

Epigenetic Mechanisms Underlying Seasonal Timing in
Nasonia vitripennis

Thesis submitted for degree in
Doctor of philosophy
at the University of Leicester

By
Akanksha Bafna
Genetics Department
University of Leicester

November 2014

Epigenetic mechanisms underlying seasonal timing in *Nasonia vitripennis*

Akanksha Bafna

Many organisms monitor the annual change in day-length, and use this information for seasonal timing of their developmental, physiological and behavioural response. The molecular mechanisms underlying this photoperiodic timing are largely unknown. The wasp, *Nasonia vitripennis*, is an emerging model organism that exhibits a strong photoperiodic response: short autumnal days experienced by females leads to the induction of developmental arrest (diapause) in their progeny, allowing winter survival of the larvae. How do the females control the developmental trajectory of their offspring is unclear. Here, I took advantage of the available complete genome sequence of the wasp, and tested the role of epigenetics in the photoperiodic response. I used reduced representation bisulfite sequencing (RRBS) to profile DNA methylation in adult females subjected to different photoperiods, and identified substantial differential methylation at the single base level. I have also found that knocking-down *DNA methyltransferase* (*Dnmt1a*, *Dnmt3*), or blocking DNA methylation pharmacologically, largely disrupts the photoperiodic diapause response of the wasps. In another set of experiments, I assessed the prevalence of 5-hydroxy methyl cytosine (5hmC), which is an intermediate in DNA demethylation in mammals. The results show that 5hmC is present in *Nasonia* although in limited amount and suggests that 5hmC-dependent demethylation may be evolutionary conserved in invertebrates. The role of microRNA (miRNA) in the photoperiodic response was also tested. I experimentally validated 32 of the computationally predicted *Nasonia* miRNA and tested their expression levels by using stem-loop real-time PCR. I identified significant differential expression in a sub-set of miRNA, which was induced by the photoperiod. To my knowledge, this is the first example uncovering the role of epigenetics in photoperiodic timing in insects.

Acknowledgements

First of all, I would like to thank my supervisor Dr Eran Tauber for guiding me throughout my research work at the University of Leicester. He has always encouraged me to develop a rational approach and undertake novel strategies that deeply enhanced my scientific outlook. He has been extremely supportive and every discussion with him has restored full enthusiasm.

Next, I'm highly thankful to Dr Mirko Pegoraro for his tremendous help and expertise in experimental design. He has always sparked the scientific curiosity within me and motivated me to face the challenging situations in a logical and organised way. During my research, I've greatly benefitted from the continuous lab discussions with him.

I'm also thankful to all the undergraduate, Masters and Erasmus exchange students who have worked with me during this project. In particular, I would like to thank James Newton and Megan Sinott for their valuable contributions in this research work. Thanks to my colleagues and friends in lab 124 for all the good times. They have immensely helped me to adjust not only in the lab but also in Leicester. Many thanks to Professor Richard Meehan and his team at the MRC Human Genetics Unit in Edinburgh for hosting my visit in his lab, and supporting my 5hmC methylation work. Ms Rita Neumann at the Department of Genetics provided extremely valuable advice about bisulphite sequencing and I am extremely grateful to her for that.

I'm truly grateful to my entire family for all their love, support and encouragement. Each one of them has contributed in special ways for completion of my studies at Leicester. I'm deeply grateful to my family for cherishing my desire to pursue PhD and making it possible despite of exigent situations. A big thank-you to Dr K. Jain & my family at Leicester for their affection, support and guidance throughout my research work. I'm highly thankful to my aunt Mrs Sadhna Jain for being a sister, friend and mother who understood and nurtured me completely.

I feel extremely lucky and thankful to my husband, Nihar Bapna, for always being on my side and encouraging me to pursue my studies with full enthusiasm and vigour. He has been the ultimate driving force at times of stress. I thank him for bearing my mood swings and always staying positive. He has gone an extra mile to understand this research work, identify mistakes and organise the thesis. Thank you.

CONTENTS

CHAPTER 1: Introduction	1
1.1 Photoperiodism.....	2
<i>Photoperiodic clock</i>	2
<i>Models of photoperiodism</i>	4
<i>Photoperiodic response in insects</i>	5
<i>Photoperiodic diapause in insects</i>	6
1.2 <i>Nasonia</i> as a model organism.....	10
<i>Photoperiodic response in N. vitripennis</i>	12
1.3 Rationale, Aims and Objectives	13
CHAPTER 2: Material and Methods.....	15
2.1 <i>Nasonia</i> handling.....	15
2.2 Diapause experiment.....	15
2.3 RNA Extraction	16
<i>DNase treatment of the extracted RNA</i>	17
2.4 cDNA synthesis	17
2.5 Gene amplification using cDNA as template	18
2.6 Real-time PCR.....	18
2.7 RNA interference (<i>RNAi</i>)	19
<i>Preparation of dsRNA</i>	20
<i>Preparation of wasps for injection</i>	23
2.8 Bisulfite Conversion	25
2.9 Data analysis: R- statistical software.....	27
<i>qpcR – Real-time PCR analysis (R environment)</i>	27
CHAPTER 3 : Genome wide analysis of DNA methylation in <i>Nasonia</i> induced by photoperiod.....	28
3.1 Introduction.....	28
<i>DNA methylation</i>	28

3.2 Material and Methods	34
<i>RRBS library preparation and sequence analysis</i>	34
<i>RRBS data analysis</i>	38
<i>RRBS data validation</i>	40
3.3 Results.....	45
<i>Photoperiodic response of AsymC strain</i>	45
<i>The Nasonia methylome</i>	45
<i>Identification of differentially methylated genes</i>	28
<i>RRBS data validation</i>	50
<i>Functional enrichment of identified differentially methylated genes</i>	52
3.4 Discussion	53
CHAPTER 4:Functional significance of DNA methylation in photoperiodism in <i>Nasonia</i>	55
4.1 Introduction.....	55
4.2 Methods	58
<i>Expression analysis of DNMTs</i>	58
<i>Knock-down of DNMTs using RNAi</i>	58
<i>Monitoring changes in CpG methylation by real-time PCR</i>	59
<i>Blocking DNA methylation by 5-aza-2-deoxycytidine</i>	59
4.3 Results.....	62
<i>Expression level of DNMTs</i>	62
<i>Effect of DNMTs knock-down on photoperiodic response</i>	63
<i>Effect of DNMTs down-regulation on genomic CpG methylation</i>	65
<i>Effect of pharmacological blocking of methylation by 5-aza-2-deoxycytidine on photoperiodic response</i>	67
4.4 Discussion	69
CHAPTER 5:Testing for the presence of 5-hydroxymethyl cytosine (5-hmC) in <i>Nasonia</i>	72
5.1 Introduction.....	72
<i>The discovery of 5hmC</i>	73

5.2 Material and methods.....	76
<i>Analysis of 5hmC mark by chemical tagging</i>	76
<i>Downstream analysis of the enriched fraction</i>	77
<i>TET knock-down</i>	79
5.3 Results:	80
<i>5hmC enrichment in Nasonia vitripennis</i>	80
<i>Effect of TET-2 knockdown in Nasonia vitripennis</i>	81
5.4 Discussion	83
CHAPTER 6: Role of <i>microRNA</i> in photoperiodism	85
6.1 Introduction.....	85
<i>The discovery of microRNA</i>	85
<i>miRNA in animals</i>	86
<i>Mode of action of miRNA</i>	87
<i>miRNA in Nasonia vitripennis</i>	88
<i>miRNA as an epigenetic mechanism</i>	88
<i>miRNA involved in phenotypic plasticity in insects</i>	90
<i>Maternal inheritance of miRNA in Drosophila and Zebrafish</i>	91
<i>Role of Dicer in photoperiodism</i>	92
6.2 Material and Methods	94
Profiling of miRNA between different photoperiods	94
<i>Target prediction of differential expressed miRNA between photoperiods</i>	100
<i>Expression level of Dicer-1 in different photoperiods</i>	102
<i>Knock-down of Dicer-1 in Nasonia</i>	102
6.3 Results.....	104
<i>Target prediction of differentially expressed miRNA</i>	106
<i>Knock-down of Dicer-1</i>	108
6.4 Discussion	111
<i>Functional characterization of the predicted targets</i>	113
<i>Knock-down of Dicer induced lethality</i>	113

CHAPTER 7 : General Discussion	115
CHAPTER 8	123
APPENDIX	123
8.1 <i>N.vitripennis</i> methylome	123
8.2 miRNA profiling by Real-time PCR	130
8.3 <i>pRNAi</i> in <i>Nasonia</i>	153
8.4 Level of diapause after 9-10 days in <i>Dnmt1a</i> and <i>Dnmt3</i> knockdown progeny	154
8.5 Standard curve quantification graphs	154
CHAPTER 9: Bibliography.....	123

LIST OF FIGURES

Figure 1.1	Examples of photoperiodic response.....	1
Figure 1.2	Photoperiodic curves in different animals.....	2
Figure 1.3	The key components of photoperiodic clock.....	3
Figure 1.4	Models of Photoperiodism.....	5
Figure 1.5	The photoperiodic response of long day insects.....	6
Figure 1.6	Ovarian diapause in <i>Drosophila</i>	8
Figure 1.7	Simplified version of molecular circadian clock in <i>D. Melanogaster</i>	9
Figure 1.8	Distribution of the four <i>Nasonia</i> species.....	10
Figure 1.9	Life cycle of <i>Nasonia vitripennis</i>	11
Figure 1.10	Photoperiodic response in <i>Nasonia vitripennis</i>	12
Figure 1.11	DNA methyltransferase genes across taxa.....	14
Figure 2.1	Scheme of diapause experiment.....	15
Figure 2.2	RNA interference system.....	20
Figure 2.3	<i>pRNAi</i> in <i>Nasonia</i>	21
Figure 2.4	Schematic representation of the primer design for <i>dsRNA</i> template synthesis....	23
Figure 2.5	The four stages of <i>Nasonia vitripennis</i> development.....	24
Figure 2.6	Effect of Sodium Bisulfite on cytosine molecules.....	25
Figure 3.1	Methylation of a cytosine residue.....	29
Figure 3.2	Action of <i>DNMTs</i> on the replicating DNA strands.....	30
Figure 3.3	Distribution of normalized CpG dinucleotide content.....	32
Figure 3.4	<i>MsPI</i> restriction digestion analysis.....	35
Figure 3.5	DNA ligation efficiency analysis.....	37
Figure 3.6	Final RRBS library amplification.....	38
Figure 3.7	An example of nucleotide sequence of the bisulfite converted sequence.....	41
Figure 3.8	Example of genomic sequence that was used as an input for primer design using MethPrimer.....	41
Figure 3.9	Analysis of qPCR data by Monte-Carlo simulation.....	43

Figure 3.10 Photoperiodic diapause response in <i>Nasonia</i> (AsymC).....	45
Figure 3.11 The <i>Nasonia</i> methylome.....	47
Figure 3.12 Differential DNA methylation associated with photoperiod in <i>Nasonia</i>	49
Figure 3.13 Validation of differential methylation by qPCR.....	51
Figure 3.14 GO analysis of differentially methylated genes in <i>Nasonia</i>	52
Figure 4.1 Multiple Sequence Alignment of <i>DNMT1</i> paralogs.....	56
Figure 4.2 Mechanism of action of cytosine analogue inhibitors.....	57
Figure 4.3 Scheme of drug experiment.....	60
Figure 4.4 Relative expression of <i>DNMTs</i> across photoperiods.....	62
Figure 4.5 Level of Diapause after <i>DNMTs</i> knock-down.....	64
Figure 4.6 Validation of knock-down of <i>DNMTs</i>	65
Figure 4.7 Reduced methylation after knock-down.....	66
Figure 4.8 Blocking DNA methylation by 5-aza-dC.....	67
Figure 4.9 Reduction in the level of CpG methylation after drug treatment.....	68
Figure 4.10 Combined action of <i>Dnmt1a</i> and <i>Dnmt3</i> in maintenance and establishment of methyl marks in the genome.....	71
Figure 5.1 DNA methylation and demethylation in mammals.....	72
Figure 5.2 Phylogenetic Tree based on <i>TET</i> orthologs.....	75
Figure 5.3 Chemical labeling of 5hmC for subsequent detection.....	76
Figure 5.4 5hmC enrichment of target genes.....	80
Figure 5.5 Level of Diapause after <i>TET-2</i> knock-down.....	81
Figure 5.6 Interaction plot diapause level after <i>TET-2</i> knockdown.....	82
Figure 6.1 miRNA processing pathway.....	86
Figure 6.2 Feedback mechanism of miRNA regulation.....	89
Figure 6.3 Comparison of the RNaseIII domain.....	92
Figure 6.4 Multiple sequence alignment of 53 putative miRNA.....	95
Figure 6.5 Quantitative and Qualitative analysis of the miRNA extractions using BioAnalyzer.....	96
Figure 6.6 Profiling of miRNA expression between photoperiods.....	96
Figure 6.7 <i>miRNA-125</i> (<i>Nasonia vitripennis</i>) sequence.....	104

Figure 6.8 Differentially expressed miRNA induced by photoperiods.....	105
Figure 6.9 GO analysis of the predicted target genes in <i>Nasonia</i>	108
Figure 6.10 Phylogenetic tree showing the prevalence of Dicer protein family	109
Figure 6.11 Relative expression level of <i>Dcr-1</i> after the knock-down.....	109
Figure 6.12 Number of progeny after <i>Dcr-1</i> knockdown.....	110
Figure 7.1 Graphical representation of the predicted Thymine DNA glycosylase in <i>N.vitripennis</i>	118
Figure 7.2 Model of epigenetic mechanisms underlying seasonal timing in <i>N.vitripennis</i> .	119
Figure 7.3. Plethora of Epigenetic mechanisms.....	120
Figure 8.1 Melting curve profiles generated by miRNA qPCR.....	152
Figure 8.2 Level of diapause after 9-10 days in <i>Dnmt1a</i> knockdown.....	154
Figure 8.3 Level of diapause after 9-10 days in <i>Dnmt3</i> knockdown.....	154
Figure 8.4 Standard graphs for target and reference genes used in quantification analysis.....	155

LIST OF TABLES

Table 2.1 List of mRNA accessions (NCBI) that were used for <i>RNAi</i> in <i>Nasonia vitripennis</i>	22
Table 3.1 Adapter sequences used for library preparation (RRBS).....	36
Table 3.2 Illumina standard forward and reverse primer used for amplification.....	38
Table 3.3 Primers used for RRBS data validation.....	44
Table 3.4 Loci selected for validation.....	50
Table 4.1 List of primers used for analysing CpG methylation after drug treatment.....	61
Table 5.1 List of specific primers used for real-time PCR for 5hmC detection.....	78
Table 6.1 List of clustered miRNA.....	106
Table 6.2 Functional annotation of predicted targets.....	107
Table 7.1 Fate of CpG sites under differential photoperiods.....	116

List of Abbreviations (acronyms)

A	Adenine
Bp	Base pair
C	Cytosine
CDL	Critical day length
CNL	Critical night length
D	Diapause
DB	Data base
DD	Constant darkness
DEPC	Diethylpyrocarbonate
DMR	Differentially methylated region
DNA	Deoxyribonucleic Acid
DNMTs	DNA methyltransferases
dNTP	Deoxyribonucleotide triphosphates
Ds	Double stranded
dsRNA	Double stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
GO	Gene ontology
hmC	Hydroxy methyl cytosine
HPLC	High performance liquid chromatography
Hr	Hours
KEGG	Kyoto encyclopedia of genes and genomes
LD	Long day
M	Molar
mC	Methyl cytosine
Min	Minutes
miRNA	Micro RNA
ml	Millilitres
mM	Millimolar
NCBI	National centre for biotechnology information
ND	Non diapauses
Ng	Nanograms
NS	Not significant
Nt	Nucleotides
NTC	No template control

Nv	Nasonia vitripennis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
piRNA	Piwi interacting RNA
RNA	Ribonucleic Acid
RNAi	RNA interference
	Reduced representation bisulfite
RRBS	sequencing
RT	Reverse transcriptase
SD	Short day
Sec	Seconds
siRNA	Small interfering RNA
T	Thymine
TBE	Tris Borate EDTA
TE	Tris-EDTA
TET	Ten-eleven-translocation elements
M	Units
M	Uracil
Ug	Micrograms
Ul	Microlitres
uM	Micromolar
UTR	Untranslated region

CHAPTER 1

INTRODUCTION

Many plant and animal species, particularly those that live in temperate regions, experience extreme seasonal changes throughout the year. The organisms thriving in such areas utilize the annual changes in day-length (photoperiod) to anticipate the coming season (Bradshaw & Holzapfel, 2001). Light, as opposed to temperature, may be considered as a reliable environmental cue for timely prediction of the coming season: while the temperature in a specific season may vary from year to year, the photoperiod remains uniform and precise. Day-length measurement allow organism to alter their behavioural, physiological, and developmental programs and overcome unfavourable environmental conditions (Bradshaw & Holzapfel, 2001).

