMJ103C* are regulated.

In order to test whether this region is PRC2 directed, in the next step the transcriptional situation at that region in *ΔE(z)* was focused on. In general, the pattern is not significantly altered in the mutant. Here, too, the whole region showed increased induction in PLT compared to T. However, in the mutant the genes *CMJ099C*, *CMJ100C* and *CMJ101C* showed elevated expression in T already, which was not seen for the WT (Figure 3.13C, D) and furthermore PLT expression rates are increased for all tested genes in the mutant. This suggests a regulatory function of E(z) in preventing hyperinduction of these genes in the WT situation. Whether these hyperinduced genes have downstream consequences remain subject for future investigations.

A link to the mode of PRC2 action at this region might be constituted by the single H3K27me3 targeted memory gene, which in the same time is centre to the commonly regulated region: *CMJ101C*. First, to ascertain that there are no other H3K27me3 peaks around that locus, RNAseq read coverage traces of control (N), triggered only (T) and primed-and-triggered (PLT) samples in WT as well as *ΔE(z)* were visualized using a genome browser (IGV, Robinson *et al.*, 2011). Then, H3K27me3 and H3 (control) reads collected in WT under control conditions (N, retrieved from (Mikulski *et al.*, 2017) were added to depict the histone mark distribution in relation to total histone (H3) occupancy. As a result, only *CMJ101C* is clearly enriched with H3K27me3 under control conditions in WT, with no further peak up- or downstream of the designated memory region (Figure 3.13C). Notably, the region which is predicted to serve as common promoter for *CmsHSP1 (CMJ100C*) and *CmsHSP2* (*CMJ101C*) (Figure 3.13C, orange arrow, Kobayashi *et al.*, 2014b) is devoid of any reads and in parallel exhibits a slight valley of H3 coverage, which could hint towards a lowered nucleosome occupancy. This assumption, however, remains to be analysed.

It can be summarized, that the commonly regulated region on chromosome 10 during heat stress memory centres around two small heat shock proteins which directly link the regulation to the biology needed during heat (memory) responses. Interestingly, those two – *CMJ100C* and *CMJ101C* – are targeted to different cellular compartments. While *CMJ100C* is targeted to the cytosol, *CMJ101C* possesses transition peptides to the mitochondrium or chloroplast. Furthermore, *CMJ101C* is decorated by H3K27me3 in WT situations. As in the mutant the whole region seems to be over-induced, which is exceptionally clear for *CMJ101C* itself, it is tempting to speculate that H3K27me3 in this process inherits an function in “opening” this region for transcription. However, the precise mechanism beyond that regional control has to be elucidated in subsequent studies.

3.4 The chloroplast as a memory hub in *C. merolae*

The previous chapters described the ability of *C. merolae* to benefit from a past heat priming experience upon recurrent heat stress and the impact of epigenetic mechanisms conferred by PRC2 on that process. Moreover, transcriptomic analyses revealed a set of heat trainable genes referred to as memory genes of which some indeed rely on an active PRC2 complex. Interestingly, a region at chromosome 10 might be particularly relevant, since there three memory genes are found in direct vicinity centred around the PRC2 target gene *CMJ101C*. The memory gene set comprises 59 genes in total, which behave differently upon the same kind of stress (i.e. the triggering stress) depending on the presence or absence of a previous priming experience. These memory genes could be divided into genes, which are encoded in the nucleus und genes which are encoded in the chloroplast genome. In detail, 50 out of 59 memory genes comprising the heat trainable gene set are encoded in the chloroplast genome (Chapter 3.3.2 and Table S2). As this project concentrates on the function of H3K27me3 in a unicellular photoautotrophic eukaryote - a histone mark which is naturally absent in the chloroplast genome - there was no obvious connection between H3K27me3 function and this plant organelle. However, as the previously mentioned H3K27me3 target gene *CMJ101C* is described to be targeted to the chloroplast (Kobayashi *et al.*, 2014), regulatory processes in the plastid genome could be controlled indirectly by nuclear epigenetics. Overall, the relevance of chloroplasts in stress responses was reported before (H.-D. Yu *et al.*, 2012; Y. Zhang *et al.*, 2020) and the here observed high proportions of chloroplast genes within the memory gene set underlines its importance in heat stress responses. Even further, it implies a substantial contribution of the chloroplast to heat stress memory formation. In the light of this observation, it was of interest to define the extent to which the chloroplast genome is involved in heat stress memory and therewith get first insights into potential mechanisms or pathways which could explain the vast memory gene contribution of the plastid genome. At the same time, the (possibly indirect) influence of H3K27me3 on the chloroplast genome should be investigated. In order to do so, first the WT situation was described for the chloroplast genome during the HSMA. Subsequently, aberrations in the *ΔE(z)* might unveil PRC2 contributions.

Notably, the 50 chloroplast memory genes made up 25% of the total number of proteins encoded in the *C. merolae* chloroplast (i.e. 206 protein coding genes). Thus, the priming effect perhaps could alter entire plastom regions rather than single genes only. To test this assumption, transcriptomic changes of the entire plastid genome during the conducted HSMA were visualized (Figure 3.14). Depicted is the expression all 200 protein coding genes of the *C. merolae* chloroplast.

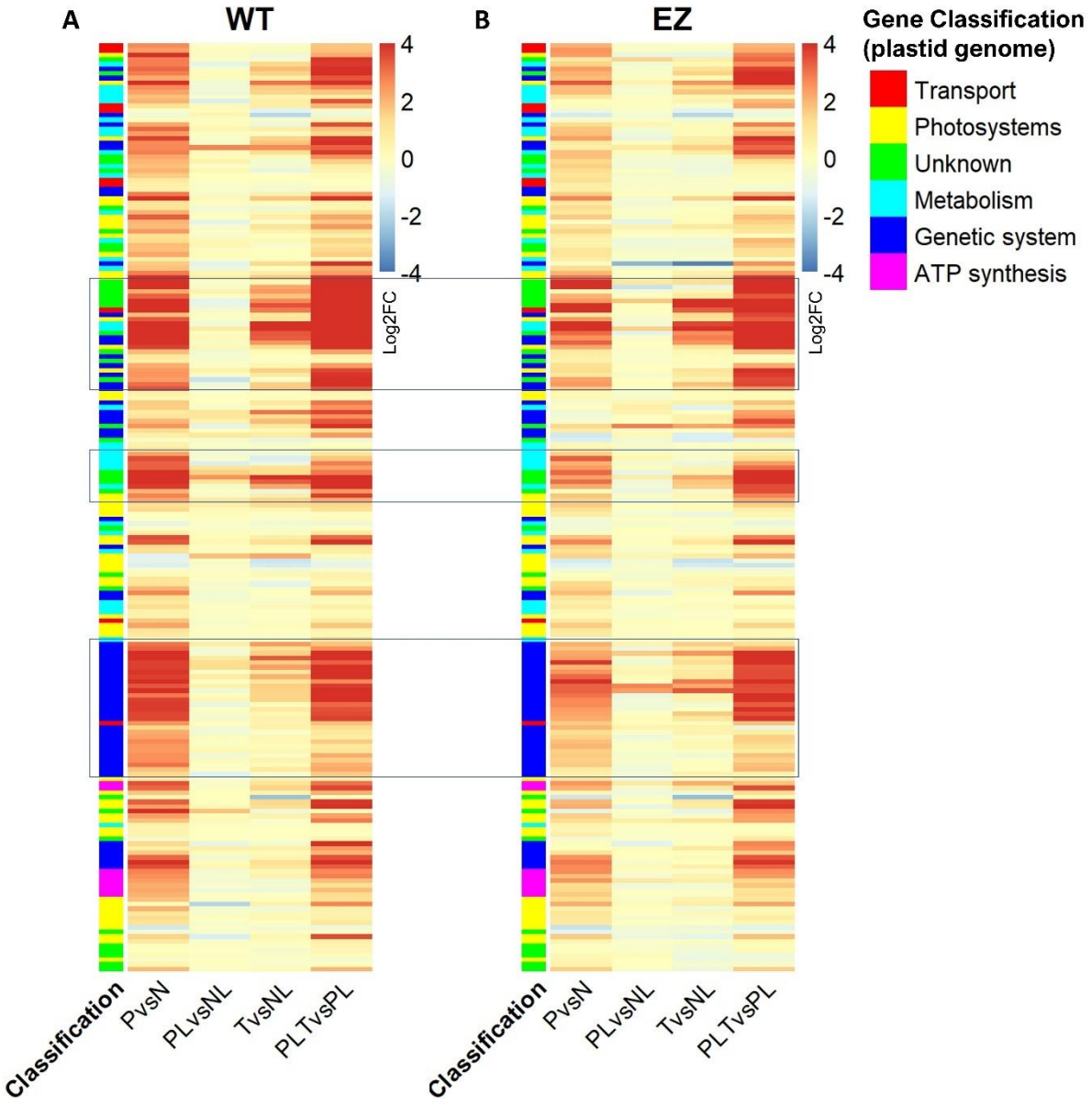


Figure 3.14: Pronounced transcriptional response of the *C. merolae* plastid genome during heat stress memory

including all 200 protein coding genes for wild-type (**A**) and ΔE(z) cells (**B**)
Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Genes are listed in plastom order starting with gene *CMV001C* allowing localization of plastomic regions. Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019). Note that the response to the priming stimulus is similar to priming-and-triggering response. Triggering-only does not lead to similar dimensions in transcriptional response. *ΔE(z)* fails to reach WT priming answer, whereas the triggering-only and priming-and-triggering responses are comparable to WT behavior. Three plastomic regions seem to be commonly regulated and are blue boxed.

Overall, a strong upregulation of the expression for the majority of chloroplastic genes could be observed upon priming heat (PvsN) and its recurrence (PLTvsPL). Yet, the triggering-only response (TvsNL) lead to sparse reactions only. This is true for both, WT and *ΔE(z)* (Figure 3.14: B). Interestingly, similar to the weaker priming responses of the entire transcriptome in *ΔE(z)* (Figure 3.7), the priming answer in the chloroplast genome of *ΔE(z)* did not fully reach

WT levels. Following the observation of such an enormous reaction to heat of the chloroplast genome, it was of interest to determine if this regulation is regional, i.e. operon-based, or whether certain functions are specifically driven. Thus, to allow for conclusions about regional regulation, genes were ordered corresponding to the plastom sequence defined by Ohta *et al.* 2003 in Figure 3.14. Remarkable are three plastomic regions (Figure 3.14, boxed regions) which displayed strong regulation during the HSMA. However, as these boxed regions only partially correspond to operons defined for the *C. merolae* plastom (Ohta *et al.*, 2003) it is not clear whether heat responses in the chloroplast are a consequence of operon-regulation. Comparing WT and *ΔE(z)* expression strengths in these three regions, the mutant showed weaker induction upon priming (PvsN), but the primed-and-triggered condition (PLTvsPL) was comparable between WT and *ΔE(z)*.

As the chloroplast genome regulation as response to heat does not seem to be conferred on the level of operon structures (only), it was asked whether gene expression was specifically driven by function instead. To allow for an initial characterization of genes within these regions, functional annotation of all chloroplast genes was included in the heatmap. Ohta *et al.* (2003) sequenced and characterized the *C. merolae* plastom and divided the plastid genes into functional classes: genetic system upkeeping, metabolism, photosynthesis, ATP synthesis, transport and RNA genes as well as a group of genes encoded in the chloroplast with uncharacterized functions to date (= Unknown). Examination of functional classification underneath the commonly regulated chloroplast genome regions revealed that genes of metabolism, transport, photosystems and genetic system are included in these three plastomic regions (Figure 3.14, boxes). However, the broadest region comprised genes involved in the maintenance of the genetic system.

Table 3.3: Distribution of identified chloroplast memory genes (n = 50) across functional classes of *C. merolae* plastid genes *

Functional Class in *C. merolae* chloroplast genome	total number of genes in this class*	memory genes within**	memory gene proportion in this class
Genetic system	61	23	38%
Photosystems	27	11	41%
ATP synthesis	8	2	25%
Metabolism	33	4	12%
Transport	9	2	22%
Unknown function	39	8	21%
RNA genes	36	0	0%

* defined by Ohta *et al.*, 2003, ** identified in this project (Chapter 3.3.2)

To further pursue the assumption of tailored functional responses of the chloroplast to heat, the 50 memory genes originating from the chloroplast were allocated to the seven functional

classes. Consequently, the 50 chloroplast memory genes were distributed to six out of the defined seven classes (Table 3.3). The absence of any memory genes within the functional class of RNA genes is due to the exclusion of RNA genes from the current annotation file. Thus, memory events in the chloroplast were distributed across all functional classes, however particular gene functional classes were affected more than others. Here, the highest proportion was found for the photosystem genes where 41 % of all photosystem genes showed trainability. The genetic system represented the second most affected class with 38% of all genetic system genes showing trainability, repeatedly drawing attention to this functional class of chloroplast genes. Metabolic genes were affected the least (Table 3.3). Next, to confer potential roles of particular chloroplast functions in the memory response of *C. merolae*, each functional class and its response to the HSMA was investigated individually.

3.4.1 Heat primes the chloroplast genetic system including transcriptional and translational regulators

Separating chloroplast genes by their functional class and examining their expression during the heat stress scheme, revealed that approx. 2/3 of all genes encoding for proteins of the chloroplast *genetic system* displayed trainability, reflected in higher PLTvsPL compared to TvsNL induction (Figure 3.15 A) .

Here, especially regulators of transcription and translation are affected. Thus, all four subunits of the chloroplast encoded DNA-directed RNA polymerase (PEP) are regulated by heat (Figure 3.16 A, E). However, only the subunit *rpoC* (*CMV217C*) exhibited significantly different expression in primed-before (PLT) cultures compared to triggered-only cultures and consequently was deemed a memory gene in the PLTvsT analysis (3.3.2). In the *ΔE(z)* mutant, in addition to *rpoC*, *rpoB* was defined as significantly higher induced in PLTvsT posing an additional memory gene in the mutant. The strong negative L2FC values in the *ΔE(z)* mutant upon triggering for the *rpoZ* subunit are due to very low amounts of reads obtained, which does not exceed 12 counts (normalized counts = 2.3 reads). By the application of lfc shrinkage (DESeq2 package) these fold changes are balanced and reveal no statistically significant differences between WT and mutantAs the *rpoZ* protein is very small (150 bp, 49 aa), it is technically challenging to capture a precise picture of its expression via RNA-sequencing. The potential difference in its transcriptional regulation upon triggering should be validated using biochemical detection experiments. In order to perform as polymerase, the PEP demands sigma factors (SIG), which provide promoter specificity to the PEP holoenzyme eventually leading to targeted gene expression (Tadini *et al.*, 2020).

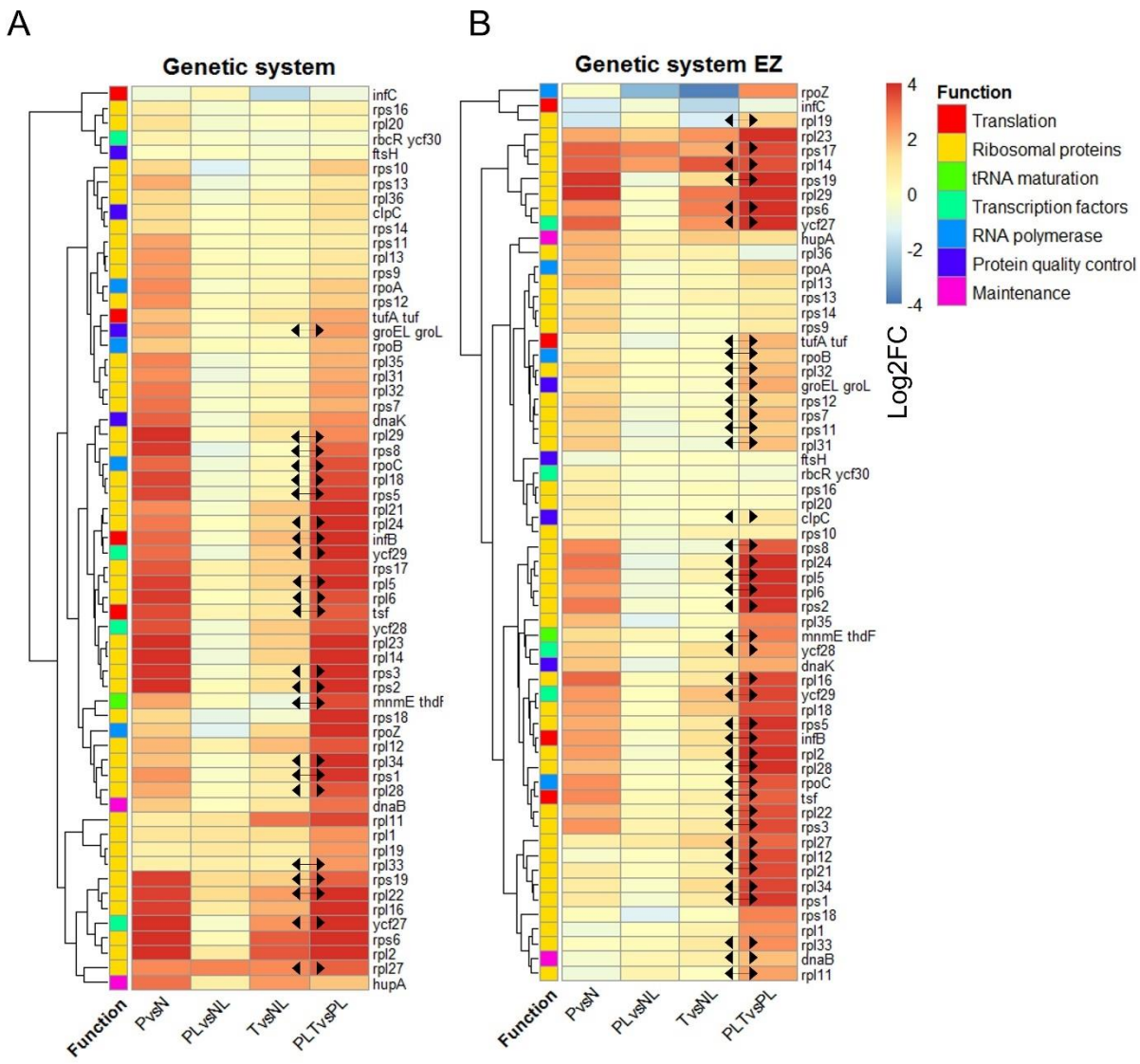


Figure 3.15: The chloroplast genetic system is vastly influenced by recurring heat stress in *Cyanidioschyzon merolae*

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL.
Genes maintaining the genetic system in the chloroplast of *C. merolae* wild type (WT, **A**) and *ΔE(z)* (**B**) cells are shown. Note that 22 and 40 out of 61 genes show memory in WT and *ΔE(z)*, respectively. Triggering-only does not lead to similar dimensions found in PLTvsPL transcriptional response. The priming response in *ΔE(z)* is considerably weaker compared to WT.

Four PEP sigma factors are described in the *C. merolae* genome: SIG1, SIG2, SIG3, SIG4 encoded by *CMK044C, CMQ213C, CMR165C*, and *CMM072C*, respectively. Figure 3.16B shows the expression changes in all HSMA conditions of all four SIGs, suggesting *SIG2* and *SIG3* being induced by heat whereas *SIG4* was repressed. In the *ΔE(z)* mutant, *SIG1* was

induced by priming heat additionally (Figure 3.16F) which might be explained by the direct neighbourhood of an pronounced H3K27me3 peak upstream of the SIG1 encoding gene *CMK044C* in WT (Figure S3).

Surprisingly, no considerable regulation of neither sigma factor encoding gene is detected upon the triggering stimulus, neither in TvsNL nor in PLTvsPL. This is true for both genotypes. Together, these results suggest *SIG3* responds strongest to heat which possibly together with the as well heat-regulated PEP might entail specific downstream gene regulation. The actual interaction between PEP and *SIG3* together with the identification of potential downstream targets remains subject to future investigations. Likewise, whether the higher expression levels of all PEP genes is reflected in elevated and functional PEP protein is unclear to date.

Besides SIGs, transcription factors ensure specific gene expression responses. In the *C. merolae* chloroplast, four transcriptional activators are present: *ycf27, ycf28, ycf29, ycf30*. Of those four are *ycf27* and *ycf29* defined as memory genes (Table S2). These two code for OmpR-like transcriptional regulators. While *ycf29* remains undescribed, *ycf27* is comprises a promoter specific protein mediating the interaction between DNA and RNA polymerase (uniport.org). However, concerning their biological function, *ycf27* and *ycf29* are both uncharacterized to date. Both show strong trainability in response to heat in both, WT and *ΔE(z)* (Figure 3.16 C,G) and thus might contribute to the revised transcriptional response of the chloroplast upon repetitive heat.

While *ycf28*, an ntcA-like transcriptional factor, is significantly regulated in PLTvsPL in both, WT (L2FC = 4, padj =7.5 $*10^{-9}$) and *ΔE(z)* (L2FC = 3.5, padj = 0.004), it showed differential expression between PLT and T directly solely in the mutant *ΔE(z)* (L2FC = 1.6, padj = 0.0046). *Ycf28* traditionally regulates the nitrogen metabolism in Cyanobacteria by activating major assimilatory proteins, like nitrogen reductase (NR) and nitrogen transporters (NTR, (Herrero *et al.*, 2001, 2004). Similar to the insignificant upregulation detected for *ycf28* in PLT no regulation of nitrogen assimilatory genes was detected (i.e. *CMG018C* - NTR, *CMG019C* - NR, *CMK129C* – nitrate ABC transporter, *CMO283C* - similar to nitrate permease). Thus a potential role of nitrogen assimilatory genes in the heat response is considered unlikely.

The last transcription factor, *ycf30*, in *C. merolae* is characterized as nuclear-independent driving force of the RuBisCo-operon in response to light (Minoda *et al.*, 2010). Similar to no obvious changes in the expression of RuBisCo subunits, i.e. *rbcL* and *rbcS*, *ycf30*, too, seems to be entirely unaffected by heat stress.

In sum, 2 out of 4 transcriptional regulators of the chloroplast seem to shape the re-occuring heat response in *C. merolae.* This happens independently of E(z) action and most probably in turn controls several downstream actors and responders. Genetic studies might reveal the importance of *ycf27* and *ycf29* in this mechanism.

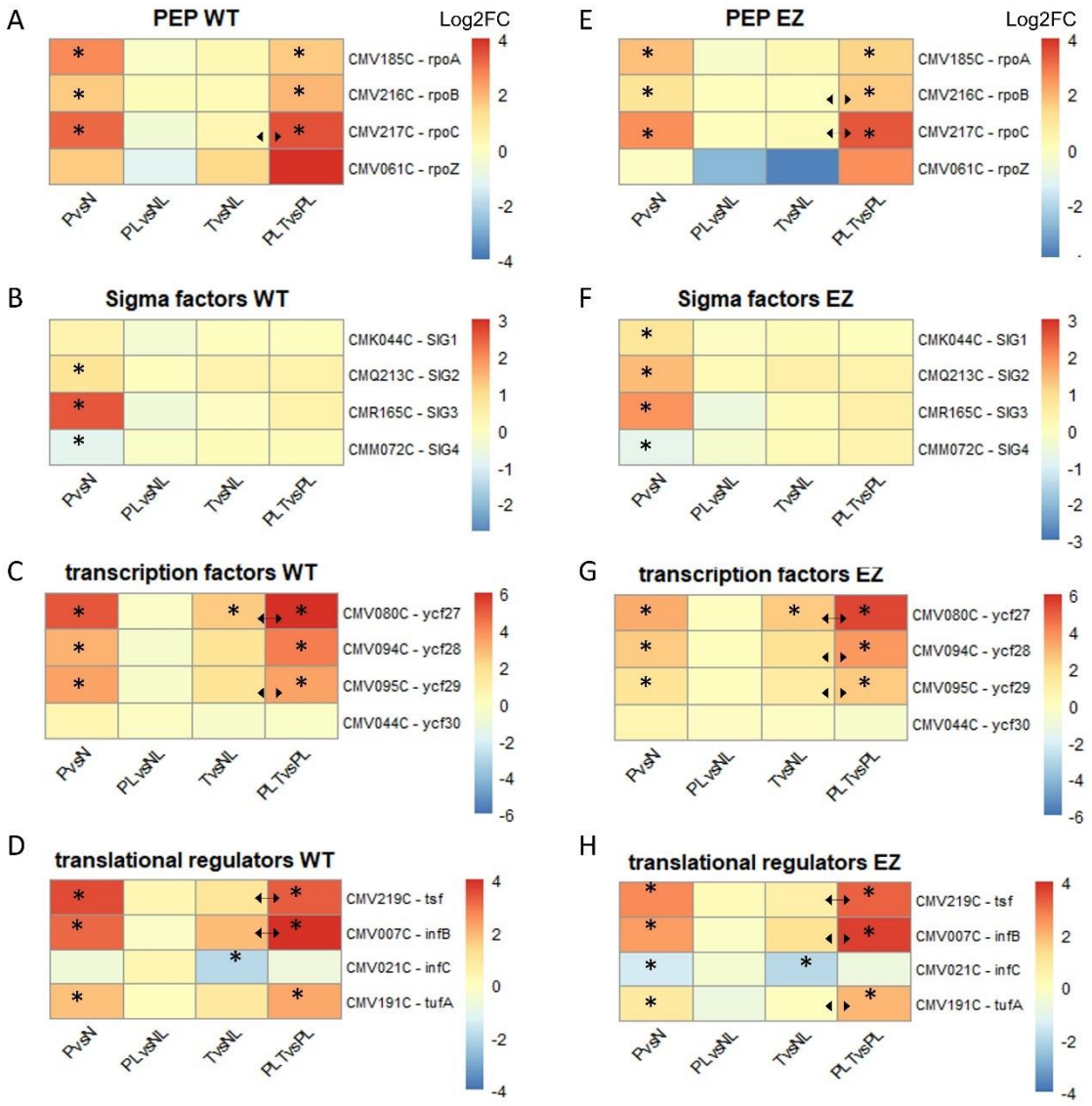


Figure 3.16: Transcriptional responses of major regulators in the genetic system of the *Cyanidioschyzon merolae* chloroplast

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Asterisks indicate significant regulation in corresponding pairwise comparison (padj < 0.05, Log2FC > |1.5|). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL.

A, E: transcriptional changes of plastid encoded RNA-directed DNA polymerase (PEP) subunit encoding genes during HSMA in WT and *ΔE(z)* cells.

B, F: transcriptional changes of nuclear encoded sigma factor (SIG) genes during HSMA.

C, G: transcriptional changes of chloroplast genome encoded transcription factors ycf 27-ycf30.

D, H: transcriptional changes of chloroplast encoded regulators of translation *tsf, infB, infC, tufA.*

(ycf: chloroplast open reading frame, rpoA: RNA polymerase alpha subunit, rpoB: RNA polymerase beta subunit, rpoC: - RNA polymerase beta´ subunit, rpoZ: - RNA polymerase omega subunit, SIG: sigma factor, tsf: elongation factor Ts, infB: – translation initiation factor IF-2, infC: translation initiation factor IF-3, tufA: elongation factor Tu)

After transcription, protein synthesis requires translation of transcribed mRNAs. The tight regulation of translation can impact the transcription reversely and should not be underestimated. Amongst others, translation factor activity as well as the alternate composition of ribosome components (reviewed in Xue & Barna, 2012) shape translation. While all four translational regulators existent in the *C. merolae* plastom are regulated by heat, only two show memory behaviour (Figure 3.16 D,H). Only *infB* and *tsf* react in a memory like manner (Table S2) with significant induction in PLTvsT. However, *tufA* is induced in priming (PvsN L2FC = 1.07, padj = 0.008) and priming-and-triggering (PLTvsPL L2FC = 1.9, padj = 2,2 $*10^{-9}$), but not upon triggering (TvsNL, L2FC = -0.03, padj = 0.89) suggesting memory-like behaviour. The difference in expression between PLT and T remains insignificant, though (PLTvsT: L2FC = 0.06, padj = 0.3). The fourths, *infC* (initiation factor 3, *CMV021C*) is repressed during heat (e.g. WT_TvsNL: L2FC = -1.7 , padj = $6*10^{-8}$).
The relevance of translational regulation in the process of heat memory is underlined by the fact, that 15 out of 46 genes encoding for the large (*rpl* genes) and small (*rps* genes) subunits of ribosomes exceed the significance threshold for differential expression between PLT and T (Table S2). Additionally, 20 more show the same behaviour in trend (Figure 3.15A) suggesting the activation of a specific set of ribosomal subunits upon heat occurrences. Interestingly, among trainable ribosome protein encoding genes was *rps1*, the ortholog of RPS1, which was recently identified as relevant in heat stress responses in *Arabidopsis* (H.-D. Yu et al., 2012).

3.4.2 Components of photosystem I and II show transcriptional memory of heat

Around half of the photosystem maintaining genes is regulated by heat (Figure 3.17). Remarkably, all genes encoding for the reaction centres of both, photosystem I as well as photosystem II, show trainable transcriptional patterns, i.e. stronger induction upon the triggering stimulus when they were primed before. Thus, *psaA*, *psaB*, *psbA* and *psbB* are significantly higher induced in PLTvsPL than in TvsNL. Interestingly, *psbA* is found among misregulated genes in the *ΔE(z)* mutant, exhibiting 1.8-fold induction (L2FC = 0.85, padj = $1.7*10^{-8}$, EZvsWT) upon the absence of H3K27me3 under control conditions. While both, PSI and PSII are reacting to heat in a trainable manner, the light harvesting machinery including phycobilisomes is barely altered. Only *ycf44*, encoding for a cytochrome biogenesis protein (*CMV075C)* and *cpcB* (phycocyanin (PC) beta chain) show memory behaviour (Figure 3.17 and Figure S4 A). However, when comparing the fold changes during the HSMA for all three phycocyanin genes (*CMV063C* – rod core linker, *CMV064C* – PC alpha chain, *CMV051C* –

PC beta chain) it is evident, that all of them are similarly regulated: induced in heat and even stronger induced, when primed before (Figure S4 A) suggesting an elevated need of light collecting phycobilisome proteins to realize a benefit from priming.

The fact that *C. merolae* possesses an thioredoxin gene in the chloroplast genome is outstanding since, opposing to the situation in red algae, in land plants this gene has migrated into the nucleus (Reynolds *et al.*, 1994). Chloroplast thioredoxin (*trxM*, *CMV012C*) is barely expressed under control conditions, but induced strongly in priming (L2FC = 3.8 , padj =3.7 $*10^{-6}$) and in the same way its expression is elevated during triggering (L2FC = 1.2, padj = 0.0004). However, the induction upon triggering is boosted, when priming occurred earlier (PLTvsPL, L2FC = 6.24, padj = 4.7 $*10^{-11}$). This pattern is found exclusively for the chloroplast thioredoxin (*CMV012C*); no considerable regulation is seen for any other nuclear protein similar to thioredoxin (Figure S4B). Thus, *CMJ096C* is induced by priming (L2FC = 1.5 , padj = 8.9 $*10^{-11}$) but not upon triggering (PLTvsPL: L2FC = 0.23, padj = 0.2; TvsNL: L2FC = 0.12, padj = 0.3). When thioredoxins reduce oxidized cysteine residues and/or cleave disulfide bonds, the corresponding enzyme itself is oxidized. Thioredoxin reductases keep the enzyme in its active state in a NADPH- or ferredoxin-dependent manner. Since it was possible that thioredoxin reductases could be similarly regulated by heat, the transcriptional patterns of those proteins was investigated during the HSMA (Figure S4).

The chloroplast-encoded thioredoxin reductase (*CMV226C*) is not considerably strong expressed (read counts of max. 111), but is induced upon priming (L2FC = 1.2, padj = 0.002). Furthermore, the nuclear-encoded thioredoxin reductase *CMP058C* is significantly upregulated after priming (L2FC = 1.4 , padj = $5.28*10^{-15}$) at high transcription rates (approx. 1000 reads), whereas no induction is detected upon the triggering stimulus, neither in TvsNL nor in PLTvsPL. The higher induction and expression of the nuclear thioredoxin reductase while nuclear thioredoxin transcription is not substantially altered, suggests the possibility of chloroplast-nuclear crosstalk where the chloroplast thioredoxin needs the nuclear thioredoxin reductase in order to fulfil its reductive function on heat-damaged proteins and/or in response to heat-generated cellular reactive oxygen species (ROS). The existence of transit peptides of the nuclear reductase to the chloroplast or *vice versa* has to be investigated in order to speculate about the interaction of those two components. It can be concluded that a type of oxidative stress co-occurs with heat stress, that it is probably met in the chloroplast and contributes to the PLT over T benefit.