In nature, various organisms exhibit distinct photoperiodic responses which are crucial for their fitness and survival. The term ‘photoperiodism’ was first coined to describe a plant’s ability to flower in response to changes in photoperiod (Garner *et al.*, 1920). In animals, vernal migration, change in skin colour or plumage, and alterations in reproductive behaviour are some of the evident photoperiodic responses (figure 1.1). In some animals long-day promotes development, while in others it leads to manifestation of obligatory/facultative developmental arrest. In the flesh fly for example, summer long days promote development, while in the silk moth, developmental arrest at pupa stage is induced (figure 1.2) (Storey & Storey, 2000).



Figure 1.1: **Examples of Photoperiodic response.**A. Induction of flowering in *Arabidopsis* occurs under long-day conditions B. Arctic terns migrate towards south during autumn.(Photo courtesy: Frans Lemmens)

1.1 Photoperiodism

Photoperiodic timing is thought to consist of two components (Saunders *et al.*, 2002): a photoperiodic timer based on the day-or-night length (Bradshaw & Holzapfel, 2008) and a photoperiodic counter that assesses the number of long-or-short day cycles experienced by the animal (Emerson *et al.*, 2008). The critical day length (CDL) is determined by the photoperiod of a certain length that triggers the particular seasonal response. The CDL is species-specific, and may also vary within species populations, residing in different latitudes. For example the butterfly *Acronycta rumicis* has a CDL of 15 hours at 45°N and 18 hours at 50°N (Saunders, 2002). The CDL usually increases with latitude (and altitude) resulting in earlier induction of the winter response (Bradshaw & Holzapfel, 2001).

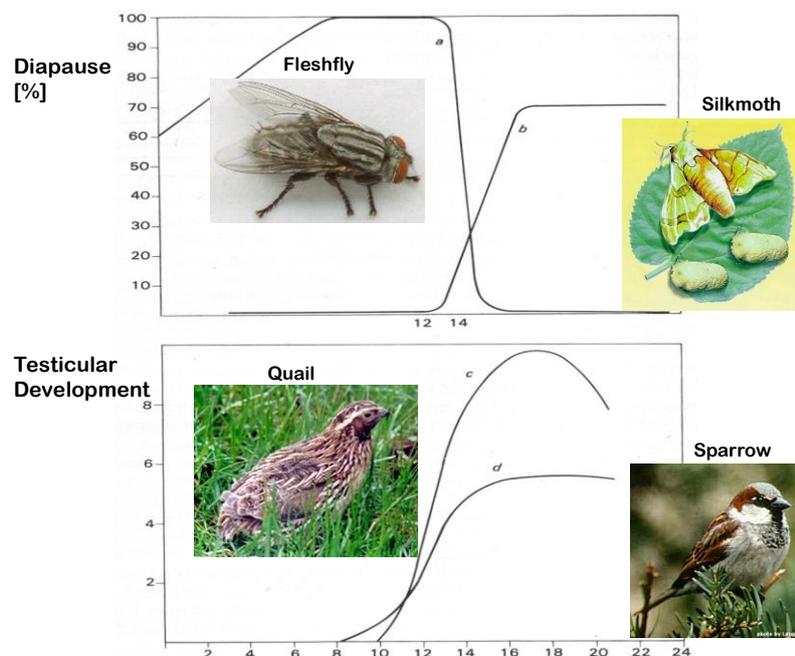


Figure 1.2: **Photoperiodic curves in different animals.** In some organisms (silkmoths, birds) the spring extension of the days acts as the critical cue resulting in the exponential increase in the seasonal response with respect to increasing day-length, reaching the maximum level at 16-17 hours of day-length. On the other hand, in flesh flies exponential increase is seen with reduced day-length, reaching the maximum at 10-12 hours and decreases with >14 hours. (Saunders, 1977)

Photoperiodic clock

The photoperiodic mechanism that enables organisms to anticipate the coming season and in turn adjust the behavioural, physiological and developmental mechanisms is vital for survival and fitness. One may expect that the clock would consist of well defined and

integrated circuits for input, interpretation, and generation of the output based on the day-length (figure 1.3).

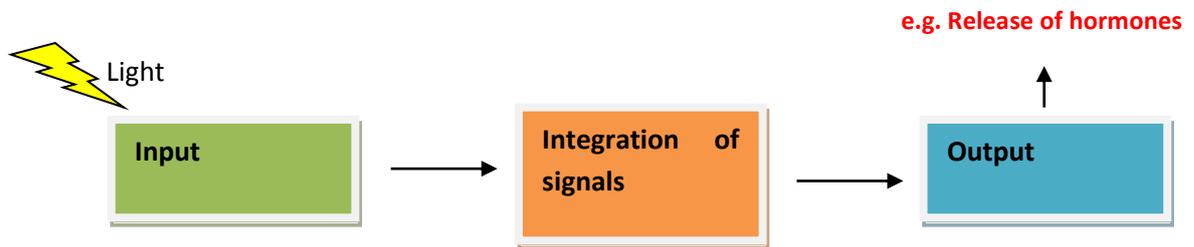


Figure 1.3: **The key components of photoperiodic clock.** The light signal is perceived through multiple channels (species dependent) and leads to release of response based on the seasonal timing in organisms (see text for details).

In mammals, the light is perceived by the retina. This results in electrical stimulation of the pineal gland via the suprachiasmatic nucleus, the paraventricular nucleus and superior cervical ganglion (Bradshaw & Holzapfel, 2010). The electrical stimulation blocks the release of melatonin from the pineal gland so that long days lead to low production of the hormone, while short days induce high levels (Goldman, 2001). During the seasonal change, the level of melatonin secretion is modulated by the day- (or night) length, and drives the associated endocrine responses (Hazlerigg, 2012). The production of melatonin results in the release of gonadotropins (Wagner *et al.*, 2008), which is responsible for sperm maturation in the testes and ovarian maturation in the ovaries. Therefore, the release of gonadotropins acts as an output that leads to synchronization of seasonal biological activity (Saenz de Miera *et al.*, 2014). Similarly in birds, light is perceived through the orbit and stimulates a cascade of events leading to the production of gonadotropins (Yoshimura, 2003; Barrett *et al.*, 2007; Hanon *et al.*, 2008).

Insects too, exhibit strong photoperiodic responses that help them to overcome the harsh climatic conditions (Saunders *et al.*, 2002; Beck, 1980; Tauber *et al.*, 1986; Danks, 2005) and have been used as a model system for studying the molecular mechanism underlying photoperiodism. The photoperiodic response in insects is described below.

Models of Photoperiodism

Various models have been conceptualized to understand the mechanism of photoperiodism in both animals and plants. Some of the models assume the involvement of the circadian clock (daily rhythm) in the photoperiodic timing resulting in the specific photoperiodic response, while others do not.

Hourglass model: The model (Yanovsky, 2002) proposes a gradual accumulation of a molecular component (e.g. hormone) with increasing (or decreasing) photoperiod (figure 1.4 a). A specific quantity of this component is essential to trigger a particular physiological response such as flowering in plants. The accumulation of the chemical substance is favoured by a certain photoperiodic condition, which then results in the specific photoperiodic response. For instance, in certain plants light may promote the accumulation of the chemical substance. In long days of summer, the substance gets accumulated beyond the threshold level and thus leads to flowering. However, during the night the substance gets degraded. In autumn, or winter, as the amount of light is not enough to meet the threshold level it fails to generate the specific physiological response. This model relies on the resetting or 'turning over ' of light cycle every day and hence the name of the model.

External coincidence model: This model, suggested by Erwin Bünning in 1936, proposes the involvement of circadian rhythm of photo-sensitivity, where most of the night phase is sensitive to light, while the day phase is insensitive. In the long days of spring or summer, light extends and illuminates the photosensitive phase, triggering the physiological response. Thus, light has a dual effect in this model: entraining the photosensitive rhythm, as well as acting as the stimulus. The coincidence of the external stimulus (light) with the light-sensitive rhythm results in generation of a specific physiological response (figure 1.4 b)

Internal coincidence model: This model was proposed by Colin Pittendrigh and Dorothea Minis (1964). This model postulates that the only role of light is to entrain the circadian rhythm. The model assumes the presence of two circadian oscillators running at slightly different frequency and consequently their phase relationship is continuously shifting. The change in photoperiods is encoded by unique internal phase relations between the two (or more) oscillators, inducing the physiological response (figure 1.4C)

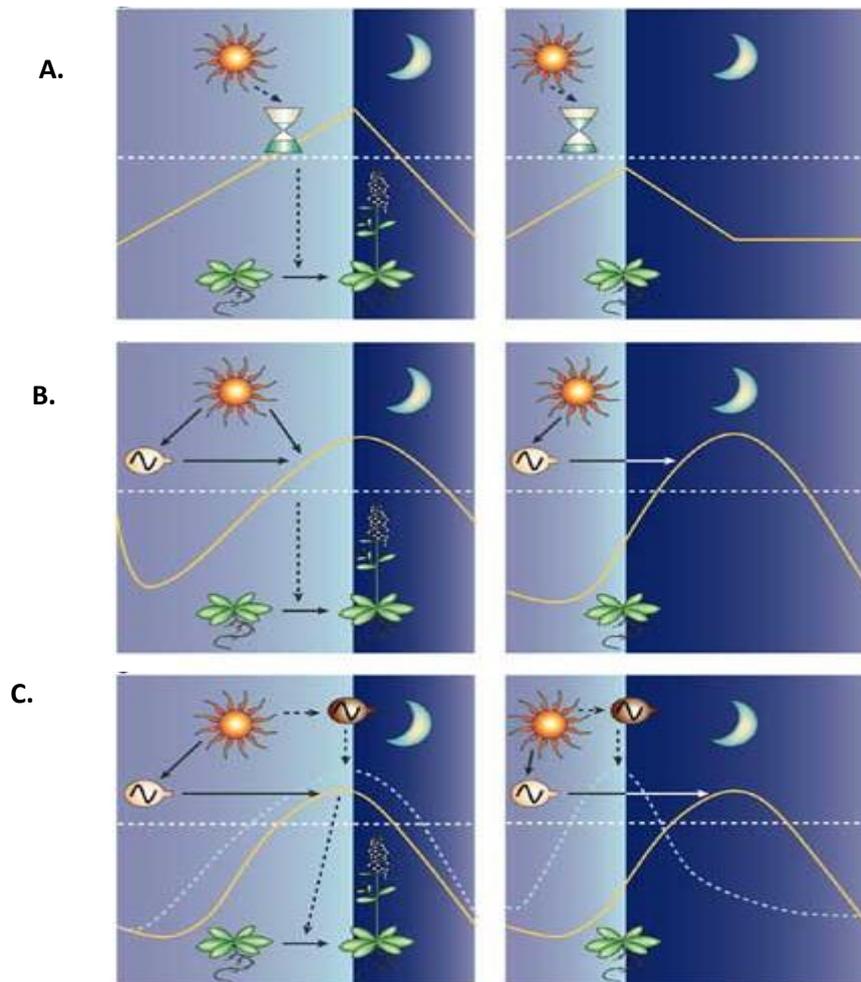


Figure 1.4: **Models of Photoperiodism.** **A.**With the hourglass model, photoperiod measurement is mediated by a chemical substance (yellow line), accumulation of which is light dependent. The photoperiodic response is triggered when this product accumulates above a certain threshold level (white dotted line). **B.**The external coincidence model proposes that circadian oscillator controls the levels of some regulatory molecule (yellow line), the activity of which is modulated by light. Photoperiodic responses are triggered when the illuminated part of the day overlaps with a phase of the cycle during which the levels of the regulatory molecule are above a certain threshold (white dotted line). **C.** In the internal coincidence model, the photoperiodic responses are induced by the coincidence of two or more rhythms that maintain a unique phase relationship under a certain photoperiod. This would be the case if some rhythms are timed at a fixed interval from dawn (yellow line) and others from dusk (light-blue dotted line), because they are driven by distinct circadian oscillators (beige and brown, respectively).Figure adapted from Marcelo, 2002.

Photoperiodic response in insects

In insects, particularly in temperate regions, diapause is a prominent photoperiodic response and is defined as a neuro-hormonally mediated developmental arrest, accompanied by a dynamic state of low metabolic activity (Tauber *et al.*, 1986). It is associated with reduced morphogenesis, increased resistance to environmental extremes, and altered or reduced behavioural activity. The photoperiod induction of diapause has

been observed in more than 500 insects species from 17 orders (Nishizuka & Masaki, 1998), and is considered widespread among the species residing at higher latitudes (Saunders *et al.*, 2002). Diapause may occur at different life stages such as embryo, larva, pupa or adult, depending upon the species (Emerson *et al.*, 2009). For example, the silkworm moth *Bombyx mori* overwinters and generates diapause at embryonic stage. On the other hand, the gypsy moth *Lymantria dispar* initiates diapause at fully formed larva stage.

In some species, diapause is facultative and occurs only when induced by environmental conditions, whereas in other species diapause has become an obligatory part of the life cycle. The CDL (defined as the day-length when 50% of the population has entered diapause) is another important species-specific attribute. CDLs are usually between 10 to 14 hours of light (figure 1.5). ‘Long-day’ insects will enter diapause when exposed to short day-autumn or winter like conditions.

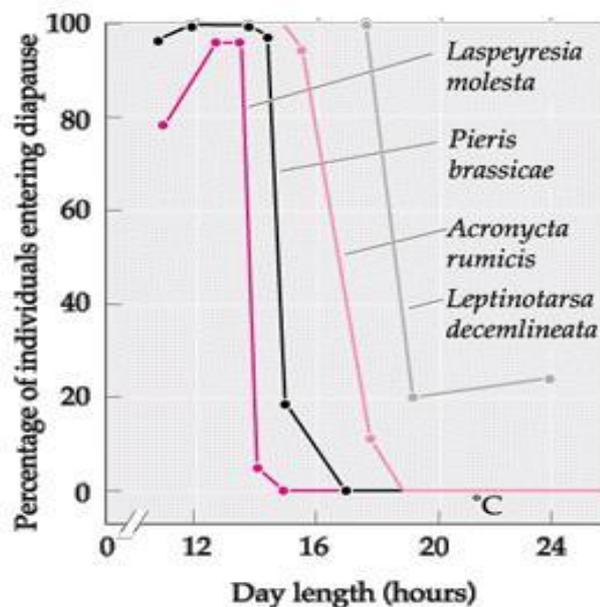


Figure 1.5: **The photoperiodic response of long-day insects**, entering diapause when the day-length falls below a certain level. The four species shown namely, *Laspeyresia molesta*, *Pieris brassicae*, *Acronycta rumicis*, and *Leptinotarsa decemlineata* leave diapause when day-length exceeds 14–17 hours. (Adopted from Danilevskii, 1965)

At higher latitudes, the majority of insects show ‘long-day’ like photoperiodic response in which continuously developing, non-diapause generations occur during the

long days of summer, but an overwintering or diapausing generation supervenes as day-length shortens (Saunders, 1982). ‘Short-day’ insects with a summer diapause (aestivation) and active growth in the autumn and winter also exist, but in more southern latitudes (Masaki, 1961; Koveos & Tzanakakis, 1989). Other environmental cues that can trigger diapause in insects include temperature, diet and age of the egg-laying mother.

Whether the photoperiodic time measurement in insects is inherently managed by the circadian clock or by some form of non-oscillatory hourglass mechanism is still under debate. One of the experimental protocols developed for testing the External coincidence model (i.e. the circadian clock is required for the photoperiodic response) is called Nanda-Hamner (NH) protocol, and involves the exposure of the animal to a relatively short day with variable night lengths (Blaney & Hamner, 1957; Nanda & Hamner, 1958; Bradshaw *et al.*, 2003). Primarily the idea was if there exists a circadian based sensitivity to light, then the long dark phase should elicit a rhythmic long-day response with increasing duration of the night cycle (Bradshaw *et al.*, 2003; Saunders & Sutton, 1969; Pittendrigh *et al.*, 1991), although it was argued that oscillation of the NH response may not necessarily be driven by circadian clock (Danks, 2005; Veerman, 2001; Bradshaw *et al.*, 2006). For example, in the pitcher-plant mosquito *Wyeomyia smithii*, the rhythm expressed via NH experiments was shown to be separated from the circadian clock (Emerson *et al.*, 2008).

In *Drosophila melanogaster*, long night photoperiods at low temperature (12°C) induce ovarian diapause (figure 1.6). In a classic set of studies (Saunders *et al.*, 1989; Saunders & Gilbert, 1990), the diapause response was tested in fly mutant of the clock gene *period*, to analyse whether the circadian clock is an essential part of photoperiodic timing. Normal photoperiodic response was observed both in short period (*per^S*), and in long period (*per^{L2}*) mutant flies, with a CDL identical to the wild type. This suggested that an important core circadian clock component *per* was not involved in the photoperiodic time measurement.

Nevertheless, several studies suggest a close association between the photoperiodic and circadian systems across various taxa (Nelson *et al.*, 2010), indicating highly evolutionary conserved link. In a separate study (Sandrelli *et al.*, 2007), the authors have shown the prevalence of a mutated clock gene *timeless* (*ls-tim*) in the European fly populations that promote diapause by attenuating the photosensitivity of

the circadian clock, and established a molecular link between diapause and circadian photoreception.

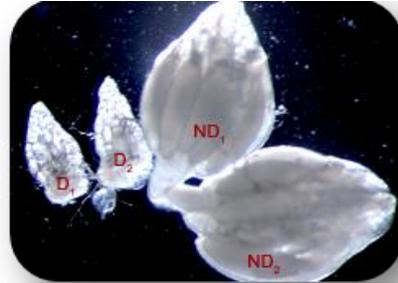


Figure 1.6: **Ovarian diapause in *Drosophila***. Pair of non-diapause (ND) and diapause (D) ovaries as observed in *Drosophila*. (Image taken from Schmidt, P 2005)

The role of circadian clock in photoperiodic time measurement has been explored in other insect orders and is briefly reviewed below.

Orthoptera: In the cricket *Modicogryllus siamensis*, exposure to short days leads to slow down in development of nymphs. In a study (Sakamoto *et al.*, 2009), following the clock gene *per* knock-down, treated nymphs entered diapause, regardless of the photoperiodic condition. Most of the treated nymphs also showed aberrant circadian locomoter activity profiles. This suggested that the normal functioning of the circadian clock controlled by *per* is indispensable for photoperiodic time measurement and corroborated Bünning hypothesis.

Heteroptera: In two Heteropteran species (*Pyrrhocoris apterus* and *Riptortus pedestris*), a sensitive nymph stage leads to ovarian diapause in the adult females exposed to long nights. Under long nights, gene expression studies in these bugs (Hodkova, 2002; Syrova *et al.*, 2003; Dolezel *et al.*, 2005) had revealed ten times higher levels of *per* mRNA, suggesting plausible role of *per* in downstream process such as translation of the photoperiodic signal into hormonal control.

BOX 1.1

Circadian clock in *Drosophila melanogaster*

The circadian clock in *Drosophila melanogaster* is composed of clock genes and their proteins acting in feedback loops (Hall, 2003; Hardin, 2005). Briefly, after the transcription of *period* (*per*) and *timeless* (*tim*), the corresponding proteins PERIOD (PER) and TIMELESS (TIM) accumulate in the cytoplasm and form the PER/TIM heterodimer in darkness. This heterodimer then enters the nucleus and acts as a negative regulator of the CYCLE/CLOCK dimer, which is responsible for the transcription of *per* and *tim*. Entrainment of this oscillation depends on the blue light photoreceptor protein CRYPTOCHROME (CRY), which subsequently promotes TIM degradation when stimulated by light. The possible functions of *per*, *tim*, *clock*, *cry* and *cycle* in the photoperiodic clock of various insects have been investigated (Bradshaw & Holzapfel, 2010; Yamada & Yamamoto, 2011).

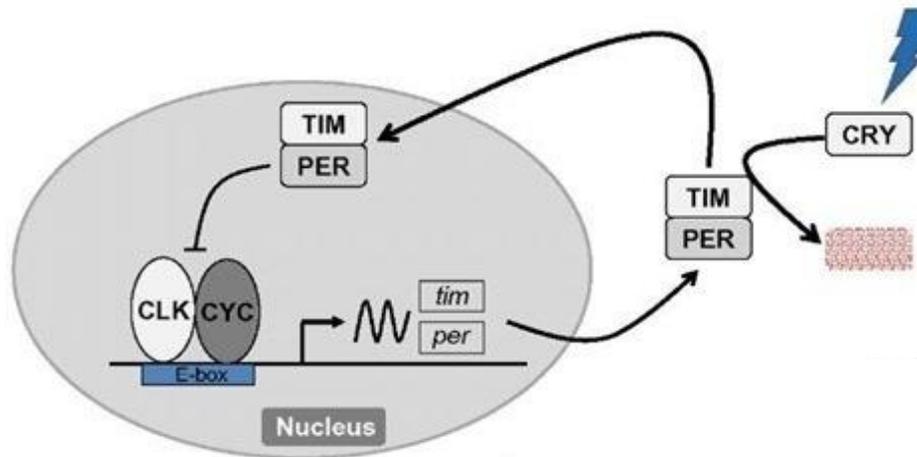


Figure 1.7: Simplified version of the molecular circadian clock in *Drosophila*

1.2 *Nasonia* as a model organism

Nasonia (the ‘jewel’ wasp’) is a parasitoid wasp, belonging to the order Hymenoptera. There are four closely related species, *N. vitripennis*, *N. giraulti*, *N. longicornis*, and *N. oneida* (Darling & Werren, 1990). *Nasonia vitripennis* is a cosmopolitan species that is found throughout the world (figure 1.8). In contrast, the other species of *Nasonia* are found only in North America.

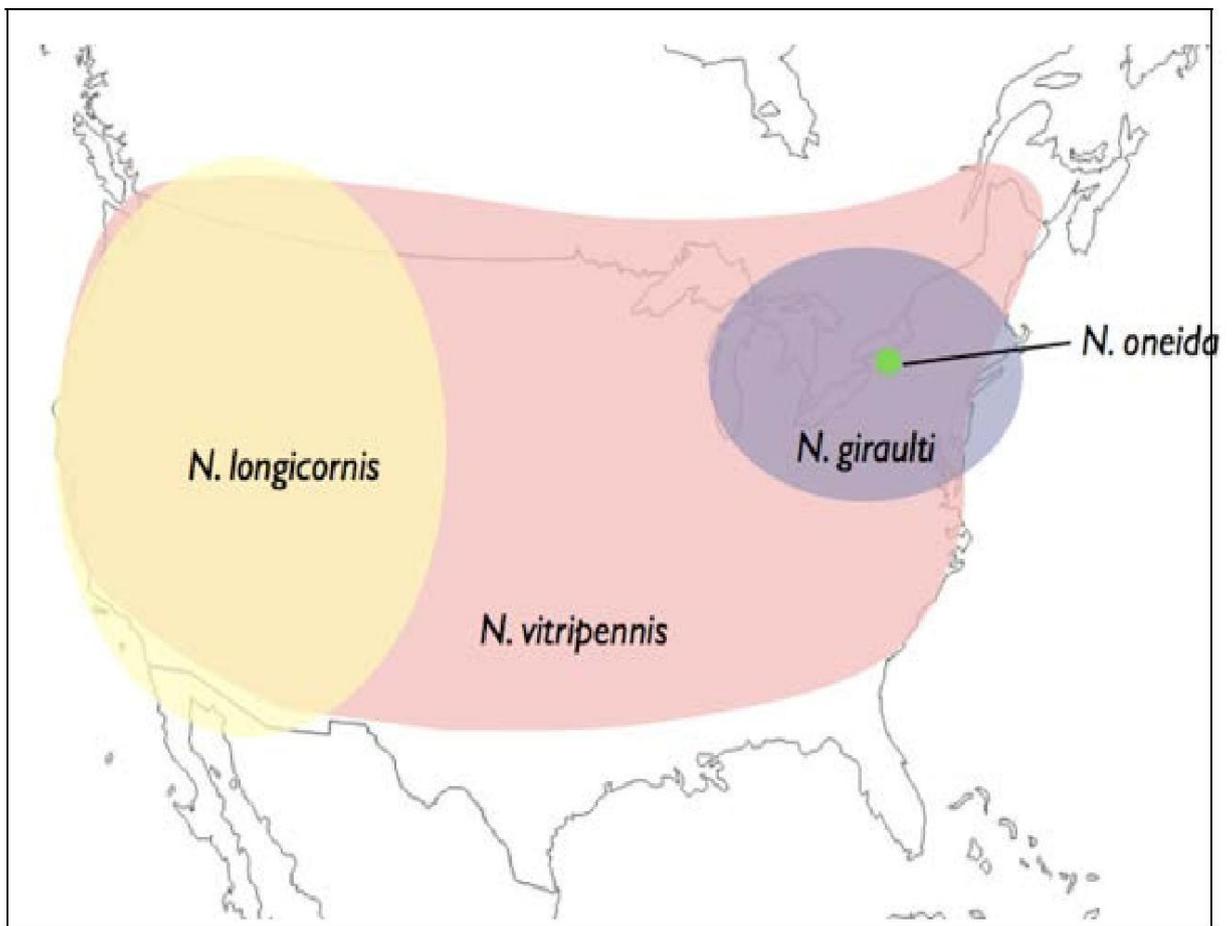


Figure 1.8. Distribution of the four *Nasonia* species. (Adapted from Werren lab resources, 2011)

These small insects, about $\frac{1}{4}$ of the size of *Drosophila*, sting and lay their eggs in the pupae of various flies (e.g. blow flies, flesh flies and house flies). *Nasonia* exhibits haplo-diploid sex system which is a characteristic feature of hymenopteran insects. The females are diploid and develop from fertilized eggs, whereas males are haploid and

develop from unfertilized eggs. The insect completes one life-cycle in approximately two weeks. When the female encounters a suitable host, it injects the venom into the pupa, and lay the eggs (figure 1.9). An individual female can typically lay 20-50 eggs per pupa (blowfly, fleshfly or housefly). Thereafter, the eggs hatch around 36 hours later at 25°C. The developing larvae further complete three instars and pupate within the host. Development of the pupae takes approximately three days. At this stage male and female can be easily distinguished. Female pupae can be distinguished on the basis of body size, fore-wing pad dimensions and presence of ovipositor at the distal end. Finally, the adults come out by puncturing a hole in the host puparium after 14 days. As a model system, *Nasonia* offers various advantages such as short-generation time, ease of handling and haplo-diploid sex determination that facilitates production of inbred lines and genetically identical recombinant individuals to study complex genetic traits. Moreover, with the recent release of the whole genome sequence (Werren *et al.*, 2010) the organism offers a great opportunity to study genetics, ecology, behaviour, development and evolution.

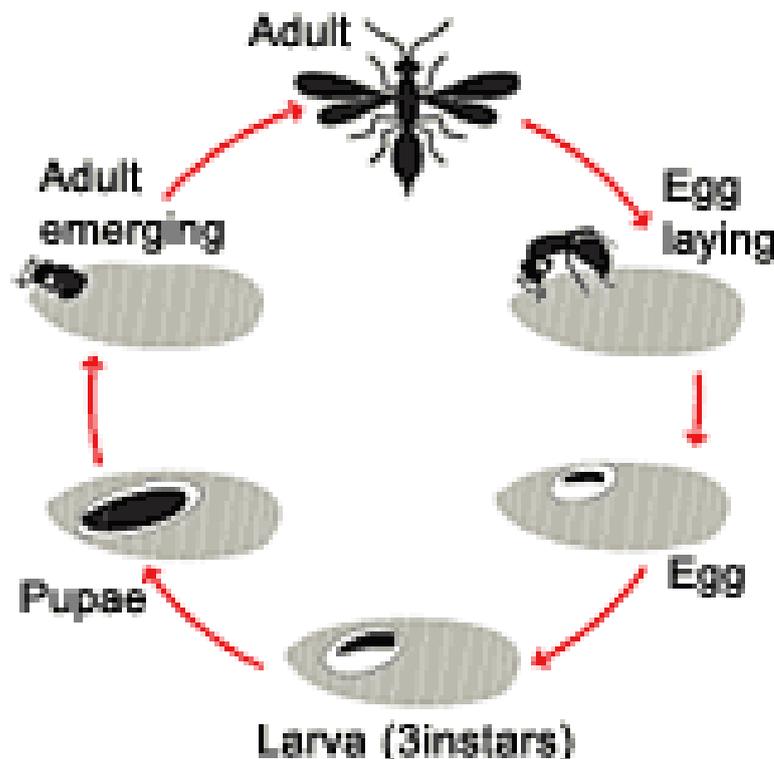


Figure 1.9: Life cycle of *Nasonia vitripennis*.

Photoperiodic response in N. vitripennis

Diapause in *Nasonia* is under the regulation of three key factors: day-length, temperature, and host availability (Saunders, 1965). Day-length is the critical cue to predict the coming season and overcome unfavourable environmental conditions (figure 1.10). *Nasonia* is a typical ‘long-day’ insect with a CDL of 15.5 hours (Saunders, 1965). It exhibits a strong photoperiodic response when exposed to differential photoperiods. The facultative diapause occurs at larval stage of development (figure 1.10). Females that are exposed to short-days of autumn (or winter) produce offspring that undergo diapause, while under the long-days of summer generate non-diapause (developing) offspring. Thus, the larval diapause in *Nasonia* is of maternal origin. The response can be effectively simulated under artificial laboratory conditions. At a constant temperature of 18°C, a female experiencing 18:6 (light:dark) LD (long-day) condition for 10 days of exposure will produce most of its offspring in non-diapause (as described in Chapter 3). On the other hand, those experiencing 6:18 (light:dark) SD (short-day) condition will generate most of its offspring in diapause (Wolschin & Gadau, 2009a).