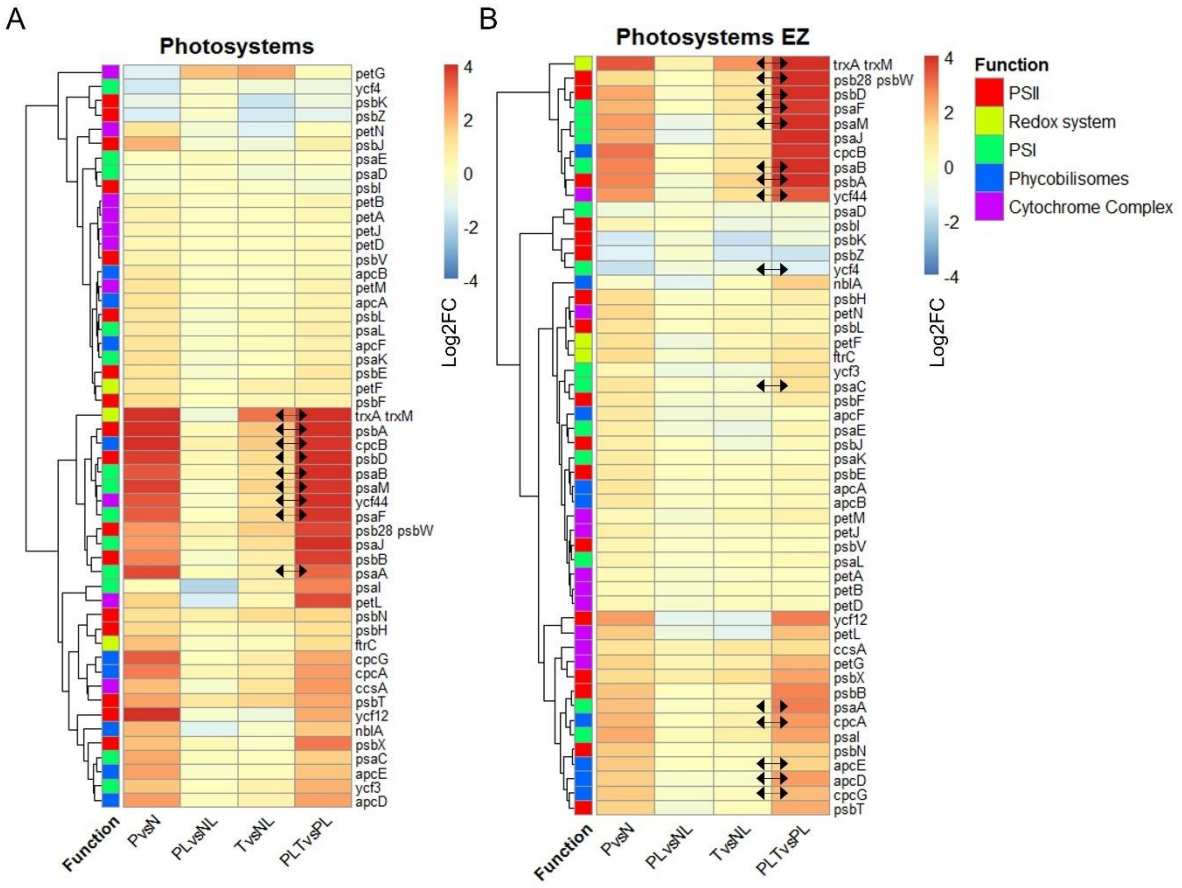


Figure 3.17: Photosystem (PS) I and II genes are transcriptionally affected by recurring heat
Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL.
Shown is the expression changes upon HSMA in WT (A) and *ΔE(z)* (B) of all genes classified as relevant to upkeep the photosystem machinery and reveal transcriptional responses of PSI and PSII components.
A In WT, 10 out of 27 genes show memory. Triggering-only response shows lack of induction of photosystem genes.
B Priming response is weaker in *ΔE(z)* compared to WT, but the number of memory genes in photosystem genes increased to 15. (Log2FC: log2 fold-change)

3.4.3 Chloroplast ATP synthase reacts to repetitive heat

In addition to the necessity to protect the cell against oxidative stress comes the cost of protection: the cell needs energy in form of the energy storage molecule ATP (adenosine triphosphate) in order to fulfil the cells energetic needs (not only) during the heat stress response. These defence strategies involve for instance the important action of heat shock proteins as ATP-dependent chaperones (Niforou *et al.*, 2014). The ATP synthase is

composed of an hydrophobic, membrane bound F_O and a water soluble F_1 multi-subunit complex, the latter one sticking out of the membrane into the chloroplast stroma and converts ADP to ATP by the addition of a inorganic phosphate P_i and the proton motive force generated in the thylakoid lumen. In the *C. merolae* thermomemory, constituents of both, F_O and F_1 subunit, are (a) affected by heat and (b) show memory behaviour meaning that the ATP supply might be enhanced during triggering stress occurrence when *C. merolae* was primed before (Figure 3.18A). The two genes designated as memory genes (Figure 3.8C, Table S2), *atpB* and *atpD* (ATP synthase F_1 delta chain) encode both for subunits of the F_1 subcomplex and show somewhat different expression patterns: Both react strongly to the priming stimulus (PvsN) and their expression is enhanced upon recurring heat (PLTvsPL). Although only slightly, *atpB* is induced upon triggering (L2FC = 0.39, padj = 0.03), but *atpD* seems to be non-responding (L2FC = -0.15, padj = 0.46). If the differential expression of the ATP synthase subunits is of functional relevance or the answer to local damage via heat stress is subject to future research. The implied primability of the entire ATP synthase shown for WT is pronounced in the *ΔE(z)* mutant. Thus, all but one subunit are significantly higher expressed upon triggering, when they have experienced priming before (PLTvsT). Since the cell survival of *ΔE(z)* cells is affected (Chapter 3.2.2) this could signify an elevated energy need of the mutant, rather than a better reaction to heat.

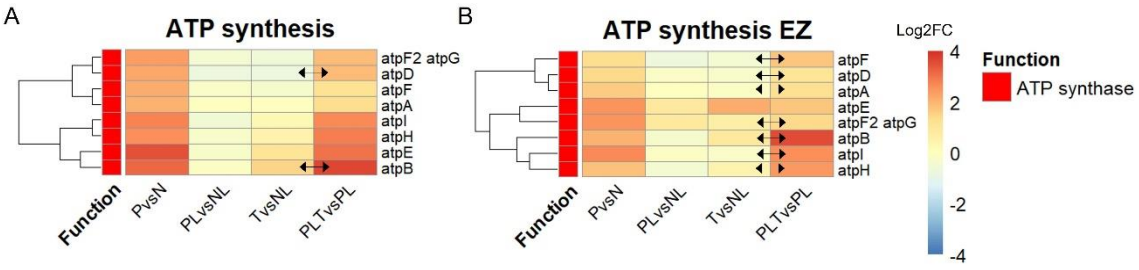


Figure 3.18: Stronger priming effect on transcription of ATP synthesis machinery in the Polycomb mutant *ΔE(z)* compared to wild type (WT) *Cyanidioschyzon merolae*

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL. The expression changes of ATP synthase component encoding genes upon the heat stress memory scheme is depicted in **A** for wild type (WT) and in **B** for the *ΔE(z)* strain.

A Only *atpD* and *atpB* are significantly expressed between PLT and T directly. Tiggering-only response shows lack of induction of ATP synthesis genes.

B All but one (*atpE*) ATP synthase subunit encoding genes are primed in the PRC2 mutant *ΔE(z)* suggesting elevated transcript levels upon repetitive heat.

3.4.4 Activated pre-protein secretory pathway and ABC transporter upon heat priming scheme

The plastid genome comprises genetic information for several active and passive transport mechanisms for proteins and metabolites. In the HSMA, components of the Twin-Arginin-transport (TAT) (*tatC*) as well as sulphur transporters (*cysW, cysT*) remain unresponsive in both, WT and *ΔE(z)* (Figure 3.19). These are possibly only relevant in late or secondary responses to heat, since they bind to and actively transport folded proteins only (uniport.org). Contrarily, the transport system of the secretory (sec) pathway, which is ATP dependent and allows the transport of nascent proteins, seems to be affected by heat. One of the secretory pathways genes, the preprotein translocase subunit *secA,* is extremely heat responsive showing nearly 20-fold induction upon priming (L2FC = 4.32, padj = 2.2 *10^{-15}).

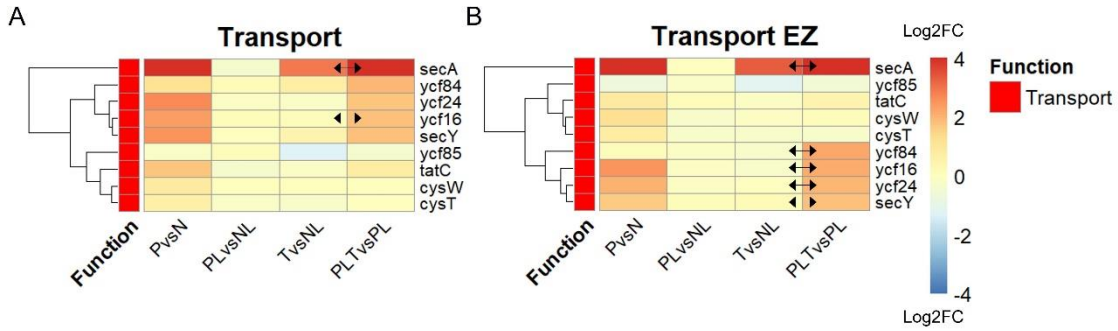


Figure 3.19: Heat primes the *Cyanidioschyzon merolae* secretory pathway relevant for the transport of nascent proteins

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL. Expression changes of genes involved in chloroplast transport mechanisms for wild type (WT, **A**) and *ΔE(z)* (**B**) strain are shown. **A** In WT, only *secA* and *ycf16* are memory genes. **B** In *ΔE(z)* cells memory is effect is enhanced showing primed secretory (*secA*, *secY*), permease (*ycf84*) and sulfur-iron cluster transport.
(*secA*: chloroplast preprotein translocase SecA subunit, *ycf84*: permease, *ycf24*: ABC transporter for iron-sulfur cluster formation sufB, *ycf16*: sufC ATP binding subunit of ABC transporter for iron-sulfur cluster formation, *ycf85*: probable ABC transporter ATP-binding protein YhbG homolog, *tatC*: sec-independent protein translocase, *cysW*: sulfate transport system permease protein)

Together with *secY*, encoding the membrane bound translocase domain, these two gene products form a preprotein secretion transport system. Like *secA*, *secY* is upregulated upon priming (L2FC = 2.1, padj = 2.1 *10^{-6}), albeit not as strong as *secA*. The difference in induction strength may be explained by the proteins location within the cell. While secY is a membrane bound complex, secA is located in the cytoplasm in order to bind transit peptides

on (nascent) proteins which then are translocated through the membrane by the interaction of secA and secY. Although both genes are stronger upregulated upon triggering when primed before (PLTvsPL) compared to non-triggered (TvsNL), only *secA* is trainable in WT. Similar to the situation of the photosystem machinery, the heat response in the *ΔE(z)* mutant is more pronounced since *secY* gains trainability (Figure 3.19B). The same is true for the passive transport conducting permease gene *ycf84* as well as the ABC transporters of sulfur-iron clusters (*ycf16*, *ycf24*) which show increased differences in *ΔE(z)* PLTvsT expression compared to WT levels, where only the ABC transporter gene *ycf16* is trainable. The more pronounced regulation of these components in the mutant could either be due to loss of nuclear repressive H3K27me3 mark which could subsequently affect downstream chloroplast transcriptional regulation. Another possibility could be given by the enhanced need of (protein) transport due to elevated damage by heat in *ΔE(z)*.
Taken together, transcriptomic analysis of the chloroplast transport system implies that the secretory pathway of nascent proteins as well as the ABC transporter subunit for iron-sulfur clusters potentially contribute to the heat stress memory in *C. merolae*.

3.4.5 Hesitant chloroplast metabolic pathway reaction to heat

Compared to the high numbers of affected genetic system genes reacting to the HSMA, the effect on metabolic genes is rather small (Figure 3.20). The genes reacting in the strongest fashion to priming and triggering include *trpG*, *thiG*, *ilvB*, *trpA*, *ilvH* and *carA*.
trpA (*CMV006C*) and *trpG* encode for enzymes involved in the tryptophan biosynthesis pathway: tryptophan and anthranilate synthase, respectively. Anthranilate synthase synthesizes one of the first intermediate products in tryptophan biosynthesis pathway (Maeda & Dudareva, 2012). It catalyzes the production of anthranilate, which in turn is an important intermediate product in the biosynthesis of indole. The two anthranilate synthase subunits are distributed to chloroplast (*trpG*, *CMV076C*, β subunit) and nuclear genome (*trpE*, *CMN228C*, α subunit). *trpG* is a chloroplast memory gene strongly reacting to heat re-occurrence (Figure 3.20). Contrarily, *trpE* shows no trainability, but induction upon priming (L2FC = 0.95, padj = 7.2 $*10^{-8}$). Sharing an operon with *trpG, thiG* (thiamine biosynthesis protein G, *CMV077C*) shows identical induction patterns. It remains unclear, which one of these two metabolic enzymes sharing an operon might be specifically activated upon heat stress. A cis-activation from one to the other is a possibility.

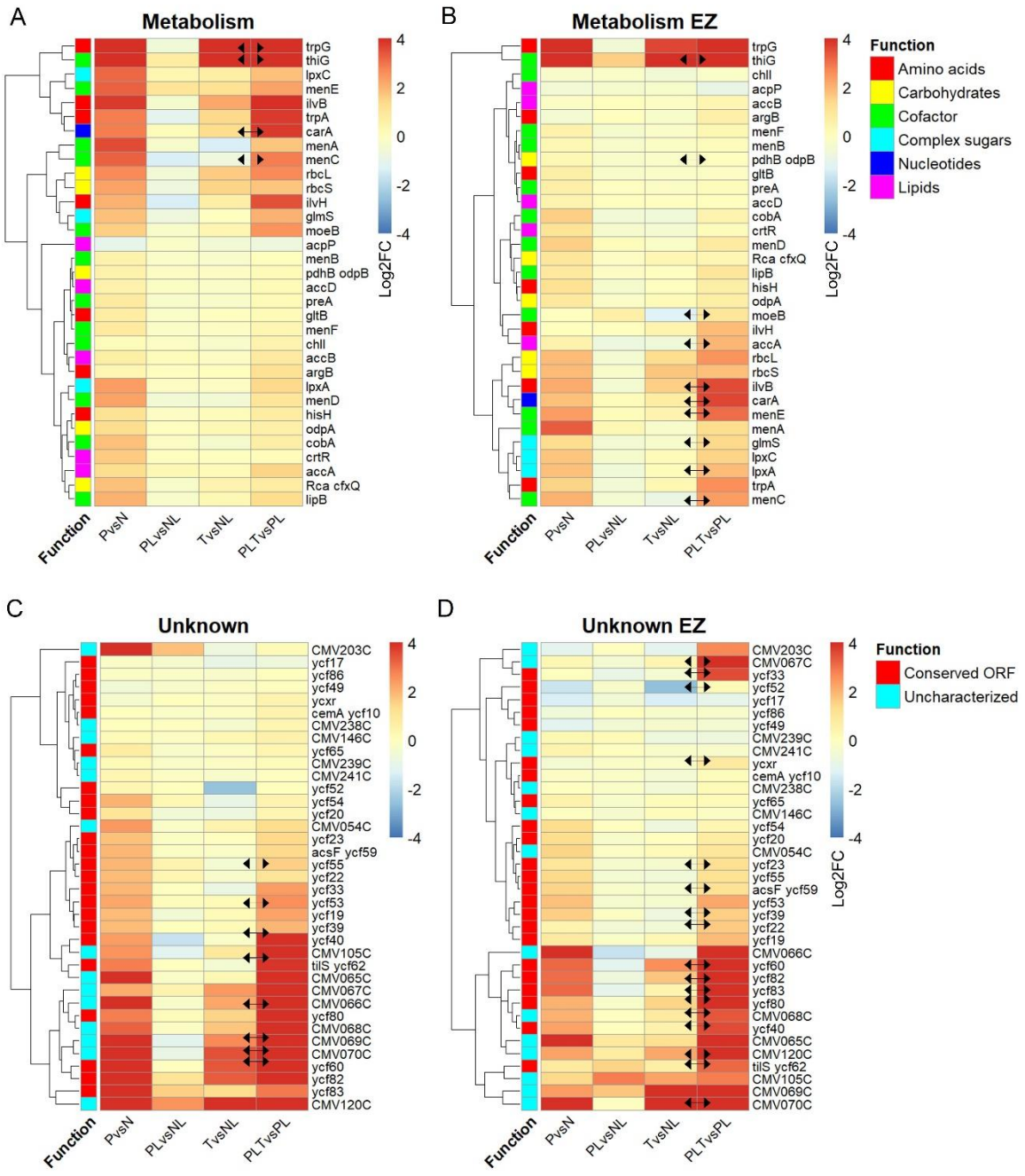


Figure 3.20: Restricted response of metabolic genes to repetitive heat stress in *Cyanidioschyzon merolae*

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL. **A** Expression changes of genes relevant for chloroplast metabolism in WT. Only four genes in metabolism are trainable, meaning differential expression between PLT and T directly. **B** Expression changes of genes relevant for chloroplast metabolism in *ΔE(z)* mutant reveal 10 memory genes. **C** Nearly half of the genes of unknown function encoded in the chloroplast of *C. merolae* WT cells react to recurring heat including a commonly regulated region ranging from CMV065C – CMV070. Many genes coding for proteins of unknown function are regulated by heat and its recurrence.
D Expression changes of metabolic genes in the *ΔE(z)* mutant.

Tryptophan synthase catalyses the last two steps in the tryptophan biosynthesis pathway. Its alpha chain encoding gene (*trpA*) is activated by heat and even stronger by recurring heat in a memory like behaviour. The tryptophan synthase beta chain (*trpB*) is nuclear encoded by *CMJ191C* and slightly induced by priming (L2FC = 0.41, padj = 0.015), but does not show trainability upon recurring heat (PLTvsT, L2FC = 0.03, padj = 0.8) (Figure S4 C). A high discrepancy exists for the expression strength for both subunits: albeit *trpA* follows a trainable expression pattern, the transcript counts do barely exceed 100 reads, whereas *trpB* does not show a change in expression upon recurring heat, but exhibits much higher basal transcription rates (approx. 3000 counts). As tryptophan synthase possesses two catalytic capabilities, namely (1) conversion of indole-3- glycerol phosphate (IGP) to glyceraldehyde-3-phosphate (G3P) and indole (α subunit) and (2) condensation of indole and serine to tryptophan, the differential expression and regulation upon heat might indicate towards step (1), the specialized upregulation of G3P and indole production.

The enzyme acetolactate synthase catalyses the biosynthesis of acetolactate, which is a precursor of the amino acids valine, leucine and isoleucine. Its small and large subunits are encoded by *ilvH* (*CMV018C*) and *ilvB* (*CMV122C*), respectively (Figure S4 D). Albeit not identified as memory genes, both subunits show high induction rates upon triggering in primed-before cultures (PLTvsPL; *ilvB*: L2FC = 4.3 , padj = 5.9 $*10^{-11}$; *ilvH*: L2FC = 0.89, padj = 0.01). Similar to the tryptophan synthase, the expression strengths of the two subunits differ strongly, in this case by factor 25. Accordingly, *ilvB* reaches normalized transcription rates of approx. 500 while the highest observed read count does not exceed 20 for *ilvH*. The differential expression strength and changes might be explained by the specialized functions of each subunit. Thus, the large subunit is responsible for the catalysis of pyruvate to acetolactate and the small subunit provides feedback inhibition. The transcriptomic results presented here suggest a memory-dependent induction of both subunits.

carA (*CMV036C*) shows strong trainability and encodes for the small subunit of carbamoyl phosphate synthetase (Figure S4 E). This cytoplasmic enzyme is involved in arginine and pyrimidine synthesis. Its expression strength is close to zero under control conditions but is elevated during priming to approx. 300 counts. In comparison, its large subunit, encoded in the nucleus by *carB* (*CML055C*), reached 10-fold higher transcript levels (basal levels around 3250 reads), whereas it is not induced by heat. This could, again, be explained by the differential function each subunits exerts. While the large subunit contains two active sites, producing carboxy phosphate and carbamoyl phosphate, the small subunit catalyses the deamination of glutamine to ammonia and glutamate. This suggests, that upon heat stress the production of ammonia and glutamate might represent an important answer.

Next to these strongly induced gene candidates the group of *men* genes is regulated to a

lesser extent. The *men* genes encode for enzymes of the menaquinone pathway. Menaquinones (or naphthaquinones) are a type of isoprenoid lipoquinones found in the electron transport chain linking this metabolic pathway to photosynthesis (Meganathan & Kwon, 2009). Four out of six *men* genes react to heat and three of them are re-induced in a trainable manner also upon heat reoccurrence (PLTvsPL) whereas no induction is detected in triggering only (TvsNL) (Figure S4 F). The latter three include *menA* (*CMV116*C, 1,4-dihydroxy 2-naphthoate octaprenyltransferase), *menC* (*CMV117C*, 4-(2'-carboxyphenyl)-4-oxybutyric acid synthase), and *menE* (*CMV118C*, o-succinylbenzoic acid-coA ligase). Although highly and heat-dependently induced, all six *men* genes unveil low transcription levels (max. 100 reads) questioning the biological relevance of this gene activation. Nevertheless, as the photosystem genes exhibit trainable behaviour, too, the directed maintenance of the electron transport might undermine the relevance of the photosystem machinery in the formation of thermomemory in *C. merolae*.

In conclusion, heat induces the metabolism of amino acid and nucleotide biosynthesis as well as of cofactors of the electron transfer chain components in the chloroplast. The metabolism of lipids, complex sugars and carbohydrates instead is nearly unaffected (RuBisCo genes *rbcL* and *rbcS*). Whether this transcriptional upregulation is due to elevated repair demands upon heat acclimation or rather is intrinsically regulated remains to be elucidated in future experiments.

Lastly, among the remaining chloroplast genome genes with no ascribed function yet (functional class: Unknown) are many highly induced trainable genes (Figure 3.20C) which repetitively exhibit even stronger memory responses in the *ΔE(z)* mutant (Figure 3.20D). Interestingly, there is another commonly regulated memory region stretching from the genes *CMV065C* to *CMV070C* which are entirely uncharacterized to date (upper part of boxed region 1 in Figure 3.14). Unravelling the function and annotation of those uncharacterized genes bears further potential to understand the mechanisms underlying the *C. merolae* thermomemory which is either generated, accomplished or maintained (?) in the chloroplast of *C. merolae*.

4 Discussion

Polycomb repressive complex 2 (PRC2) has been previously shown to silence genes and repetitive elements in eukaryotes. Hence, PRC2 and its repressive histone mark H3K27me3 substantially contribute to epigenetic regulation of developmental processes in higher eukaryotes. The functional consequences of PRC2-mediated silencing are well described in multicellular species, and PRC2 orthologues have been identified in various unicellular and primitive organisms, including ciliates, filamentous fungi, diatoms, and green and red algae. However, the relevance of PRC2 in unicellular species is less understood. Recent studies suggest, that PcG proteins preferentially repress repetitive elements in more primitive organisms while in more complex organisms PRC2 favours genes as targets (Schubert, 2019; Vijayanathan *et al.*, 2022). Yet, this shift likely happened gradually and studies concerning PRC2 function in primitive non-seed plant systems might significantly contribute to the understanding of genome regulation in the course of land plant evolution. The red alga *Cyanidioschyzon merolae* (*C. merolae*) represents an excellent model system to study the function of the major chromatin regulator PRC2 in a unicellular photoautotrophic eukaryote, since PRC2 existence and activity was determined previously (Mikulski *et al.*, 2017). The present study contributes to the sparse knowledge of epigenetics in unicellular photoauthotrophic cells by elucidating potential functions of PRC2 and its histone mark H3K27me3 in *C. merolae*. This is achieved by examining the ability of the H3K27me3-depleted *C. merolae* strain Δ*E(z)* to silence repetitive elements and maintain genome stability as well as Δ*E(z)*-performance in repetitive heat stress situations. Dissecting the roles of major chromatin regulators such as PRC2 in unicellular plant organisms will substantially contribute to understand the archaic functions of epigenetics, especially in scope of the green lineage evolution.

4.1 PRC2 regulates the transcription of repetitive elements and, to a lesser extent, the transcription of genes

4.1.1 PRC2 transcriptional regulation beyond H3K27me3 targeting

Former work from our lab on *C. merolae* described H3K27me3 to target both, genes and repetitive elements and by analysing the wild type transcriptome the authors concluded, that H3K27me3 occupancy coincides with lower transcription rates (Mikulski *et al.*, 2017). To draw conclusions about a direct regulatory function of H3K27me3 in silencing its targeted genomic elements this study analysed the transcriptome of the mutant strain Δ*E(z)*, which is devoid of genomic H3K27me3 levels. Comparison of WT and Δ*E(z)* transcriptome revealed that both, expression rates of genes and repetitive elements are regulated by H3K27me3 (Figure 3.1 A) which is in accordance with the finding of H3K27me3 enrichment on genes and repetitive elements (Mikulski *et al.*, 2017). The affected H3K27me3 targets were nearly entirely upregulated in the mutant which is in agreement with what is known for many organisms: H3K27me3 correlates with transcriptional quiescence (Mozgova *et al.*, 2015; Schuettengruber *et al.*, 2007). However, not all genes decorated with H3K27me3 in WT were de-regulated in the mutant, suggesting that the mere absence of H3K27me3 does not suffice to regulate (activate) its targets. Similarly, studies in human, *Drosophila*, *Neurospora crassa* and *Arabidopsis* reported only fractional upregulation of H3K27me3 targets upon its absence and the need of additional, activating factors are discussed as second regulatory layer (Bouyer *et al.*, 2011; Bracken *et al.*, 2006; Jamieson *et al.*, 2013; Kirmizis *et al.*, 2004; Lafos *et al.*, 2011; Schwartz *et al.*, 2010). In addition to transcriptional activators, and developmental or environmental cues, an appropriate chromatin environment might be needed to successfully start transcription including H3K4me3, a histone mark set by Trithorax group regulators and generally associated with transcriptional activity (Kouzarides, 2007). It will be interesting to test, how the constitution of chromatin modifications changes in response to depleted H3K27me3 levels. Interestingly, increased H3K4me3 and H4 acetylation levels have been observed in *Chlamydomonas* E(z) mutants (Shaver *et al.*, 2010). First of all, investigation of genomic H3K4me3 and acetylation levels should reveal whether genomic loss of H3K27me3 directly entails increased levels of activating histone marks which could favour transcription of previously H3K27me3 silenced genes in *C. merolae*, too. Moreover, this investigation could distinguish between H3K27me3 targets which do and do not show upregulation to see whether the regulation correlates with a switch to an active chromatin state for the set of H3K27me3 targets showing transcriptional regulation.

Notably, one H3K27me3 target (*CMI116C*) was found highly repressed upon the marks absence (Table S4) which might represent an example of such multi-layer regulation. Here, it is possible that a transcriptional repressor of this gene is subject to activation when H3K27me3 is depleted. Consequently, the repressive action of this regulator on the gene *CMI166C* might overrun the transcriptional activation gained by H3K27me3 absence. As a result, albeit losing its repressive mark this gene is downregulated in the mutant. Yet, none of the upregulated genes in the Δ*E(z)* is described as transcriptional regulator. Although the genome of *C. merolae* is completely annotated, not all genes are characterized which is reflected in several genes with unknown function found in the Δ*E(z)* DEG list (Table S4). The possibility remains, that within the fraction of genes with unknown function are such transcriptional regulators, which might be unique to red algae - or even specific to *C. merolae*. Further improvement of the annotation will aid in elucidating potential transcription modulating candidates.

4.1.2 PRC2 regulates repetitive elements with preferential localization at the subtelomere

Within the entity of deregulated elements in the Δ*E(z)* mutant were by far not only H3K27me3 targets. Indeed, the majority of DEGs is composed of non-targets (Figure 3.1A). Thus, regulators acting in upstream regulatory pathways of these genes must react to the loss of genomic H3K27me3 levels which subsequently affect the expression patterns of these genes. Similar to the non-regulated H3K27me3 targets the regulated non-targets might demand multilayer regulation, which to some extent is shaped by H3K27me3. How precisely these regulatory pathways are composed is an intricate question, but here, too, the annotation of the uncharacterized gene fraction might serve as first strategy to tackle this problem.

By the removal of genomic H3K27me3 in *C. merolae*, increased expression of ~30% of the genomic REs could be observed. Roughly half of the REs which showed increased expression in the mutant are covered by H3K27me3 in WT (Figure 3.1B). This result indicates that PRC2 has a role in RE silencing in *C. merolae* and poses an additional example of the ancestral function of PRC2 in lower eukaryotes in silencing genome threatening elements. Likewise, H3K27me3 regulates REs in representative species of ciliates, green algae, and also in mammals (Frapporti *et al.*, 2019; Leeb *et al.*, 2010; Shaver *et al.*, 2010; X. Zhao *et al.*, 2020). However, as not all targeted REs are regulated in the mutant, additional regulatory mechanisms in silencing or activating RE transcription/transposition are conceivable. Conventionally, in eukaryotes transposable elements are kept silent in tightly condensed heterochromatic regions associated with methylation of DNA and heterochromatic histone marks such as H3K9me2 (reviewed in Xu &

Jiang, 2020). Interestingly, an interaction between H3K27me3 and H3K9me2 was shown for the ciliates *Tetrahymena thermophila* and *Paramecium tetraurelia*. In those two species, a dual substrate specificity of the *E(z)* homologs *EZL1* was proposed: Here, H3K27me3 as well as H3K9me2 were deposited by the same enzyme (Frapporti *et al.*, 2019; J. Xu *et al.*, 2021). Further, both marks were found in association with heterochromatin to repress transposable elements in *Podospora anserina*. In this ascomycete fungus PRC2 seems to exhibit a link between facultative and constitutive heterochromatin, as H3K27me3 is mutually exclusive with H3K9me in gene-rich regions (either one or the other), whereas the two marks overlap and are independent in repeats (Carlier *et al.*, 2021). Thus, in lower eukaryotes both histone marks, H3K27me3 and H3K9me2 seem to harbour the function in silencing repeats. While CG-specific DNA methylation could not be detected in *C. merolae* (Huff & Zilberman, 2014), homologs of the main *Arabidopsis* methyltransferase responsible for H3 lysine 9 dimethylation, KRYPTONITE (Du *et al.*, 2015) exists in the *C. merolae* genome (*CMN313C*). However, H3K9me2 genomic contents have not been determined to date. The possibility remains, that the KRYPTONITE homolog is active and that thus, next to H3K27me3, H3K9me2 bears the potential to be involved in silencing repetitive elements in *C. merolae* which could explain the only fractional regulation of repeats in the Δ*E(z)* mutant. As a first result, H3K9me1 as well as H3K27me1 levels could be detected in *C. merolae* (Schubert laboratory, unpublished). Notably, H3K27me1 levels were *E(z)*-dependent. Yet, the finding of H3K9me1 is promising and it will be interesting to determine, whether H3K9me2 genomic levels can be detected in addition and whether these two contribute to RE silencing in *C. merolae*. Moreover, as unrestrained activity of genome threatening elements could accumulate over time, and the Δ*E(z)* mutant is cultured in our lab for more than two years now, southern blotting for extrachromosomal DNA (or qPCR) could reveal whether the activated RE fraction indeed transposed within the *C. merolae* Δ*E(z)* genome.