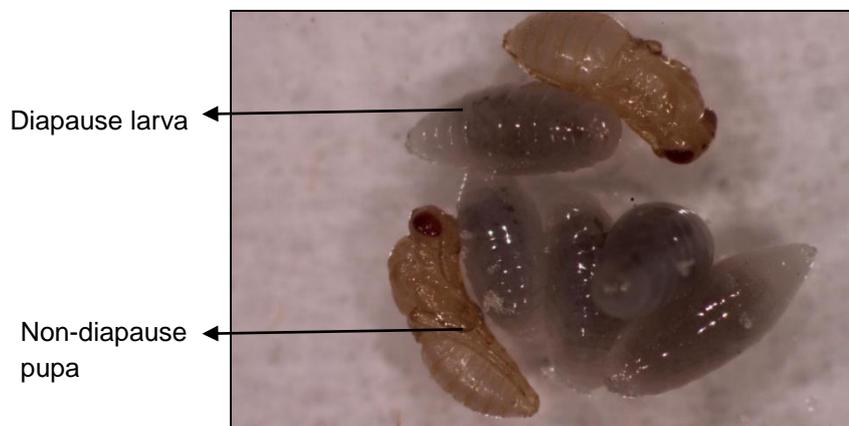


Figure 1.10: **Photoperiodic response in *Nasonia vitripennis***. Cluster of non-diapause pupae (towards the sides) and 3rd instar diapause larvae (center) are shown after isolation from the host puparium.

1.3 Rationale, Aims and Objectives

Drosophila has been a major model system for studying the circadian clock, but the photoperiodic response (ovarian diapause) of the fly is rather shallow and not easy to distinguish (Schiesari *et al.*, 2011). In contrast, the photoperiodic response in *Nasonia* (Fig 1.10) is robust, making it an ideal organism for studying this process. As mentioned above, the recent sequencing of the *Nasonia* genome has paved the way for molecular analysis of the photoperiodic clock.

In both plants and animals, epigenetic mechanisms have been shown to play an important role in photoperiodism as described below. ‘Epigenetics’ was defined for the first time by Conrad Waddington (1942) “as the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being.” Generally speaking, epigenetics is considered as the link between genotype and phenotype -- the phenomenon that changes the gene expression without altering the underlying DNA sequence (Goldberg *et al.*, 2007). The chief epigenetic mechanisms that could modulate the gene expression include DNA methylation, histone modifications and non-coding RNA (Bird, 2007). In particular, DNA methylation has been suggested to be important for seasonal timing in plants (Burn *et al.*, 1993; Dennis *et al.*, 1998; Sheldon *et al.*, 1999). Burn and colleagues (1993) have clearly shown the involvement of DNA methylation in early induction of flowering in plants exposed to cold conditions. In addition, DNA methylation was also noticed to control the expression of genes (such as *dio-3*) involved in seasonal reproductive response, in hamsters (Stevenson & Prendergast, 2013). These studies will be described in more detail in Chapter 4.

One of the key findings of the *Nasonia* genome sequence analysis was the presence of several DNA methyltransferase genes including *DNMT1*, *DNMT2* and *DNMT3* (Werren *et al.*, 2010). In fact, three paralogs of *DNMT1*, *DNMT1a*, *DNMT1b* and *DNMT1c* were identified (figure 1.11). Indeed, a few studies have suggested that DNA methylation in *Nasonia* is functional (Park *et al.*, 2011; Zwier *et al.*, 2012; Wang *et al.*, 2013b). This is in a significant contrast to *Drosophila*, where DNA methylation was thought to be non-existing (Raddatz *et al.*, 2013), although recent observations suggest that it may be present at extremely low levels (Takayama *et al.*, 2014).

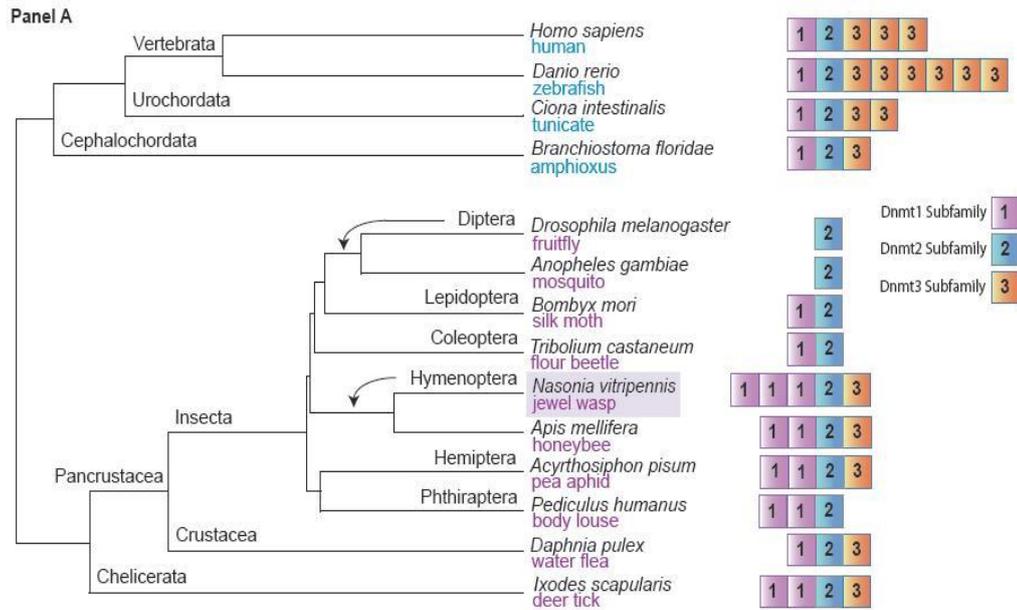


Figure 1.11: **DNA methyltransferase genes across taxa.** *Nasonia* has a mammalian-like set of methyltransferase genes. Worthy of note is *Drosophila* having only *DNMT2*. (Adapted from Carol Trent, 2009).

Thus, the combination of strong photoperiodic response and functional DNA methylation provided a unique opportunity to test the role of various epigenetic mechanisms in insect photoperiodism. In the case of *Nasonia*, where the photoperiodic information is maternally transferred to the progeny, one may speculate that the developmental program of the progeny may be regulated by a specific (or multiple) epigenetic mechanisms.

In Chapter 3, I describe the use of high-throughput bisulphite sequencing for mapping differential DNA methylation induced by the photoperiod, and in the following Chapter 4, I report the results of various functional assays that I have carried to demonstrate the causative role of DNA methylation in the photoperiodic response. For DNA methylation to be an effective regulating system, one would expect to find a concomitant de-methylation system. Since in mammals, DNA demethylation was shown to involve 5-hydroxymethylation and mediated by Ten Eleven Translocation (TET) proteins, I tested whether a similar system is utilized by *Nasonia* and the results are described in Chapter 5. Chapter 6 describes experiments exploring another epigenetic mechanism that involves small, non-coding RNA (microRNA). A General Discussion concludes the thesis.

CHAPTER 2

MATERIAL AND METHODS

2.1 *Nasonia* handling

Nasonia vitripennis strain *AsymC* was reared in 12:12 h light: dark cycle at 25°C. The wasps were fed on commercially available blowfly (*Phaenicia cuprina*) pupae. The wasps have a generation time of approximately two weeks at 25°C. They were kept in glass bottles with fresh host blowfly pupae (preferably less than a month old). The blowfly maggots were purchased (Worms Direct) and kept at 25°C to mature. After four days, when the maggots matured into pupae, they were transferred and stored at 4°C. During parasitization, the adult females laid eggs (20-30 per host) inside the host pupae and left it to mature. The developing embryo underwent subsequent larval and pupal stages inside the host pupae and eclosed after 14 days (as explained in Chapter 1). Once the new generation emerged, the wasps were transferred into clean glass bottles with fresh hosts. Subsequently, new glass bottle was placed in inverted position over the old bottle, and the mouth of the bottles was sealed with the help of adhesive tape. The wasps are negatively geotaxic and consequently moved into the new bottle. Fresh hosts were then added and the bottle was closed with a sterilized bung. The stocks were kept at 25°C and re-cultured after two weeks; soon after emergence of new generation.

2.2 Diapause experiment

Each newly eclosed female wasp was placed in a glass vial with two blowfly pupae referred as Host 1 (here and elsewhere). The vials were then placed for a total of 10 days (figure 2.1) inside an incubator set at 18°C under either short day conditions (SD; 6 hours of light and 18 hours of darkness) or in long day conditions (LD; 18 hours of light and 6 hours of darkness) generated by LED (light emitting diode) illumination.

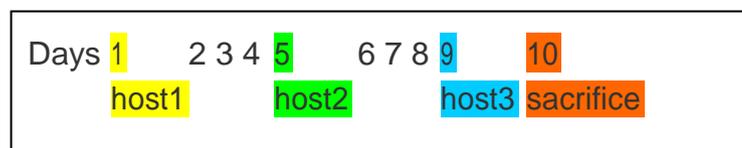


Figure 2.1: **Scheme of diapause experiment** (see text for details).

After 5 days the host pupae were replaced with new blowfly pupae (host2). The old host pupae were collected and placed at 18°C in DD (continuous darkness). Subsequently, after 9 days the host pupae (host2) were again replaced again with 2 new blowfly pupae (host3) and the old host pupae were kept at 18°C in DD. After 10 days the female wasps were sacrificed in liquid nitrogen (2 hours after light on) and stored individually in -80°C. The recovered hosts (host1, host 2 and host3) were dissected after 20 days in order to observe and record the proportion of diapausing progeny.

2.3 RNA Extraction

The working bench area, micropipettes, and other pieces of equipment were cleaned with RNaseZAP (Sigma) prior to the extraction to minimize the degradation of the RNA by *RNases*. For total RNA isolation, a guanidiniumthiocyanate-phenol-chloroform (Jackson & Linsley, 2010; Chomczynski & Sacchi, 1987) based extraction method was adopted. In RNase free 1.5 ml microfuge tube, 10 newly eclosed wasps were collected. The collected wasps were homogenized in 500 µl TRizol (Invitrogen) reagent and incubated for 5 minutes at room temperature to permit the complete dissociation of nucleo-protein complex. After the incubation, 100 µl of chloroform was added. The samples were vortexed for 15 seconds and incubated for 10 minutes at room temperature. The samples were then centrifuged at 12,000g for 20 minutes at 4°C. Following centrifugation the mixture was separated into lower red-phenol-chloroform phase, an interphase and a colourless upper aqueous phase. The aqueous phase was transferred into a new tube with 500 µl of isopropanol and 1 µl of RNase-free glycogen. Thereafter, the samples were incubated at room temperature for 10 minutes and re-centrifuged at 12,000g for 10 minutes at 4°C. The supernatant was carefully removed and the RNA pellet was washed with 1000 µl of 75% ethanol by centrifugation at 7,500g at 4°C for 5 minutes. Finally, the pellet was air-dried for 2 minutes and suspended in 44ul of DEPC (diethylpyrocarbonate) treated water.

DNase treatment of the extracted RNA

The residual genomic DNA was removed from the RNA samples using *rDNaseI* (AMBION®). One μl of *rDNaseI* and 5 μl of 10X *DNaseI* buffer were added and followed by incubation at 37°C for 30 minutes. At the end of the incubation, 5 μl DNase inactivation reagent (supplied with the kit) was added and the samples that were kept at room temperature for another 2 minutes with occasional mixing. Samples were then spun at maximum speed for 1.5 minutes and the supernatant was transferred in to a new tube. Finally, the concentrations of the RNA samples were measured with the spectrophotometric based Nanodrop (Thermo Scientific).

2.4 cDNA synthesis

The first strand of the cDNA molecule was synthesized using the *SuperScriptII Reverse Transcriptase* (Invitrogen). In a nuclease free-microcentrifuge tube, a 12 μl final volume reaction was set up for each of the RNA samples. The following components were added sequentially: total RNA 1 μg , 1 μl of 10mM dNTP mix, 50-250ng random primers, or gene specific primer (2 pmoles), *Aequorin* mRNA (250pg were spiked, to be used as an exogenous reference in real-time PCR analysis), and finally DEPC water was added up to 12 μl . The mixture was heated at 65°C for 5 minutes followed by quick chill on ice. Then in each reaction tube 4 μl of 5X first strand buffer, 2 μl of 0.1 M (molar) DTT and 1 μl of *RNaseOUT*TM (40 units/ μl) (Life technologies) was added. The contents of the tube were mixed gently and incubated at 42°C for 2 minutes for synthesis using gene specific primer, or 25°C for 2 minutes when using random primers. Further, 1 μl of *SuperScriptII Reverse Transcriptase* was added and mixed gently by pipetting up and down. In addition, a minus reverse transcriptase (-RT) control was set up, where the enzyme was replaced by DEPC treated water. The samples were incubated at 42°C for 50 minutes. The reaction was inactivated by heating at 70°C for 15 minutes. To completely remove the template RNA from the samples, 1 μl of *RNaseH* (2 units) (New England Biolabs) was added and incubated at 37°C for 20 minutes. Finally, the synthesized cDNA was diluted (10 times) and used for amplifying target gene fragment.

2.5 Gene amplification using cDNA as template

Target gene (for example *DNMTs*) and green fluorescent protein (*GFP*) was amplified from cDNA extracted using the above procedure and *GFP* construct respectively by polymerase chain reaction (PCR). The 10 times diluted cDNA (5 μ l) was used in a 50 μ l total reaction volume with 1.5 μ l *Taq polymerase* (5U/ μ l) (KAPA Biosystems) and 0.3 μ M gene-specific forward and reverse primers, each. The thermal cycling conditions for PCR began with 92°C for 2 minutes for initial denaturation, followed by 35 cycles of 92°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds, plus a final extension step at 72°C for 10 minutes. The amplified product was analyzed by agarose gel electrophoresis (1%) and extracted from the gel using gel extraction kit (E.Z.N.A) according to the manufacturer instructions. After the amplification of the desired template, Sanger's DNA sequencing was done on the purified PCR product to validate the target genomic sequence. For this purpose the sequenced gene fragment was aligned (Chromas Lite 2.1.1) and matched using BLAST to the source genomic (mRNA) sequence.

2.6 Real-time PCR

Real-time PCR (qPCR) was employed to quantify the expression level of the specific target gene. SYBR green based qPCR was adopted and the transcript level of the target gene was assessed by standard curve method (Gilsbach *et al.*, 2006) using either external reference gene *Aequorin* which is a jellyfish photoreceptor or the internal reference gene *RPL-32* (ribosomal protein). Briefly, RNA was extracted using TRIzol reagent as explained above from the adult females. The extracted RNA was treated with *rDNase I* to minimize the genomic DNA contamination. Total RNA was used to synthesize cDNA by *SuperScriptII Reverse transcriptase* kit (Invitrogen).

Each cDNA sample was analyzed by using 2X SYBR green mix (Agilent Technologies) in duplicates by target and reference gene primer pair along with the reaction controls (minus RT and no template controls (NTC)). For each reaction, 5 μ l diluted cDNA (10 times), 0.4 μ M specific forward and reverse primer pair (target/reference) and 12.5 μ l 2X SYBR green mix were added to the 25 μ l total reaction volume. Two-step PCR with 95°C for 15 seconds and 60°C for 45 seconds was carried for 45 cycles. After the amplification the melting curve analysis was performed for each reaction between 50.0 °C to 95.1°C for every 0.2 °C. The distinct melt profiles between

the samples amplified with different reaction primers ensured specific amplification of the gene of interest.

2.7 RNA interference (RNAi)

Background

In order to knock-down specific genes and carry out loss-of-function experiments, parental RNA interference (*pRNAi*) approach was adopted (Lynch & Desplan, 2006). *RNAi* was first studied in *Caenorhabditis elegans* in 1988 by Andrew Fire and Craig Mello as a response to double stranded RNA (dsRNA). When long double stranded RNA (dsRNA) was injected into the worm gonads by a standard way of introducing transgenes into worms, it blocked the expression of endogenous gene in the sequence specific manner. The dsRNA is recognised and cleaved by *Dicer* (RNase III family dsRNA specific endonucleases). Cleavage by *Dicer* creates short dsRNA (siRNA) that are characterized by 2 nucleotide long 3' end overhangs (figure 2.2). In worms, flies and mammals siRNAs (21-23 nucleotides long) has been shown to form a ribonucleoprotein complex called RISC (RNA Induced Silencing Complex). RISC recruits one of the siRNA strand to bind with the target mRNA in a sequence-specific manner and further degrades the mRNA and leads to translational repression.

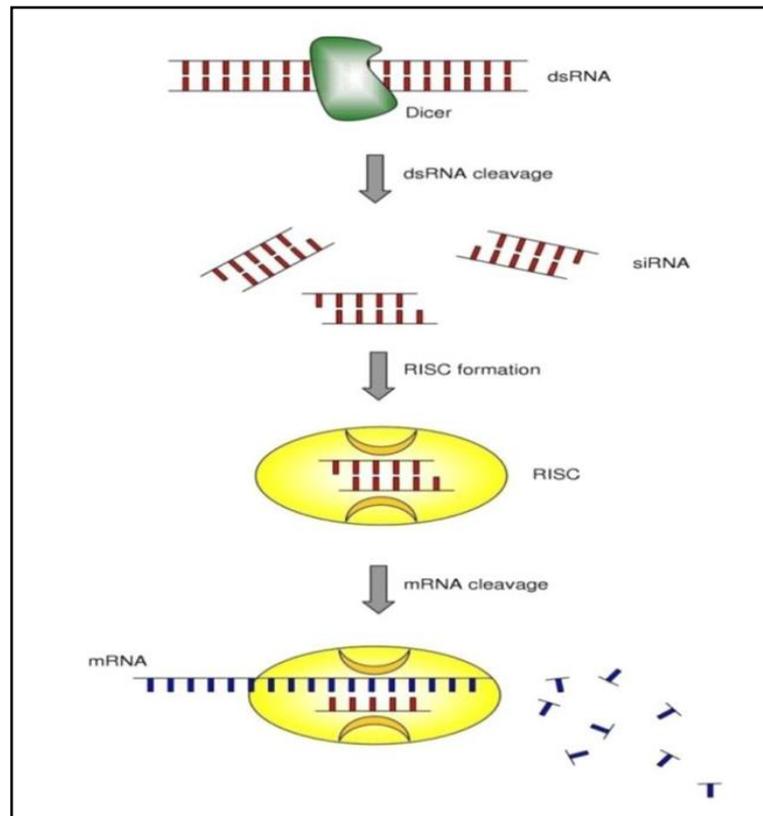


Figure 2.2. **The RNA interference system** (Mocellin&Provenzano, 2004).

Preparation of dsRNA

The dsRNA for the knock-down of the specific target genes in *Nasonia* was synthesized using the steps explained below in a sequential fashion (fig. 2.3).

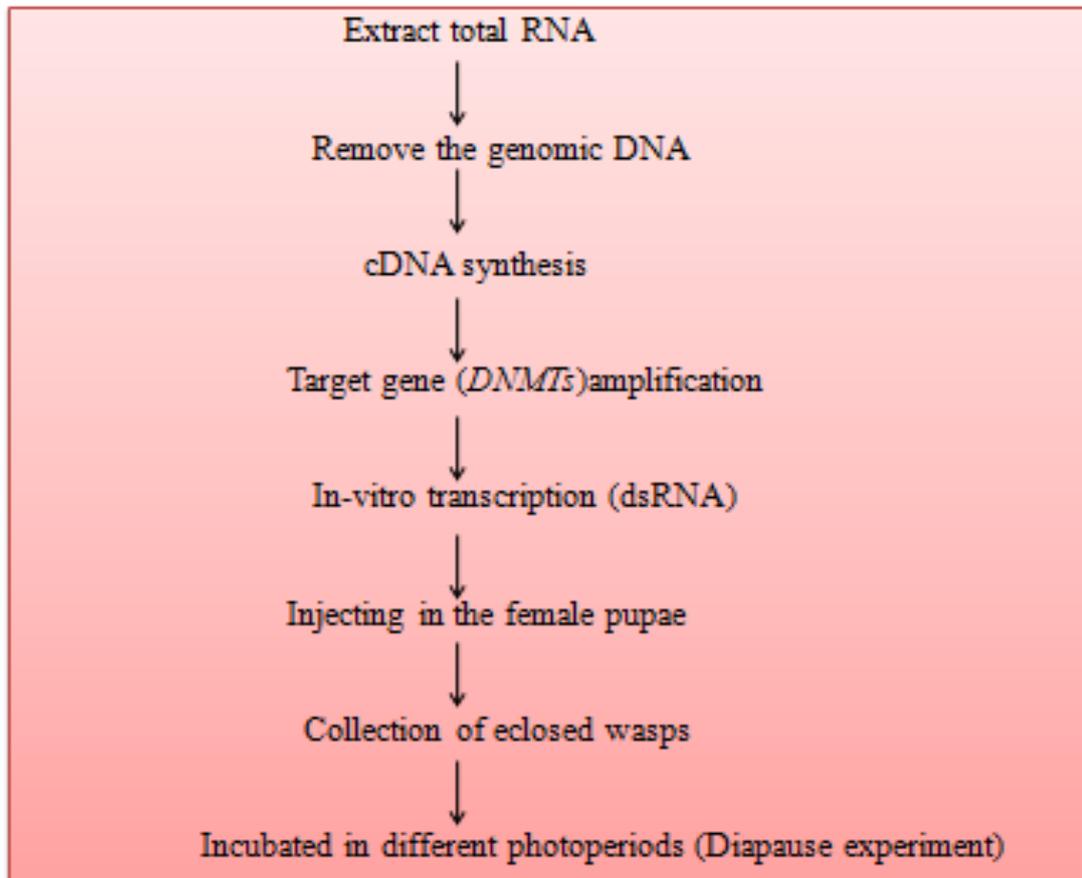


Figure 2.3: *pRNAi* in *Nasonia*: main steps (see text for details).

Off-targets scan for RNAi

One of the crucial steps for the execution of *RNAi* is to achieve specificity, and mitigate the off-target effect of the siRNAs (Jackson & Linsley, 2010). To encounter this, I performed an off-target scan based on the chosen region from the gene of interest (table 2.1) and zeroed in on the genomic region free of non-specific knock-down (off-target) for the dsRNA synthesis. Firstly, the corresponding mRNA sequence of the target gene was retrieved from NCBI database (<http://www.ncbi.nlm.nih.gov>). Then, 450-550 bases long region was selected for the synthesis of dsRNA (double-stranded RNA) preferably from the 3'end of the transcript. This region was primarily scanned for the prevalence of off-target effects. Briefly, all the possible 20-22 base fragments (hypothetical siRNA) were generated from the selected region (using the R-script) and scanned for the probable off-targets across the *Nasonia* RefSeq (Reference RNA sequences from NCBI) database. Each computationally predicted siRNA (from a particular genomic region)

was matched against the *Nasonia* RefSeq genome using the BLAST program and any possible match ≥ 18 mer based on sequence complementarity (one gap permitted) was designated as a potential off-target. Only mRNA sequence region with no potential off-targets was selected for the synthesis of dsRNA.

Gene	NCBI Accession	Coordinates	Size (bp)
<i>nvDnmt1a</i>	NM_001171050.1	4166-4644	479
<i>nvDnmt1b</i>	XM_001600125.2	143-625	483
<i>nvDnmt1c</i>	XM_001607286.1	2893-3349	457
<i>nvDnmt3</i>	XM_001599173.2	706-1236	531
<i>nvDcr-1</i>	XM_001605237.2	1010-1681	671
<i>nvTET-2</i>	XM_008205119.1	4790-5299	510

Table 2.1: List of mRNA accessions (NCBI) that were used for RNAi in *Nasonia vitripennis*. Each row gives the corresponding NCBI accession, sequence coordinates and dsRNA construct size for the specific gene knock-down.

Synthesis of dsRNA

Total RNA was isolated from the adult females using phenol-chloroform extraction procedure as described above. The purified RNA was used as a template for first-strand cDNA synthesis by *SuperScriptII Reverse Transcriptase* (Invitrogen). Target gene such as *Dnmt1a* was amplified using the synthesized cDNA by PCR (see above). For each gene amplification, specific forward and reverse primers were used as mentioned in appendix.

In order to synthesize dsRNA (target and control) the template (PCR product) was incorporated with T7 polymerase binding site at the 5' ends (figure 2.4).

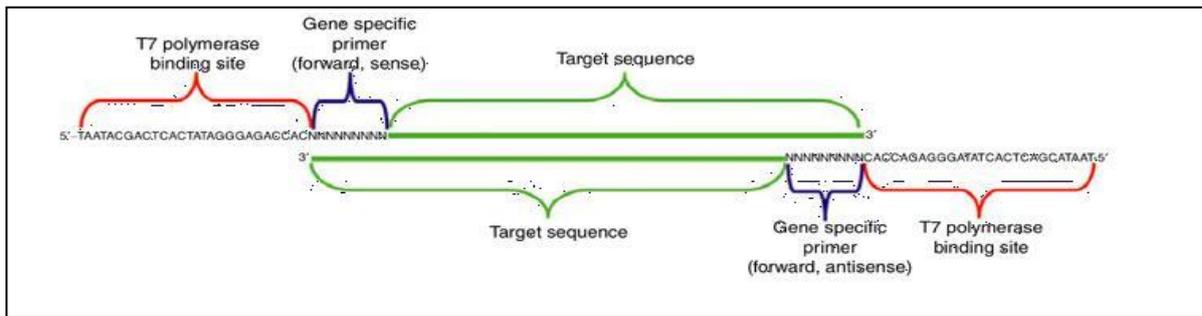


Figure 2.4: **Schematic representation of the primer design for dsRNA template synthesis.** The T7 polymerase binding site (red brackets) incorporated at the 5' ends of 20 nt specific forward and reverse primer (blue brackets) was used to amplify the target gene sequence (marked in green). From Desplan & Lynch, 2006.

The in-vitro transcription reaction was assembled at room temperature using MegaScript T7 RNAi kit (Ambion, cat no.1333) according to the manufacturer instructions with the exception of overnight incubation (instead of 2-3 hours) at 37°C. On the following day, the template DNA was denatured using *TURBO DNase* (supplied with the kit). Thereafter, dsRNA was recovered by phenol-chloroform purification method. To the reaction mixture 115 µl of nuclease free water and 15 µl of ammonium acetate stop solution was added. Then equal volume of phenol: chloroform: isoamylalcohol [125:124:1] (ph 4.7, Life Technologies) was added followed by centrifugation at 12000 x g for 2 minutes. The upper aqueous layer was carefully transferred to a fresh tube. After that equal volume of chloroform: isoamylalcohol [24:1] (Life technologies) was added, vortexed and centrifuged at 12000 x g for 2 minutes. After centrifugation the aqueous phase was again transferred to a fresh tube and mixed with equal volume of isopropanol and 1µl RNase free glycogen. The mixture was chilled for at least 15 minutes at 20°C and centrifuged for 15 minutes at 4°C at maximum speed to pellet the RNA. The dsRNA was re-suspended in phosphate buffered saline (PBS). Finally, the concentration of dsRNA was determined on spectrophotometric based nanodrop and assessed on 1% agarose gel.

Preparation of wasps for injection

Female pupae of appropriate age were selected and aligned with their ventral surface facing upwards. The females were identified by their larger size relative to males, the disproportionately larger wing pads and by the presence of the ovipositor along the ventral midline of the abdomen. The optimal stage for dsRNA injections was the yellow-body pupae (figure 2.5) in which the pupae either lacked or had light-orange eye

pigmentation with yellow coloured body. The pupae containing any traces of dark pigmentation in the thorax were discarded due to compromised *RNAi* efficiency (Lynch & Desplan, 2006). The female wasp pupae of the appropriate age were collected and aligned after 8-9 days from egg-laying, when incubated at 25 °C. A small drop of water was placed on a microscope slide onto which the 22 × 22 mm coverslip was positioned. This prevented the coverslip from moving while the female pupae were affixed in the same orientation with their ventral side up. A maximum of 40 pupae were affixed to each coverslip to prevent overcrowding after eclosion.

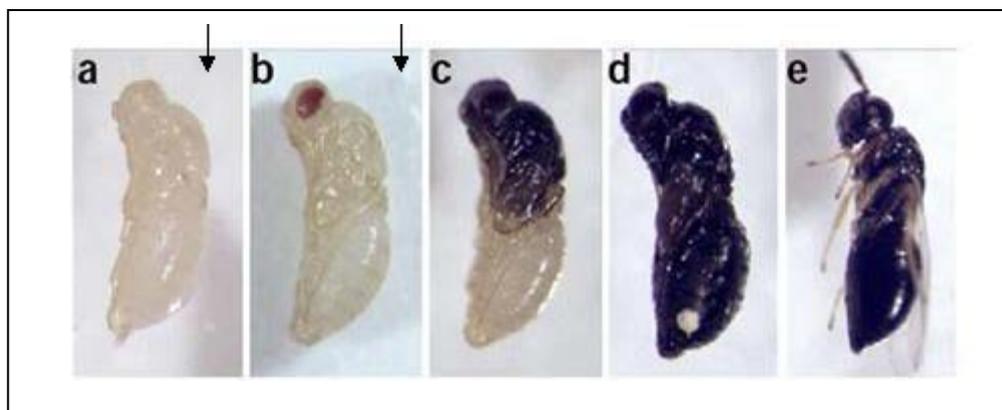


Figure 2.5: **The four stages of *Nasonia vitripennis* development.** During the (a) stage the pupa is yellow and without red eyes, the (b) stage is the red – eyed stage where the eyes pigmentation start, during the (c) stage the body pigmentation starts and (d) the stage of the body pigmentation completed (e) stage the development finishes and the adult wasp is shown. The stages marked by black arrows were appropriate and chosen for injection. (From Desplan & Lynch, 2006)

Microinjection in *Nasonia*

Fine pointed needles were made using needle puller (model p80/PC) from Sutter Instrument Co. as per the settings; Heat: 700, Pull: 150, Velocity: 080, Time: 100. The tip of the needle was broken before the injection by rubbing against the microscope slide. With the help of the microinjector which was connected with nitrogen cylinder from one side and the needle on the other side, the injection was performed. After injection, the wasp pupae were placed in a petridish containing sucrose (10%) soaked cotton at 25°C for 4-5 days. The surviving injected female wasps were incubated either in LD (18 hours of light per 24 hour period) or SD (6 hours of light per 24 hour period) at moderate temperature conditions (18°C) and the proportion of diapausing progeny was recorded.

2.8 Bisulfite Conversion

Bisulfite conversion is considered as a 'gold standard' method to assess the methylation status of cytosine at single nucleotide resolution. This methodology was developed by Hayatsu *et al.* in 1970 (Hayatsu *et al.*, 1970). During the bisulfite treatment, sodium bisulfite deaminates unmethylated cytosine to produce uracil under high temperature conditions and disrupts the complementarity between the DNA strands. In contrast, methylated cytosines are protected from the conversion to uracil and are retained as cytosines. By direct sequencing and comparison to a reference sequence (figure 2.6), one can determine the methylation status of cytosines at single-nucleotide level.