Formerly, preferential occupancy of H3K27me3 towards chromosomal ends was detected for *C. merolae* (Mikulski *et al.*, 2017). In the present study, comparison of transcriptomic data at subtelomeres between WT and Δ*E(z)* revealed that H3K27me3 in addition to occupying it, also silences subtelomere expression (Figure 3.2). The transcriptional regulation via E(z) affects 26/40 subtelomeres, whereas its occupancy is much broader concerning 36 chromosomal ends. The fact that H3K27me3 occupancy does not fully reflect its transcriptional regulation implies that E(z) is necessary yet not sufficient to regulate all its targets. Further, among subtelomeric loci all combinations of H3K27me3-targets and non-targets, repetitive elements as well as genes are misexpressed. Thus, the subtelomere regulation via H3K27me3 seems to be locally specified, rather than genomic element-directed. The observation of subtelomere-directed H3K27me3 occupancy (and regulation) was also made in *Neurospora* (Jamieson *et al.*, 2013) and *Fusarium graminearum* (Connolly

et al., 2013). Moreover, a recent report described the relevance of specific H3K27me3-demethylation of transposable elements to maintain the *Arabidopsis* genome integrity (Antunez-Sanchez *et al.*, 2020). This is supported by the finding that PRC2-dependent H3K27me3 loading on telomeres seems to be essential for the establishment of heterochromatic marks H3K9me3 and H4K20me3 which are vital for transcriptional silence of the telomere (Montero *et al.*, 2018). All in all, transcriptional regulation conferred by H3K27me3 at/around chromosomal ends might not be imperative, but given its involvement in (sub)telomere regulation in several species ranging from human to slime molds - and with this study also red algae - PRC2 seems to be involved in the upkeeping of genome integrity in eukaryotes similar to what was discussed recently (Vijayanathan *et al.*, 2022).

4.1.3 H3K27me3 relevance in transcriptional control of genes only arises under stress conditions concerning inducible genes?

The majority of DEGs found in the Δ*E(z)* (Figure 3.1) was either found in repetitive elements (Figure 3.1 B) and/or they localized to subtelomeric regions (Figure 3.2). Under control conditions the mutant showed limited numbers of misregulated genes which could potentially confer biological consequences, e.g. be involved in stress responses. Moreover, by far not all genes which are decorated with H3K27me3 in WT are misregulated upon the marks absence (Figure 3.1 B). As discussed above, this fractional regulation might hint towards the need of additional, regulatory processes in order to entail changes in expression. Overlap analyses using the Δ*E(z)* DEG list and comparing it to other transcriptomic data collected under stress situations revealed similar regulated genes in limiting nitrogen conditions as well as TOR regulation, respectively. In line with this finding, a recent study proposes the protein kinase TOR acting as repressor of stress responsive genes through global maintenance of H3K27me3 in *Arabidopsis* (Dong *et al.*, 2021). In *C. merolae*, TOR acts as an amino acid and nutrient sensor - a fact which may explain TOR as link between the Δ*E(z)* DEG data set and the limiting nitrogen response (Imamura *et al.*, 2015). It is worth noting, that the limiting nitrogen dataset comprises 30x the DEG number of Δ*E(z)* and the overlap with the rapamycin treatment (TOR) dataset yielded only 12 genes commonly regulated which both limits the reliability of any biological meaning. Although the overlap analysis has limitations, it provides first indications towards a link between PRC2 mediated gene regulation and stress responses in *C. merolae*. Similarly, PRC2 targets stress-responsive genes in *Arabidopsis* and further regulates them via associated proteins (Kleinmanns & Schubert, 2014). This hypothesis is additionally strengthened by the finding that the *Arabidopsis* PcG mutant *clf* showed upregulation of many stress responsible genes only under dehydration stress, whereas they remained unresponsive under control conditions (N. Liu *et al.*, 2014). To further

tackle the H3K27me3 involvement in stress responses, it appeared crucial to expose the Δ*E(z)* mutant line to stress since H3K27me3 could act as regulatory mark only when the need to react to certain environmental changes arises, e.g. under stress.
As a result of the exposition of *C. merolae* cells to heat stress significant differences between WT and the Δ*E(z)* strain could be observed. When encountering a 57°C heat stress for 30 min (which was designated *priming stimulus* within this project), chlorophyll fluorescence measurements revealed that mutant cells showed significantly lower PSII quantum yields compared to WT (Figure 3.6 A). Lowered chlorophyll fluorescence levels traditionally indicate heat stress-induced damages and correlate with disruptions of photochemical reactions in thylakoid lamellae of chloroplasts (Allakhverdiev *et al.*, 2008; Wahid *et al.*, 2007; Wise *et al.*, 2004; Yamada *et al.*, 1996). Thus, the finding of significantly stronger loss of chlorophyll fluorescence indicates increased heat-induced photosystem damages in *C. merolae* when genomic H3K27me3 levels are lost. This suggests that the active PRC2 complex contributes to the basal heat tolerance of *C. merolae*.

The finding of an altered heat response in the Δ*E(z)* mutant was further undermined by lowered number of genes reacting to priming (Figure 3.7, PvsN) indicating that the mutant line is not capable to establish a proper response to heat stress. Decreased PSII fitness (Figure 3.6 C) and lowered DEGs (Figure 3.7, TvsN) in the mutant could be additionally observed when *C. merolae* cells were exposed to 60°C for 15 min (short *triggering* stress). It is worth noting, that a temperature upshift of 12-15°C from optimum growth is defined as heat stress for photosynthetic processes (Allakhverdiev *et al.*, 2008; Wahid *et al.*, 2007). Accordingly, the application of 60°C as triggering stress in the HSMA (which is 18°C above the optimum temperature (i.e. 42°C) for *C. meroale*) is considered as severe heat stress and hence it is not unexpected that it leads to the absence of growth on solidified medium. The decreased DEG numbers upon heat in the mutant bear the possibility that genes needed in the immediate response to heat are unresponsive, thus leading to worse repair mechanisms which eventually could lead to decreased PSII fitness. It is well established that the photosystem machinery is among the first suffering from heat stress (Allakhverdiev *et al.*, 2008; Wahid *et al.*, 2007) and accordingly, already in the WT the PSII fitness decreases somewhat upon the priming and dramatically upon triggering stress (Figure 3.4 C). The constitution of the unresponsive genes under heat stress reflected in the difference between WT and Δ*E(z)* DEGs in PvsN as well as TvsNL might reveal potential candidates responsible for the inefficient heat response in the mutant. As a first attempt to do so, the heat transcriptomes of priming and triggering, respectively, were compared between WT and Δ*E(z)* directly (PvsP, TvsT). Transcriptomic heat responses are principally mediated by heat shock factors (HSF), which entail the rapid activation of specific transcription factors leading to activation or repression of heat-responsive genes (Ohama *et al.*, 2017). Of the four

C. merolae HSFs, the expression levels of *CMR316C* detected after priming resp. after triggering were significantly lower in the mutant compared to WT, rendering this HSF homolog as potential candidate responsible for the worse heat response in Δ*E(z)*. However, as it is not heat responsive as such, it does not seem to be essential in the direct heat response but might act redundantly with *CML277C* and *CMT597C*. Noteworthy, the number of genes showing altered expression rates between WT and ΔE(z) only in the situation of heat did not exceed 30 DEGs, and among them were only three directly targeted H3K27me3 genes. The effect of H3K27me3 on regulating the expression of heat-sensitive genes is therefore estimated to be small, reflecting the lower, but not strongly different, thermotolerance of the mutant. The slightly reduced heat resistance in Δ*E(z)* might be caused by the incomplete response of the HSF *CMR316C* - a question that could be clarified in the future by genetic studies. Similar to the reduced transcriptomic heat response of Δ*E(z)* which may involve the insufficient action of an potential HSF, a heat sensitive cultivar of Chinese date (Jujube) showed lowered DEG numbers in the response to heat stress compared to its heat-resistant control. Here, too, an incomplete reaction of one of the heat-relevant HSFs was identified in the thermosensitive cultivar (Jin *et al.*, 2020).

Taken together, it can be concluded that although no phenotypic aberrance was detected in the Δ*E(z)* mutant under control conditions, differences arose under non-control conditions designating the Δ*E(z)* as less thermotolerant than WT cells. Thus, H3K27me3 seemingly - by shaping specific parts of the transcriptomic heat shock response - contributes to PSII fitness when heat stress occurs. However, whether the observed reduced heat stress tolerance is a specific response to elevated temperature or whether PRC2 instead acts as general stress capacitor in *C. merolae* should be addressed in exposing the Δ*E(z)* strain to various stress regimes. Lastly and additionally, as (a) stress induces RE transcription, which is well described for the *Arabidopsis* heat stress response (Cavrak *et al.*, 2014), and (b) RE´s regulation in Δ*E(z)* was detected, it would be interesting to know how repetitive element expression is affected upon heat stress. This question will be easily answered, once the new annotation file including RE´s will be available for *C. merolae*.

4.2 *Cyanidioschyzon merolae* is able to memorize heat stress

PcG protein complexes regulate developmental phase transitions in plants (Kleinmanns *et al.*, 2017). Previously, several publications suggested additional contributions of one of the PcG complexes, namely PRC2 and its histone mark H3K27me3 to memory mechanisms of environmental cues, including the flowering initiation process (vernalization) as well as memory of heat stress in *Arabidopsis* (Bastow et al., 2004; Shen et al., 2021). *C. merolae* is devoid of major developmental phases as one cell cycle corresponds to one lifecycle. In this project it was therefore hypothesized that instead of developmental phases, PRC2 may regulate putative memory mechanisms of environmental stress in *C. merolae*. Yet, as stress memory was not described for *C. merolae* before, first of all the existence of such mechanism had to be proven to define whether *C. merolae* is able to acquire improved tolerance to stress. Only then the PRC2 mutant ∆*E(z)* might be included and its putative influence could be dissected. As one of the most heat tolerant eukaryotic species worldwide *C. merolae* inhabits volcanic hot springs (Ciniglia *et al.*, 2004; Yoon *et al.*, 2006). In consequence, heat may represent the most prevalent natural stressor *C. merolae* will be exposed to. Therefore, an assay using sequential heat stimuli was established. As a result the heat stress memory assay (HSMA) was established. Accordingly, this study describes the ability of *C. merolae* to memorize a given stress information on physiological and transcriptional level for the first time. In detail, primed *C. merolae* cells show survival upon a lethal heat stress while their un-primed counterparts were unable to grow on solidified medium. Furthermore, primed cells show significantly higher PSII fitness indicating an elevated photosynthetic energy conversion ability of primed cells compared to the non-primed control (Figure 3.5 B). The priming benefit observed for primed-before cells could be linked to a set of heat stress trainable genes, identified by differential gene expression analysis of the primed-before and non-primed transcriptomes upon a short triggering stress episode (Figure 3.8). Although information exists for *C. merolae* responses to stress including heat (Kobayashi *et al.*, 2014), limiting nitrogen (personal communication S. Rader), light (Abram *et al.*, 2022; Tardu *et al.*, 2016), and limited CO_2 (Rademacher *et al.*, 2017) amongst others, the knowledge about reactions to sequential stress occurrences were undescribed so far. In general, there is very little information about memory phenomena in unicellular photoautotrophic eukaryotes. A rare example is the recent study of Mikulski and colleagues who described the priming ability of *Chlamydomonas* to repetitive nitrogen deprivation (Mikulski & Santos-Aberturas, 2022). Heat stress memory in particular is described for instance in the multicellular red alga *Bangia* and the grass tall fescue (Bi *et al.*, 2021; Kishimoto *et al.*, 2019). Most information about heat stress memory and underlying mechanisms, however, is available for *Arabidopsis* (Balazadeh, 2022; Nishad & Nandi, 2021;

Xie *et al.*, 2021). Taken together, a robust and reproducible heat stress priming and memory scheme could be established, which now will serve the intent to dissect the mechanism leading to lasting memories of previously experienced stress events, referred to as thermomemory, in *C. merolae*.

4.2.1 *C. merolae* thermomemory may be mitotically stable

After the establishment of a working HSMA, it was of interest to determine how long the priming benefit could be maintained by gradually extending the time period between priming and triggering stress (i.e. lag phase). The maximal thermomemory preservation in *C. merolae* lies between 24h and 48h, or better, between 2 and 3.5 cell divisions suggesting that the priming information is likely inherited to the next generation of cells (Figure 3.5). However, division rate determination based on total cell numbers in a culture like applied in this study does not allow to discriminate between parental and daughter cells. As it is likely, that the originally primed cells are still existent in the culture when triggering occurs, accurate statements about heritability of the priming information are hampered. To precisely address the question of mitotic heritability of the priming stimulus, two scenarios are possible.
First, synchronized cultures could be used in the HSMA, meaning that all cells are at the same stage of cell cycle during priming/lag/triggering. Thus, precise number(s) of cell divisions could be tested in the HSMA. This includes optimization, since the priming duration of 30 min is very short compared to a cell division cycle of 24h (in a synchronized scenario of 12h dark/12h light cycle). As a consequence the priming stimulus would affect only certain mitotic states of the cell and the possibility remains, that *C. merolae* cells cannot be primed in all cell cycle stages. Thus, all mitotic phases have to be tested. As a second option to determine the maximal stability of priming benefit, microfluidic devices allow for single cell monitoring, since individual cells are attached to surfaces on chips and conditions are precisely controlled (Z. Yu *et al.*, 2018). Thus, in a future approach, cells originating from those fixed (and formerly primed) cells could be collected. This fraction of cells - newly developed during lag phase - then should be exposed to triggering in order to assess their priming benefit. Besides, microfluidics might contribute to the answer of the generally unsolved question of the age one *C. merolae* cell can achieve.

Taken together, the non-synchronized scenario used in this study allows for the general description of heat stress memory in *C. merolae* independent from cell cycle stage with an estimated memory preservation of more than one mitotic division. Whether primability depends on cell cycle state as well as the maximal duration of memory is subject to future investigations.

4.2.2 HS trainable genes as memory determinants

After heat stress memory has been determined on physiological level and its stability was proven for mitosis-including lengths of lag phases, transcriptomic analyses of primed and non-primed samples originating from the 24h lag phase experiment lead to the identification of 59 trainable genes for *C. merolae* thermomemory (Figure 3.8). As a *proof of concept* experiment, candidate memory genes identified in the 24h lag phase were tested in samples collected during different HSMA set ups, including a short lag phase of 2h and a prolonged lag phase of 48h to see whether the expression of these genes can be linked to physiological memory. As a result, when 2h parted priming from triggering stress (2h lag) none of the memory genes showed significant trainability. Thus, a duration of 2h of time does not suffice to significantly consolidate the priming information on transcriptional level while other, faster memory processes cause the priming benefit on physiological level already after 2h. In case of 24h lag phase, the gene expression of all four tested genes reflected physiological memory occurrence, *CMP040C* and *CMF092C* in a significant manner defining a duration of 24h as sufficient for the generation of transcriptional memory. Both, *CMP040C* and *CMF092C* are uncharacterized proteins specific to rhodo- and/or onchrophyta (red and brown algae) and in WT not targeted by H3K27me3. After 48h, a general absence of trainability was detected on memory gene expression reflecting the situation determined on physiological level. However, except for *CMJ099C*, a slight trend towards elevated expression rates in the PLT condition was still visible in the 48h lag phase. Consequently, the possibility remains that after 48h the information about the previous priming was not completely “forgotten”. This repeatedly poses the question of the absolute length of memory, which should be tested e.g. in single-cell analysis as described above.

When wisely chosen, the expression of single genes can precisely describe emergence and decay of memory underscoring the potential of the here described set of memory genes. Linking memory gene expression directly to physiological memory defines the tested genes, in particular *CMP040C* and *CMF092C*, as valuable tool to rapidly and sensitively determine whether or not memory occurred in a given condition. Therefore, using memory gene expression will aid in the identification of major regulators of memory.

4.3 Potential mechanisms underlying thermomemory in *C. merolae*

4.3.1 Trainable genes are prevalently hyperinduced upon a second heat stress

The composition of the memory gene set can aid in the disclosure of memory-driven processes and/or mechanisms leading to memory. In addition, they can serve as molecular markers to determine memory. Analysis of the transcriptional behaviour of memory genes revealed that the majority of trainable genes were hyperinduced in previously primed cells compared to their non-primed control. Hence, the priming benefit can be linked to a general need in gene activation rather than targeted repression. As they were upregulated already during the first heat stimulus, the majority of *C. merolae* heat trainable genes exhibit [+/+] trainability (according to Ding *et al.*, 2012). Yet, the current project does not provide the response pattern of this [+/+] trainability leaving the question open, whether it is a stronger or faster reinduction after triggering. This question should be addressed in a time course experiment with later sampling points during the triggering stress since the possibility remains, that these trainable genes gain comparable induction rates in triggered-only conditions at slower paces.

In addition to its unsolved kinetics, potential mechanisms leading to transcriptional memory remain to be addressed. Clustering of all 59 memory genes resulted in five clusters with similar expression behaviour during the HSMA (Figure 3.9). Clusters 1 and 2 include genes with are generally heat responsive and in addition trainable, showing even higher induction rates when facing heat stress repetitively. Cluster 3 and 5 comprise genes which are induced in priming and barely react to the triggering stimulus, however reach high induction rates upon a second stress trigger. The mRNA levels of cluster 4 genes, however, responded only vaguely to the priming stress and nearly no reaction was detected to the sole triggering stimulus (Figure 3.9). Interestingly, this group of genes showed slightly elevated expression rates in PL cells compared to the non-primed control NL cells, hinting towards incomplete recovery of expression rates during lag phase. This observation is reproduced on single gene level for *CMB107C*, *CMP040C* and *CMF092C* since all of them show slight priming inductions followed by a further increase during lag (Figure 3.10). Interestingly, such sustained induced memory genes in *Arabidopsis* thermomemory were found associated with elevated H3K4me3 levels governed during priming heat (Lämke *et al.*, 2016). Hence, chromatin modifications might represent a possible way to govern the putative sustained induction. In general are chromatin modifications potent candidates for storing past information about transcription as their presence or absence directly correlate with

transcription rates. Remodelling the chromatin at site upon a first heat stress might ensure the rapid re-induction upon the recurrent one, the triggering stimulus. In line with this assumption, the trainability of *CMB107C*, *CMP040C,* and *CMF092C* was identified as H3K27me3-dependent (Figure 3.10, Table S2) suggesting a potential involvement of H3K27me3 in both, reaching appropriate induction rates resulting in [+/+] trainability as well as the implied sustained induction during lag phase. In *Arabidopsis*, [+/+] trainability was repeatedly linked to chromatin modifications in several stress regimes. Accordingly, in repetitive dehydration stress the hyperinduction of trainable genes was associated with a gain in the activating H3K4me3 modification as well as elevated Ser5P levels (Ding *et al.*, 2012). Sani *et al.* reported persistent priming-induced changes of H3K4me3 upon hyperosmotic priming, which was associated with transcriptional activation as well as shortening and fractioning of H3K27me3 islands (Sani *et al.*, 2013). A similar study of cold stress, investigating gene expression in *Arabidopsis* seedlings upon single stress and subsequent recovery, could link the recently cold-activated genes with lastingly low levels of H3K27me3 even after the cold has passed. As H3K4me3 levels were not monitored in this study, the possibility remains that this activating mark might be elevated at those genes (Kwon *et al.*, 2009). Moreover, upon heat exposure, especially promoter regions of heat-relevant transcription factors remained occupied by heat-induced elevated levels of H3K4me3. The latter were maintained even after the stress subsided and correlated with expression (H.-C. Liu *et al.*, 2018). Furthermore, the important role of sustained H3K27me3 demethylation on heat-responsive genes by active demethylation upon heat stress could be shown. This demethylation was needed for a faster re-induction of those genes upon reoccurrence of the heat stress trigger (Yamaguchi & Ito, 2021). Albeit not all of these examples included repetitive stress exposures, these studies identify chromatin modifications as prominent cause of lastingly-changed transcriptional patterns generated upon a first stress exposure which eventually can lead to the [+/+] trainability (and in some cases sustained induction) detected for *C. merolae* thermomemory genes. As exciting additional possibility, the concept of bivalency, i.e. the co-existence of activating and repressive histone marks, gains increasing attention as a poising mechanism for inducible genes. Holding the chromatin modifications for both, activation as well as repression, bivalent genes might be silent under non-stress conditions however rapidly activated when needed, also in the context of stress inducibility in plants (Faivre, 2022). Thus, future research should dissect the chromatin environment of these genes, to proof the hypothesis whether the here observed trainability might be encoded on chromatin level with potential changes in histone methylation patterns. While H3K27me3 occupancy in *C. merolae* is known for control conditions, its behaviour upon heat stress is unclear. Regarding a potential contribution of sustainedly induced memory genes it has to be pointed out, that the induction rates found for

CMB107C, *CMP040C* and *CMF092C* were not significant compared to the non-primed controls. Therefore, the suggestion of sustained induction during lag phase for cluster 4 memory genes remains an hypothetical possibility and demands further validation - first and foremost on single gene expression level via qRT-PCR. Alternatively, as the memory gene *CMJ101C (CmsHSP2)* is directly H3K27me3-targeted under control conditions and furthermore detected as H3K27me3-dependent in the RNAseq analysis, this gene might serve as potent first candidate to dissect its chromatin environment including both, H3K27me3 as well as H3K4me3.

4.3.2 Targeted RE repression following priming heat

Notably, the only two negatively trainable genes are among cluster 4. *CMT010C* and *CMJ308C* encode for retrotransposons and are induced when a single heat stress (priming or triggering) occurs. In addition, increasing age lead to elevated transcription rates reflected in higher expression levels in NL compared to N cells. However, when these genes have been activated during priming already, a triggering stimulus resulted in lower mRNA levels in primed-before cells compared to the non-primed control (Figure 3.11). Heat-activated repetitive sequences are well described in the *Arabidopsis* heat shock response. It is believed that some REs exploit the plant´s defense response to become active (Ito *et al.*, 2011; L. Sun *et al.*, 2020). Exemplarily, the retrotransposon *ONSEN* has evolved several features which lead to the evasion of the plant´s silencing mechanisms when heat occurs, such as non-methylatable promoter regions and instead the integration of heat response elements (HSE) which couples the *ONSEN* activation to an important stress response. Upon re-integration these sequences bear the potential to convey heat responsiveness to neighbouring genes (Ito *et al.*, 2011), which by increasing the genetic variability might be of advantage for the plant in the long run (Cavrak *et al.*, 2014). A detailed analysis of the promoter regions for HSE sequences should provide information whether a similar mechanism is underlying *CMT010C* and *CMJ308C* heat activation eventually contributing to the heat stress memory. Strikingly, after priming induction, the expression of both is strongly repressed during lag phase resembling the re-silencing of *Arabidopsis* REs after heat stress (L. Sun *et al.*, 2020). However, opposing to what was described for *Arabidopsis, CMT010C* and *CMJ308C* expression rates even undercut the NL age control levels which suggests an active heat-induced silencing of *CMT010C* and *CMJ308C*. Conventionally, TEs are silenced by DNA methylation, heterochromatin formation or localization to silent chromatin sections and/or decoration with repressive histone marks. As DNA methylation could not be verified in *C. merolae* (Huff & Zilberman, 2014), the contribution of chromatin modifications to the enhanced repression during lag phase is probable. This is reminiscent of the *Arabidopsis*

heat response, where a fraction of REs remained active after the heat stress subsided, which correlated with the addition of active histone marks as well as lowered nucleosome density (Pecinka *et al.*, 2010). The post-priming repression of *CMT010C* depends on H3K27me3 in the genome, as *CMT010C* is upregulated in the Δ*E(z)* mutant and not re-silenced during lag phase nor is it further downregulated in Δ*E(z)* PLT. In fact, the *CMT010C* locus is surrounded by some levels of H3K27me3, raising the possibility of post-priming spreading of H3K27me3 eventually covering *CMT010C* as well entailing transcriptional repression. The pronounced intergenic H3K27me3 peak found in an putative promoter region upstream of *CMT010C* (Figure 3.11B) could explain (a) the re-silencing of *CMT010C* in WT after priming via H3K27me3 spreading and (b) the upregulation in the H3K27me3 depleted Δ*E(z)* mutant under control levels. The additional H3K27me3 peak found on the repetitive element (*CMT011X*) downstream to *CMT010C* could reinforce this effect. For *CMJ308C*, the situation is similar but different at the same time. Thus, *CMJ308C*, too, has an adjacent H3K27me3 peak, however this peak is downstream of the gene and furthermore the distance between *CMJ308C* and the intergenic region upstream of *CMJ306C* is greater which might explain its transcriptional independency of H3K27me3. Nonetheless, the possibility remains, that H3K27me3 spreads from this intergenic region down to *CMJ308C* adding to the suppression after the heat stimulus. To verify the contribution of chromatin marks to the repression of REs after heat stress, genome wide analysis of the chromatin environment during the HSMA could contribute to understand the observed RE repression following priming. Here, H3K27me3 itself should be included as well as other representative silencing and activating marks, such as H3K9me2 and H3K4me3, respectively, coupled to H3 levels to additionally draw conclusions about potential nucleosome density alterations.
It has to be pointed out, that this priming-repressive effect on REs could be described for two genes only. Consequently, it is hard to draw general conclusions about priming-induced RE repression as mechanism underneath the *C. meroale* thermomemory. The majority of REs is not included in the HSMA analysis since the current annotation file lacks this genomic element class, leaving open the interesting possibility that more than these two genes demonstrate priming-based repression.

In sum, the observed repressive memory effect shown here for the retrotransposons *CMT010C* and *CMJ308C* bears high potential to investigate memory-related silencing mechanisms of repetitive elements which might depend on chromatin regulation. It has to be emphasized, that a potential contribution of RE´s to memory mechanisms is only fairly investigated to date. Thus, many questions remain to be asked such as: When including all repetitive elements in the analysis, is the priming-based activation followed by enhanced repression during lag phase a general feature of repetitive sequences, or more specific, of retrotransposons? Do *CMT010C* and *CMJ308C* produce extrachromosomal DNA which

consequently integrates elsewhere in the genome bearing the risk of essential gene mutations or, in turn, does reintegration lead to heat responsiveness of additional (potentially relevant) genes? And in connection to the latter: What, if any, is the biological relevance of active RE repression in PLT: does it actively contribute to the priming benefit in *C. merolae*? As soon as the new genome annotation is finalized, future research should address these questions.

4.3.3 Combination of nuclear and chloroplast genome encoded memory genes suggest inter-organellar regulation during *C. merolae* thermomemory

a. Trainable heat shock locus on chromosome 10 and its putative memory-regulation

Knowledge of trainable genes identified for distinct stresses bear great potential in determining memory events and, moreover, in elucidating which mechanisms exactly lead to or contribute to memory. The heat stress memory gene set for *C. merolae* thermomemory included nine nuclear encoded genes, of which three were located in a jointly regulated locus on chromosome 10. Interestingly, this affects in total five genes (*CMJ099C – CMJ103C,* Figure 3.13) and raises the question how this common regulation is achieved. Annotation analysis revealed that the region´s centre is formed by the only two small heat shock protein coding genes of the *C. merolae* genome: *CMJ100C (CmsHSP1)* and *CMJ101C (CmsHSP2)*. The small amount of sHSP genes in *C. merolae* is in contrast to genomes of higher plants, where sHSP´s are highly diverse. Exemplarily, sHSP´s in *Arabidopsis* can be divided into six subclasses, defined by different localizations of the sHSP (Scharf *et al.*, 2001; Waters & Rioflorido, 2007). Belonging to the sHSP family, *CmsHSP1* and *CmsHSP2*, are known responders to heat in *C. merolae*, with fast and high induction rates of gene and protein expression upon heat shock (Kobayashi *et al.*, 2014) which is in line with the finding of strong heat responses of both genes in the present study. They are encoded at opposite strands and the intergenic region between their transcriptional start sites (TSS) is thought to act as common promoter driving the transcription of both genes (Kobayashi *et al.*, 2014). Moreover, the same study identified a putative heat shock element (HSE) as binding site for heat shock transcription factors (HSF) in both, the forward and reverse direction, reinforcing the notion of co-regulation of *CmsHSP1/2* via a single promoter. Interestingly, the promoter region is associated with lowered H3 levels (Figure 3.13C), possibly indicating decreased nucleosome occupation which may allow immediate accessibility of the promoter under heat stress to the transcriptional machinery. The usage of a single promoter sequence for two genes in inverted orientation would mean that activation of either one will entail the parallel activation

of the other representing a way of how transcription at the whole memory-region may be activated. Furthermore, maintaining or extending low nucleosome levels after the priming heat subsided might facilitate re-induction in case heat returns. Consistent with this assumption, chromatin remodeling has been implicated in the transcriptional memory of heat in *Arabidopsis* (Brzezinka *et al.*, 2016; Mozgová *et al.*, 2015). For instance, the *Arabidopsis* defence gene activation was found associated with reduced nucleosome occupancy close the TSS (Mozgová *et al.*, 2015). Studying nucleosome occupancy in context of biotic stress responses in CHROMATIN ASSEMBLY FACTOR 1 (CAF-1) mutants, the same study linked lowered nucleosome occupancy to the primed state of stress defense genes (Mozgová *et al.*, 2015). Moreover, the Strawberry notch (Sno) orthologue in *Arabidopsis*, FORGETTER 1 (FGT1), associates globally with promoter regions of actively expressed genes in a heat-dependent manner and interacts with chromatin remodellers leading to low nucleosome occupancy (Brzezinka *et al.*, 2016). In the light of the collected results, it cannot be excluded that the region on chromosome 10 (comprising the heat shock locus) may be similarly regulated. By maintaining or extending the lowered nucleosome occupancy at the TSS, chromatin remodeling complexes might lastingly lead to stronger responses of *CmsHSP1/2* by facilitating e.g. transcription factor (TF) binding. As potential candidate TF, two out of the four genomic heat shock transcription factors (HSF) were induced by priming (Figure S3), but do not show trainability. This priming induced expression might result in elevated HSF levels in the cell, which may, in their inactive form, endure the lag phase. Upon recurrent heat, the HSFs may be quickly activated and via rapid gene activation substantially contribute to trainability. It will be interesting to study, whether one of those two HSF bear the potential to bind to the HSE in the *CmHSP1/2* promoter. In addition to chromatin remodelling, H3K27 methylation at *CmsHSP2* (Mikulski *et al.*, 2017) might serve as an anchor for spreading or erasure of itself and other histone marks after the *CmsHSP1/2* locus was activated. Together these processes might lead to lasting changes of the chromatin environment and therewith altered expression. However, this study only refers to H3 levels collected in ChIP-sequencing (Mikulski *et al.*, 2017) and thus provides only indirect indication to lowered nucleosome occupancy as mechanism to induce transcription of those five genes. Methods precisely measuring the amount of nucleosome-depleted DNA as well as chromatin modifications, such as formaldehyde-assisted isolation of regulatory elements (FAIRE, Giresi *et al.*, 2007) and chromatin-immunoprecipitation followed by qRT-PCR, respectively, should provide first indications about whether or not such scenario may be reality.

Independent of the mechanism leading to transcriptional memory at that nuclear region, the finding of commonly regulated sHSP in the memory of heat opens the possibility to speculate about the biological consequences once these genes are translated to proteins. sHSP are found in all three domains of life (De Jong *et al.*, 1993; Franck *et al.*, 2004; Plesofsky-Vig *et*

al., 1992; Waters, 1995) and encode for chaperone proteins that can prevent the irreversible denaturation of other proteins in an adenosine triphosphate (ATP)- independent manner (Ehrnsperger *et al.*, 2000). This function might gain central biological relevance of CmsHSP1/2 in the heat memory response. This suggestion is undermined by the fact, that CmsHSP1/2 have been induced to highest extent within the region, with gradual decrease towards left and right region ends, respectively, emphasizing their superior relevance within the memory region on chromosome 10. Intriguingly, for the two *C. merolae* sHSP different localizations have been proposed, whereas CmsHSP1 is predicted to localize to the cytosol (Kobayashi *et al.*, 2014; Waters & Rioflorido, 2007) and for CmsHSP2 a chloroplast or mitochondrial transit peptide was predicted (Kobayashi *et al.*, 2014). Subsequently, activation of one genomic locus might supply sHSPs to several cellular compartments at once. Here, additional experimental verification should provide evidence as to whether or not CmsHSP2 is mitochondria and/or chloroplast-targeted *in vivo*, for instance via fusion of the *CmsHSP2 (CMJ101C)* with a fluorescent tag followed by microscopic localization of the protein. It cannot be ruled out that CmsHSP2 is activated by heat, translated in the cytosol and then exerts its protein-protective role in the chloroplast.

b. Chloroplast thermomemory

The putative migration of a nuclear encoded (trainable) protein to the chloroplast (CP) is particularly interesting in the light of the vast CP genome trainability upon sequential heat in *C. merolae* (Figure 3.8, Figure 3.14). Not only is the majority of trainable genes encoded in the CP, but moreover, the CP genome as a whole is significantly regulated by recurrent heat and shows trainability for 1/3 of all CP protein coding genes. These strong CP transcriptomic reactions to recurring heat are not unexpected, since (a) heat sensitivity of the chloroplast is a well described phenomenon (reviewed in Hu *et al.*, 2020) and (b) heat is a known trigger for transcriptomic changes in the plant CP (Danilova *et al.*, 2018). Consequently the degree of CP regulation found in this study corroborates the known importance of the chloroplast in the heat stress context and extends it to a potential contribution to thermomemory.

In anticipation of the following discussion on probable chloroplast responses to recurring heat, it has to be clarified, that the overall read counts observed for CP genes are considerably lower compared to nucleus encoded genes. This can be explained by the application of poly(A) selection prior to sequencing in the present study since poly(A) is found more irregularly on chloroplast mRNA. Accordingly, nuclear mRNA relies on polyadenylation at the 3´- OH^- terminus which is a highly complex regulatory network necessary for stability, nuclear export, and the ability to be transcribed (reviewed in Stewart, 2019). In contrast, chloroplast mRNA polyadenylation is diverse and predominantly (however not only) localized within the aa coding region rather than at the mature 3´end with major functions in regulating

mRNA degradation (Rorbach *et al.*, 2014). In general RNA sequencing measures the steady-state level of mRNA in the cell which results from a combination of both, transcription and mRNA as decay. Together, this does however not question the validity of the obtained DEGs, but rather explains the lowered counts for CP genes and in the same time reminds on mRNA dynamics, which could display yet another mechanism forming thermomemory in *C. merolae*. Accordingly, the chance remains that under certain conditions, actual transcription rates remain same but mRNA degradation (or stabilization) is elevated.

As just introduced, the chloroplast genome strongly and broadly reacts to the applied HSMA. This applies for nearly all functional gene classes in the CP. This is in line with the finding that in higher plants chloroplast number, size and activity highly depend on the environment (Y. Li *et al.*, 2013). A particular regulation could be described for the photosystem machinery as well as for the bulk of the genetic system maintaining genes (Figure 3.15, Figure 3.17, Table 3.3). Among memory genes was *rpoC*, which encodes for the beta´ chain of the plastid-encoded RNA polymerase (PEP). Furthermore, as all four PEP subunits were considerably upregulated in previously primed (PLT) cells compared to the non-primed control (T) trainability of the whole enzyme is proposed. The absence of any significances might be the result of the very low read counts sequenced for CP genes. For instance, the detection is even aggravated in case of *rpoZ*, as it is a very small gene (147 bp). Hence, it is imperative to validate the common upregulation of all four subunits as well as confirm whether the higher expression levels correlate with similar high and functional PEP protein levels. Yet, PEP trainability opens interesting possibilities in thermomemory regulation. Gene transcription is much more complicated than the existence of the PEP enzyme itself. Hence, PEP activity is mediated by regulating factors. In order to target specific genes, this enzyme of bacterial origin demands a sigma factor, which provides promoter specificity (Tadini *et al.*, 2020). Furthermore, transcription factors (TF) confer specificity to DNA transcription by attaching to specific DNA sites and therewith promoting PEP binding which results in transcription initiation. Candidate genes for both, sigma and TFs, could be identified for *C. merolae* thermomemory in the present study. While non-trainable, *SIG3* (*CMR165C*) shows high mRNA levels upon heat stress and thus poses a possible candidate to ensure targeted gene transcription by PEP during priming. Interestingly, SIG3 does not accumulate under the two investigated conditions: not in changing light conditions nor upon nitrogen de- and repletion (Fujii *et al.*, 2013, 2015) rendering its precise function unknown. It will be interesting to analyse SIG3 accumulation under heat and recurring heat to see whether the protein accumulates upon a priming stimulus in protein blots using (flag/his-)tagged *SIG3* lines. It is possible, that the SIG3 is maintained during lag phase, possibly even at its corresponding promoter binding sites and thus is rapidly available when triggering occurs. In parallel, genetic mutant strains could determine the absolute relevance of *SIG3* in heat

responses and furthermore, test its contribution to *C. merolae* thermomemory. Finally, it remains elusive, if and to what extent SIG3 and PEP actually interact (pull down assays). As a potential additional regulatory layer of CP gene transcriptional regulation, two TF encoding genes showed trainability: *ycf27* and *ycf29*. Both encode for OmpR-like TFs, without specific target genes nor functions described to date. Subsequently, here, too, further experiments are needed to proof the definite role of *ycf27/29* in *C. merolae* thermomemory.

Following transcription, proteins are synthesized by ribosomes which translate the mRNA sequence to amino acids in a highly complex manner. This intricacy is mirrored in the diversity of ribosome genes as well as translational regulators. The translation factors *infB* and *tsf* were identified as trainable genes, adding further potential candidates to the list of *C. merolae* thermomemory regulators. Similar to the TFs, to date these two translational regulators are not specifically described for red algae rendering their preferred mRNA sequences elusive. In order to specify translation, not all ribosome genes are used at all times in the chloroplast and hence the ribosome composition favors the translation of specific mRNA sequences according to the cell´s needs (Xue & Barna, 2012). This situation might also apply for the *C. merolae* thermomemory, since - albeit a considerable number (15 out of 46) is defined as trainable - not all ribosome genes are involved.
Taken together, the specific activation of ribosome subunits as well as sigma-, transcription- , and translation factors detected in this study strongly suggest a tailored response in CP gene expression upon recurring heat in *C. merolae.* However, much more research is demanded to answer the questions of how these components are interconnected. Finally, the overarching question remains: Which proteins are required for the formation of thermomemory in *C. merolae* and what is the signal transduction pathway of regulation?

While the observed regulation of the genetic system could represent a higher-level regulatory system, the affected (and consequently primed) photosystem machinery should be considered a direct heat response. Upon heat stress the photosystem machinery is among the first responders as high temperature alters thylakoid membrane fluidity leading to the dissociation of PSII light harvesting complexes which in turn results in lowered PSII integrity (Baker & Rosenqvist, 2004; Janka *et al.*, 2013; Mathur *et al.*, 2011). Indeed, due to this fast and severe PSII reaction to thermal damage the chloroplast was proposed as sensor for (not only) heat stress (A. Z. Sun & Guo, 2016) which is in line with the decreased chlorophyll fluorescence detected upon heat (Figure 3.6). Moreover, both the PSI as well as PSII reaction centre genes reacted in trainable manner to recurring heat stress in *C. merolae* (Chapter 3.4.2). In line with this finding, the differential reaction of photosystem components to stress was reported for *Arabidopsis* high light acclimation, where PSII components were unaffected on protein level while PSI protein levels decreased. The authors concluded that stress acclimation leads to increased abundances of PSII repair cycle machinery (Flannery

et al., 2021). Alike to *Arabidopsis* high light response, *C. merolae* cells exposed to low temperature (25°C) rapidly modulated their PS apparatus on protein level, which predominantly affects the light harvesting phycobilisomes at PSII (Nikolova *et al.*, 2017). Accordingly, the transcriptional regulation of photosystem genes appears to be temperature dependent. The difference in PS regulation between *Arabidopsis* and *C. merolae* upon temperature stress might either represent evolutionary distances (see Chapter 1.4 for differences in photosystem architecture between green and red lineage). Alternatively, different kinds of temperature stress may entail specified responses. Together, this indicates the possibility that the photosystem-containing chloroplast is involved in temperature sensing.

The present project described essentially the entire photosystem machinery as trainable which may result in the energetic housekeeping of the whole cell. Trainability includes reaction centres of PSI/II, the ATP synthase complex, the light harvesting complexes (phycobilisomes), components of the thioredoxin system as well as metabolites involved in the maintenance of the electron transport chain (menaquinones). By maintaining the photosystems intact, the risk of secondary light-driven damages are minimized additionally, as photosystems are principal sites of heat-induced reactive oxygen species (ROS) generation (reviewed in J. Zhao *et al.*, 2020).

A major cellular regulator in the optimization of photosynthesis under fluctuating environmental conditions is the thioredoxin system, which aids in balancing the activities between light and carbon-fixation reactions in chloroplasts (Gelhaye *et al.*, 2005). Notably, among CP trainable genes was *trxM* (Figure 3.17) Generally, m-type thioredoxin represents the most conserved thioredoxin and regulates the redox state of Calvin-Benson-Bassham cycle enzymes in *Cyanobacteria* as well as in chloroplasts, defining these as potential targets also in *C. merolae*. Furthermore, *trxM* mutants in the Cyanobacterium *Synetocystis* showed elevated ROS levels and was more sensitive to stress than the WT (Mallén-Ponce *et al.*, 2021). Thus, the trainablility of *trxM* might indicate that oxidative stress occurs during heat stress also in *C. merolae*. In consequence, oxidative stress resistance might illustrate an additional regulatory pathway for the *C. merolae* thermomemory. In addition to its contribution to the memory process, *trxM* might represent one system which connects nuclear and CP actions. Thus, the higher induction and expression of the nuclear thioredoxin reductase described in the present study while nuclear thioredoxin transcription is not substantially altered, suggests the necessity of chloroplast-nuclear crosstalk. It cannot be excluded that the CP thioredoxin needs the nuclear thioredoxin reductase in order to fulfil its reductive function on heat-damaged proteins and/or the defeat of heat-generated ROS. The existence of transit peptides of the nuclear reductase to the chloroplast or *vice versa* has to be investigated in order to speculate about the interaction of the two components. It is

possible that a type of oxidative stress co-occurs with heat stress, that it is probably met in the chloroplast and contributes to the PLT over T fitness.

Among CP metabolic trainable genes are enzymes involved in the biosynthesis of the amino acids (aa) tryptophane, valin/leucin/isoleucine, arginine and pyrimidine (Figure 3.20). In *C. merolae*, the aa research is not well studied, but the anti-stress contribution of some aa metabolites is known in plants. Accordingly, the tryptophane metabolites and potent antioxidants melatonin and serotonin play important roles in the plant´s life cycle including the response to thermal stress (Erland *et al.*, 2018; Erland & Saxena, 2019). Thus, training the tryptophane biosynthesis pathway may have broad beneficial downstream effects for *C. merolae* heat stress memory, too, which could eventually lead to superior physiological responses. Yet, as the presented results are considered as first indications, proteomic and metabolomic studies might aid in further reconciling any contribution of tryptophan metabolites to thermomemory in *C. merolae*.

Taken together, this project defined a vast trainable reaction of the CP genome, which includes broad parts of the photosystem machinery as well as of the genetic system, comprising subunits of the RNA polymerase PEP, transcriptional (*ycf27/29*) and translational regulators (*infB/tsf/tufA*). Furthermore, a heat induction, but no trainability on gene level was detected for the sigma factor *SIG3* suggesting this so far undescribed sigma factor to be involved in transcriptomic heat responses. However, the SIG3 protein accumulated during priming might still be available in the cell over lag phase. Consequently, it may contribute to faster gene expression responses upon recurring heat. Together, these components may be responsible for the targeted response to recurring heat which is necessary to survive a recurring heat stress of lethal magnitude. In addition, the plastid thioredoxin gene *trxM*, as major redox housekeeper, might particularly contribute to the protection against imbalances in cellular homeostasis created by recurring heat. Notably, components of the same pathways are distributed to nuclear as well as chloroplast genome implying inter-organellar communication. The initial heat stress memory characterization provided by this project serves as basis for exciting further research which should concentrate on major regulators of heat stress memory and include the question about which responses are inherent to general stress (memory) reactions and which are specific to heat stress (memory). It will be interesting to know, whether the transcriptional upregulation seen in the context of thermomemory is due to elevated repair demands upon heat acclimation or rather is intrinsically regulated. Furthermore, the order of events happening are unsolved and will potentially also reveal the so-called first responders, or better, heat sensors in *C. merolae*.

c. Possibilities of nucleus and chloroplast crosstalk during thermomemory

The discovery of trainable genes distributed to both, the nucleus as well as to the chloroplast (e.g. for the tryptophane biogenesis pathway to name just one) poses the question how memory processes in both organelles might be interconnected. Before transferring the information about heat stress, it has to be perceived and recognized. Temperature sensing mechanisms are still largely unknown across species and must be activated directly by temperature without the need of upstream activating signals. It is important to stress, that heat affects the cell as a whole. Thus, it is likely that heat is sensed at multiple locations in the cell, including the nucleus and chloroplast. In general, CPs can be viewed as “highly important decision making platform in the cell under heat stress“ (A. Z. Sun & Guo, 2016) as fast biochemical processes are affected, suggesting that heat might be prominently perceived in the CP and transferred to the nucleus, i.e. retrograde signaling pathways. In the retrograde pathway, intermediate signaling molecules/proteins are used to perceive information e.g. in the chloroplast and to trigger reactions in the cell nucleus. Recently, potential candidates for heat stress retrograde signalling pathways were comprehensively discussed and include reactive oxygen species (ROS), tetrapyrroles, carotenoids, and organellar gene expression (A. Z. Sun & Guo, 2016). As component of the latter, the ribosome subunit RPS1 was shown to induce nuclear HSFA2 transcription and subsequently entail the transcription of heat responsive genes in *Arabidopsis* (Danilova *et al.*, 2018). A putative RPS1 orthologue exhibits trainability in *C. merolae*, too, which defines the *rps1* gene as exciting regulatory candidate which may confer chloroplast information to the nucleus as well as be responsible for the activation of a proper heat response (H.-D. Yu *et al.*, 2012).

Proteins like the CP encoded elongation factor TU (EF-Tu, tufA) represent an additional and even more complex way in sensing heat, responding to it, and transferring information from one organelle to the other. Thus, EF-Tu is directly affected by heat as it aggregates and furthermore it affects, like RPS1, directly nuclear HSF transcriptional responses (X. Li *et al.*, 2018). Albeit not reaching significant trainability (i.e. difference directly between PLT and T expression), *tufA* is significantly induced in the PLT condition, but not in the T condition designating it potentially-trainable. The possibility remains, that the tufA aggregation triggered by the priming stimulus (and leading to tufA activity) is sustained during lag phase thus ensuring a quick second nuclear response upon triggering. After verifying its trainability in qRT-PCR, it would be interesting to proof the possibility of tufA and RPS1 migration to the nucleus via nuclear localization sequence (NLS) analysis. In the long-term genetic mutants might indicate, whether these proteins are needed for *C. merolae* thermotolerance and memory.

While the current project does not provide any information about produced ROS levels during thermomemory, the contribution of these signaling molecules to transfer information from CP

to nucleus is a possibility. The finding of *trxM* thioredoxin as memory gene emphasizes this possibility. ROS are constantly produced as result from the highly energetic photosynthesis reaction and their production increases under stress situations. After generation they can diffuse from their site of generation and consequently elicit a different set of signaling events under a wide range of biotic and abiotic stress conditions (Miller *et al.*, 2008; Mittler *et al.*, 2004; Neill *et al.*, 2002; Suzuki *et al.*, 2012). Albeit their generalized effect might be considered harmful, ROS represent potent downstream acting molecules. Accordingly, a link has been reported between ROS and HSF-HSP activation in response to oxidative stress: H_2O_2 as most stable ROS molecule induces the expression of HSPs via HSF activation early during heat (Volkov *et al.*, 2006). Both, HSF and sHSP activation could be identified for the thermomemory in *C. merolae*, yet the activation processes remain unknown. Thus, whether the observed trainability of the CmsHSP´s depends on ROS (H_2O_2) dependent HSF activation in *C. merolae* could be investigated by the usage of ROS production inhibitors or ROS scavenging agents during the HSMA in *C. merolae*.

Notably, the three discussed retrograde candidates - RPS1, tufA, and ROS (H_2O_2), - have a common nuclear encoded and located target class: Heat shock transcription factors (HSF). Intriguingly, HSFs themselves represent promising candidates to sense heat, too, since HSFs are activated by hyperphosphorylation and/or dissociation from HSPs as response to elevated temperature (Sorger & Nelson, 1989; Sorger & Pelham, 1988). Following their activation, they initiate the expression of genes necessary for a successful stress response, such as small heat shock proteins (sHSP). As mentioned earlier, mRNA levels of two *C. merolae* HSF´s were induced during priming (Figure S3). Thus, they represent potential candidates for triggering the *C. merolae* heat response as well as representing possible targets of retrograde signaling. Intriguingly, the circle of chloroplast and nuclear responses might close here with the fact, that the trainable heat shock locus around *CmsHSP1/2* possesses a heat shock element (HSE) in the promoter site. Hence, it remains possible that the heat information perceived in the chloroplast is transferred to nucleus and reinforces HSF heat activation leading to the expression of CmsHSP2. This protein in turn migrates back to the chloroplast and may confer heat-protective functions. Biologically, in the chloroplast CmsHSP2 may stabilize components of the genetic and photosystem via its chaperone activity (Waters & Rioflorido, 2007), two functional classes of genes showing strong trainability in the *C. merolae* thermomemory (Table 3.3). In line with this result, the small heat shock protein 21 could be linked to the regulation of chloroplast transcription regulation by stabilizing PEP polymerase via the nucleoid protein pTAC5 (Zhong *et al.*, 2013). Furthermore, sHSPs act as molecular chaperones on thylakoid membranes via stabilizing the thermolabile PSII and electron transport during heat stress contributing to thermotolerance in plants (Heckathorn *et al.*, 1998; X. Hu *et al.*, 2014). Accordingly the

trainability of *CmsHSP2* might execute similar functions essentially contributing to *C. merolae* chloroplast thermomemory. However, as this mechanism of action is highly speculative it should be confirmed by the proof of interaction between PEP and CmsHSP2 to test whether the products of those genes interact in *C. merolae* under heat conditions, too. Interestingly CmsHSP2 is an H3K27me3 target, notably the only one among memory genes. Hence, the contribution of chromatin modifications to the memory at this commonly regulated heat-shock locus at chromosome 10 might pose another layer of regulation.

Taken together, transcriptomic analysis in combination with literature review opens precise possibilities how the information about heat could be exchanged and interconnected between the two trainable genomes in *C. merolae*. Chloroplast translation, including the translation factor EF-Tu (*tufA*) as well as the ribosome protein S1 (*rps1*) might induce nuclear responses via HSF activation, reinforcing parallel heat perception directly by HSF in the nucleus. Subsequently, heat protection might be transferred to the chloroplast via the small heat shock protein CmsHSP2. Interestingly the CmsHSP2 gene locus is a target of the H3K27me3 histone modification, and since *C. merolae* cells lacking genomic H3K27me3 levels are compromised in thermomemory formation, special attention is drawn to *CmsHSP2* which might be central not only to the commonly regulated region on chromosome 10, but also central to thermomemory formation in *C. merolae*.

4.4 On the roles of PRC2 in *C. merolae* thermomemory

The ability of *C. merolae* to memorize a priming heat stress leading to superior survival even if the stress reaches lethal doses could be described for the first time in this study. The initial interest in such memory response was prerequisite to answer the hypothesis, that a working PRC2 system in *C. merolae* may shape the regulation of stress memory mechanisms like it was detected in higher plants (Sani *et al.*, 2013; Yamaguchi & Ito, 2021). As a result, testing the PRC2 mutant Δ*E(z)* in the established HSMA revealed aberrant memory capabilities. The following Chapter makes an attempt to specify critical processes or points which might be regulated by H3K27me3 during thermomemory of *C. merolae*.

4.4.1 H3K27me3 may participate in the transduction of heat information to daughter cells

The inclusion of the H3K27me3 histone methylase *E(z)* homolog deletion strain in *C. merolae*, i.e. Δ*E(z)*, in the HSMA resulted in aberrant memory when a cell division-including lag phase was applied, i.e. 24h. Both, PSII fitness and survival growth were altered in Δ*E(z)* compared to WT (Figure 3.5) revealing the need of a working PRC2 system to achieve full thermomemory over this lag phase length. Contrarily, when administering a short lag phase of 2h the memory was not changed by the *E(z)* mutation specifying that *E(z)* is not necessary to establish the priming benefit but rather involved in maintaining it over longer periods of time. Assuming that H3K27me3 regulates transcription, this result might be reasonable, because it is likely that the much faster metabolomic and proteomic responses precede transcriptional ones. Thus, it is possible that the memory mechanisms underlying the 2h and 24h lag phase differ in composition, while the short (2h) memory depends more on fast biochemical responses and the long-term (24h) memory relies on transcriptional memory partially generated by H3K27me3. In line with this assumption, the memory genes identified for the 24h lag phase did not reach substantial trainability during the 2h lag phase suggesting that indeed 2h is not long enough to establish transcriptional memory.
While the Δ*E(z)* mutant showed significantly worse memory during a 24-hour lag period at the level of PSII fitness, the result was less clear on survival plates. There, PLT culture growth differed in only one replicate between WT and Δ*E(z)* reflected in growth for WT_PLT whereas nothing grew for the Δ*E(z)*_PLT sample (Figure 7, 24h_B). Experimentally, this could either be a consequence of varying triggering temperature in the culture (probably elevated) or due to different amounts of cells within the culture volume applied on survival plate (caused by non-homogenous cell densities). It is tempting to assume, that a critical cell concentration needed for growth was not reached in E(z)_PLT, which could be due to an

overall lower number of primed cells in the loaded sample and that this difference only becomes apparent at higher selective pressure, i.e. higher triggering temperature. The contribution of H3K27me3 dynamics to transgenerational stress adaptation has been described previously (J. Liu et al., 2019) and supports the suggestion that the Δ*E(z)* strain has limited abilities in passing priming information to newly developed daughter cells. As a result, in the situation of 24h_A the response of daughter cells might be hidden under the still existing directly primed Δ*E(z)* cells and is only detectable when cell numbers are decreased or the selective pressure stronger. To prove this assumptions, future HSMA´s should increase the triggering temperature and include dilution series of cultures applied on survival plates to confirm differences in PLT survival growth between WT and Δ*E(z)* strain.
In summary, these results strongly suggest that PRC2 is indeed required for a complete memory of heat stress – as soon as the duration of the lag phase includes mitotic divisions. It will be interesting to proof whether *E(z)* specifically acts on cells of the next generation or whether it contributes to the long term revision of the transcriptome. Evidence for this question might only be provided by the unambiguous identification of daughter cells at the end of the HSMA like it was proposed in Chapter 4.2.1.

4.4.2 H3K27me3 acts preferentially on cluster 4 thermomemory genes

In line with alterations in physiological memory responses in Δ*E(z)*, a set of WT-trainable genes remained un-trainable in the mutant upon a 24h lag phase. The finding, that 8 out of 59 WT-memory genes could not establish transcriptional memory in Δ*E(z)* (Figure 3.8, Figure 3.10) may explain the aberrant physiological memory found for the mutant. Interestingly, four of these genes were allocated to cluster 4 memory genes. Genes within this cluster are (only slightly) induced upon priming and point to sustained induction, as their transcription levels did not fully return to control levels during lag phase reflected in higher levels for primed cells (PL) compared to the non-primed control cells (NL). For two of these genes, *CMF092C* as well as *CMP040C*, H3K27me3-dependency could be verified in qRT-PCR experiments for the 24h lag phase. Here, the trainability seen for WT could not be detected in Δ*E(z)* any longer. Even further, the trainable trend detected in the WT for these two genes in the short (2h) and prolonged (48h) lag phase was vanished in the Δ*E(z)*. Thus it can be speculated, that PRC2 may act specifically on memory genes which depend on the sustained induction during lag phase. Yet, transcriptomic analyses did not significantly describe sustained induction of any memory gene as none of the genes was differentially expressed between PL and NL cells in neither genotype. Subsequently, the assumption that PRC2 might regulate sustained induced memory genes should be validated in qRT-PCR experiments specifically describing mRNA levels for *CMP040C* as well as *CMF092C* during the fulfactorial HSMA to

finally conclude whether they maintain elevated induction rates throughout the lag phase of 24h. Only then further steps towards understanding the causes of the sustained induction can be made. Although this project does not allow to draw conclusions about mechanisms leading to sustained induction literature provides information about potential processes underneath. In *Arabidopsis,* heat exposure triggered the transcriptional induction of HSPs which was accompanied by a gain in the activating histone mark H3K4me3. Both, transcriptional induction and H3K4me3 occupancy persisted during lag phase (Lämke *et al.*, 2016). Accordingly, monitoring the histone mark H3K4me3 in response to heat stress on *CMP040C* and *CMF092C* could provide further insights on this topic. The analysis of H3K27me3 might be included, too, since the trainability of both genes depends on genomic H3K27me3 levels, although neither of those two genes is defined as direct target. Yet, the chromatin environment of the up- and downstream regions around both genes should be observed for any intergenic H3K27me3 peaks which might contribute to the regulation of *CMP040C* and *CMF092C* in *C. merolae*. The relevance of H3K27me3 in stress responses was reinforced by a recent study describing H3K27me3-removal accomplished by demethylases of the *jumonji* family as essential mechanisms in order to gain the activating mark H3K4me3 at heat stress responsive loci in *Arabidopsis* (Yamaguchi & Ito, 2021). The *C. merolae* genome comprises five putative demethylase genes of which currently three are deleted in the Schubert lab. As a promising first result, one of those (Δ*CMO204C*) shows aberrant thermomemory responses (Schubert laboratory, S. Reinhard). Hence, it cannot be excluded that, similar to what was found for *Arabidopsis*, active H3K27me3 demethylation might contribute to *C. merolae* thermomemory additionally.

The retrotransposon *CMT010C* represents one of two lastingly repressed memory genes following priming. The finding that its repression as well as memory behaviour depends on H3K27me3 in *C. merolae* designates H3K27me3 as potent contributor to the defence against uncontrolled transcriptional events of REs in the context of repetitive heat stress responses. In general, the transcriptional control of genomic RE transcription governed by H3K27me3 under control conditions in *C. merolae* (Chapter 4.1) might gain further importance when heat-stress occurs, since heat is a known trigger for RE activation (Cavrak *et al.*, 2014; Pecinka *et al.*, 2010; L. Sun *et al.*, 2020). As stated earlier, the current annotation file does not include all repetitive elements and thus the analysis of RE behaviour upon repetitive heat is incomplete. Hence, the conjecture/chance remains, that many more repetitive sequences are regulated in a memory dependent manner (and via H3K27me3 occupancy). By high induction upon priming it is possible, that they remain active over the course of time when a major controlling histone mark like H3K27me3 is absent. This uncontrolled RE transcription could highly interfere with a regulated and targeted memory response eventually leading to a worse thermomemory response like it was shown for Δ*E(z)*. Thus, it can be concluded that

H3K27me3 is involved in targeted and lasting RE repression during thermomemory in *C. merolae*. After this sequence class has been integrated into the annotation file further statements will be drawn about the scope of H3K27me3 contribution to trainable repression of REs and answer the question whether H3K27me3 is key or assistant to temper RE expression in response to heat.

4.4.3 The chloroplast-directed small heat shock protein CmsHSP2 as central H3K27me3-targeted and regulated memory gene?

During the course of this project the small heat shock protein CmsHSP2 encoding gene *CMJ101C* repeatedly attracted attention. First, it is the only thermomemory gene in *C. merolae* which is directly targeted by H3K27me3 (Figure 3.8). Furthermore, it is centre to the region showing joint regulation on chromosome 10, the heat shock locus (Figure 3.13). Lastly, it is among the identified trainable genes losing trainability in $\Delta E(z)$ (Figure 3.10). This gene is not only rapidly and strongly induced (and translated) upon a heat shock and even stronger when the heat reoccurs, but in parallel its gene product CmsHSP2 probably harbours the ability to be transferred to the chloroplast (Kobayashi *et al.*, 2014) and possibly stabilizes genetic system and photosystem components (Heckathorn *et al.*, 1998; X. Hu *et al.*, 2014; Zhong *et al.*, 2013). In connection, these results open the intriguing possibility that H3K27me3 by regulating the expression of *CMJ101C* influences (a) the transcription of the entire heat shock locus on chromosome 10 and (b) might shape the heat memory response detected in the chloroplast (Chapter 3.4). Given its central role in the thermomemory process in combination with being enriched with H3K27me3 under WT control conditions, an important next step will be to dissect the H3K27me3 dynamics at *CmsHSP2* during the HSMA. As discussed in the context of potential mechanisms leading to transcriptional memory of the heat shock locus at chromosome 10, chromatin modification and - remodeling are known mechanisms to confer transcriptional memory. Accordingly, chromatin-immunoprecipitation (ChIP) experiments will reveal whether the *CmsHSP2* locus in a similar manner is subject to histone modification changes during the HSMA in presumingly gaining activating histone modification marks (H3K4me3, H3K36me3) and possibly losing or attenuating its repressive H3K27me3 peak. To extent the hypothesis, it is also possible that this region is poised for transcription in a bivalent chromatin state after priming heat.

Whatever the precise mode of action by which CmsHSP2 is involved in the thermomemory of *C. merolae*, its contribution to it is more than likely. Yet, as stated above, further research should aim in proving that CmsHSP2 indeed localizes to the chloroplast. By functional disruption of *CmsHSP2* its precise contribution to heat stress (memory) responses should be

monitored. Finally, chromatin modifications likely directly contribute to the thermomemory at the *CmsHSP2* containing heat shock locus. Dissecting the chromatin modification composition and dynamics at site (ChIP) will contribute to the molecular understanding of a central trainable locus eventually verifying the direct contribution of epigenetics in the thermomemory response of *C. merolae.*

5 Conclusion

The aim of this project was to delineate the role of the major epigenetic regulator PRC2 and its hallmark H3K27me3 in the unicellular red alga *Cyanidioschyzon merolae*. In order to test whether H3K27me3 occupancy directly entails transcriptional regulation, transcriptomic analyses of the novel PRC2 loss-of-function mutant Δ*E(z)* and wild type cells revealed a substantial role of H3K27me3 in the repression of repetitive sequences (RE), as 28% of all REs were upregulated in the mutant. Since the affected REs were predominantly found at subtelomeric regions, the silencing function of PRC2 may contribute significantly to maintain *C. merolae* genome integrity. Furthermore, as the mutant exhibited decreased DEG numbers as well as poorer photosystem II fitness upon heat exposure, PRC2 seems to aid in establishing an appropriate heat response.

In the context of answering the question of whether stress memory processes are shaped by H3K27me3 in *C. merolae*, this work demonstrated the ability of a unicellular photoautotrophic eukaryote to remember heat stress, reflected by superior survival of a lethal (triggering) heat stress when cells have experienced a preparatory heat stimulus (priming). This phenomenon is termed thermomemory and in *C. merolae* persisted for 24h which included at least one cell division. However, when approx. 3 mitotic divisions parted priming from triggering stimulus the priming benefit was lost in *C. merolae*. Moreover, thermomemory was accompanied by the revised transcriptional response of 59 genes, coined trainable. The trainable gene set originated from both, the chloroplast and nuclear genome and included a nuclear heat shock locus on chromosome 10. This region comprises five genes surrounding the only two genomic small heat shock proteins CmsHSP1/2 (encoded by *CMJ100C* and *CMJ101C*). The trainability of these two might substantially contribute to the revised heat shock response as those two proteins inhabit the potential to spread the heat response throughout the cell: to cytosol and mitochondrium/chloroplast, respectively. Consistent with the heat sensitivity of the chloroplast, the plastid genome contributed extensively to the trainable gene fraction with dominant responses of photo- as well as genetic system components. Together, analyses suggest a contribution of chloroplast transcription and translation to *C. merolae* thermomemory. Moreover, possible candidates for linking nuclear and chloroplast thermomemory reactions were discussed. Accordingly, the small heat shock protein *CmsHSP2* might possibly transfer heat shock responses generated in the nucleus to chloroplast via its putative transit peptide. On the other hand, two trainable genes encoded in the plastid - homologs of the translational regulator *tufA* as well as ribosome protein encoding *rps1* orthologue - have been implicated in retrograde signalling processes in relation to heat stress and thus bear the potential to confer similar pathways in *C. merolae*, too. As potent third mechanism, a preparatory heat stimulus negatively and lastingly primed RE expression, leading to sustained repression of potentially genome threatening elements.