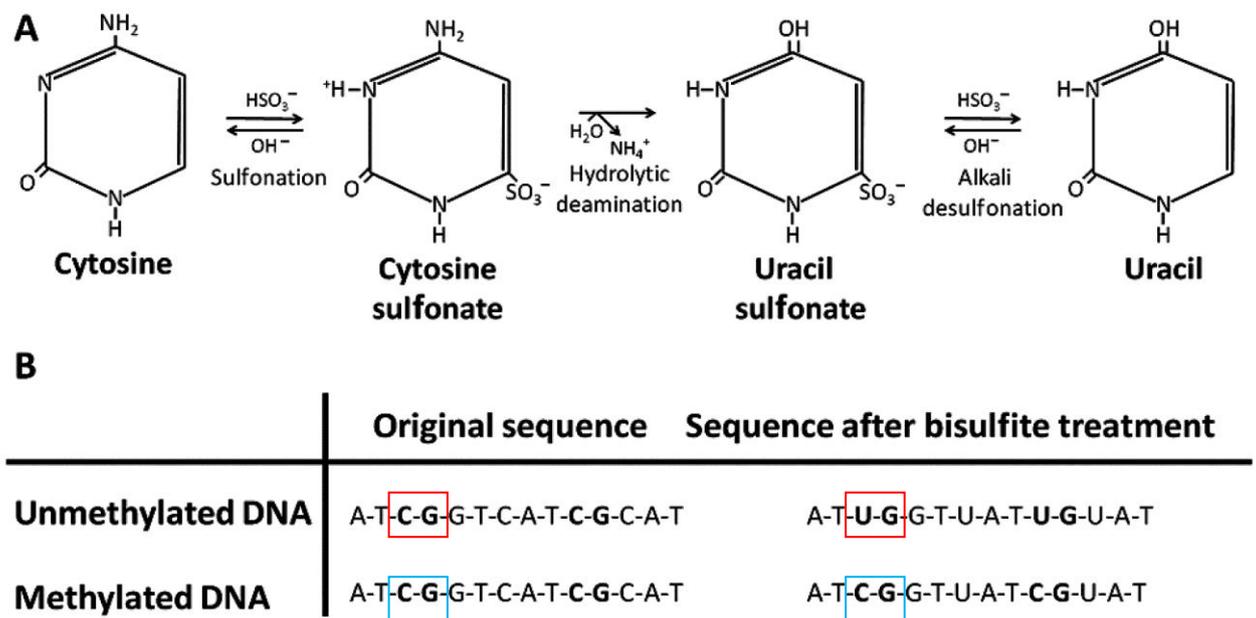


Figure 2.6: **Effect of Sodium Bisulfite on cytosine molecules.** In the figure **A**) shows the subsequent conversion of unmethylated cytosine to uracil after sodium bisulfite treatment and **B**) the genomic sequence with unmethylated cytosine (red box) and methylated cytosine (blue box) gets converted to uracil and cytosine respectively. (Kristensen & Hansen, 2009)

In the current study, bisulfite conversion was performed using Epiect kit (Qiagen, cat. no. 59104). The bisulfite conversion protocol was initiated by sodium bisulfite-mediated conversion of unmethylated cytosines under high temperature and pressure. Thereafter the converted single-stranded (ss) DNA was bound to the column (supplied in the kit) with the help of optimized buffers, desulfonated and purified. Finally, the purified ssDNA was eluted from the spin column and used for downstream analysis (e.g. PCR, cloning). From the experimental point of view, complete conversion of unmethylated cytosines was the most critical step for the accurate determination of

methylation state of cytosines. This was broadly achieved by incubating the DNA in high bisulfite salt concentrations at high temperature (as mentioned below) and low pH. These harsh conditions led to a high degree of DNA fragmentation and subsequent loss of DNA during purification. Purification was, however, necessary to remove bisulfite salts and chemicals used in the conversion process which could otherwise inhibit downstream sequencing procedures. Excessive DNA degradation during the conversion process was prevented by the use of DNA protect buffer (supplied in the kit) which had a pH indicator dye as a mixing control in reaction setup.

The genomic DNA (500-1000ng) was extracted from adult female wasps and digested with *EcoRI* restriction enzyme (New England Biolabs). The fragmented DNA was then purified and used as a template for bisulfite conversion. Bisulfite conversion reaction was set up as described in the kit protocol (see above) and subjected to modified thermocycling conditions facilitating high rate of conversion of unmethylated cytosines into uracil. The cycling conditions included an initial denaturation step at 99°C for 3 minutes, followed by 25 cycles at 96°C for 30 seconds and 55°C for 20 minutes with final hold at 4°C. Using these cyclic conditions more than 97% conversion efficiency (as tested in parallel using synthetic double stranded DNA containing pre-determined methylated and unmethylated CpG sites) was achieved repeatedly. The bisulfite treated DNA was then purified using one of the purification methods (as per the downstream applications) mentioned in the kit protocol and was stored at -20°C for future use.

2.9 Data analysis: the R statistical software

qpcR – Real-time PCR analysis (R environment)

qpcR is an 'R' package for analyzing the difference in gene expression between two or more conditions such as control and treated samples (Ritz & Spiess, 2008). It utilizes the raw fluorescent data (based on the SYBR Green real-time PCR) acquired by the real-time PCR detection system. The underlying approach for quantification between samples relies on the calculation of fold change by means of PCR efficiency and threshold cycle (commonly termed as Ct).

The data obtained from the DNA Engine Opticon2 System (BioRad) were copied in a MS-excel file in the required format. The software calculated the PCR efficiencies of each reaction, and these values were used to calculate the expression ratio between the different conditions that were compared (normalized by the expression of the reference gene) as per the below equation. The software employs a permutation approach, and permutes the threshold cycles (cp) and PCR efficiency (E) amongst replicates within treatment and control group. Ratios were calculated based on each permutation and compared with ratios obtained if samples were randomly reallocated from the treatment to control group. Three statistical p- values were calculated from all permutations that gave a higher/equal/lower ratio. A significant threshold value of p = 0.05 was used to reject each of the three possible scenarios (as explained in the package manual (Ritz & Spiess, 2008)).

$$\text{Ratio} = \frac{E(\text{gc})^{\text{cp}(\text{gc})}}{E(\text{gs})^{\text{cp}(\text{gs})}} \bigg/ \frac{E(\text{rc})^{\text{cp}(\text{rc})}}{E(\text{rs})^{\text{cp}(\text{rs})}}$$

Here, E(gc) = efficiency of target gene for control group , cp(gc) = cycle threshold of target gene for control group, E(gs) = efficiency of target gene for treatment group, cp(gc) = cycle threshold of target gene for treatment group, E(rc) = efficiency of reference gene for control group , cp(rc) = cycle threshold of reference gene for control group, E(rs) = efficiency of reference gene for treatment group, cp(rc) = cycle threshold of reference gene for treatment group.

CHAPTER 3

GENOME-WIDE ANALYSIS OF DNA METHYLATION IN *NASONIA* INDUCED BY PHOTOPERIOD

3.1 Introduction

Epigenetic mechanisms such as DNA methylation, histone modifications and non-coding RNA can effectively regulate transcription or translation of a gene (Bird, 2007). Various factors such as nutrition, toxins, drugs, infection, disease state, and exposure to toxic agents are known to affect DNA and histone modifications (Zhang & Meaney, 2010) and consequently, the epigenome constitutes a crucial interface between the environment and the genome.

In a recent study, the epigenetic process has been shown to play a dynamic role in regulating flowering time in response to environmental stress such as cold, drought and high salinity (Yaish *et al.*, 2011). Environmental cues have been found to modulate the expression level of a set of genes involved in flowering induction and this regulation was facilitated by the dynamic change in the epigenetic state, particularly DNA methylation, of the target genes. In addition, the epigenetic marks can be propagated through mitosis (from parent to descendant cells), or transmitted by means of non-genetic inheritance such as genomic imprinting. There is direct evidence for the transmission of chromatin marks through meiosis in *Drosophila* (Cavalli & Paro, 1998; Siomi & Siomi, 2011) and mammals (Morgan *et al.*, 1999; Sutherland *et al.*, 2000).

DNA methylation

DNA methylation refers to the addition of a methyl group to the 5th position of the cytosine pyrimidine ring or the number 6 nitrogen of the adenine purine ring by DNA methyltransferases (DNMTs; Figure 3.1). DNA methyltransferases catalyze the transfer of a methyl group from a donor molecule such as S-adenosyl-L-methionine (SAM) to the recipient cytosine nucleotide. This covalently bound methyl group does not affect base pairing, and the methyl-bound cytosine will still pair with guanine in the regular fashion. DNA methylation is involved in broad range of processes including regulation of gene expression, X-chromosome inactivation in mammals (Augui *et al.*, 2011), parental imprinting and protection of the genome against selfish DNA (Smith & Meissner, 2013).

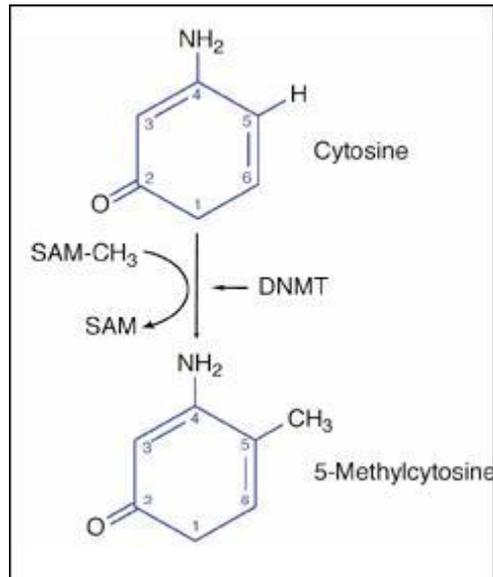


Figure 3.1. **Methylation of a cytosine residue** at the 5' position by the action of DNMTs (Walsh & Xu, 2006).

Current knowledge about the structure and function of DNMTs is largely based on mammalian models, but many of the key motifs are conserved across the taxa. Broadly, three distinct types of methyltransferases are involved in the process: DNMT1 which is a maintenance methylase, DNMT2, and DNMT3 which is a *de novo* methylase (Klose & Bird, 2006). DNMT1 adds methyl marks to the newly synthesized (hemimethylated) DNA (figure 3.2) to retain pre-established patterns across cell generations (Bestor, 2000; Chenet *et al.*, 2003). In contrast, DNMT3 adds novel methyl groups to unmethylated DNA (Okano *et al.*, 1999; Aapola *et al.*, 2002; Hata *et al.*, 2002; Kato *et al.*, 2007). The precise role of DNMT2 is still under investigation, but a few studies have indicated its involvement in tRNA methylation (Goll *et al.*, 2006). Broadly, the substrates of DNMT1 and DNMT3 are not random cytosines but those that precede guanine on the DNA strand (5'—CpG—3'), with a symmetric pattern on the opposite strand. This catalysis results in palindromic daughter strand to the parental strand, also known as symmetric methylation. However, non-CpG methylation conferred by either DNMT3a or DNMT3b has also been reported in mammals (Guo *et al.*, 2013). Methyl-CpG-binding domain proteins (MBDs) represent another vital component of DNA methylation toolkit, as MBDs contain a methyl CpG recognition motif that allows the selective binding of methylated DNA (Klose & Bird, 2006; Clouaire *et al.*, 2010) and can potentially affect epigenetic modifications at multiple levels (Jones & Takai, 2001).

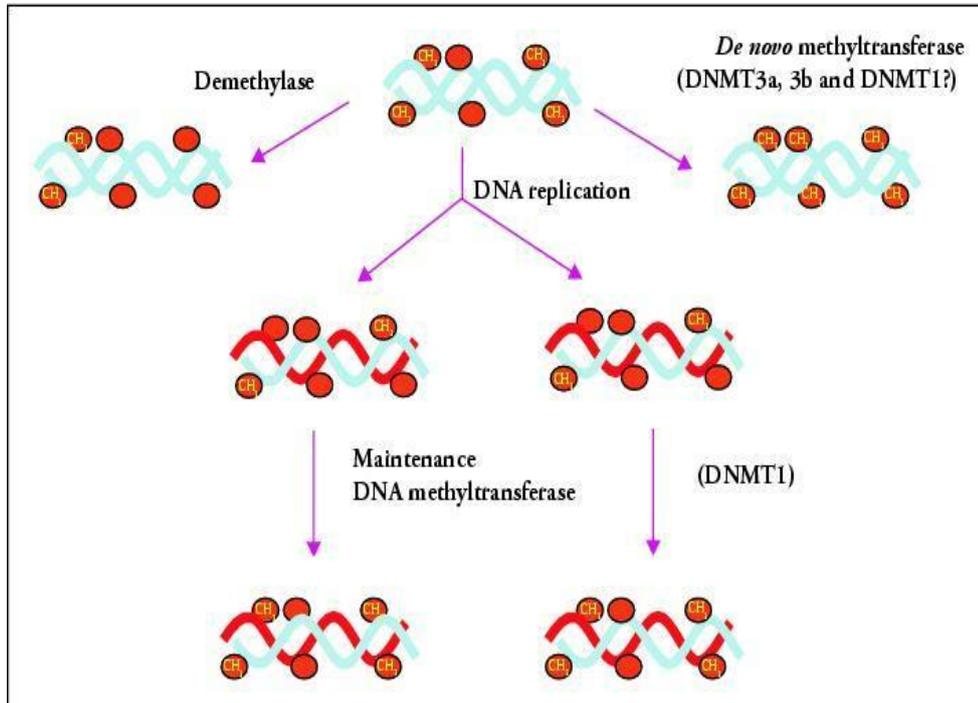


Figure 3.2: **Action of DNMTs on the replicating DNA strands.** DNMT1 (maintenance methylase) passes methyl marks from hemi-methylated parental strand to daughter strand. DNA Demethylase (top left) removes the methyl group from 5-methyl cytosine while DNMT3 (top right) can add novel methyl marks to the cytosine molecule (Szyf, 2006).

In vertebrates, DNA methylation occurs throughout the genome (Suzuki & Bird, 2008; Okamura *et al.*, 2010), with 60-90% of all CpG dinucleotides subjected to methylation in most mammals (Li *et al.*, 2009a). Importantly, methylation of CpG dinucleotides at promoter region (CpG islands) leads to transcriptional repression and subsequent down-regulation or silencing of the target gene. Additionally, DNA methylation in vertebrates could also lead to reduced or repressed activity of transposable elements (Yoder *et al.*, 1997).

In contrast, DNA methylation in invertebrates was found at relatively sparse level (Suzuki & Bird, 2008; Bird *et al.*, 1979b) and highly distinct from mammals and plants. Negligible DNA methylation was detected in the model invertebrate organisms *Drosophila melanogaster* (Raddatz *et al.*, 2013; Rae & Steele, 1979) and *Caenorhabditis elegans* (Simpson *et al.*, 1986) which initially was thought to reflect DNA methylation in invertebrates as a whole. But lately a sporadic level of DNA

methylation has been noticed in *Drosophila* (Takayama *et al.*, 2014). In addition, few genome-wide studies have revealed the presence of DNA methylation, at intermediate (to sparse) level, in many other invertebrates such as *Strongylocentrotus purpuratus* (Bird *et al.*, 1979a), *Ciona intestinalis* (Simmen & Bird, 2000; Suzuki *et al.*, 2007b), *Apis mellifera* (Lyko *et al.*, 2010), *Crassostrea gigas* (Gavery & Roberts, 2010), and importantly in *Nasonia* (Werren *et al.*, 2010; Suzuki *et al.*, 2007a; Feng *et al.*, 2010; Zemach *et al.*, 2010; Wang *et al.*, 2013a), suggesting an evolutionary conservation and a functional role of DNA methylation in these organisms.

In invertebrates, DNA methylation was found to be confined to CpG dinucleotides, particularly in intragenic regions (Suzuki & Bird, 2008; Feng *et al.*, 2010; Simmen *et al.*, 1999). Interestingly, though moderate level of methylation was found to be present in the repeat and transposable elements in basal invertebrates, in insects it is almost non-existent (Feng *et al.*, 2010; Zemach *et al.*, 2010; Regev *et al.*, 1998; Schaefer & Lyko, 2007). In insects, the normalized CpG content (CpG observed/expected) is widely used as a proxy measure for DNA methylation. As conserved methylated cytosines are prone to undergo spontaneous deamination to thymine with high frequency (Glastad *et al.*, 2011), highly methylated CpG regions often exhibit a marked reduction in the CpG dinucleotides, represented by low CpG content (low CpG_{O/E}) cluster, while sparsely methylated regions exhibit high CpG dinucleotide content represented by high CpG genes (high CpG_{O/E}). The genes with high CpG dinucleotide content have increased capacity for differential gene expression, whereas genes with dense methylation exhibit a greater propensity for static level of expression (Hunt *et al.*, 2010). This may underlie the bimodal distribution of normalized CpG dinucleotide content that was observed in *Apis mellifera* and in *Acyrothosiphon pisum* (figure 3.3), which have also previously suggested the involvement of environmentally induced epigenetic modifications in generation of phenotypic plasticity (see below). In contrast, *Drosophila melanogaster*, which does not exhibit substantial level of CpG methylation, clearly showed unimodal distribution of CpG_{O/E} (Fig. 3.3).

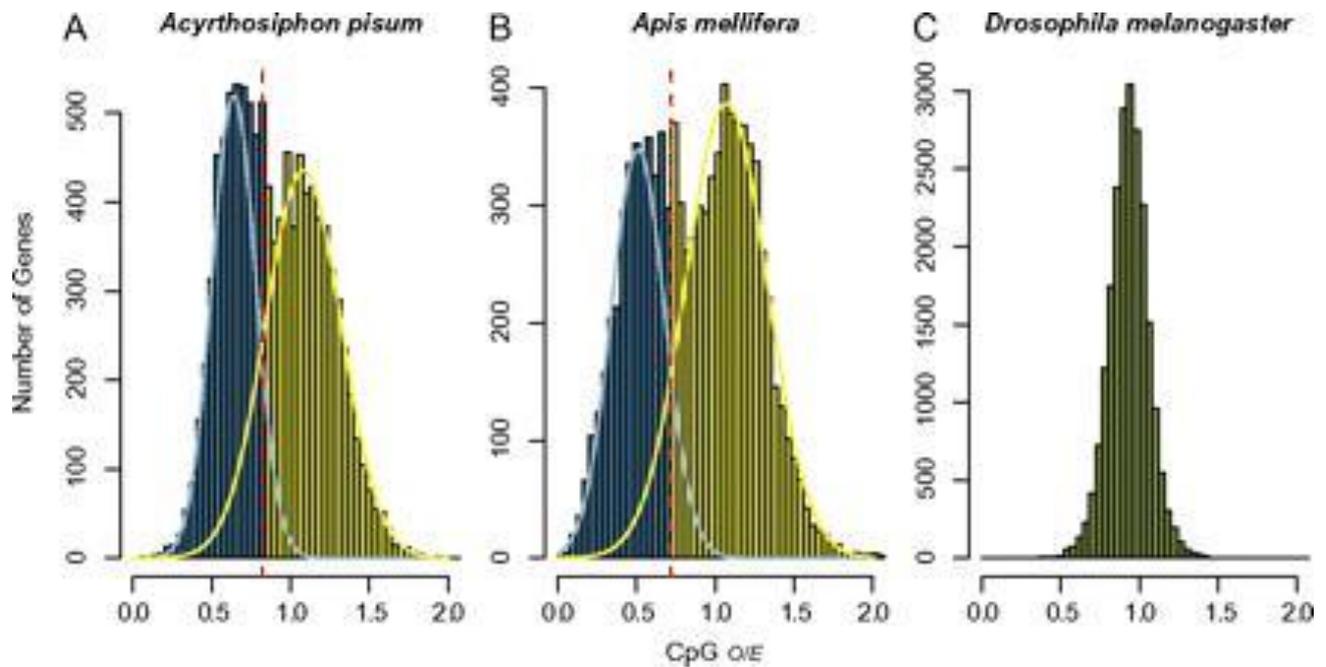


Figure 3.3. Distribution of normalized CpG dinucleotide content (CpG observed/expected) in (A)*Acyrthosiphon pisum* (B)*Apis mellifera* (C) *Drosophila melanogaster*. (Hunt *et al.*, 2010)

Although *Nasonia* possesses a complete DNA methylation tool-kit consisting of DNA methyltransferases DNMT1 (1a, 1b and 1c), DNMT2 and DNMT3 (Werren *et al.*, 2010), the functional significance of DNA methylation in *Nasonia* remains unknown. However, studies in the closely related honeybees demonstrated that silencing of *DNMT3* in the larval stage by RNA interference results in the development of destined adult worker bees with developed ovaries, a phenomenon generally associated with the development of the queen bees (Kucharski *et al.*, 2008). Thus, DNA methylation seems to be involved in caste switching in honeybees. Moreover, DNA methylation is also involved in the behavioural division of labour in the adult honey bees (Lockett *et al.*, 2012), as the observed levels of methylation differed between adult nurse bees and foragers in response to the external social stimuli. Similarly, difference in DNA methylation was also observed between non-reproductive and reproductive worker bumble-bees (Amarsinghe *et al.*, 2014).

Differences in DNA methylation levels were also associated with developmental stages and division of labor in ants (Smith *et al.*, 2012). Although DNA methylation was similar amongst workers and gynes (the primary reproductive female) at larval and pupal stages, the methylation levels differed between adult workers and virgin queens

(lower in adult workers). In the virgin queens, the increased level of methylation led to controlled protein and carbohydrate metabolism.

Another example of encoding environmental information by DNA methylation is the phenotypic plasticity in the aphid, *Acyrtosiphon pisum* (Dombrovsky *et al.*, 2009). This aphid produces asexual male (winged and wingless) morphs (parthenogens) in spring and sexual female morphs in the fall and winter. Clonal individuals of a founder mother produces different types of progeny (pink, green, and white aphids) when exposed to different environmental conditions and this variation is associated with the level of CpG methylation, which differs considerably between the three types of aphids.

In the experiments described here, I investigated whether DNA methylation has a similar role in encoding environmental information underlying phenotypic plasticity in *Nasonia*. I have tested whether exposure to long or short day are accompanied by differential DNA methylation. To that end, I have adopted an advanced, highly efficient technique known as Reduced Representation Bisulfite Sequencing (RRBS) to map genome-wide 5-methylcytosine at single base level (Gu *et al.*, 2011). RRBS is a next-generation sequencing approach, in which the genome subjected to the sequencing gets enriched for CG rich regions.

3.2 Material and Methods

RRBS library preparation and sequence analysis

RRBS promotes mapping genome wide 5-methylcytosine at a single nucleotide resolution. Briefly, to construct RRBS sequencing libraries, purified genomic DNA is digested with methylation insensitive restriction enzyme (*MspI*) to generate fragments that contain CpG dinucleotides at their ends. After end-repair and A-tailing, the fragments are ligated with methylated Illumina adapters. At this stage size selection (40-220 bp) is performed prior to bisulfite conversion. Finally, the libraries are amplified, purified and sequenced on an Illumina platform. The RRBS library was constructed following a previously published protocol (Gu *et al.*, 2011). Adult females (AsymC strain), were incubated at either long (16 hours) or short (8 hours) photoperiod for 10 days at 18°C and used for the assay.

Genomic DNA was isolated from the whole body of 10 wasps, using the DNeasy Blood and Tissue kit (Qiagen). The samples were thawed on ice and homogenized with 180µl PBS (50mM potassium phosphate, 150mM NaCl, pH-7.2) for 5 minutes followed by *Proteinase K* treatment. To minimise potential DNA degradation *DNase*-free *RNase* was used in the lysis buffer. The samples were then incubated at 56°C for 15 minutes and spun at 13,400 rpm for 30 seconds to pellet cell debris. The supernatant was collected in fresh tube and treated using the kit protocol as per the manufacturer instructions. To elute DNA, 40 µl elution buffer (supplied in the kit) was added directly on column, placed in 1.5 ml microcentrifuge tube and incubated at room temperature for 1 minute. Finally the column was centrifuged at 6000 x g for 1 minute. The concentration of the eluted DNA was determined by Qubit fluorometer and dsDNA BR (broad range) assay kit.

The extracted genomic DNA (300 ng) from long and short day samples was digested using *MspI* (C|CGG), a methylation-insensitive restriction enzyme, to generate short fragments with CpG dinucleotides at the ends (10 units of *MspI* [New England Biolabs] were added with 10X NEB Buffer 4 to the samples and incubated at 37°C for 2 hours). To stop the digestion reaction 0.5M EDTA was used. An aliquot equivalent of 50ng DNA was taken from each sample and assessed on 4-20% pre-cast polyacrylamide TBE gels (BioRad) (figure 3.4). The digested DNA samples were purified, and the DNA pellet was dissolved in 15 µl elution buffer.

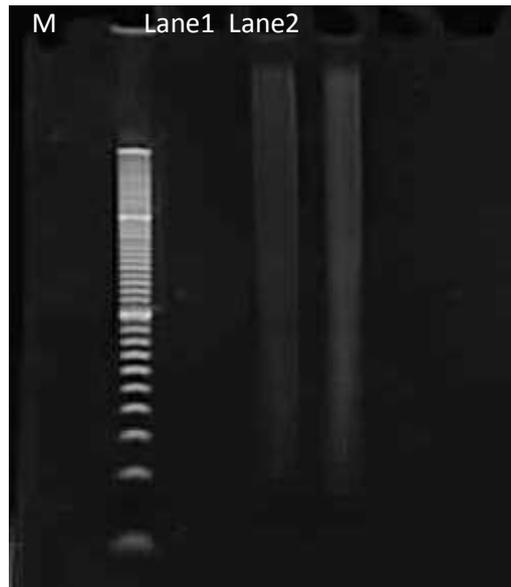


Figure 3.4: ***MspI* restriction digestion analysis.** 4-20% pre-cast polyacrylamide TBE gel, from the left: M represents 20bp extended molecular weight marker (Lonza, 20-500 bp), Lane 1: *MspI* digested long day (LD) sample and Lane 2: *MspI* digested short day (SD) sample.

The purified digested DNA was end-repaired and 3' A-overhang was attached using 5 units Klenow fragment (3'-5' exonuclease; New England Biolabs, cat. no. M0212L). The end repaired and A-tailed DNA samples were purified following a previously published protocol (Gu *et al.*, 2011). Methylated adapter oligonucleotides, in which all cytosines (C's) were replaced with 5-methyl cytosines (5-mC) were used to maintain the integrity of double stranded DNA during subsequent bisulfite conversion reaction. The blunt ended fragments were ligated to methylated standard Illumina adaptors (15 μ M, double HPLC purified, Integrated DNA Technologies, Inc) which also included a 5 nucleotide unique barcode at the 3' end (Table 3.1), allowing sequencing the samples on a single Illumina lane. Ligation reaction was performed using 2000 units of *T4 DNA Ligase* (New England Biolabs, cat. no. M0202M), followed by overnight incubation at 16°C. The adapter ligated fragments were then purified and eluted in 15 μ l TE (Tris-EDTA) buffer.

Oligonucleotide	Sequence (5'-3')
Adapter 1 Bottom	/5Phos/TTTTTAGATmCGGAAGAGmCTmCGTATGmCmCGTmCTTmCTGmCTTG -100nmole DNA oligo (Double HPLC purification)
Adapter 1 Top	AmCAmCTmCTTTmCmCmCTAmCAmCGAmCGmCTmCTTmCmCGATmCTAAAAA*T-100nmole DNA oligo (Double HPLC)
Adapter 2 bottom	/5Phos/AATTAAGATmCGGAAGAGmCTmCGTATGmCmCGTmCTTmCTGmCTTG-100nmole DNA oligo (Double HPLC Purification)
Adapter 2 Top	AmCAmCTmCTTTmCmCmCTAmCAmCGAmCGmCTmCTTmCmCGATmCTTAATT*T-100nmole DNA oligo (Double HPLC purification) 5-Phos : Phosphate, m: Methylated, *: phosphorothiate linkage

Table 3.1: **Adapter sequences used for library preparation (RRBS)**. The top and bottom oligonucleotide sequences of the respective adapters (1, 2) were ligated before use.

Based on in-silico digest, 40-200 bp fraction of the *Nasonia* genome was isolated from the adapter-ligated DNA using a 3% Nusieve gel (3:1 agarose gel; Lonza , cat. no. 50084). The size selected libraries were purified using Min Elute Qiagen Gel extraction kit (cat. no. 28604). The end-repair and adapter ligation efficiency was determined by using semi-quantitative PCR. In the PCR reaction, 1 µl size selected and purified DNA was used as template in 50 µl total reaction volume. For amplification, 2.5 units PfuTurbo Cx Hotstart polymerase (Agilent technologies, cat. no. 600412), 2.5µM of each standard Illumina forward and reverse primers, and 25mM each dNTPs (deoxynucleotide triphosphates) were used per reaction. The amplification program with initial denaturation at 95°C for 2 minutes, followed by 10-20 cycles at 95°C for 30 seconds, 65°C for 30 seconds, 72°C for 45 seconds and final extension step at 72°C for 7 minutes, was followed. The PCR products (from different PCR cycles) were viewed on a 4-20% precast polyacrylamide TBE gel (BioRad) (figure 3.5). The increase in DNA smear with varying number of amplification cycles confirmed the successful ligation of adapters.

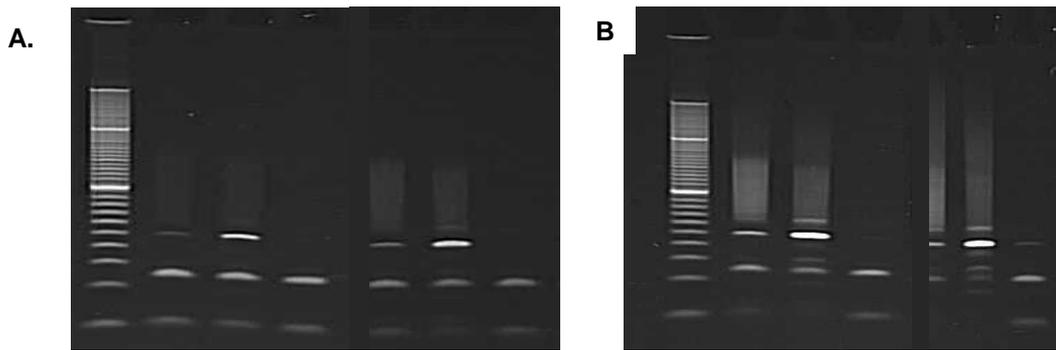


Figure 3.5: **DNA ligation efficiency analysis.** Lanes on panel A (4-20% pre-cast polyacrylamide TBE gel stained with SYBR Green I), from the left: M represents 20bp extended molecular weight marker (Lonza), Lane 1 and 2 represents LD and SD samples respectively after 10 amplification cycles, Lane 4 and 5 represents LD and SD samples respectively after 13 amplification cycles. On Panel B (4-20% pre-cast polyacrylamide TBE gel), from the left: M represents 20bp extended molecular weight marker (Lonza), Lane 1 and 2 represents LD and SD samples respectively after 16 amplification cycles, Lane 4 and 5 represents LD and SD samples respectively after 19 amplification cycles. In both the panels Lane 3 and Lane 6 represents no template controls (NTC) for the respective amplification (10, 13, 16 and 19 cycles).