The description of thermomemory served to address the question of whether PRC2 plays a role in regulating stress memory mechanisms in *C. merolae*. To test PRC2 influence on memory fomration, the PRC2 mutant Δ*E(z)* was exposed to the established memory scheme. As a result, the absence of genomic H3K27me3 levels lead to a reduced ability of *C. merolae* to maintain memory of previous heat exposure beyond mitosis. In consequence, these results indeed identify PRC2 as contributor to thermomemory processes in *C. merolae*. When cells did not divide during lag phase (2h lag) the memory was not altered in the mutant, specifying the *E(z)* function as being relevant in the maintenance of memory rather than establishing it. The loss of physiological memory could be linked to a set of genes losing trainability in the mutant. Furthermore, opposing to the WT situation, in the mutant candidate RE encoding genes could not be repressed following priming heat, suggesting a potent function of H3K27me3 in preventing enhanced transposition of these genome threatening elements in response to heat stress. As for the heat stress memory transcriptome analysis the published annotation file was used which does not include the entirety of REs in *C. merolae*, the interesting possibility remains, that more than the described REs are subject to H3K27me3-dependent repression following heat stress.

Given the wealth of information collected for potential thermomemory mechanisms in *C. merolae*, the main results of this project were related to known heat shock responses and summarized in a cell model (Figure 5.1). On the one hand, this model aims to characterize general thermomemory mechanisms of *C. merolae*. On the other hand, an attempt is made to elaborate the influence of PRC2 on memory formation. In summary, H3K27me3 is indeed necessary to establish a complete thermomemory response in *C. merolae*.

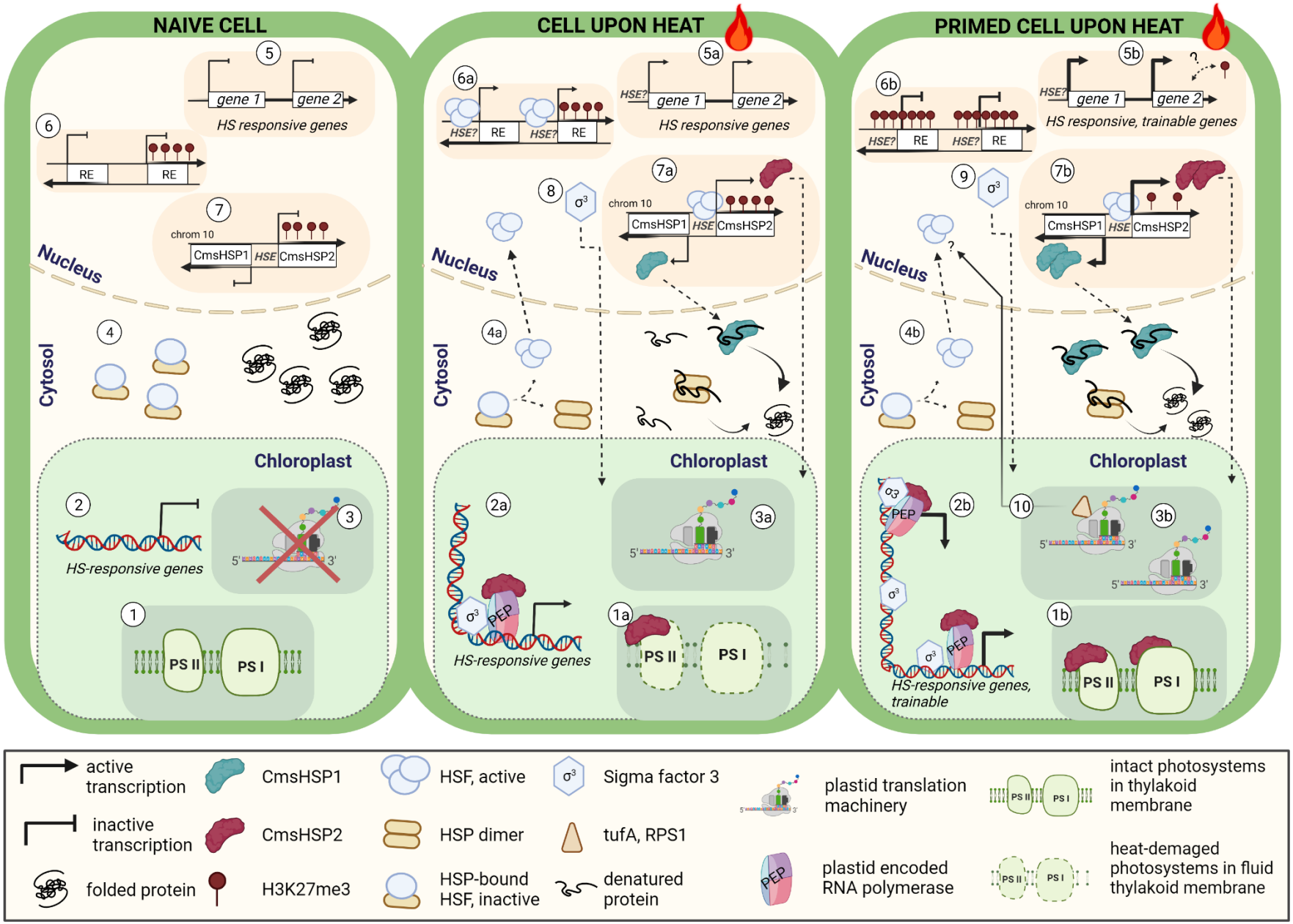

NAIVE CELL
CELL UPON HEAT
PRIMED CELL UPON HEAT
Nucleus
Cytosol
Chloroplast
gene 1
gene 2
HS responsive genes
HS responsive, trainable genes
RE
HSE?
chrom 10
CmsHSP1
HSE
CmsHSP2
HS-responsive genes
HS-responsive genes, trainable
PEP
PS II
PS I
active transcription
inactive transcription
folded protein
CmsHSP1
CmsHSP2
H3K27me3
HSF, active
HSP dimer
HSP-bound HSF, inactive
Sigma factor 3
tufA, RPS1
denatured protein
plastid translation machinery
plastid encoded RNA polymerase
intact photosystems in thylakoid membrane
heat-demaged photosystems in fluid thylakoid membrane

Figure 5.1 Schematic model of *Cyanidioschyzon merolae* heat, and thermomemory responses
NAÏVE CELL: In the non-stressed cell, photosystems are intact and show control PSII fitness (Figure 3.4C, N condition) (**1**), plastid heat shock (HS) responsive genes are not induced (**2**), and in consequence translation is inactive (**3**). In the cytosol, proteins are folded, and inactive heat shock transcription factors (HSFs) bound to heat shock proteins (HSPs) (Sorger & Nelson, 1989; Sorger & Pelham, 1988) (**4**). Nuclear HS responsive genes show basal transcription levels prior to heat stress (**5**). Genomic repetitive elements (REs) are silenced. 50% of REs carry the H3K27me3 under control conditions (Mikulski *et al.*, 2017, Figure 3.1) (**6**). The small HSPs CmsHSP1/2 are encoded next to each other on opposite strands and share a common promoter which has an identified heat shock element (HSE). *CmsHSP1/2* show low expression rates under control conditions (Kobayashi *et al.*, 2014) CmsHSP2 is enriched with H3K27me3 under control conditions (Mikulski *et al.*, 2017) (**7)**.
CELL UPON HEAT: When heat stress (60°C for 15 min) occurs photosystems (PS) suffer damages (Biswal *et al.*, 2011; Murata *et al.*, 2007; Yamamoto *et al.*, 2008) entailing decreased PSII fitness (Figure 3.4C) (**1a**). Elevated repair demands initiate expression of plastid HS-responsive genes (**2a**). Photosystem genes are upregulated, as well as the genetic machinery, including RNA polymerase PEP, transcriptional and translational regulators, as well as the ribosome machinery (Figure 3.15). In the nucleus, sigma factor 3 (SIG3) is expressed (Figure 3.16) which is needed to specify transcription governed by PEP (**8**). The translational machinery is upregulated to gain protein levels which replenish heat damaged cellular components (**3a).** HSFs are activated by dissociation from HSPs. In parallel, nuclear expression of HSFs is upregulated by heat (Figure S3). While the HSPs fulfil their protein-protective roles in the cytosol, HSF trimers migrate to the nucleus (Sorger & Nelson, 1989; Sorger & Pelham, 1988) (**4a**). HSFs recognize target genes by HSEs in promoters and contribute to transcriptional activation of HS-responsive genes (von Koskull-Döring *et al.*, 2007) (**5a**). Some REs show high induction rates upon heat (Figure 3.11) possibly via HSF-mediated transcription and putative HSEs in promoter sites (Cavrak *et al.*, 2014). RE heat-induction seems H3K27me3-independent (**6a**). *CmsHSP1/2* locus is rapidly and strongly unregulated during HS probably mediated by HSF. CmsHSP1 stabilizes denatured proteins in the cytosol and promotes refolding. Contrarily, for CmsHSP2 protective roles of in the plastid are suggested like shown for other sHSPs (Kobayashi *et al.*, 2014; Zhong *et al.*, 2013) Together, if the heat lasts too long (1 hour) cells enter growth arrest (Figure 3.4).
PRIMED CELL UPON HEAT: In a cell which has previously experienced a priming heat stimulus (57°C, 30min) and which 24h later is repeatedly exposed to heat (60°C, 15 min), sHSP, like CmsHSP2, could remain in the cell associated with photosystems (Heckathorn *et al.*, 1998) (**1b**). SIG3 might stay associated with its target promoter sequences in the chloroplast genome contributing to faster PEP transcription (**9**). Notably, *SIG3* is induced upon heat, yet the induction rate is independent of a former priming event (non-trainable, Figure 3.16). The chloroplast genome shows significant trainability (Figure 3.15) (**2b**). Main photosystem genes entail higher levels of photosystem proteins which in turn contribute to photosystem stability. The genetic system shows significant trainability (Figure 3.17), including PEP, transcription factors, and most of the translation machinery (Figure 3.15) (**3b**). The elongation factor *Tu* (EF-Tu, tufA) which represents trainability and is associated with plastid translation under heat stress, may shape retrograde signalling pathways, similar to the ribosome subunit RPS1 orthologue (rps1) (X. Li *et al.*, 2018; H.-D. Yu *et al.*, 2012). However, it is unclear whether and how these two candidates indeed induce retrograde signals (**10**). HSFs are activated and may stay associated with HSE of nuclear HS responsive genes (**4b**). A set of 9 nuclear genes show trainability for heat. In specific cases, this trainability depends on H3K27me3 genomic levels (Figure 3.10) (**5b**). Trainable repression was detected for both, REs decorated with and without H3K27me3 (Figure 3.11). This may be achieved by addition of repressive chromatin modifications (H3K27me3) on RE bodies (**6b**). Transcription of both sHSP genes (*CmsHSP1/2*) is trainable spreading the information to cytosol and chloroplast. Here, trainability might be gained by lasting HSFs levels. Furthermore, H3K27me3 levels of *CmsHSP2* may be lost and activating histone marks added (**7b**). Created with BioRender.com

6 Future Perspectives

The novel description of *C. merolae* thermomemory process raises many questions about its regulation. Using unsynchronized cell cultures, in this project the thermomemory response could be described as PRC2-dependent for the 24h lag phase. However, although mitotic division has happened during the 24h lag phase, the used set-up does not allow for a precise statement about the mitotic inheritance of memory. After all, the possibility remains, that cells, which have directly experienced the priming stimulus, are still existing in the culture upon triggering stimulus. The most elegant way to determine, whether the primed *C. merolae* cell is indeed capable of transferring the information about experienced heat to daughter cells demands **live imaging** allowing for the monitoring of daughter cells originating from a fixed and marked parental cell. At the same time, definite proof about whether or not PRC2 is contributing to this transgenerational inheritance can be tested. In addition to the question about heritability, it cannot be excluded that the susceptibility for priming may be cell-cycle stage dependent and that the sensitivity to a heat stress differs during the mitotic cycle. **Synchronized growth** of *C. merolae* cells and the exposure to priming heat at single cell cycle stages might provide first insights into this topic.

Although thermomemory was described as partially H3K27me3-dependent, it is unclear how H3K27me3 contributes to this mechanisms. Accordingly, the exact relationship between genomic levels of H3K27me3 and trainability of both, *CMP040C* and *CMF092C* remains unclear. Whether the WT trainability of these two genes depends on **sustained activation** during lag phase can be clarified by monitoring the expression dynamics during and/or after lag phase. In addition, ChIP experiments, either for individual genes (ChIP-RT qPCR) or genome-wide (ChIP-seq) will reveal whether the priming heat information is stored on chromatin level, possibly mediated by H3K27me3 levels. It would be interesting to compare the dynamics of H3K27me3 and that of H3K4me3 during stress responses, as their coexistence is a known feature of genes poised for activation. Furthermore, this project could not fully describe the response of RE to repetitive heat, as the current annotation file does not include the entirety of this gene element class. Thus, once the new annotation file is reviewed, the heat stress memory transcriptome analyses shall include the transcriptional response and dynamic of all **REs during the HSMA** to confer whether the finding of H3K27me3 contributing to targeted RE repression upon heat stress can be generalized for the *C. merolae* genome. The question about regulatory mechanisms of heat stress involves the question about **heat sensory systems**. Potential candidates have been discussed in this project, such as HSFs and tufA. Still, further research into this direction might proof promising as heat sensory systems are described in other organisms already. In addition, the results of this study point to **SIG3, as a putative sigma factor relevant for heat stress**. A first step of further analysis in this regard is to verify possible SIG3 protein accumulation after

(priming) heat and to determine whether SIG3 levels are stably maintained across the lag phase.

Finally, the overarching **regulators of thermomemory** remain obscure. The contribution of PRC2 to thermomemory suggests the involvement of epigenetics to the maintenance of memory. In extension to this finding, other chromatin modifying actors may additionally influence this response, such as histone demethylases or chromatin remodellers.

Overall, this project contributes to the sparse knowledge on the function of epigenetics in basal photoautotrophic eukaryotes by defining the role of PRC2 in the red algal cell *C. merolae*. First, it substantially contributes to genomic repetitive element silencing with preferential action at the subtelomere which may contribute to genome stability. Second, PRC2 is involved in the basal heat response as well as in the maintenance of acquired thermotolerance, defined as thermomemory. Interestingly, a putative link between both PRC2 functions may be found in the lasting repression of REs following heat-induction. In the context of addressing PRC2 relevance in *C. merolae*, its inherent ability to memorize heat was described, and an initial characterization revealed a contribution of both, nuclear and plastid genome to the revised transcriptional response upon recurrent heat. In the context of evolution, it will be interesting whether the observed thermomemory is specific to *C. merolae* or whether it can be similarly be observed under the given conditions in its sister species ***Galdieria sulphuraria***. Moreover, it will be interesting whether PRC2 contributes to such phenomenon in *G. sulphuraria*, too.

Knowledge of PRC2 function in *C. merolae* contributes to our understanding of the functional evolution of Polycomb group proteins. The postulated functions of PRC2 in *C. merolae* may reflect the phylogenetic position of this red alga, which is located at the boundary between plants and lower eukaryotes. Thus, *C. merolae* combines the known PRC2 functions in lower/unicellular and higher/multicellular organisms, as it shapes genome stability and simultaneously (or thereby?) regulates environmental responses.

7 Supplementary Chapter

List of supplementary tables

List of supplementary figures

Table S1: Read counts and general statistics of Assignment (featureCounts) and Alignment (STAR) rates of reads

Scores are listed for each sequenced mRNA sample (forward and reverse read combined). Sample name is given as “SequencingNumber_replicateGenotype_condition”. (M: million reads, EZ: Δ*E(z) C. merolae* mutant, WT: 10D *C. merolae*, conditions as in experimental scheme of Figure 3.7)

Sample Name	M reads	% Assigned	M Assigned	% Aligned	M Aligned
10_1EZ_PL	14,80	68.3%	10,60	94.8%	14,10
11_1EZ_T	10,10	68.2%	7,20	93.9%	9,50
12_1EZ_PLT	10,00	66.9%	7,00	93.7%	9,40
13_2WT_N	10,40	67.2%	7,20	94.0%	9,80
14_2WT_P	13,70	66.6%	9,40	95.0%	13,00
15_2WT_NL	11,10	68.4%	7,90	94.3%	10,50
16_2WT_PL	10,00	65.6%	6,80	93.3%	9,30
17_2WT_T	11,70	69.8%	8,50	94.2%	11,00
18_2WT_PLT	13,10	69.1%	9,40	94.7%	12,40
19_2EZ_N	10,70	63.3%	7,10	92.5%	9,90
1_1WT_N	11,60	67.1%	8,10	93.5%	10,90
20_2EZ_P	10,60	61.4%	6,70	94.8%	10,00
21_2EZ_NL	12,10	68.7%	8,60	94.7%	11,40
22_2EZ_PL	10,10	68.3%	7,20	94.5%	9,60
23_2EZ_T	20,80	69.7%	15,10	95.4%	19,90
24_2EZ_PLT	10,20	67.4%	7,10	94.6%	9,60
25_3WT_N	9,90	67.8%	7,00	94.5%	9,40
26_3WT_P	14,50	57.7%	8,50	95.9%	13,90
27_3WT_NL	12,80	69.7%	9,20	94.7%	12,10
28_3WT_PL	21,60	68.2%	15,30	94.3%	20,30
29_3WT_T	10,60	70.2%	7,70	94.1%	10,00
2_1WT_P	11,30	56.5%	6,50	95.6%	10,80
30_3WT_PLT	14,60	68.9%	10,50	94.2%	13,80
31_3EZ_N	12,70	67.2%	8,90	94.6%	12,10
32_3EZ_P	10,30	59.7%	6,30	95.7%	9,80
33_3EZ_NL	11,10	68.5%	7,90	95.0%	10,50
34_3EZ_PL	10,90	68.4%	7,70	94.6%	10,30
35_3EZ_T	11,20	68.6%	8,00	94.5%	10,60
36_3EZ_PLT	9,90	67.0%	6,90	93.2%	9,20
3_1WT_NL	13,80	69.0%	9,90	94.7%	13,10
4_1WT_PL	11,30	69.4%	8,10	93.9%	10,60
5_1WT_T	10,60	67.5%	7,40	92.7%	9,80
6_1WT_PLT	14,80	68.0%	10,50	94.5%	14,00
7_1EZ_N	13,60	66.3%	9,40	93.4%	12,70
8_1EZ_P	10,90	61.8%	6,90	93.0%	10,10
9_1EZ_NL	11,50	68.4%	8,10	94.2%	10,80

Figure S1: Mean quality values across each base position in each read determined by FastQC for 36 samples in forward and reverse reads.
Sample details are listed in Table S1. (bp: base pair, Phred score: indicates the measure of base quality. Phred ranges are color coded and represent quality: red = QC not passed, yellow = acceptable, green = QC passed**)**

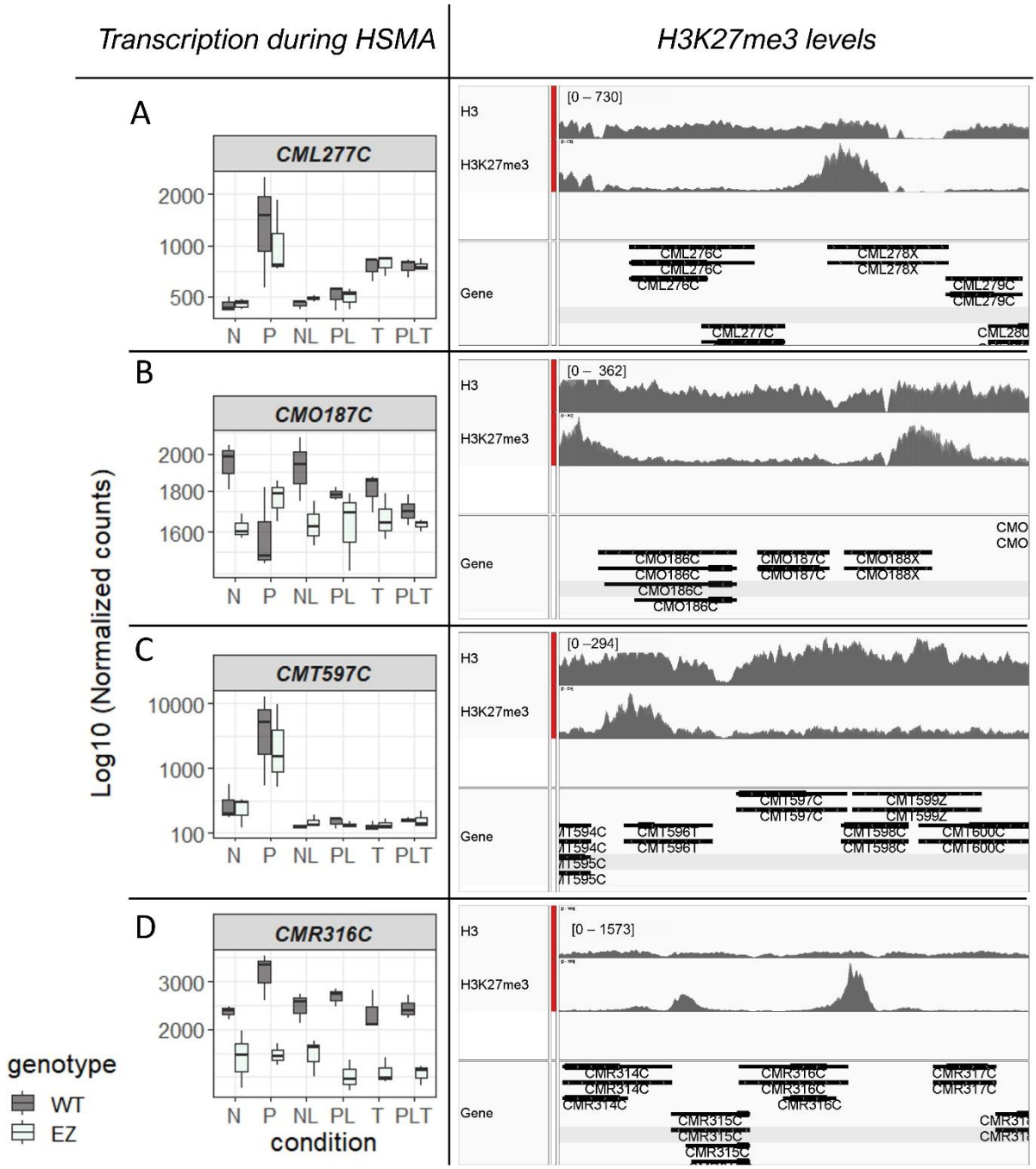


Figure S2: Expression rates during HSMA and histone mark H3K27me3 levels of *C. merolae* heat shock transcription factors (CmHSF)

Transcription during HSMA (right): Gene expression in boxplots. Y-axis depicts relative expression given as log10 transformed normalized counts in different conditions collected during the heat stress memory assay (HSMA): naïve (N), primed (P), naïve – after lag phase (NL), primed – after lag phase (PL), triggered-only (T) and primed-and-triggered (PLT). Wild type (WT) and *ΔE(z)* counts are depicted in black and mint green, respectively. ***H3K27me3 levels (left)*** Genome browser view of H3 and H3K27me3 read coverage at the loci of the four CmHSFs. Depicted are histone H3 and H3K27me3 ChIP-seq signals in naïve WT cells (Mikulski *et al.*, 2017). Both tracks exhibit three merged biological replicates. Numbers in brackets above each track denote the scale read count. ChIP-seq data are retrieved form (Mikulski *et al.*, 2017). The "Gene" track indicates gene position as given by the "new" reference annotation file (see Table 2.5).

A *CML277C* (annotated as similar to heat shock transcription factor) is significantly upregulated upon priming (padj < $7*10^{-9}$). *CML277C* shows moderate expression in control conditions, and high expression in the heat context (P, T, PLT). WT and *ΔE(z)* show similar expression pattern. *CML277C*

is neighbor to *CML278X*, an repetitive element showing a slight H3K27me3 peak in WT control cells. Note, that *CML278X* was not identified as H3K27me3 target in Mikuslski *et al.*, 2017.
B *CMO187C* (annotated as heat shock transcription factor) shows downregulation upon priming heat in WT cells and seems (insignificantly) affected in its transcription upon loss of genomic H3K27me3 levels in *ΔE(z)* mutant. *CMO187C* is neighbor to *CMO188X*, an repetitive element with weak H3K27me3 enrichment.
C *CMT597C* (annotated as heat shock transcription factor) shows pronounced induction upon priming (padj < $7*10^{-8}$), but is not induced upon triggering, independent of the preceding priming. *CMT597C* is barely surrounded by H3K27me3 in WT control cells, with a very low H3K27me3 peak at its neigboring gene *CMT596T* (hypothetical transcript).
D *CMR316C* (annotated as hypothetical protein, HSF-domain containing protein (uniport.org)) is significantly lower expressed in primed ∆*E(z)* cells compared to primed WT cells suggesting insufficient expression rates. The same is true for the triggering stimulus: the mutant fails to reach WT expression levels upon triggering. *CMR316C* is surrounded by pronounced H3K27me3 peaks at its up- and downstream genes which might explain the aberrant transcription rates in the *ΔE(z)* during the HSMA.

Table S2: Heat trainable genes in *Cyanidioschyzon merolae* thermomemory

Genes were identified as differentially expressed between PLT and T in WT. Table depicts DESeq2 results (normalized counts, log2fold change, adjusted p-values) for 59 WT-trainable genes in WT and Δ*E(z)* mutant. Genes were designated trainable, when their expression differed more than 1.5 fold (L2FC > |0.58|) and the adjusted p-value was below 0.05 (padj < 0.05) in the WT. Genes indicated in red are WT-memory genes, which lost trainability in Δ*E(z)*.

		10D Wild type			E(z)		
GeneID	**Gene name**	**Normalized counts_WT**	**log2Fold Change WT**	**padj_WT**	**Normalized counts_EZ**	**log2Fold Change EZ**	**padj_EZ**
CMB107C	CYME_*CMB107C*	326,9	0,73	8,06E-08	171,6	0,05	0,28008736
CMF092C	CYME_*CMF092C*	1497,8	0,85	0,00218479	1137,7	0,08	0,16076905
CMI250C	CYME_CMI250C	1959,7	0,83	0,00220329	1613,2	1,42	1,42E-09
CMJ099C	CYME_*CMJ099C*	41586	1,98	7,47E-11	43698	2,16	1,59E-12
CMJ101C	CYME_*CMJ101C*	504484,2	0,9	0,0244145	302261,6	0,26	0,148379
CMJ102C	CYME_CMJ102C	7441,1	3,16	8,28E-20	7565,7	3,79	7,00E-37
CMJ308C	CYME_CMJ308C	123,3	-0,93	0,00218479	80,6	-1,49	7,47E-05
CMP040C	CYME_*CMP040C*	478,1	0,65	1,38E-06	314,8	0,01	0,47975476
CMT010C	CYME_CMT010C	105,4	-1,05	0,00465425	429,5	0	1
CMV002C	ycf16	59,7	0,65	0,02950888	51,6	2,18	5,82E-07
CMV004C	tilS ycf62	52,1	3,75	3,19E-05	42,4	3,71	1,68E-10
CMV007C	infB	94,5	1,5	0,00640157	81,1	2,33	3,22E-07
CMV010C	ycf80	325,5	3,01	7,32E-05	313,7	3,29	1,33E-11
CMV011C	rpl28	82,6	2,75	6,70E-05	74,6	3,47	1,59E-12
CMV012C	trxA trxM	244,6	1,39	0,02042435	198,8	3,44	1,11E-08
CMV025C	mnmE thdF	26,7	3,85	0,00012985	19,9	2,83	3,00E-05
CMV028C	psaM	42,3	1,39	0,01137684	32,8	1,15	0,048155
CMV035C	rpl27	30,9	0,93	0,03416341	19,9	1,56	0,02061393
CMV036C	carA	107,5	1,48	0,01402875	83,6	2,78	2,45E-09
CMV037C	ycf53	40,5	1,99	0,00495155	32,1	1,94	0,00237286

CMV038C	ycf55	34,4	2,19	0,00172267	28,2	0,49	0,07080177
CMV047C	psbA	2565,4	3,16	0,00027198	1440,1	2,45	2,63E-11
CMV064C	cpcB	1034	1,06	0,02541305	367,8	2,84	9,47E-07
CMV071C	secA	2037,1	2,7	6,70E-05	2311,4	2,36	7,19E-07
CMV072C	rpl34	93,8	2,03	0,00772871	81,2	1,18	0,03395878
CMV075C	ycf44	170,9	2,03	0,00580441	155,9	1,76	1,34E-05
CMV076C	trpG	218,3	1,89	0,00640157	159,6	2,87	6,60E-19
CMV077C	thiG	207,3	1,81	0,00640157	159,4	2,64	8,13E-05
CMV078C	ycf60	130	0,9	0,0277034	97,1	2,49	0,00056227
CMV080C	ycf27	170,7	2,65	0,00220329	119,3	2,84	1,30E-16
CMV081C	psbD	1204,6	2,91	0,0004164	794,8	2,74	9,59E-10
CMV086C	groEL groL	118,1	1,48	0,00677716	101,4	2,19	9,25E-10
CMV091C	rps1	80,9	2,34	0,00313956	64,9	2,55	2,25E-09
CMV092C	ycf40	13,1	0,68	0,04675685	8,1	1,23	0,04870105
CMV094C	ycf29	52,2	0,78	0,02923084	45,9	1,64	0,00467902
CMV108C	rpl19	32,4	0,68	0,04287736	20,9	2,7	0,00193484
CMV116C	menA	23,2	0,78	0,04042151	19,6	0,48	0,12734401
CMV121C	ycf82	84,5	1,83	0,00640157	98,9	1,87	7,71E-07
CMV124C	psbB	600,5	2,63	0,00050396	394,6	2,31	4,46E-15
CMV135C	psaA	4126,4	1,31	0,01871469	2042,3	2,32	1,46E-15
CMV136C	psaB	1428,6	2,65	0,0013557	1005,6	3,02	7,02E-17
CMV169C	rpl22	25,1	1,22	0,0244145	13,6	2,63	0,00279649
CMV170C	rps3	77,2	2	0,00640157	41,7	2,96	1,51E-06
CMV171C	rpl16	45,5	1,58	0,01173748	26,3	2,68	7,17E-06
CMV175C	rpl24	25,4	1,65	0,01465396	17,3	3,29	7,77E-06
CMV176C	rpl5	59,4	2,63	0,00207551	36,8	3,43	3,19E-09
CMV178C	rpl6	53,5	0,98	0,0277034	29,8	3,29	2,88E-08

CMV179C	rpl18	25,6	0,88	0,03475167	15,9	0,22	0,29803621
CMV196C	atpB	159,6	0,96	0,02149882	107,9	2,3	1,59E-12
CMV201C	psaF	87,8	1,85	0,00640157	48,3	2,79	9,47E-07
CMV214C	rpl33	40,3	0,74	0,03915377	40,1	1,41	0,0155203
CMV217C	rpoC	216,5	0,98	0,02743258	164,3	3,07	2,15E-16
CMV218C	rps2	59,4	1,32	0,02058679	41	3,44	1,05E-11
CMV219C	tsf	56,6	2,05	0,00541034	42,4	2,66	5,90E-07
CMV224C	atpD	92,2	1,16	0,0125649	73,5	1,14	0,00099985
CMV009C	rps4	176,5	2,22	0,00640157	164,5	2,87	0,00044277
CMV082C	psbC	408,4	2,36	0,00043674	319,9	2,13	4,02E-40

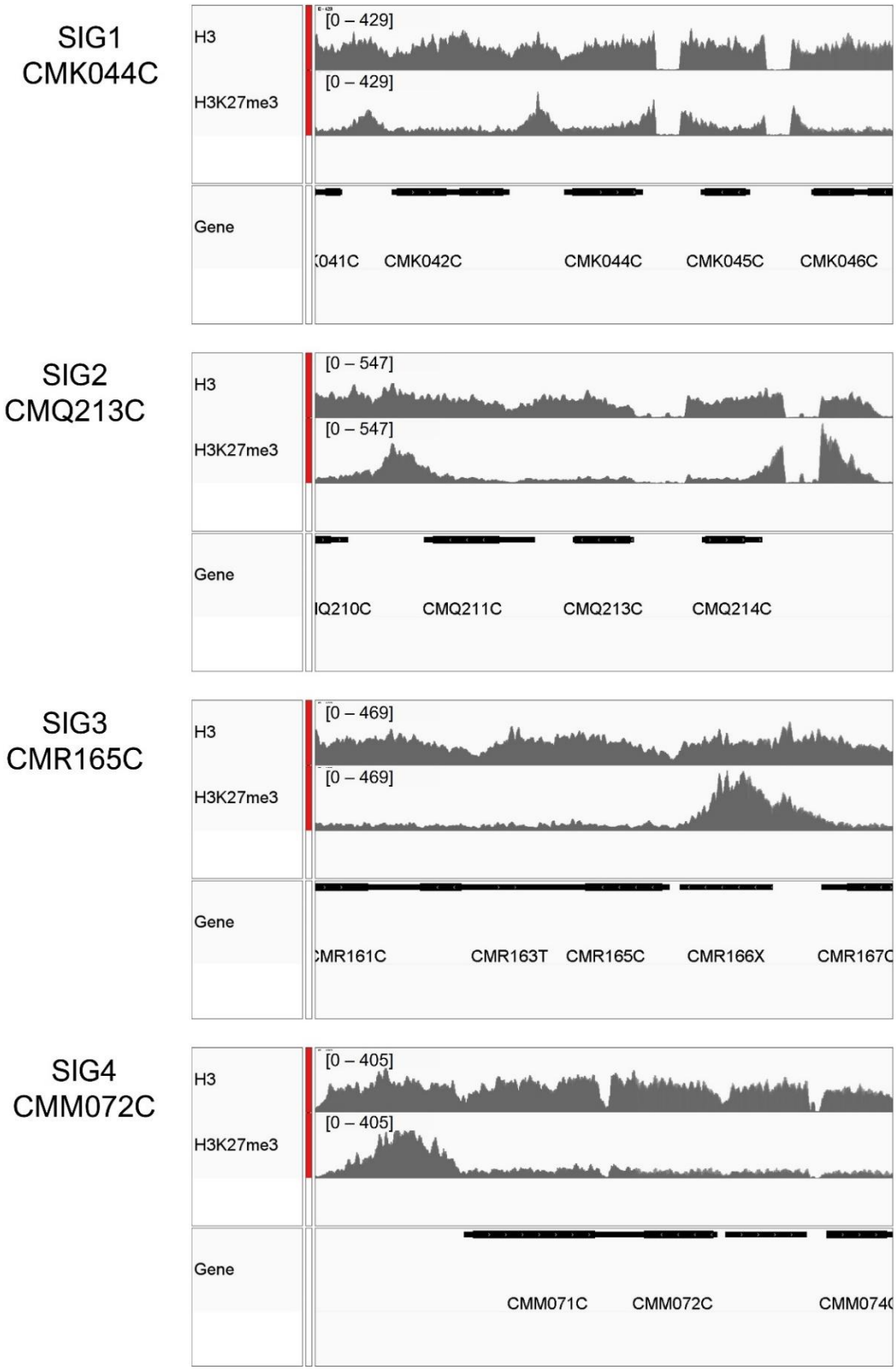


Figure S3: Genome browser view of H3 and H3K27me3 read coverage at the loci of the four nucleus encoded sigma factors SIG1-4

(*CMK044C*, *CMQ213C*, *CMR165C*, *CMM072C*). Depicted are histone H3 and H3K27me3 ChIP-seq signals in naïve WT cells (Mikulski *et al.*, 2017). Both tracks exhibit three merged biological replicates. Numbers in brackets above each track denote the scale read count. ChIP-seq data are retrieved form (Mikulski *et al.*, 2017). The “Gene” track indicates gene position as given by the “new” reference annotation file.

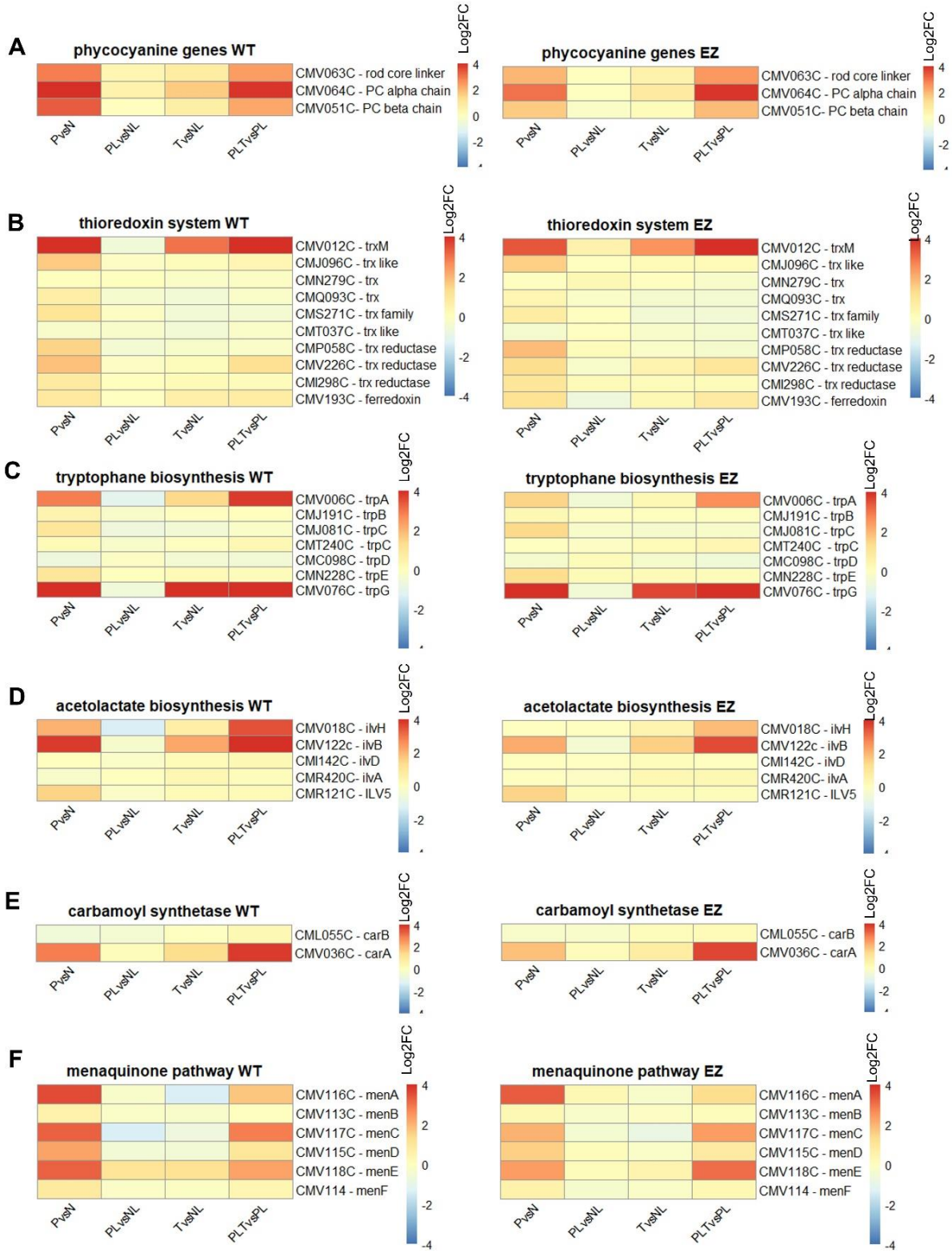


Figure S4: Expression changes of genes involved in specific regulations and/or pathways during heat shock memory assay (HSMA) in *Cyanidioschyzon merolae*

Annotated heatmap showing log2 fold-changes (Log2FC, red – upregulated, blue - downregulated) after priming (PvsN), priming and lag phase (PL), triggering (TvsNL) and priming-lag phase-triggering (PLT). Functional annotation of genes is in accordance with Ohta *et al.*, 2003 (see legend in figure). Heatmap was generated using pheatmap package (Kolde, 2019) and color coding shows log2 fold changes (Log2FC) of single genes in each condition (red = high, blue = low). Significantly expressed genes between PLT and T directly (= memory genes) are indicated by 2-sided arrows between TvsNL and PLTvsPL.

Expression pattern of phycocyanin genes (**A**), thioredoxin and thioredoxin reductase genes (**B**), genes involved in tryptophan biosynthesis pathway (**C**), acetolactate synthase (**D**), carbamoyl synthetase (**E**), and menaquinone pathway (**F**) are shown for WT and *ΔE(z)* (EZ).

Table S3: Primer sequences used for genotyping ∆*E(z) C. merolae* strain.

Primer name	***Primer Sequence (5´→ 3´)***
Cm_trafo_D	cgtcgcgactggttggtgact
Cm_trafo_L	gacttgtatagtgacccagtcgg
Cm_trafo_C	aagaagacgcgctccagggc

Table S4 Differentially expressed genes in ∆*E(z)* compared to 10D wild type (WT) *C. merolae* cells

DESeq2 analysis results using published (=old) and new annotation, merged. Genes under old(new) category in ANNO column are genes, which are detected by old and new anno, but the values of L2FC & padj are retrieved from the old analysis, considering this one more reliable. Genes are considered differentially expressed, when L2FC < |1.5| and padj < 0.05. GeneID: as given by czon.jp, EZ: ∆E(z)

GeneID	Normalized Counts	L2FC EZvsWT	padj	ANNO detected in	Protein names	K27me3target? (Mikulski *et al.* 2017)	subtelomeric? (Nozaki *et al.* 2017)	genomic element
CMA003X	3206,0	8,0	5E-273	new	repetitive	yes	yes	repetitive
CMK034C	1538,6	-3,1	8E-196	old(new)	BHLH domain-containing protein			gene
CMD191C	1042,5	4,9	3E-176	old(new)	Similar to trefoil factor	yes	yes	gene
CMT003C	743,1	5,3	2E-172	old(new)	Amine oxidase (EC 1.4.3.-)		yes	gene
CMS404T	1470,5	-2,9	1E-143	new	hypothetical trascsript			gene
CMA142C	1193,5	3,0	8E-143	old(new)	Similar to trefoil factor	yes	yes	gene
CMT005X	974,3	4,6	8E-137	new	repetitive		yes	repetitive
CMR001C	685,9	3,9	4E-110	old(new)	Similar to trefoil factor		yes	gene
CMC188C	1703,3	3,1	1E-95	old	L-lactate dehydrogenase		yes	gene
CMB162C	918,5	3,8	1E-79	old(new)	Similar to trefoil factor	yes	yes	gene
CMN314X	660,3	3,9	2E-75	new	repetitive	yes		repetitive
CMM069C	1919,1	-2,7	1E-72	old(new)	Uncharacterized protein			gene
CMI009T	837,0	3,1	6E-61	new	hypothetical transcript	yes		gene
CME003X	2236,6	7,3	1E-58	new	repetitive	yes	yes	repetitive
CMO185C	1266,5	-2,2	2E-56	old(new)	Similar to hedgehog protein			gene
CMH002Z	467,6	2,9	1E-50	new	pseudogene	yes	yes	gene
CMT007X	387,8	7,5	3E-48	new	repetitive		yes	repetitive
CMK201T	2298,3	-2,5	4E-47	new	hypothetical transcript			gene
CMD004X	231,0	5,8	7E-46	new	repetitive	yes	yes	repetitive
CMR153C	1864,4	1,7	2E-44	old	Similar to BEL1-related homeotic			gene

					protein			
CMK215C	1339,9	-1,8	7E-43	old(new)	Uncharacterized protein			gene
CMP358C	2841,2	4,8	1E-37	old(new)	Similar to hedgehog protein		yes	gene
CMI002C	926,6	-2,4	5E-36	old(new)	Amine oxidase (EC 1.4.3.-)		yes	gene
CMS003X	151,0	5,3	1E-35	new	repetitive	yes	yes	repetitive
CMC004Z	123,7	4,2	1E-34	new		yes	yes	gene
CMD149X	3207,2	1,6	3E-34	new	repetitive			repetitive
CMP121C	308,9	2,2	4E-33	new	Retroelement, dead			repetitive
CMI033X	160,5	3,2	7E-33	new	repetitive	yes		repetitive
CMI003X	144,8	3,5	1E-31	new	repetitive		yes	repetitive
CMA001C	673,2	2,6	3E-31	old(new)	Similar to trefoil factor	yes	yes	gene
CMI297T	5399,5	-1,8	1E-30	new				gene
CMC002C	7023,5	1,7	8E-28	old(new)	Similar to hedgehog protein		yes	gene
CMS002C	124,2	5,6	9E-25	old(new)	Amine oxidase (EC 1.4.3.-)		yes	gene
CMQ156C	895,7	-13,1	2E-24	old(new)	Similar to Polycomb-group developmental gene, enhancer of zeste			gene
CMN025X	200,3	3,4	1E-20	new	repetitive			repetitive
CMR114C	138,1	2,5	5E-20	old(new)	Purple acid phosphatase (EC 3.1.3.2)	yes		gene
CMP055T	1088,1	-1,2	1E-19	new				gene
CML337X	112,8	4,2	2E-19	new	repetitive	yes	yes	repetitive
CMN339Z	211,8	2,1	2E-19	new			yes	gene
CMM107X	179,3	2,5	3E-19	new	repetitive			repetitive
CMN282Z	1018,0	1,3	4E-19	new				gene
CMK072C	4342,7	1,2	2E-18	old	Uncharacterized protein			gene

CMO184C	1912,8	-0,9	4E-18	old(new)	Similar to hedgehog protein			gene
CMT002X	69,7	4,8	4E-18	new	repetitive		yes	repetitive
CMJ002C	1184,6	1,5	9E-18	old(new)	L-lactate dehydrogenase (EC 1.1.1.27)		yes	gene
CMJ005C	3348,5	0,9	1E-17	old(new)	Similar to cell surface glycoprotein		yes	gene
CMT547C	2310,6	1,0	1E-17	old(new)	Starch associated protein R1			gene
CMN310C	6866,9	1,3	3E-17	old(new)	RWP-RK domain-containing protein			gene
CMJ020C	1098,9	1,2	4E-17	new	CYTH domain-containing protein			gene
CMQ357T	4104,6	0,8	6E-17	new				gene
CMT641C	277,0	1,6	7E-17	new				gene
CMD002C	2069,1	3,3	7E-17	old(new)	Similar to trefoil factor	yes	yes	gene
CMH141T	888,3	1,3	2E-16	new				gene
CMI015C	1627,9	0,8	2E-16	old(new)	L-asparaginase			gene
CMO156X	661,7	1,1	6E-16	new	repetitive			repetitive
CML004C	552,0	1,2	2E-15	old(new)	Similar to high-affinity iron permease		yes	gene
CMI286C	480,3	1,4	2E-15	old(new)	Uncharacterized protein	yes		gene
CMR036C	2965,6	-0,9	2E-14	old(new)	Inositol-3-phosphate synthase (EC 5.5.1.4)			gene
CMI166C	55,0	-4,3	2E-14	old(new)	Uncharacterized protein	yes		gene
CMF184X	56,2	3,3	5E-14	new	repetitive	yes		repetitive
CMG089C	742,2	-1,0	7E-14	new	Similar to respiratory burst oxidase protein			gene
CMK076C	2721,0	0,8	9E-14	old(new)	HAP2-GCS1 domain-containing protein			gene
CMC153T	103,2	-2,1	1E-13	new				gene

CMD138T	778,5	1,0	2E-13	new				gene
CMR192T	8879,5	0,8	2E-13	new				gene
CME001C	1060,9	3,1	3E-13	old(new)	Similar to trefoil factor	yes	yes	gene
CMH217X	97,9	2,2	3E-13	new	repetitive			repetitive
CMI130C	345,1	-1,4	4E-13	old(new)	Probable zinc protease			gene
CMD148C	3330,0	0,9	4E-13	old(new)	Probable ATP-dependent transporter ycf16			gene
CMR502C	476,0	1,2	1E-12	old(new)	Cupin type-1 domain-containing protein	yes	yes	gene
CMK310X	79,9	3,3	1E-12	new	repetitive	yes	yes	repetitive
CMS104T	359,2	1,3	2E-12	new				gene
CMA022C	1404,0	-0,8	3E-12	new	Uncharacterized protein			gene
CMP318C	2133,4	-0,7	3E-12	old(new)	Sphingolipid delta 4 desaturase protein DES-1			gene
CMB129T	1177,8	-0,8	4E-12	new				gene
CMP082X	344,2	1,2	7E-12	new	repetitive			repetitive
CMT189X	113,1	1,8	8E-12	new	repetitive	yes		repetitive
CMP048X	139,1	1,7	1E-11	new	repetitive			repetitive
CMT163C	890,5	-1,0	1E-11	old	Uncharacterized protein			gene
CMI001X	88,1	-2,3	2E-11	new	repetitive	yes	yes	repetitive
CMK277C	1149,1	-0,7	3E-11	old(new)	Uncharacterized protein			gene
CMG018C	75,7	2,6	3E-11	old(new)	Nitrate transporter	yes		gene
CML233X	100,3	2,1	3E-11	new	repetitive			repetitive
CMA144Z	47,0	3,6	4E-11	new			yes	gene
CMR155C	1490,3	1,0	7E-11	old	Similar to TATA element modulatory factor			gene
CMQ158C	583,4	-0,9	8E-11	old(new)	Hist_deacetyl domain-containing			gene

					protein			
CMR505Z	1117,9	0,7	3E-10	new			yes	gene
CMV013C	1098,9	-0,8	3E-10	old(new)	Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit) (EC 4.1.1.39)			gene
CMS371C	2404,4	0,6	4E-10	old	MYB-related protein			gene
CMR300X	410,9	1,2	5E-10	new	repetitive	yes		repetitive
CMO049Z	219,3	1,7	5E-10	new				gene
CMK048X	47,1	3,3	1E-09	new	repetitive	yes		repetitive
CMG055X	46,7	2,5	1E-09	new	repetitive	yes		repetitive
CMT236C	797,8	-1,0	2E-09	old(new)	Uncharacterized protein			gene
CMJ287Z	392,3	1,1	2E-09	new				gene
CMO050Z	82,2	1,9	2E-09	new		yes		gene
CMQ067C	673,4	0,7	2E-09	old	Phosphomannomutase			gene
CMM327X	40,7	2,7	4E-09	new	repetitive	yes		repetitive
CMN090C	315,1	1,2	4E-09	new	Similar to syntaxin binding protein			gene
CMT010C	151,2	1,9	4E-09	old(new)	Retroelement, alive			repetitive
CMN309C	2408,1	1,0	5E-09	old(new)	Fe2OG dioxygenase domain-containing protein	yes		gene
CMI213C	3984,4	-0,7	7E-09	old(new)	Calnexin			gene
CMI085X	35,2	2,8	8E-09	new	repetitive			repetitive
CMC166X	46,8	3,1	1E-08	new	repetitive			repetitive
CMT505X	241,1	1,2	1E-08	new	repetitive	yes		repetitive
CMV047C	566,1	0,9	2E-08	old(new)	Photosystem II protein D1 (PSII D1 protein) (EC 1.10.3.9) (Photosystem II Q(B) protein)			gene

CMO087C	78,8	2,0	2E-08	old(new)	Uncharacterized protein			gene
CMO213X	147,7	1,6	2E-08	new	repetitive			repetitive
CMJ103C	1215,6	-0,6	2E-08	old	Uncharacterized protein			gene
CMO193C	312,2	-1,0	5E-08	new	HEAT repeat-containing protein 1			gene
CMV122C	183,4	1,3	5E-08	old(new)	Acetolactate synthase (EC 2.2.1.6)			gene
CMQ398T	301,2	1,0	6E-08	new				gene
CMT142T	864,1	0,7	7E-08	new				gene
CMD030C	722,6	-1,2	8E-08	old(new)	Uncharacterized protein			gene
CMN137C	869,5	-1,1	1E-07	old(new)	Uncharacterized protein			gene
CMJ006Z	28,2	5,1	1E-07	new		yes	yes	gene
CMR431C	1238,7	0,8	1E-07	new	Retroelement			repetitive
CMI071Z	139,2	1,4	2E-07	new				gene
CMI005X	131,2	5,8	2E-07	new	repetitive	yes	yes	repetitive
CMT164T	253,5	-1,2	2E-07	new				gene
CMI069Z	192,8	1,2	3E-07	new				gene
CMF058C	857,5	-0,6	3E-07	new	Probable zinc transporter			gene
CMR004C	1110,3	0,6	3E-07	old(new)	Uncharacterized protein		yes	gene
CMH053X	42,1	2,7	3E-07	new	repetitive	yes		repetitive
CMN335X	22,6	3,4	4E-07	new	repetitive	yes		repetitive
CMR503X	22,3	4,2	4E-07	new	repetitive	yes	yes	repetitive
CMO011X	43,5	2,3	4E-07	new	repetitive			repetitive
CMJ309C	70,8	-2,0	5E-07	old				gene
CMI233C	21807,0	-0,6	5E-07	old(new)	Glutamine synthetase (EC 6.3.1.2)			gene
CMC073X	37,9	2,3	6E-07	new	repetitive	yes		repetitive

CMO002X	27,9	4,1	7E-07	new	repetitive	yes	yes	repetitive
CMJ001X	44,4	2,1	7E-07	new	repetitive	yes	yes	repetitive
CMP163X	51,6	1,8	9E-07	new	repetitive			repetitive
CMT412C	558,0	-0,7	1E-06	old(new)	Isocitrate dehydrogenase subunit 1, mitochondrial			gene
CMI307X	116,7	1,3	1E-06	new	repetitive		yes	repetitive
CMB094C	445,4	-0,8	1E-06	old(new)	Rhomboid domain-containing protein			gene
CMD189X	25,3	5,8	1E-06	new	repetitive	yes	yes	repetitive
CMH004X	14,7	5,9	1E-06	new	repetitive	yes	yes	repetitive
CMI305X	19,9	4,1	1E-06	new	repetitive	yes	yes	repetitive
CMK292C	431,1	-0,8	1E-06	old(new)	Uncharacterized protein			gene
CMS062C	1007,5	-0,7	2E-06	old(new)	Similar to chromosome segregation protein SepB			gene
CMF031C	409,1	0,8	2E-06	new	Uncharacterized protein			gene
CMD177X	41,8	2,2	2E-06	new	repetitive			repetitive
CMF074Z	268,5	-1,0	2E-06	new				gene
CME032X	98,7	1,3	2E-06	new	repetitive			repetitive
CMI287T	39,7	2,2	3E-06	new		yes		gene
CMG219Z	28,8	2,4	3E-06	new			yes	gene
CMM169C	482,6	-0,7	4E-06	old(new)	Probable calcium-binding mitochondrial carrier protein Aralar			gene
CMJ308C	76,5	-1,4	4E-06	old(new)	Retroelement, alive			repetitive
CMM156C	1498,8	-0,7	4E-06	old(new)				gene
CMF189Z	20,5	5,5	4E-06	new		yes	yes	gene
CMK071C	2678,5	0,7	5E-06	old(new)	Uncharacterized protein			gene

CMC131C	500,8	-0,7	6E-06	old(new)	Uncharacterized protein			gene
CMR494X	98,6	1,3	6E-06	new	repetitive	yes		repetitive
CML151X	6433,2	0,6	7E-06	new	repetitive			repetitive
CMN021C	256,4	-0,9	7E-06	old	Similar to UDP-sugar transporter			gene
CMP170C	188,7	1,0	7E-06	new	DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18)			gene
CMN049C	56,5	1,8	9E-06	old(new)	Similar to formin			gene
CMA145C	3278,4	0,8	9E-06	old(new)	L-lactate dehydrogenase (EC 1.1.1.27)		yes	gene
CMM302X	19,5	3,1	9E-06	new	repetitive			repetitive
CMR359C	540,0	-0,8	9E-06	new	Uncharacterized protein			gene
CMA146X	10,7	5,5	1E-05	new	repetitive	yes	yes	repetitive
CME173C	31,9	2,5	1E-05	old(new)	Retroelement	yes		repetitive
CMJ282C	1354,4	-0,8	1E-05	old(new)	MYB-related protein			gene
CMN281X	25,6	2,6	1E-05	new	repetitive	yes		repetitive
CMS244T	307,4	0,9	2E-05	new				gene
CMT011X	54,2	2,3	2E-05	new	repetitive	yes		repetitive
CMJ071X	41,1	1,9	2E-05	new	repetitive			repetitive
CMK057X	54,9	1,6	2E-05	new	repetitive			repetitive
CMF190Z	304,2	0,8	2E-05	new		yes	yes	gene
CME180C	1117,7	-0,6	2E-05	old(new)	Uncharacterized protein			gene
CMP329C	477,9	0,6	2E-05	old(new)	Putrescine aminopropyltransferase			gene
CMR185C	736,8	-0,7	2E-05	new	Mitochondrial division protein Mda1 WD40 repeat protein	yes		gene
CMQ026C	469,7	-0,6	2E-05	old(new)	Ubiquitinyl hydrolase 1 (EC 3.4.19.12)			gene

CMM057C	608,8	-0,7	3E-05	old(new)	Gamma-glutamylcyclotransferase (EC 4.3.2.9)			gene
CMT231C	1204,2	-0,7	3E-05	old(new)	Uncharacterized protein			gene
CMT333C	139,8	1,0	3E-05	old(new)	Similar to inorganic phosphate transporter			gene
CMR198C	499,8	-0,6	3E-05	new	3'(2'),5'-bisphosphate nucleotidase (EC 3.1.3.7)			gene
CMP357X	19,4	2,8	3E-05	new	repetitive	yes	yes	repetitive
CMQ120C	2189,9	-0,6	3E-05	new	Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 2.5.1.54)			gene
CMR293C	76,3	1,3	3E-05	old(new)	Uncharacterized protein			gene
CMS412C	67,5	1,5	3E-05	new	Actin-like protein	yes		gene
CMI298C	906,4	-0,6	3E-05	old	Thioredoxin reductase (EC 1.8.1.9)			gene
CME002Z	9,2	5,2	4E-05	new		yes	yes	gene
CMM171C	540,6	-0,7	4E-05	new	Gamma-glutamylcyclotransferase (EC 4.3.2.9)			gene
CMT162C	330,6	-1,1	4E-05	old(new)	APS reductase			gene
CMT120C	1360,4	-0,6	5E-05	new	ABC transporter, LktB-related			gene
CMI011X	16,6	2,9	6E-05	new	repetitive	yes		repetitive
CMQ155C	15,3	3,2	7E-05	new	Uncharacterized protein	yes		gene
CMV014C	284,1	-0,7	7E-05	old(new)	Ribulose bisphosphate carboxylase small subunit (RuBisCO small subunit)			gene
CMF130C	579,8	-0,7	8E-05	old(new)	Uncharacterized protein			gene
CMW058R	59,1	-1,3	8E-05	new				gene
CML220X	25,2	2,0	9E-05	new	repetitive			repetitive
CML050C	800,2	-0,8	0,0001	new				gene

CMP348X	12,9	3,6	0,0001	new	repetitive	yes		repetitive
CMM323C	2264,7	-0,6	0,0001	old(new)	Ribonucleoside-diphosphate reductase (EC 1.17.4.1)			gene
CMH196X	61,8	1,4	0,0001	new	repetitive			repetitive
CMJ053C	1761,6	-0,6	0,0002	old(new)	Uncharacterized protein			gene
CMR358C	1784,9	-0,7	0,0002	old	Uncharacterized protein			gene
CMS001X	13,3	4,3	0,0002	new	repetitive	yes	yes	repetitive
CMN227C	619,8	-0,6	0,0002	old	S-adenosylmethionine synthase (EC 2.5.1.6)			gene
CMP030C	271,8	-0,7	0,0002	old(new)	Probable chromatin assembly factor 1 subunit B			gene
CMD025C	423,0	-0,7	0,0002	new	DNA replication licensing factor MCM2 (EC 3.6.4.12) (Minichromosome maintenance protein 2)			gene
CMR479C	1311,9	-0,6	0,0002	old(new)	p68 RNA helicase			gene
CMT252C	192,9	-0,8	0,0003	new	Phosphoserine aminotransferase			gene
CMN138C	449,2	-0,6	0,0003	old	Probable AAA protein spastin			gene
CMQ438C	593,5	-0,6	0,0003	new	Uncharacterized protein			gene
CMR237C	1145,1	-0,6	0,0003	new	Threonine 3-dehydrogenas			gene
CMT632X	11,8	3,4	0,0003	new	repetitive	yes		repetitive
CMC050X	16,1	2,6	0,0003	new	repetitive	yes		repetitive
CMT127C	24,6	1,8	0,0003	old(new)	Uncharacterized protein	yes		gene
CMO017Z	38,8	1,4	0,0003	new				gene
CMW062R	79,0	-1,2	0,0003	new				gene
CMQ188X	200,3	0,9	0,0003	new	repetitive	yes		repetitive
CMJ331C	8085,9	-0,6	0,0003	new				gene

CMQ402R	172,5	0,9	0,0004	new				gene
CMS306C	13579,1	-0,7	0,0004	old	Uncharacterized protein			gene
CMT558C	587,1	-0,7	0,0004	old(new)	T-complex protein 1 subunit eta (TCP-1-eta) (CCT-eta)			gene
CMF147C	554,4	-0,6	0,0004	old(new)	Uncharacterized protein			gene
CMS317C	223,8	-0,7	0,0005	new	Uncharacterized protein			gene
CMN277X	30,1	1,6	0,0005	new	repetitive			repetitive
CMR234C	317,7	-0,9	0,0005	new	DNA replication licensing factor MCM7 (EC 3.6.4.12)			gene
CMJ148C	6028,6	-0,6	0,0005	old(new)	Uncharacterized protein			gene
CMI306C	4786,3	0,6	0,0005	new	L-lactate dehydrogenase (EC 1.1.1.27)		yes	gene
CMQ173C	513,5	-0,7	0,0005	old(new)	Tousled-like kinase			gene
CMG221C	813,3	1,5	0,0006	new			yes	gene
CMF005X	20,5	1,9	0,0006	new	repetitive	yes	yes	repetitive
CMR218C	206,5	-0,8	0,0006	old(new)	Uncharacterized protein			gene
CMJ286C	424,9	-0,8	0,0007	new	Polyadenylate-binding protein			gene
CMD178X	30,3	1,6	0,0007	new	repetitive			repetitive
CMM120Z	230,3	-0,6	0,0007	new				gene
CMV068C	53,7	1,1	0,0012	old(new)	Uncharacterized protein			gene
CMR464X	33,6	1,3	0,0013	new	repetitive			repetitive
CMK230C	108,9	0,8	0,0014	new	Uncharacterized protein			gene
CMS363T	85,0	-0,9	0,0014	new				gene
CMP249C	511,0	-0,6	0,0017	new	Probable endothelin converting enzyme-1			gene
CMS196X	155,6	0,7	0,0017	new	repetitive			repetitive
CMG021C	91,6	0,8	0,0018	old	Assimilatory sulfite reductase			gene

					(ferredoxin) (EC 1.8.7.1)			
CMP304C	365,0	-0,7	0,0018	old	Protein-synthesizing GTPase (EC 3.6.5.3)			gene
CMB104X	10,5	2,5	0,0019	new	repetitive	yes		repetitive
CMT237Z	106,6	-0,9	0,002	new				gene
CMC187X	20,9	1,9	0,002	new	repetitive		yes	repetitive
CMR342X	20,4	1,6	0,002	new	repetitive			repetitive
CMN016C	34,2	1,1	0,0021	old(new)	P-type domain-containing protein			gene
CMI184C	132,1	-0,7	0,0022	old	H/ACA ribonucleoprotein complex subunit			gene
CMT087C	277,3	-0,7	0,0022	old(new)	DNA helicase (EC 3.6.4.12)			gene
CMH003Z	18,9	1,9	0,0025	new		yes	yes	gene
CME183C	7525,9	-0,9	0,0025	old(new)	Uncharacterized protein			gene
CME174X	19,6	1,5	0,0026	new	repetitive	yes		repetitive
CMF037X	66,4	0,9	0,0026	new	repetitive			repetitive
CMQ108X	30,4	1,4	0,0026	new	repetitive			repetitive
CMH184Z	25,9	1,5	0,0026	new		yes		gene
CMN194C	78,4	0,8	0,003	old(new)	Cathepsin D	yes		gene
CMS252Z	69,4	-0,8	0,0031	new				gene
CMN009C	137,6	1,0	0,0034	old				gene
CMP153X	4,5	3,1	0,0037	new	repetitive			repetitive
CMT540C	258,6	-0,6	0,0038	new	Uncharacterized protein			gene
CMJ291X	93,4	0,9	0,0039	new	repetitive			repetitive
CMP264X	9,4	2,4	0,0041	new	repetitive			repetitive
CMQ126Z	45,6	1,0	0,0042	new		yes		gene
CMV067C	49,7	0,9	0,0045	old(new)	Uncharacterized protein			gene

CMJ290C	159,0	0,6	0,0049	new	Retroelement, dead			repetitive
CMR207R	107,6	0,7	0,0052	new				gene
CMB063X	124,5	0,7	0,0054	new	repetitive	yes		repetitive
CMJ127X	17,8	1,4	0,0055	new	repetitive			repetitive
CMD003Z	8,7	2,1	0,0057	new		yes	yes	gene
CMT446X	9,9	2,0	0,0058	new	repetitive			repetitive
CMQ109X	8,7	2,0	0,0062	new	repetitive			repetitive
CMN065C	46,2	1,0	0,0073	new	Uncharacterized protein			gene
CMH026C	113,0	0,6	0,0075	old		yes		gene
CMS378Z	135,3	-0,9	0,0077	new				gene
CMS016X	39,3	2,3	0,008	new	repetitive			repetitive
CMM187Z	30,3	1,0	0,0082	new				gene
CML084Z	36,6	0,9	0,0084	new				gene
CMJ147X	11,2	1,5	0,0092	new	repetitive			repetitive
CMM251X	23,0	1,0	0,0093	new	repetitive			repetitive
CMH085X	23,2	1,1	0,0096	new	repetitive			repetitive
CMI223C	27,6	0,7	0,0097	old(new)	Retroelement			repetitive
CMS071C	5,4	1,2	0,0101	old(new)	Retroelement			repetitive
CMO368Z	21,8	-1,1	0,0105	new				gene
CMN209X	18,6	1,1	0,0107	new	repetitive	yes		repetitive
CMJ283Z	73,6	0,7	0,0108	new				gene
CMN144C	87,4	0,6	0,0115	new				gene
CMM306C	26,7	1,0	0,0118	new				gene
CMS127C	85,3	-0,7	0,0118	new	Similar to nucleolar phosphoprotein			gene
CMC003X	6,8	1,8	0,012	new	repetitive	yes	yes	repetitive

CMM244Z	85,2	-0,6	0,0123	new				gene
CMH192C	569,3	-0,7	0,0135	old(new)	Uncharacterized protein			gene
CMI269X	39,9	0,9	0,014	new	repetitive			repetitive
CMP120X	15,3	1,1	0,0141	new	repetitive			repetitive
CMA069C	9,7	1,2	0,0167	new	Uncharacterized protein			gene
CMT455X	4,6	1,7	0,0174	new	repetitive			repetitive
CMJ003X	9,8	1,3	0,0174	new	repetitive		yes	repetitive
CMM242C	9,3	0,7	0,0177	old(new)	Uncharacterized protein			gene
CMQ352C	35,0	-0,7	0,0181	new	Uncharacterized protein			gene
CMT514C	23,1	0,8	0,0187	new	Uncharacterized protein			gene
CMG168C	26,5	0,8	0,02	new	Probable purple acid phosphatase protein			gene
CMN015X	38,9	0,7	0,02	new	repetitive			repetitive
CMN099X	25,5	0,7	0,0205	new	repetitive			repetitive
CMT509X	11,7	1,0	0,0225	new	repetitive			repetitive
CMQ025X	24,7	0,7	0,0228	new	repetitive	yes		repetitive
CMR104Z	21,0	-0,8	0,0229	new				gene
CMI093X	26,8	0,7	0,0232	new	repetitive	yes		repetitive
CMV104C	43,7	0,6	0,0233	new	50S ribosomal protein L12, chloroplastic			gene
CML132X	7,6	1,1	0,0242	new	repetitive	yes		repetitive
CMQ419X	21,1	0,8	0,0246	new	repetitive			repetitive
CMV121C	24,4	0,7	0,0256	new	Probable glycosyl transferase			gene
CMV095C	21,7	0,7	0,0256	new	NtcA-like transcriptional regulator			gene
CMG127C	24,8	0,7	0,0263	new	Uncharacterized protein			gene
CMP132X	4,1	1,3	0,0268	new	repetitive	yes		repetitive

CMJ275C	7,6	0,6	0,0269	old(new)	Retroelement			repetitive
CML005C	9,5	1,0	0,0279	new	Similar to dTDP-4-dehydrorhamnose reductase		yes	gene
CMJ035X	6,4	1,1	0,0287	new	repetitive	yes		repetitive
CMN333C	11,3	-0,8	0,0296	new	Phosphatidylinositol-4-phosphate 5-kinase	yes		gene
CMA143X	7,1	1,0	0,0305	new	repetitive		yes	repetitive
CMT310X	6,5	1,0	0,0348	new	repetitive	yes		repetitive
CMO001C	3,7	1,0	0,0375	new	Uncharacterized protein	yes	yes	gene
CMN294C	6,8	-0,9	0,0378	new	Uncharacterized protein			gene
CMS497C	5,5	0,8	0,0477	new	Uncharacterized protein	yes		gene

Table S5: Differential expression analysis of 59 memory genes in different conditions.
Differential expression is reached, when L2FC < |1.5| and padj < 0.05. PvsN: priming induction, PLvsNL: primed after lag phase(recovery after priming), TvsN: triggering-only induction, PLTvsPL: induction upon triggering in primed-before culture, L2FC: log(2) fold change, padj: adjusted pvalue by Benjamini-Hochberg procedure (DESeq2)

		PvsN		PLvsNL		TvsN		PLTvsPL	
GeneID	Normalized counts	L2FC	padj	L2FC	padj	L2FC	padj	L2FC	padj
CMB107C	327	0,40	0,0036	0,39	0,0023	-0,03	0,9139	0,32	0,0070
CMF092C	1498	1,03	0,0000	0,20	0,0914	0,55	0,0014	0,64	0,0008
CMI250C	1960	4,27	0,0000	0,00	0,9900	0,27	0,1229	1,15	0,0000
CMJ099C	41586	5,06	0,0000	-0,09	0,3996	0,35	0,0556	2,71	0,0000
CMJ101C	504484	14,20	0,0000	0,02	0,8422	8,51	0,0000	8,90	0,0000
CMJ102C	7441	7,19	0,0000	-0,02	0,9148	0,58	0,0017	3,77	0,0000
CMJ308C	123	1,02	0,0000	-0,47	0,0007	0,07	0,7589	-0,10	0,6658
CMP040C	478	0,63	0,0000	0,24	0,0498	0,14	0,4672	0,46	0,0001
CMT010C	105	1,37	0,0000	-0,27	0,0100	0,05	0,8351	0,00	0,9932
CMV002C	60	1,85	0,0001	0,01	0,9740	-0,08	0,7191	0,77	0,0044
CMV004C	52	2,03	0,0004	0,00	0,9932	0,12	0,4706	3,63	0,0000
CMV007C	94	2,97	0,0000	-0,01	0,9554	0,35	0,0451	3,97	0,0000
CMV009C	177	1,19	0,0100	0,01	0,9591	0,32	0,0267	5,04	0,0000
CMV010C	325	2,43	0,0000	0,00	0,9900	0,44	0,0115	4,47	0,0000
CMV011C	83	1,58	0,0003	0,01	0,9758	0,04	0,8551	3,28	0,0000
CMV012C	245	3,82	0,0000	-0,01	0,8833	1,27	0,0004	6,24	0,0000
CMV025C	27	1,17	0,0097	0,00	0,9833	0,02	0,9378	3,19	0,0001
CMV028C	42	3,45	0,0000	-0,01	0,9682	0,07	0,6128	3,74	0,0000
CMV035C	31	1,39	0,0077	0,03	0,2765	0,17	0,1725	1,44	0,0049
CMV036C	107	2,11	0,0002	0,00	0,9902	0,23	0,2047	3,64	0,0000
CMV037C	40	1,52	0,0007	0,00	0,9831	-0,07	0,7191	2,10	0,0002
CMV038C	34	1,43	0,0005	0,02	0,8451	-0,20	0,2952	0,58	0,0104
CMV047C	2565	4,28	0,0000	0,00	0,9764	0,79	0,0002	4,87	0,0000
CMV064C	1034	4,54	0,0000	0,01	0,8864	0,20	0,2272	3,59	0,0000
CMV070C	99	4,35	0,0000	-0,01	0,8175	3,79	0,0000	6,74	0,0000
CMV071C	2037	4,32	0,0000	-0,02	0,8689	3,22	0,0000	6,02	0,0000
CMV072C	94	0,88	0,0341	0,00	0,9833	0,22	0,1355	3,73	0,0000
CMV075C	171	3,08	0,0000	-0,01	0,9690	0,50	0,0071	3,79	0,0000
CMV076C	218	5,71	0,0000	-0,01	0,8619	3,54	0,0000	7,28	0,0000
CMV077C	207	5,20	0,0000	0,02	0,7975	4,20	0,0000	6,45	0,0000
CMV078C	130	4,43	0,0000	0,00	0,9959	1,05	0,0009	5,47	0,0000
CMV080C	171	4,64	0,0000	-0,01	0,9646	2,39	0,0000	5,58	0,0000
CMV081C	1205	3,38	0,0000	0,01	0,9542	0,35	0,0502	4,10	0,0000

CMV082C	408	2,80	0,0000	0,00	0,9932	0,63	0,0000	3,73	0,0000
CMV086C	118	1,66	0,0001	0,00	0,9833	-0,05	0,8470	2,14	0,0000
CMV091C	81	1,73	0,0007	-0,01	0,9690	0,27	0,1384	3,71	0,0000
CMV092C	13	1,02	0,0263	-0,02	0,5883	0,04	0,7869	3,89	0,0003
CMV094C	52	2,68	0,0000	-0,02	0,8371	0,48	0,0069	4,09	0,0000
CMV108C	32	0,24	0,5397	0,02	0,7582	-0,17	0,2020	0,79	0,0101
CMV116C	23	2,82	0,0001	-0,01	0,9690	-0,05	0,8007	0,29	0,0933
CMV121C	84	4,11	0,0000	0,03	0,6374	1,07	0,0000	4,49	0,0000
CMV124C	601	2,28	0,0000	-0,01	0,9474	0,18	0,3881	3,56	0,0000
CMV135C	4126	3,08	0,0000	0,00	0,9832	0,20	0,3089	2,81	0,0001
CMV136C	1429	3,08	0,0000	0,01	0,9495	0,43	0,0194	3,82	0,0000
CMV169C	25	2,85	0,0002	0,01	0,7810	0,05	0,7444	3,71	0,0002
CMV170C	77	3,52	0,0000	-0,01	0,9483	0,15	0,3457	3,76	0,0000
CMV171C	45	3,40	0,0000	0,02	0,8285	0,14	0,3569	3,73	0,0000
CMV175C	25	1,90	0,0013	0,00	0,9995	0,04	0,7806	4,28	0,0000
CMV176C	59	3,34	0,0000	0,01	0,9335	0,08	0,6602	3,99	0,0000
CMV178C	53	3,36	0,0000	0,00	0,9961	0,00	0,9908	3,11	0,0001
CMV179C	26	2,94	0,0001	-0,01	0,9554	0,17	0,1667	2,76	0,0006
CMV180C	47	3,23	0,0000	-0,01	0,9354	0,16	0,3155	3,52	0,0000
CMV196C	160	2,92	0,0000	0,00	0,9833	0,39	0,0335	3,52	0,0000
CMV201C	88	3,02	0,0000	0,00	0,9908	0,19	0,2463	3,69	0,0000
CMV214C	40	0,27	0,4884	0,02	0,8389	0,15	0,3500	0,96	0,0065
CMV217C	216	2,54	0,0001	-0,01	0,8672	0,08	0,7099	3,22	0,0001
CMV218C	59	3,77	0,0000	0,00	0,9908	0,00	0,9932	3,80	0,0000
CMV219C	57	3,25	0,0000	0,01	0,9145	0,14	0,4276	3,14	0,0000
CMV224C	92	1,74	0,0001	-0,04	0,6060	-0,15	0,4645	0,99	0,0017

8 References

Abram, M., Kaňa, R., Olofsson, D., Pniewski, F., Šedivá, B., Stark, M., Fossl, D., Slat, V., Neumann, A., Rader, S., & Kargul, J. (2022). Molecular mechanisms of long-term light adaptation of an extremophilic alga Cyanidioschyzon merolae. *BioRxiv*, 2022.03.02.482653. https://doi.org/10.1101/2022.03.02.482653

Albertano, P., Ciniglia, C., Pinto, G., & Pollio, A. (2000). The taxonomic position of Cyanidium, Cyanidioschyzon and Galdieria: an update. *Hydrobiologia 2000 433:1*, *433*(1), 137–143. https://doi.org/10.1023/A:1004031123806

Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., & Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynthesis Research*, *98*(1–3), 541–550. https://doi.org/10.1007/S11120-008-9331-0

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.*, 403–410.

Altschul, S. F., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., & Dj, L. (1997). Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data* (p. http://www.bioinformatics.babraham.ac.uk/projects/).

Antunez-Sanchez, J., Naish, M., Ramirez-Prado, J. S., Ohno, S., Huang, Y., Dawson, A., Opassathian, K., Manza-Mianza, D., Ariel, F., Raynaud, C., Wibowo, A., Daron, J., Ueda, M., Latrasse, D., Slotkin, R. K., Weigel, D., Benhamed, M., & Gutierrez-Marcos, J. (2020). A new role for histone demethylases in the maintenance of plant genome integrity. *ELife*, *9*, 1–32. https://doi.org/10.7554/ELIFE.58533

Baker, N. R., & Rosenqvist, E. (2004). Applications of chlorophyll fluorescence can improve crop production strategies: an examination of future possibilities. *Journal of Experimental Botany*, *55*(403), 1607–1621. https://doi.org/10.1093/JXB/ERH196

Balazadeh, S. (2022). A 'hot' cocktail: The multiple layers of thermomemory in plants. *Current Opinion in Plant Biology*, *65*, 102147. https://doi.org/10.1016/J.PBI.2021.102147

Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. A., & Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature*, *427*(6970), 164–167. https://doi.org/10.1038/NATURE02269

Bäurle, I., & Trindade, I. (2020). Chromatin regulation of somatic abiotic stress memory. *Journal of Experimental Botany*, *71*(17), 5269–5279. https://doi.org/10.1093/JXB/ERAA098

Bennett, L., Melchers, B., & Proppe, B. (2020). *Refubium - Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin*. https://doi.org/http://dx.doi.org/10.17169/refubium-26754

Berke, L., & Snel, B. (2015). The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. *BMC Evolutionary Biology*. https://doi.org/10.1186/s12862-015-0319-z

Bi, A., Wang, T., Wang, G., Zhang, L., Wassie, M., Amee, M., Xu, H., Hu, Z., Liu, A., Fu, J., Chen, L., & Hu, T. (2021). Stress memory gene FaHSP17.8-CII controls thermotolerance via remodeling PSII and ROS signaling in tall fescue. *Plant Physiology*, *187*(3), 1163–1176. https://doi.org/10.1093/PLPHYS/KIAB205

Biswal, B., Joshi, P. N., Raval, M. K., & Biswal, U. C. (2011). Photosynthesis, a global sensor of environmental stress in green plants: stress signalling and adaptation on JSTOR. *Current Science*, *101*(1). https://www.jstor.org/stable/24077862

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2021). The InterPro

protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., Goodrich, J., Renou, J. P., Grini, P. E., Colot, V., & Schnittger, A. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genetics*, *7*(3). https://doi.org/10.1371/JOURNAL.PGEN.1002014

Bowman, J. L., Floyd, S. K., & Sakakibara, K. (2007). Green Genes—Comparative Genomics of the Green Branch of Life. *Cell*, *129*(2), 229–234. https://doi.org/10.1016/j.cell.2007.04.004

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., & Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development*, *20*(9), 1123–1136. https://doi.org/10.1101/GAD.381706

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J. C., & Calonje, M. (2010). Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Current Biology : CB*, *20*(20), 1853–1859. https://doi.org/10.1016/J.CUB.2010.09.046

Brzezinka, K., Altmann, S., Czesnick, H., Nicolas, P., Gorka, M., Benke, E., Kabelitz, T., Jähne, F., Graf, A., Kappel, C., & Bäurle, I. (2016). Arabidopsis FORGETTER1 mediates stress-induced chromatin memory through nucleosome remodeling. *ELife*, *5*. https://doi.org/10.7554/ELIFE.17061

Carlier, F., Li, M., Maroc, L., Debuchy, R., Souaid, C., Noordermeer, D., Grognet, P., & Malagnac, F. (2021). Loss of EZH2-like or SU(VAR)3–9-like proteins causes simultaneous perturbations in H3K27 and H3K9 tri-methylation and associated developmental defects in the fungus Podospora anserina. *Epigenetics & Chromatin 2021 14:1*, *14*(1), 1–28. https://doi.org/10.1186/S13072-021-00395-7

Cavrak, V. V, Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., & Mittelsten Scheid, O. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. *PLoS Genetics*, *10*(1). https://doi.org/10.1371/journal.pgen.1004115

Chammas, P., Mocavini, I., & Di Croce, L. (2020). Engaging chromatin: PRC2 structure meets function. *British Journal of Cancer*, *122*(3), 315–328. https://doi.org/10.1038/S41416-019-0615-2

Chang, Y. N., Zhu, C., Jiang, J., Zhang, H., Zhu, J. K., & Duan, C. G. (2020). Epigenetic regulation in plant abiotic stress responses. *Journal of Integrative Plant Biology*, *62*(5), 563–580. https://doi.org/10.1111/JIPB.12901

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y. H., Sung, Z. R., & Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development (Cambridge, England)*, *131*(21), 5263–5276. https://doi.org/10.1242/DEV.01400

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, *12*(1), 1–7. https://doi.org/10.1186/1471-2105-12-35

Chouard, P. (1960). Vernalization and its Relations to Dormancy. *Annual Review of Plant Physiology*, *11*(1), 191–238. https://doi.org/10.1146/ANNUREV.PP.11.060160.001203

Ciniglia, C., Yoon, H. S., Pollio, A., Pinto, G., & Bhattacharya, D. (2004). Hidden biodiversity of the extremophilic Cyanidiales red algae. *Molecular Ecology*, *13*(7), 1827–1838. https://doi.org/10.1111/J.1365-294X.2004.02180.X

Connolly, L. R., Smith, K. M., & Freitag, M. (2013). The Fusarium graminearum Histone H3 K27 Methyltransferase KMT6 Regulates Development and Expression of Secondary Metabolite Gene Clusters. *PLOS Genetics*, *9*(10), e1003916. https://doi.org/10.1371/JOURNAL.PGEN.1003916

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/NAR/GKAB1049

Danilova, M. N., Kudryakova, N. V., Andreeva, A. A., Doroshenko, A. S., Pojidaeva, E. S., & Kusnetsov, V. V. (2018). Differential impact of heat stress on the expression of chloroplast-encoded genes. *Plant Physiology and Biochemistry : PPB*, *129*, 90–100. https://doi.org/10.1016/J.PLAPHY.2018.05.023

De Jong, W. W., Leunissen, J. A. M., & Voorter, C. E. M. (1993). Evolution of the alpha-crystallin/small heat-shock protein family. *Molecular Biology and Evolution*, *10*(1), 103–126. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A039992

De Luca, P., Taddei, R., & Varano, L. (1978). « Cyanidioschyzon merolae »: a new alga of thermal acidic environments. *Webbia*, *33*(1), 37–44. https://doi.org/10.1080/00837792.1978.10670110

de Mendiburu, F., & Yaseen, M. (2020). *agricolae: Statistical Procedures for Agricultural Research. R package* (1.3-5).

Derkacheva, M., & Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *Journal of Experimental Botany*, *65*(10), 2769–2784. https://doi.org/10.1093/jxb/ert410

Ding, Y., Fromm, M., & Avramova, Z. (2012). Multiple exposures to drought “train” transcriptional responses in Arabidopsis. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms1732

Ding, Y., Liu, N., Virlouvet, L., Riethoven, J. J., Fromm, M., & Avramova, Z. (2013). Four distinct types of dehydration stress memory genes in Arabidopsis thaliana. *BMC Plant Biology*, *13*(1), 1–11. https://doi.org/10.1186/1471-2229-13-229

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, Y., Uslu, V. V., Berr, A., Singh, G., Papdi, C., Steffens, V. A., Heitz, T., & Ryabova, L. (2021). TOR represses stress responses through global regulation of H3K27 trimethylation in plants. *BioRxiv*, 2021.03.28.437410. https://doi.org/10.1101/2021.03.28.437410

Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews. Molecular Cell Biology*, *16*(9), 519–532. https://doi.org/10.1038/NRM4043

Duan, R., Du, W., & Guo, W. (2020). EZH2: a novel target for cancer treatment. *Journal of Hematology & Oncology*, *13*(1). https://doi.org/10.1186/S13045-020-00937-8

Ehrnsperger, M., Gaestel, M., & Buchner, J. (2000). Analysis of Chaperone Properties of Small Hsp’. *Methods in Molecular Biology (Clifton, N.J.)*, *99*, 421–429. https://doi.org/10.1385/1-59259-054-3:421

Erland, L. A. E., & Saxena, P. (2019). Auxin driven indoleamine biosynthesis and the role of tryptophan as an inductive signal in Hypericum perforatum (L.). *PLoS ONE*, *14*(10). https://doi.org/10.1371/JOURNAL.PONE.0223878

Erland, L. A. E., Yasunaga, A., Li, I. T. S., Murch, S. J., & Saxena, P. K. (2018). Direct visualization of location and uptake of applied melatonin and serotonin in living tissues and their redistribution in plants in response to thermal stress. *Journal of Pineal Research*, *66*(1), e12527. https://doi.org/10.1111/JPI.12527

Finch, J. T., Lutter, L. C., Rhodes, D., Brown, R. S., Rushton, B., Levitt, M., & Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature*, *269*(5623), 29–36. https://doi.org/10.1038/269029A0

Flannery, S. E., Hepworth, C., Wood, W. H. J., Pastorelli, F., Hunter, C. N., Dickman, M. J., Jackson, P. J., & Johnson, M. P. (2021). Developmental acclimation of the thylakoid proteome to light intensity in Arabidopsis. *The Plant Journal*, *105*(1), 223. https://doi.org/10.1111/TPJ.15053

Franck, E., Madsen, O., Van Rheede, T., Ricard, G., Huynen, M. A., & De Jong, W. W. (2004). Evolutionary diversity of vertebrate small heat shock proteins. *Journal of Molecular Evolution*, *59*(6), 792–805. https://doi.org/10.1007/S00239-004-0013-Z

Frapporti, A., Miró Pina, C., Arnaiz, O., Holoch, D., Kawaguchi, T., Humbert, A., Eleftheriou, E., Lombard, B., Loew, D., Sperling, L., Guitot, K., Margueron, R., & Duharcourt, S. (2019). The

Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-10648-5

Friedrich, T., Faivre, L., Bäurle, I., & Schubert, D. (2018). Chromatin-based mechanisms of temperature memory in plants. *Plant, Cell & Environment*, *42*(3), 762–770. https://doi.org/10.1111/PCE.13373

Fujii, G., Imamura, S., Era, A., Miyagishima, S. Y., Hanaoka, M., & Tanaka, K. (2015). The nuclear-encoded sigma factor SIG4 directly activates transcription of chloroplast psbA and ycf17 genes in the unicellular red alga Cyanidioschyzon merolae. *FEMS Microbiology Letters*, *362*(10). https://doi.org/10.1093/femsle/fnv063

Fujii, G., Imamura, S., Hanaoka, M., & Tanaka, K. (2013). Nuclear-encoded chloroplast RNA polymerase sigma factor SIG2 activates chloroplast-encoded phycobilisome genes in a red alga, Cyanidioschyzon merolae. *FEBS Letters*, *587*(20), 3354–3359. https://doi.org/10.1016/j.febslet.2013.08.031

Fujiwara, T., Hirooka, S., Ohbayashi, R., Onuma, R., & Miyagishima, S. Y. (2020). Relationship between cell cycle and diel transcriptomic changes in metabolism in a unicellular red alga1[OPEN]. *Plant Physiology*, *183*(4), 1484–1501. https://doi.org/10.1104/pp.20.00469

Fujiwara, T., Kanesaki, Y., Hirooka, S., Era, A., Sumiya, N., Yoshikawa, H., Tanaka, K., & Miyagishima, S. Y. (2015). A nitrogen source-dependent inducible and repressible gene expression system in the red alga Cyanidioschyzon merolae. *Front Plant Sci*, *6*(August). https://doi.org/10.3389/fpls.2015.00657

Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M., Morris, S. A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B. D., Allis, C. D., & Hunt, D. F. (2007). Organismal differences in post-translational modifications in histones H3 and H4. *The Journal of Biological Chemistry*, *282*(10), 7641–7655. https://doi.org/10.1074/JBC.M607900200

Gelhaye, E., Rouhier, N., Navrot, N., & Jacquot, J. P. (2005). The plant thioredoxin system. *CMLS Cellular and Molecular Life Sciences*, *62*, 24–35. https://doi.org/10.1007/s00018-004-4296-4

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, *17*(6), 877. https://doi.org/10.1101/GR.5533506

Graham, L. E., & Wilcox, L. W. (2000). *Algae*. Prentice Hall.

Green, B. R., & Durnford, D. G. (1996). THE CHLOROPHYLL-CAROTENOID PROTEINS OF OXYGENIC PHOTOSYNTHESIS. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*(1), 685–714. https://doi.org/10.1146/ANNUREV.ARPLANT.47.1.685

Gross, W. (1999). Revision of Comparative Traits for the Acido- and Thermophilic Red Algae and. *Enigmatic Microorganisms and Life in Extreme Environments*, 437–446. https://doi.org/10.1007/978-94-011-4838-2_34

Guitton, A. E., & Berger, F. (2005). Control of reproduction by Polycomb Group complexes in animals and plants. *The International Journal of Developmental Biology*, *49*(5–6), 707–716. https://doi.org/10.1387/IJDB.051990AG

Guo, L., & Yang, G. (2015). Predicting the Reproduction Strategies of Several Microalgae Through Their Genome Sequences. *J. Ocean Univ. China (Oceanic and Coastal Sea Research*, *14*(3), 491–502. https://doi.org/10.1007/s11802-015-2442-2

Heckathorn, S. A., Downs, C. A., Sharkey, T. D., & Coleman, J. S. (1998). The small, methionine-rich chloroplast heat-shock protein protects photosystem II electron transport during heat stress. *Plant Physiology*, *116*(1), 439–444. https://doi.org/10.1104/PP.116.1.439

Herrero, A., Muro-Pastor, A. M., & Flores, E. (2001). Nitrogen control in cyanobacteria. *Journal of Bacteriology*, *183*(2), 411–425. https://doi.org/10.1128/JB.183.2.411-425.2001

Herrero, A., Muro-Pastor, A. M., Valladares, A., & Flores, E. (2004). Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. *FEMS Microbiology Reviews*, *28*(4), 469–

487. https://doi.org/10.1016/J.FEMSRE.2004.04.003

Hilker, M., Schwachtje, J., Baier, M., Balazadeh, S., Bäurle, I., Geiselhardt, S., Hincha, D. K., Kunze, R., Mueller-Roeber, B., Rillig, M. C., Rolff, J., Romeis, T., Schmülling, T., Steppuhn, A., van Dongen, J., Whitcomb, S. J., Wurst, S., Zuther, E., & Kopka, J. (2016). Priming and memory of stress responses in organisms lacking a nervous system. *Biological Reviews*, *91*(4), 1118–1133. https://doi.org/10.1111/brv.12215

Hu, S., Ding, Y., & Zhu, C. (2020). Sensitivity and Responses of Chloroplasts to Heat Stress in Plants. *Frontiers in Plant Science*, *11*, 375. https://doi.org/10.3389/FPLS.2020.00375/BIBTEX

Hu, X., Yang, Y., Gong, F., Zhang, D., Zhang, L., Wu, L., Li, C., & Wang, W. (2014). Protein sHSP26 improves chloroplast performance under heat stress by interacting with specific chloroplast proteins in maize (Zea mays). *Journal of Proteomics*, *115*, 81–92. https://doi.org/10.1016/J.JPROT.2014.12.009

Huff, J. T., & Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell*, *156*(6), 1286–1297. https://doi.org/10.1016/j.cell.2014.01.029

Hugues, A., Jacobs, C. S., & Roudier, F. (2020). Mitotic Inheritance of PRC2-Mediated Silencing: Mechanistic Insights and Developmental Perspectives. *Frontiers in Plant Science*, *11*, 262. https://doi.org/10.3389/FPLS.2020.00262/BIBTEX

Imamura, S., Kawase, Y., Kobayashi, I., Sone, T., Era, A., Miyagishima, S., Shimojima, M., Ohta, H., & Tanaka, K. (2015). Target of rapamycin (TOR) plays a critical role in triacylglycerol accumulation in microalgae. *Plant Molecular Biology 2015 89:3*, *89*(3), 309–318. https://doi.org/10.1007/S11103-015-0370-6

Imamura, S., Nomura, Y., Takemura, T., Pancha, I., Taki, K., Toguchi, K., Tozawa, Y., & Tanaka, K. (2018). The checkpoint kinase TOR (target of rapamycin) regulates expression of a nuclear-encoded chloroplast RelA-SpoT homolog (RSH) and modulates chloroplast ribosomal RNA synthesis in a unicellular red alga. *Plant Journal*, *94*(2), 327–339. https://doi.org/10.1111/tpj.13859

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., & Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, *472*(7341), 115–120. https://doi.org/10.1038/NATURE09861

Jamieson, K., Rountree, M. R., Lewis, Z. A., Stajich, J. E., & Selker, E. U. (2013). Regional control of histone H3 lysine 27 methylation in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 6027–6032. https://doi.org/10.1073/PNAS.1303750110

Janka, E., Körner, O., Rosenqvist, E., & Ottosen, C. O. (2013). High temperature stress monitoring and detection using chlorophyll a fluorescence and infrared thermography in chrysanthemum (Dendranthema grandiflora). *Plant Physiology and Biochemistry : PPB*, *67*, 87–94. https://doi.org/10.1016/J.PLAPHY.2013.02.025

Jaskiewicz, M., Conrath, U., & Peterhälnsel, C. (2011). Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Reports*, *12*(1), 50–55. https://doi.org/10.1038/EMBOR.2010.186

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science*, *293*(5532), 1074–1080. https://doi.org/10.1126/science.1063127

Jin, J., Yang, L., Fan, D., Liu, X., & Hao, Q. (2020). Comparative transcriptome analysis uncovers different heat stress responses in heat-resistant and heat-sensitive jujube cultivars. *PloS One*, *15*(9). https://doi.org/10.1371/JOURNAL.PONE.0235763

Kim, J. H. (2021). Multifaceted Chromatin Structure and Transcription Changes in Plant Stress Response. *International Journal of Molecular Sciences*, *22*(4), 1–25. https://doi.org/10.3390/IJMS22042013

Kim, J. M., Sasaki, T., Ueda, M., Sako, K., & Seki, M. (2015). Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Frontiers in Plant Science*, *6*(MAR), 114.

https://doi.org/10.3389/FPLS.2015.00114/BIBTEX

Kirmizis, A., Bartley, S. M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., & Farnham, P. J. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes & Development*, *18*(13), 1592–1605. https://doi.org/10.1101/GAD.1200204

Kishimoto, I., Ariga, I., Itabashi, Y., & Mikami, K. (2019). Heat-stress Memory is Responsible for Acquired Thermotolerance in Bangia fuscopurpurea. *Journal of Phycology*, *55*(5), 971–975. https://doi.org/10.1111/JPY.12895

Kleinmanns, J. A., Schatlowski, N., Heckmann, D., & Schubert, D. (2017). BLISTER regulates polycomb-target genes, represses stress-regulated genes and promotes stress responses in Arabidopsis thaliana. *Frontiers in Plant Science*, *8*, 1530. https://doi.org/10.3389/FPLS.2017.01530/BIBTEX

Kleinmanns, J. A., & Schubert, D. (2014). Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biological Chemistry*, *395*(11), 1291–1300. https://doi.org/10.1515/hsz-2014-0197

Kobayashi, Y., Harada, N., Nishimura, Y., Saito, T., Nakamura, M., Fujiwara, T., Kuroiwa, T., & Misumi, O. (2014). Algae sense exact temperatures: Small heat shock proteins are expressed at the survival threshold temperature in Cyanidioschyzon merolae and Chlamydomonas reinhardtii. *Genome Biology and Evolution*, *6*(10), 2731–2740. https://doi.org/10.1093/gbe/evu216

Kobayashi, Y., Ohnuma, M., Tanaka, K., & Hanaoka, M. (2010). The basics of cultivation and molecular genetic analysis of the unicellular red alga Cyanidioschyzon merolae. *Cell*, 53–61.

Kolde, R. (2019). *pheatmap: Pretty Heatmaps. R package version 1.0.12.* https://cran.r-project.org/package=pheatmap

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705.

Kuroiwa, T., Miyagishima, S., Matsunaga, S., Sato, N., Nozaki, H., Tanaka, K., & Misumi, O. (2018). Cyanidioschyzon merolae: A new model eukaryote for cell and organelle biology. In *Cyanidioschyzon merolae: A New Model Eukaryote for Cell and Organelle Biology*. Springer Singapore. https://doi.org/10.1007/978-981-10-6101-1

Kwon, C. S., Lee, D., Choi, G., & Chung, W. Il. (2009). Histone occupancy-dependent and -independent removal of H3K27 trimethylation at cold-responsive genes in Arabidopsis. *The Plant Journal*, *60*(1), 112–121. https://doi.org/10.1111/J.1365-313X.2009.03938.X

Lafos, M., Kroll, P., Hohenstatt, M. L., Thorpe, F. L., Clarenz, O., & Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS Genetics*, *7*(4). https://doi.org/10.1371/journal.pgen.1002040

Lämke, J., & Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology 2017 18:1*, *18*(1), 1–11. https://doi.org/10.1186/S13059-017-1263-6

Lämke, J., Brzezinka, K., Altmann, S., & Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *The EMBO Journal*, *35*(2), 162–175. https://doi.org/10.15252/EMBJ.201592593

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., & Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development*, *24*(3), 265–276. https://doi.org/10.1101/GAD.544410

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature 1978 276:5688*, *276*(5688), 565–570. https://doi.org/10.1038/276565a0

Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., & Duharcourt, S. (2014). Local Effect of Enhancer of Zeste-Like Reveals Cooperation of Epigenetic and cis-Acting Determinants for Zygotic Genome Rearrangements. *PLOS Genetics*, *10*(9), e1004665. https://doi.org/10.1371/JOURNAL.PGEN.1004665

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16),

2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, X., Cai, C., Wang, Z., Fan, B., Zhu, C., & Chena, Z. (2018). Plastid Translation Elongation Factor Tu Is Prone to Heat-Induced Aggregation Despite Its Critical Role in Plant Heat Tolerance. *Plant Physiology*, *176*(4), 3027. https://doi.org/10.1104/PP.17.01672

Li, Y., Ren, B., Ding, L., Shen, Q., Peng, S., & Guo, S. (2013). Does Chloroplast Size Influence Photosynthetic Nitrogen Use Efficiency? *PLOS ONE*, *8*(4), e62036. https://doi.org/10.1371/JOURNAL.PONE.0062036

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, S., Offner, G. D., & Troxler, R. F. (1990). Studies on Cyanidium caldarium Phycobiliprotein Pigment Mutants. *Plant Physiology*, *93*(2), 772. https://doi.org/10.1104/PP.93.2.772

Liu, H.-C., Lämke, J., Lin, S.-Y., Hung, M.-J., Liu, K.-M., Charng, Y.-Y., & B€ Aurle, I. (2018). Distinct heat shock factors and chromatin modifications mediate the organ-autonomous transcriptional memory of heat stress. *The Plant Journal*, *95*(3), 401–413. https://doi.org/10.1111/TPJ.13958

Liu, J., Feng, L., Gu, X., Deng, X., Qiu, Q., Li, Q., Zhang, Y., Wang, M., Deng, Y., Wang, E., He, Y., Bäurle, I., Li, J., Cao, X., & He, Z. (2019). An H3K27me3 demethylase-HSFA2 regulatory loop orchestrates transgenerational thermomemory in Arabidopsis. *Cell Research*, *29*(5), 379. https://doi.org/10.1038/S41422-019-0145-8

Liu, N., Fromm, M., & Avramova, Z. (2014). H3K27me3 and H3K4me3 Chromatin Environment at Super-Induced Dehydration Stress Memory Genes of Arabidopsis thaliana. *Molecular Plant*, *7*(3), 502–513. https://doi.org/10.1093/MP/SSU001

Liu, S. L., Chiang, Y. R., Yoon, H. S., & Fu, H. Y. (2020). Comparative Genome Analysis Reveals Cyanidiococcus gen. nov., A New Extremophilic Red Algal Genus Sister to Cyanidioschyzon (Cyanidioschyzonaceae, Rhodophyta). *Journal of Phycology*, *56*(6), 1428–1442. https://doi.org/10.1111/jpy.13056

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., & Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes & Development*, *21*(12), 1530–1545. https://doi.org/10.1101/gad.1544207

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/S13059-014-0550-8

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature 1997 389:6648*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Maeda, H., & Dudareva, N. (2012). The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annual Reviews*, *63*, 73–105. https://doi.org/10.1146/ANNUREV-ARPLANT-042811-105439

Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., Watkins, K. P., Barkan, A., & van Wijk, K. J. (2012). Nucleoid-Enriched Proteomes in Developing Plastids and Chloroplasts from Maize Leaves: A New Conceptual Framework for Nucleoid Functions. *Plant Physiology*, *158*(1), 156–189. https://doi.org/10.1104/PP.111.188474

Malik, S.-B. (2007). The early evolution of meiotic genes. *Theses and Dissertations*. http://ir.uiowa.edu/etd/275

Mallén-Ponce, M. J., Huertas, M. J., Sanchez-Riego, A. M., & Florencio, F. J. (2021). Depletion of m-type thioredoxin impairs photosynthesis, carbon fixation, and oxidative stress in cyanobacteria. *Plant Physiology*, *187*(3), 1325–1340. https://doi.org/10.1093/PLPHYS/KIAB321

Margueron, R., & Reinberg, D. (2011). The Polycomb Complex PRC2 and its Mark in Life. *Nature*, *469*(7330), 343–349. https://doi.org/10.1038/nature09784.The

Mathur, S., Allakhverdiev, S. I., & Jajoo, A. (2011). Analysis of high temperature stress on the

dynamics of antenna size and reducing side heterogeneity of Photosystem II in wheat leaves (Triticum aestivum). *Biochimica et Biophysica Acta*, *1807*(1), 22–29. https://doi.org/10.1016/J.BBABIO.2010.09.001

Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., … Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*, *428*(6983), 653–657. https://doi.org/10.1038/nature02398

Mayer, B. F., & Charron, J. B. (2021). Transcriptional memories mediate the plasticity of cold stress responses to enable morphological acclimation in Brachypodium distachyon. *The New Phytologist*, *229*(3), 1615. https://doi.org/10.1111/NPH.16945

Meganathan, R., & Kwon, O. (2009). Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone(Coenzyme Q). *EcoSal Plus*, *3*(2). https://doi.org/10.1128/ECOSALPLUS.3.6.3.3

Merola, A., Castaldo, R., De Luca, P., Gambardella, R., Musacchio, A., & Taddei, R. (1981). Revision of Cyanidium caldarium. Three species of acidophilic algae. *Giornale Botanico Italiano*, *115*(4–5), 189–195. https://doi.org/10.1080/11263508109428026

Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A. P. M., & Schubert, D. (2017). Characterization of the Polycomb-Group Mark H3K27me3 in Unicellular Algae. *Frontiers in Plant Science*, *8*(April), 1–17. https://doi.org/10.3389/fpls.2017.00607

Mikulski, P., & Santos-Aberturas, J. (2021). Chlamydomonas reinhardtii exhibits stress memory in the accumulation of triacylglycerols induced by nitrogen deprivation. *BioRxiv*, 2021.07.23.453471. https://doi.org/10.1101/2021.07.23.453471

Mikulski, P., & Santos-Aberturas, J. (2022). Chlamydomonas reinhardtii exhibits stress memory in the accumulation of triacylglycerols induced by nitrogen deprivation. *Plant-Environment Interactions*, *3*(1), 10–15. https://doi.org/10.1002/PEI3.10069

Miller, G., Shulaev, V., & Mittler, R. (2008). Reactive oxygen signaling and abiotic stress. *Physiologia Plantarum*, *133*(3), 481–489. https://doi.org/10.1111/J.1399-3054.2008.01090.X

Minoda, A., Nagasawa, K., Hanaoka, M., Horiuchi, M., Takahashi, H., & Tanaka, K. (2005). Microarray profiling of plastid gene expression in a unicellular red alga, Cyanidioschyzon merolae. *Plant Molecular Biology*, *59*(3), 375–385. https://doi.org/10.1007/s11103-005-0182-1

Minoda, A., Weber, A. P. M., Tanaka, K., & Miyagishima, S. ya. (2010). Nucleus-independent control of the rubisco operon by the plastid-encoded transcription factor Ycf30 in the red alga Cyanidioschyzon merolae. *Plant Physiology*, *154*(3), 1532–1540. https://doi.org/10.1104/PP.110.163188

Mittler, R., Vanderauwera, S., Gollery, M., & Van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends in Plant Science*, *9*(10), 490–498. https://doi.org/10.1016/J.TPLANTS.2004.08.009

Miyagishima, S. Y., & Tanaka, K. (2021). The Unicellular Red Alga Cyanidioschyzon merolae-The Simplest Model of a Photosynthetic Eukaryote. *Plant & Cell Physiology*, *62*(6), 926–941. https://doi.org/10.1093/PCP/PCAB052

Mohammed, A. R., & Tarpley, L. (2009). Impact of High Nighttime Temperature on Respiration, Membrane Stability, Antioxidant Capacity, and Yield of Rice Plants. *Crop Science*, *49*(1), 313–322. https://doi.org/10.2135/CROPSCI2008.03.0161

Montero, J. J., López-Silanes, I., Megías, D., Fraga, M., Castells-García, Á., & Blasco, M. A. (2018). TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin. *Nature Communications 2018 9:1*, *9*(1), 1–14. https://doi.org/10.1038/s41467-018-03916-3

Mozgova, I., Köhker, C., & Hennig, L. (2015). Keeping the gate closed : functions of the polycomb repressive complex PRC2 in development. *The Plant Journal*, *83*, 121–132. https://doi.org/10.1111/tpj.12828

Mozgová, I., Wildhaber, T., Liu, Q., Abou-Mansour, E., L'Haridon, F., Métraux, J. P., Gruissem, W., Hofius, D., & Hennig, L. (2015). Chromatin assembly factor CAF-1 represses priming of plant defence response genes. *Nature Plants 2015 1:9*, *1*(9), 1–8. https://doi.org/10.1038/nplants.2015.127

Müller, J., & Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Current Opinion in Genetics and Development*, *19*(2), 150–158. https://doi.org/10.1016/j.gde.2009.03.001

Murata, N., Takahashi, S., Nishiyama, Y., & Allakhverdiev, S. I. (2007). Photoinhibition of photosystem II under environmental stress. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1767*(6), 414–421. https://doi.org/10.1016/J.BBABIO.2006.11.019

Nakayama, Y., Goebl, M., O'Brine Greco, B., Lemmon, S., Pingchang Chow, E., & Kirchhausen, T. (1991). The medium chains of the mammalian clathrin-associated proteins have a homolog in yeast. *European Journal of Biochemistry*, *202*(2), 569–574. https://doi.org/10.1111/J.1432-1033.1991.TB16409.X

Neill, S. J., Desikan, R., Clarke, A., Hurst, R. D., & Hancock, J. T. (2002). Hydrogen peroxide and nitric oxide as signalling molecules in plants. *Journal of Experimental Botany*, *53*(372), 1237–1247. https://doi.org/10.1093/JEXBOT/53.372.1237

Neilson, J. A. D., & Durnford, D. G. (2010). Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynthesis Research 2010 106:1*, *106*(1), 57–71. https://doi.org/10.1007/S11120-010-9576-2

Niforou, K., Cheimonidou, C., & Trougakos, I. P. (2014). Molecular chaperones and proteostasis regulation during redox imbalance. *Redox Biology*, *2*(1), 323. https://doi.org/10.1016/J.REDOX.2014.01.017

Nikolova, D., Weber, D., Scholz, M., Bald, T., Scharsack, J. P., & Hippler, M. (2017). Temperature-Induced Remodeling of the Photosynthetic Machinery Tunes Photosynthesis in the Thermophilic Alga Cyanidioschyzon merolae. *Plant Physiology*, *174*(1), 35–46. https://doi.org/10.1104/PP.17.00110

Nishad, A., & Nandi, A. K. (2021). Recent advances in plant thermomemory. *Plant Cell Reports*, *40*(1), 19–27. https://doi.org/10.1007/S00299-020-02604-1

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., Takio, S., Tamura, K., Chung, S. J., Nakamura, S., Kuroiwa, H., Tanaka, K., Sato, N., & Kuroiwa, T. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biology*, *5*, 1–8. https://doi.org/10.1186/1741-7007-5-28

Oberkofler, V., Pratx, L., & Bäurle, I. (2021). Epigenetic regulation of abiotic stress memory: maintaining the good things while they last. *Current Opinion in Plant Biology*, *61*, 102007. https://doi.org/10.1016/J.PBI.2021.102007

Oesterhelt, C., Schnarrenberger, C., & Gross, W. (1999). European Journal of Phycology Characterization of a sugar/polyol uptake system in the red alga Galdieria sulphuraria Characterization of a sugar/polyol uptake system in the red alga Galdieria sulphuraria. *European Journal of Phycology*, *34*(3), 271–277. https://doi.org/10.1080/09670269910001736322

Ohama, N., Sato, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2017). Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science*, *22*(1), 53–65. https://doi.org/10.1016/J.TPLANTS.2016.08.015

Ohnuma, M., Yokoyama, T., Inouye, T., Sekine, Y., & Tanaka, K. (2008a). Polyethylene Glycol (PEG)-Mediated Transient Gene Expression in a Red Alga, Cyanidioschyzon merolae 10D. *Plant and Cell Physiology*, *49*(1), 117–120. https://doi.org/10.1093/pcp/pcm157

Ohnuma, M., Yokoyama, T., Inouye, T., Sekine, Y., & Tanaka, K. (2008b). Polyethylene Glycol (PEG)-Mediated Transient Gene Expression in a Red Alga, Cyanidioschyzon merolae 10D. *Plant and Cell Physiology*, *49*(1), 117–120. https://doi.org/10.1093/pcp/pcm157

Ohta, N., Matsuzaki, M., Misumi, O., Miyagishima, S. y., Nozaki, H., Tanaka, K., Shin-i, T., Kohara, Y.,

& Kuroiwa, T. (2003). Complete sequence and analysis of the plastid genome of the unicellular red alga Cyanidioschyzon merolae. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *10*(2), 67–77. https://doi.org/10.1093/DNARES/10.2.67

Pecinka, A., Dinh, H. Q., Baubec, T., Rosa, M., Lettner, N., & Scheid, O. M. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *The Plant Cell*, *22*(9), 3118–3129. https://doi.org/10.1105/TPC.110.078493

Pfaffl, M. W. (2001). A New Mathematical Model for Relative Quantification in Real-Time RT-PCR. *Nucleic Acids Research*, *29*(9). https://doi.org/10.1093/nar/29.9.e45

Plesofsky-Vig, N., Vig, J., & Brambl, R. (1992). Phylogeny of the alpha-crystallin-related heat-shock proteins. *Journal of Molecular Evolution*, *35*(6), 537–545. https://doi.org/10.1007/BF00160214

Powikrowska, M., Oetke, S., Jensen, P. E., & Krupinska, K. (2014). Dynamic composition, shaping and organization of plastid nucleoids. *Frontiers in Plant Science*, *5*(SEP). https://doi.org/10.3389/FPLS.2014.00424

Probst, A. V., Dunleavy, E., & Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nature Reviews. Molecular Cell Biology*, *10*(3), 192–206. https://doi.org/10.1038/NRM2640

Rademacher, N., Wrobel, T. J., Rossoni, A. W., Kurz, S., Bräutigam, A., Weber, A. P. M., & Eisenhut, M. (2017). Transcriptional response of the extremophile red alga Cyanidioschyzon merolae to changes in CO2 concentrations. *Journal of Plant Physiology*, *217*, 49–56. https://doi.org/10.1016/J.JPLPH.2017.06.014

Ramakrishnan, V. (1997). Histone H1 and chromatin higher-order structure. *Critical Reviews in Eukaryotic Gene Expression*, *7*(3), 215–230. https://doi.org/10.1615/CRITREVEUKARGENEEXPR.V7.I3.20

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/NAR/GKW257

Reynolds, A. E., Chesnick, J. M., Woolford, J., & Cattolico, R. A. (1994). Chloroplast encoded thioredoxin genes in the red algae Porphyra yezoensis and Griffithsia pacifica: evolutionary implications. *Plant Molecular Biology*, *25*(1), 13–21. https://doi.org/10.1007/BF00024194

Rigano, C., Aliotta, G., Di Martino Rigano, V., Fuggi, A., & Vona, V. (1977). Heterotrophic growth patterns in the unicellular alga Cyanidium caldarium. *Archives of Microbiology 1977 113:3*, *113*(3), 191–196. https://doi.org/10.1007/BF00492024

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology 2011 29:1*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rorbach, J., Bobrowicz, A., Pearce, S., & Minczuk, M. (2014). Polyadenylation in Bacteria and Organelles. *Methods in Molecular Biology*, *1125*, 211–227. https://doi.org/10.1007/978-1-62703-971-0_18

Salganik, R. I., Dudareva, N. A., & Kiseleva, E. V. (1991). Structural organization and transcription of plant mitochondrial and chloroplast genomes. *Electron Microscopy Reviews*, *4*(2), 221–247. https://doi.org/10.1016/0892-0354(91)90004-V

Sani, E., Herzyk, P., Perrella, G., Colot, V., & Amtmann, A. (2013). Hyperosmotic priming of Arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biology*, *14*(6), 1–24. https://doi.org/10.1186/GB-2013-14-6-R59/TABLES/3

Scharf, K. D., Siddique, M., & Vierling, E. (2001). The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins). *Cell Stress & Chaperones*, *6*(3), 225. https://doi.org/10.1379/1466-1268(2001)006<0225:tefoat>2.0.co;2

Schubert, D. (2019). Evolution of polycomb-group function in the green lineage [version 1; referees: 2

approved]. *F1000Research*, *8*, 1–6. https://doi.org/10.12688/f1000research.16986.1

Schuettengruber, B., Bourbon, H. M., Di Croce, L., & Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, *171*(1), 34–57. https://doi.org/10.1016/j.cell.2017.08.002

Schwartz, Y. B., Kahn, T. G., Stenberg, P., Ohno, K., Bourgon, R., & Pirrotta, V. (2010). Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genetics*, *6*(1). https://doi.org/10.1371/JOURNAL.PGEN.1000805

Schwartz, Y. B., & Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Current Opinion in Cell Biology*, *20*(3), 266–273. https://doi.org/10.1016/J.CEB.2008.03.002

Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: eunexpected partnerships and emerging functions. *Nature Reviews*, *14*, 853–864. https://doi.org/10.1038/nrg3603

Shaver, S., Casas-Mollano, J. A., Cerny, R. L., & Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics*. https://doi.org/10.4161/epi.5.4.11608

Shen, Q., Lin, Y., Li, Y., & Wang, G. (2021). Dynamics of H3K27me3 Modification on Plant Adaptation to Environmental Cues. *Plants (Basel, Switzerland)*, *10*(6). https://doi.org/10.3390/PLANTS10061165

Smith, K. M., Kothe, G. O., Matsen, C. B., Khlafallah, T. K., Adhvaryu, K. K., Hemphill, M., Freitag, M., Motamedi, M. R., & Selker, E. U. (2008). The fungus Neurospora crassa displays telomeric silencing mediated by multiple sirtuins and by methylation of histone H3 lysine 9. *Epigenetics & Chromatin*, *1*(1), 5. https://doi.org/10.1186/1756-8935-1-5

Sorger, P. K., & Nelson, H. C. M. (1989). Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell*, *59*(5), 807–813. https://doi.org/10.1016/0092-8674(89)90604-1

Sorger, P. K., & Pelham, H. R. B. (1988). Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell*, *54*(6), 855–864. https://doi.org/10.1016/S0092-8674(88)91219-6

Stephens, M. (2017). False discovery rates: A new deal. *Biostatistics*, *18*(2), 275–294. https://doi.org/10.1093/BIOSTATISTICS/KXW041

Stewart, M. (2019). Polyadenylation and nuclear export of mRNAs. *Journal of Biological Chemistry*, *294*(9), 2977–2987. https://doi.org/10.1074/JBC.REV118.005594

Sun, A. Z., & Guo, F. Q. (2016). Chloroplast retrograde regulation of heat stress responses in plants. *Frontiers in Plant Science*, *7*(MAR2016), 398. https://doi.org/10.3389/FPLS.2016.00398/BIBTEX

Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z., Wang, D., He, H., & Qian, W. (2020). Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in Arabidopsis. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15809-5

Sun, Y. J., Peffer, N., & Meier, I. (2004). Phosphorylation by protein kinase CKII modulates the DNA-binding activity of a chloroplast nucleoid-associated protein. *Planta*, *219*(2), 298–302. https://doi.org/10.1007/S00425-004-1215-8

Suzuki, N., Koussevitzky, S., Mittler, R., & Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress. *Plant, Cell & Environment*, *35*(2), 259–270. https://doi.org/10.1111/J.1365-3040.2011.02336.X

Tadini, L., Jeran, N., Peracchio, C., Masiero, S., Colombo, M., & Pesaresi, P. (2020). *The plastid transcription machinery and its coordination with the expression of nuclear genome: Plastid-Encoded Polymerase, Nuclear-Encoded Polymerase and the Genomes Uncoupled 1-mediated retrograde communication*. https://doi.org/10.1098/rstb.2019.0399

Tardu, M., Dikbas, U. M., Baris, I., & Kavakli, I. H. (2016). RNA-seq analysis of the transcriptional response to blue and red light in the extremophilic red alga, Cyanidioschyzon merolae. *Functional and Integrative Genomics*, *16*(6), 657–669. https://doi.org/10.1007/s10142-016-0521-0

Terui, S., Suzuki, K., Takahashi, H., Itoh, R., & Kuroiwa, T. (1995). Synchronization of chloroplast division in the ultramicroalga Cyanidioschyzon metolae (Rhodophyta) by treatment with light and aphidicolin. *Journal of Phycology*, *31*(6), 958–961. https://doi.org/10.1111/J.0022-3646.1995.00958.X

Uckelmann, M., & Davidovich, C. (2021). Not just a writer: PRC2 as a chromatin reader. *Biochemical Society Transactions*, *49*(3), 1159–1170. https://doi.org/10.1042/BST20200728

van Buer, J., Cvetkovic, J., & Baier, M. (2016). Cold regulation of plastid ascorbate peroxidases serves as a priming hub controlling ROS signaling in Arabidopsis thaliana. *BMC Plant Biology 2016 16:1*, *16*(1), 1–20. https://doi.org/10.1186/S12870-016-0856-7

Van Zanten, M., Perrella, G., & Bäurle, I. (2022). Epigenetic regulation of thermomorphogenesis and heat stress tolerance. *New Phytologist*. https://doi.org/10.1111/NPH.17970

Veerappan, C. S., Avramova, Z., & Moriyama, E. N. (2008). Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. *BMC Evolutionary Biology*, *8*(1), 190. https://doi.org/10.1186/1471-2148-8-190

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., Dingli, F., Rivarola, M., Ott, S., Liu, X., Sun, Y., Rabinowicz, P. D., McCarthy, J., Allen, A. E., Loew, D., Bowler, C., & Tirichine, L. (2015). An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biology*, *16*(1), 1–18. https://doi.org/10.1186/s13059-015-0671-8

Vijayanathan, M., Trejo-Arellano, M. G., & Mozgová, I. (2022). Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective. *Epigenomes*, *6*(1), 3. https://doi.org/10.3390/EPIGENOMES6010003/S1

Volkov, R. A., Panchuk, I. I., Mullineaux, P. M., & Schöffl, F. (2006). Heat stress-induced H(2)O (2) is required for effective expression of heat shock genes in Arabidopsis. *Plant Molecular Biology*, *61*(4–5), 733–746. https://doi.org/10.1007/S11103-006-0045-4

von Koskull-Döring, P., Scharf, K. D., & Nover, L. (2007). The diversity of plant heat stress transcription factors. *Trends in Plant Science*, *12*(10), 452–457. https://doi.org/10.1016/J.TPLANTS.2007.08.014

Wahid, A., Gelani, S., Ashraf, M., & Foolad, M. R. (2007). Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, *61*(3), 199–223. https://doi.org/10.1016/J.ENVEXPBOT.2007.05.011

Waters, E. R. (1995). The molecular evolution of the small heat-shock proteins in plants. *Genetics*, *141*(2), 785–795. https://doi.org/10.1093/GENETICS/141.2.785

Waters, E. R., & Rioflorido, I. (2007). Evolutionary analysis of the small heat shock proteins in five complete algal genomes. *Journal of Molecular Evolution*, *65*(2), 162–174. https://doi.org/10.1007/S00239-006-0223-7

Wedler, M. (2019). *Charakterisierung einer Deletion des Enhancer of Zeste Gens in Cyanidioschyzon merolae*. FU Berlin.

Wise, R. R., Olson, A. J., Schrader, S. M., & Sharkey, T. D. (2004). Electron transport is the functional limitation of photosynthesis in field-grown Pima cotton plants at high temperature. *Plant, Cell & Environment*, *27*(6), 717–724. https://doi.org/10.1111/J.1365-3040.2004.01171.X

Xie, W., Tang, Q., Yan, F., & Tao, Z. (2021). Transcriptional memory and response to adverse temperatures in plants. *Journal of Zhejiang University-SCIENCE B*, *22*(10), 791–804. https://doi.org/10.1631/JZUS.B2100287

Xu, J., Zhao, X., Mao, F., Basrur, V., Ueberheide, B., Chait, B. T., David Allis, C., Taverna, S. D., Gao, S., Wang, W., & Liu, Y. (2021). A Polycomb repressive complex is required for RNAi-mediated heterochromatin formation and dynamic distribution of nuclear bodies. *Nucleic Acids Research*, *49*(10), 5407–5425. https://doi.org/10.1093/NAR/GKAA1262

Xu, L., & Jiang, H. (2020). Writing and Reading Histone H3 Lysine 9 Methylation in Arabidopsis. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/FPLS.2020.00452

Xue, S., & Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature Reviews. Molecular Cell Biology*, *13*(6), 355. https://doi.org/10.1038/NRM3359

Yamada, M., Hidaka, T., & Fukamachi, H. (1996). Heat tolerance in leaves of tropical fruit crops as measured by chlorophyll fluorescence. *Scientia Horticulturae*, *67*(1–2), 39–48. https://doi.org/10.1016/S0304-4238(96)00931-4

Yamaguchi, N., & Ito, T. (2021). JMJ Histone Demethylases Balance H3K27me3 and H3K4me3 Levels at the HSP21 Locus during Heat Acclimation in Arabidopsis. *Biomolecules 2021, Vol. 11, Page 852*, *11*(6), 852. https://doi.org/10.3390/BIOM11060852

Yamamoto, Y., Ryota, A. E., Ae, A., Yoshioka, M., Mahbuba, A. E., Ae, K., Komayama, K., Daichi, A. E., Ae, T., Yamashita, A., Nijo, A. N., Kayo, A. E., Ae, I., Morita, N., Takayuki, A. E., Ae, S., Yamamoto, Y., Yamamoto, Y., Aminaka, Á. R., … Yamamoto, Á. Y. (2008). Quality control of photosystem II: impact of light and heat stresses. *Photosynthesis Research 2008 98:1*, *98*(1), 589–608. https://doi.org/10.1007/S11120-008-9372-4

Yoon, H. S., Ciniglia, C., Wu, M., Comeron, J. M., Pinto, G., Pollio, A., & Bhattacharya, D. (2006). Establishment of endolithic populations of extremophilic Cyanidiales (Rhodophyta). *BMC Evolutionary Biology*, *6*(1), 1–12. https://doi.org/10.1186/1471-2148-6-78/TABLES/3

Yu, H.-D., Chen, Y. X.-F., Wang, S.-T., & Li, Y.-T. (2012). Downregulation of Chloroplast RPS1 Negatively Modulates Nuclear Heat-Responsive Expression of HsfA2 and Its Target Genes in Arabidopsis. *PLoS Genet*, *8*(5), 1002669. https://doi.org/10.1371/journal.pgen.1002669

Yu, Z., Boehm, C. R., Hibberd, J. M., Abell, C., Haseloff, J., Burgess, S. J., & Reyna-Llorens, I. (2018). Droplet-based microfluidic analysis and screening of single plant cells. *PLOSone*. https://doi.org/10.1371/journal.pone.0196810

Zervudacki, J., Yu, A., Amesefe, D., Wang, J., Drouaud, J., Navarro, L., & Deleris, A. (2018). Transcriptional control and exploitation of an immune-responsive family of plant retrotransposons. *The EMBO Journal*, *37*(14), e98482. https://doi.org/10.15252/EMBJ.201798482

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., & Jacobsen, S. E. (2007). Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis. *PLoS Biology*, *5*(5), e129. https://doi.org/10.1371/journal.pbio.0050129

Zhang, Y., Zhang, A., Li, X., & Lu, C. (2020). The Role of Chloroplast Gene Expression in Plant Responses to Environmental Stress. *International Journal of Molecular Sciences*, *21*(17), 1–16. https://doi.org/10.3390/IJMS21176082

Zhao, J., Lu, Z., Wang, L., & Jin, B. (2021). Plant Responses to Heat Stress: Physiology, Transcription, Noncoding RNAs, and Epigenetics. *International Journal of Molecular Sciences*, *22*(1), 1–14. https://doi.org/10.3390/IJMS22010117

Zhao, X., Rastogi, A., Deton Cabanillas, A. F., Ait Mohamed, O., Cantrel, C., Lombard, B., Murik, O., Genovesio, A., Bowler, C., Bouyer, D., Loew, D., Lin, X., Veluchamy, A., Vieira, F. R. J., & Tirichine, L. (2020). *Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in Phaeodactylum tricornutum*.

Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L., & Lu, C. (2013). Chloroplast Small Heat Shock Protein HSP21 Interacts with Plastid Nucleoid Protein pTAC5 and Is Essential for Chloroplast Development in Arabidopsis under Heat Stress. *The Plant Cell*, *25*(8), 2925. https://doi.org/10.1105/TPC.113.111229

Zienkiewicz, M., Krupnik, T., Drożak, A., Wasilewska, W., Golke, A., & Romanowska, E. (2018). Deletion of psbQ' gene in Cyanidioschyzon merolae reveals the function of extrinsic PsbQ' in PSII. *Plant Molecular Biology*, *96*(96), 135–149. https://doi.org/10.1007/s11103-017-0685-6