The size selected DNA was treated with sodium bisulfite and purified using the Epitect kit (Qiagen, cat. no. 59104). The conversion was done as previously described (Chapter 2), and the bisulfite treated DNA was purified using the specified kit protocol for formalin-fixed paraffin-embedded [FFPE] tissues as per the manufacturer instructions to increase the yield. The DNA was finally eluted in 20 μ l elution buffer. The final libraries were generated by PCR amplification using PfuTurbo Cx Hotstart polymerase (2.5 U/ μ l; Agilent Technologies, Inc), in a 50 μ l PCR reaction. 1 μ l of bisulfite converted and purified DNA sample was used. The amplification was performed using 2.5 units of PfuTurbo Cx Hotstart polymerase, 2.5 μ M each standard Illumina forward and reverse primer (table 3.2) and 25mM each dNTP. The PCR programme (detailed above) was run for 13, 16, 19 and 20 cycles, in order to determine the least number of cycles required to amplify the libraries. The PCR products were viewed on a 4-20% precast polyacrylamide TBE gel (BioRad) (figure 3.6). The increase in DNA smear with the varying number of amplification cycles confirmed the successful construction of RRBS libraries. PCR amplified RRBS libraries (13 cycles) were cleaned by using AMPure magnetic beads (Agencourt, cat. no. 000130) and quantified on Agilent 2100 Bioanalyzer (Agilent Technologies). The purified RRBS libraries (25 ng) were sequenced for Illumina HiSeq 2000 platform.

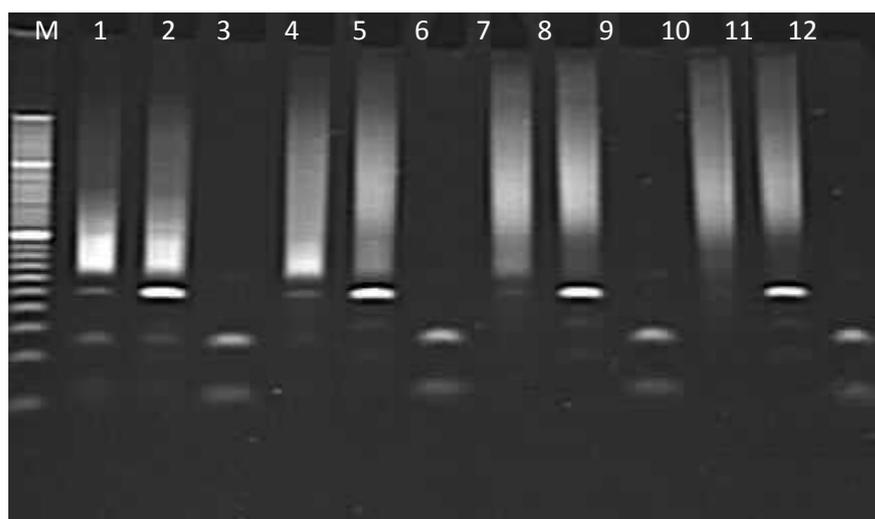


Figure 3.6: **Final RRBS library amplification.** The gel (4-20% pre-cast polyacrylamide TBE gel stained with SYBR Green I), shows from the left: M represents 20bp extended molecular weight marker (Lonza), Lane 1 and 2 represents LD and SD samples respectively after 13 amplification cycles, Lane 4 and 5 represents LD and SD samples respectively after 16 amplification cycles, Lane 7 and 8 represents LD and SD samples respectively after 19 amplification cycles, Lane 10 and 11 represents LD and SD samples respectively after 20 amplification cycles. Lane 3, 6, 9 and 12 represent no-template control (NTC) for each amplification (13,16, 19 and 20 cycles) respectively.

Primer	Sequence (5'-3')
PCR Primer 1.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T-100nmole DNA oligo (Double HPLC purified)
PCR Primer 2.1	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC*T- 100nmole DNA oligo (Double HPLC purified)

Table 3.2, Illumina standard forward (PCR Primer 1.1) and reverse primer (PCR Primer 2.1) used for amplification.

RRBS data analysis

The libraries were sequenced using Illumina Hiseq2000 analyzer (Beijing Genomic Institute, Hong Kong). Two Illumina sequencing runs yielded 117,646,358 raw 50bp reads in total. FastQC version 0.10.1 was used to provide data quality information at each step of the quality control process. The data were taken through several pre-processing steps before being aligned to the *Nasonia* genome (version Nvit_2.0) using bismark v.0.7.7 (Krueger & Andrews, 2011). These steps included de-multiplexing and barcode trimming, as well as trimming of low quality bases and the Illumina adapter sequences from the 3' end of the reads, using the Trim galore v0.2.5 package. Low

quality reads (cutoff of 20 Phred score) were removed using FASTX-Toolkit (0.0.13; http://hannonlab.cshl.edu/fastx_toolkit/), that also removed potential noise (i.e. reads that did not begin with YGG (where Y = C or T), as all RRBS reads are expected to begin with either CGG or TGG, in the non-converted and converted cases, respectively). To facilitate downstream analysis, the resulting *bismark* SAM output was sorted by chromosome and position using *Samtools* version 0.1.18 (Li *et al.*, 2009b).

Once aligned to the genome, methylation sites were mapped by identifying C to T conversions produced by the bisulfite conversion. The error rate of methylation calling was estimated for each of the samples. During the RRBS library construction, unmethylated Cs was added as a part of end-filling at the *MspI* restriction site. Cs at this site was expected to appear as Ts in the sequencing data (under high bisulfite conversion efficiency). If Cs appeared in the sequences at these position, it was indication for an incomplete bisulphite conversion, or T>C sequencing error. Only reads which contained the Illumina adapter and barcode at their 3' end with an overlap of 11 bases or more were used to estimate the error, as this was judged to be a good indication that the filled-in position could be located within these reads. The error estimate (E) due to incomplete bisulphite conversion and sequencing error was given by $E = n_{Cf} / (n_{Cf} + n_{Tf})$, where n_{Cf} is the number of C's found at the filled-in position and n_{Tf} is the number of T's found at the filled-in position. The error estimates for each sample were computed and found to be 0.0068 (LD) and 0.00622 (SD).

To classify a cytosine as methylated (mC), the binomial distribution was used to calculate the probability of getting K successes (methylated Cs) in N trials (read depth) with probability E of a success (Lister *et al.*, 2009). The p-values that had been generated in this way for each site were adjusted using the Benjamini-Hochberg method, setting the FDR at 1%.

Differentially methylated Cs were identified using methylKit (Akalin *et al.*, 2012) that uses the raw *bismark* output data as an input. For each CpG, the proportions of methylated reads in the two photoperiods were compared using the Fisher's exact test and the p-values were corrected to genomic-wide false discovery rate (FDR) based q-values, using the SLIM method (Wang *et al.*, 2011). The methylated sites were then mapped to genes using custom Perl scripts.

RRBS data validation

To validate the RRBS data and confirm the differential site-specific CpG methylation, a sensitive and highly robust real-time PCR based approach, MethylQuant was used (Dugast-Darzacq & Grange, 2009). MethylQuant is particularly useful when the CpG density is low, or the methylation state differs between neighboring CpG molecules. Based on the real-time PCR approach, a discriminative (D) primer was used which amplified only if the cytosine to be tested, present at the 3' end was methylated, and a non-discriminative (ND) primer amplified regardless of the methylation state of the cytosine at the 3' end. This enabled to quantify the proportion of cytosine methylated molecules at a single specific locus in the genome between different samples such as long and short day.

The procedure starts with the extraction of genomic DNA and sodium bisulfite treatment. Next, a pre-amplification is done spanning the region that possesses CpG of interest using conventional PCR with up-downstream primers (Box 3.1), preferably containing no methylated cytosines. This is followed by a real-time PCR to quantify the total amount of amplified material (using non-discriminative primer) and the fraction corresponding to the methylation status of the specific cytosine (using discriminative primer). The pre-amplification minimise the introduction of PCR bias from the use of small amounts of starting material or heterogeneous samples (mixture of methylated and unmethylated molecules). This amplification step is ideally performed with a buffer containing tetramethylammonium chloride (TMAC), which has proved specifically useful for the amplification of the difficult sequences lacking 5-methyl cytosines (Thomassin *et al.*, 1999) and AT rich primer (Melchior & Von Hippel, 1973).

BOX 3.2: Primer Design using MethPrimer for RRBS validation

The CpG of interest (figure 3.7) was identified and primers for the first round of pre-amplification were picked up-and-down stream of the CpG of interest using MethPrimer as exemplified for the locus *LOC100117390* (see figure below). For the qPCR, primers were designed using MethPrimer, as discriminate (with C at 3' end) and non-discriminate forward and reverse primer pair.

```

TTGGTGTTAATTTTTTTTTTATTTAAGGTACGTAATTTAAATGACGTTTAAAAACGTTTGT
ATTTGAAATTAGAAAGTGAAAGGTTTTATTTAATCGGTATCGTTATATTTCGTGGTTAT
TTAATAGATTTTTTATTCGTTTGTATTTGAGAAATTAATGTAATTATTGTAATAAATGTATT
TTTTTTTTTGTAGTTGATAGATACGATTTTATAAGTATGGTAAGTATTTTATTAATTGG
TTTATCGGTTATTAGCGTTTTGTTTAACTTTGCGAATTAGTTGGATTGGTTGATGTTTGA
TTAAATTGAGTAGTATCGGATGTTTTAGAATTAGTATATGTTTTTGAATAAATTCGGGT
AAGTGTTATTAATAAGTTATCGTAGAGTGAGTAGTGGTCGAGTTTTATTTTCGGTCCGAGTA
GAATTAATTGGTGTATAAATACGTTAGATCGATTCGATTTTCGAATTTGCGGATTTGTCGTC
GTTATAGTCGTTTGTGATTTT
  
```

Figure 3.7. An example of nucleotide sequence of the bisulfite converted sequence spanning the test CpG (highlighted in red). The forward and reverse primers (highlighted in yellow) were picked for pre-amplification using MethPrimer. The qPCR primers (highlighted in blue) were picked as forward (discriminate and non-discriminate) and reverse primers amplifying a region between 80-150 base pairs.

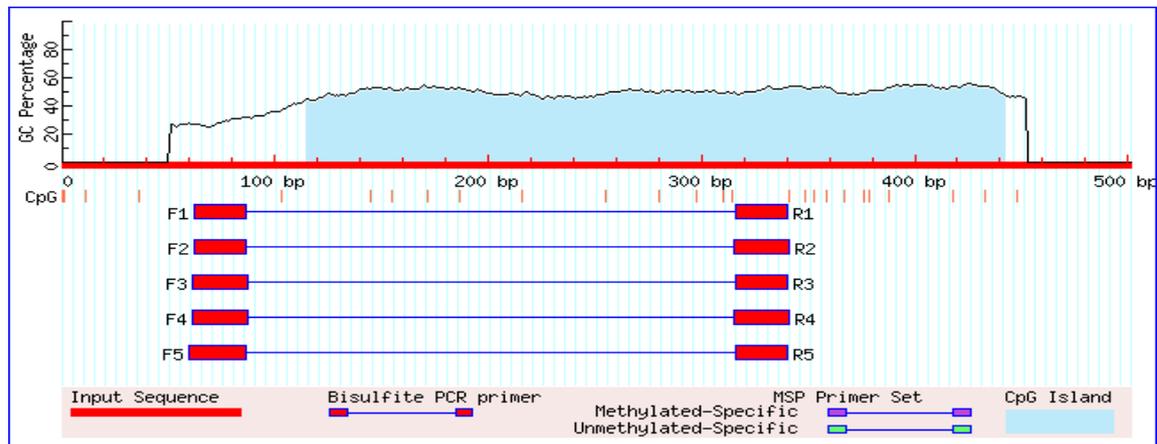


Figure 3.8. Example of genomic sequence that was used as an input for primer design using MethPrimer. Specific parameters were defined for primer design with CpG island size > 100 bp, GC Percent > 50.0, CpG Obs/Exp ratio > 0.6. One CpG island(s) was found in the sequence (shaded in blue) and primers were picked in the up-downstream region as mentioned above.

For each sample, two types of real-time PCR reactions were performed using either one of the two primer sets: (1) a non-discriminative (ND) primer set that permitted quantification of the PCR product irrespective of the methylation status of the CpG of interest, and acted as a reference for the quantification analysis in real-time PCR and (2) a discriminative (D) primer set that enabled quantification of the PCR product corresponding to the specific methylation state of the CpG of interest (present at the 3' end). The presence of a locked nucleic acid (LNA) at the extreme 3' end of the discriminate primer further ensured efficient discrimination during amplification

between respective methylated and unmethylated test cytosine (Thomassin *et al.*, 2004). LNA is a nucleic acid analog with a 2'-O, 4'-C methylene-bridge that locks the ribose moiety into a C3'-endo conformation, thus increasing the hybridization specificity of primers (Latorra *et al.*, 2003). In principle, four different primers could be designed to analyze a single specific CpG of interest, two for the methylated cytosine molecule and two for the unmethylated cytosine molecule on either of the two strands of the amplified product. Here, the less abundant molecule (in our case methylated cytosine molecules in contrast to unmethylated cytosines) was measured. In situations where the CpG of interest was flanked by another CpG that might undergo changes in methylation levels, primers incorporated with inosine at the corresponding cytosine position were used (Thomassin *et al.*, 2004). In principle, inosine can base pair with any of the four conventional bases with approximately equal strength (Martin *et al.*, 1985).

Genomic DNA was extracted from 10 females kept in either long or short-day conditions for 10 days using the All-in-one purification kit (Norgen, cat. no. 24200). The extracted genomic DNA was digested using *EcoRI* (New England Biolabs, cat. no. R0101S) and purified (see Chapter 2). The fragmented DNA was then treated with sodium bisulfite and purified using the Epiect kit (Qiagen, cat. no. 59104). The conversion was done as previously described (Chapter 2) and purified using the protocol for formalin-fixed paraffin-embedded [FFPE] tissues as per the manufacturer instructions to increase the yield. The DNA was finally eluted in 20 µl volume and pre-amplified individually for each of the selected gene (highly differentially methylated genes: *LOC100117390*, *LOC100117821*, *WDR36*) with primers up-downstream to the test specific CpG of interest, designed using MethPrimer (Li & Dahiya, 2002). The amplification was performed using KAPA Taq polymerase (KAPA Biosystems, BK1000), 0.4µM specific forward and reverse primer (Table 3.3) each with the thermocycler conditions: 92°C for 2 minutes, followed by 15 cycles of 30 seconds at 92°C, 1 minute at 55°C for locus *WDR36* and 50°C for *LOC10117821* and *LOC10117390*, 1 minute at 72°C, then 7 minutes at 72°C. The amplicon was further purified and quantified using Qubit dsDNA High Sensitivity reagent. The purified pre-amplified product was then assessed using Agilent 2X SYBR green mix by real-time PCR on BioRad Opticon light cycler. The real-time PCR was performed individually in 25µl reaction volume with 0.4µM specific forward (discriminative or non-discriminative) and reverse primer. For qPCR, three-step cycling conditions with initial denaturation at 95°C for 15 minutes, followed by 45 cycles of 15 seconds at 94°C, 30 seconds at 49°C for

LOC10117821, 53°C for *LOC10117390* and 51°C for *WDR6*, and 30 seconds at 68°C was performed. The PCR was followed by melting curve analysis (settings: 50°C to 95.1°C, read every 0.2°C hold for 1 sec). The data were analysed using the R- statistical software, with the qpcR package (Ritz & Spiess, 2008b).

The analysis used a Monte-Carlo simulation, which is based on repeated random sampling followed by a statistical analysis. The input variables such as Efficiency and Cycle threshold (in real-time PCR) are drawn at random from the continuous distributions (normal distribution) or non-continuous distribution (binomial distribution). The mathematical modeling is applied on the input variables to generate the output. This generates hundreds of probabilistic outcomes (such as fold change in the level of expression of the target gene between the sample and the control group in real-time PCR) which are then tested for statistical significance. The histograms based on the output values can be plotted to compute the probability distribution (De neve *et al.*, 2013).

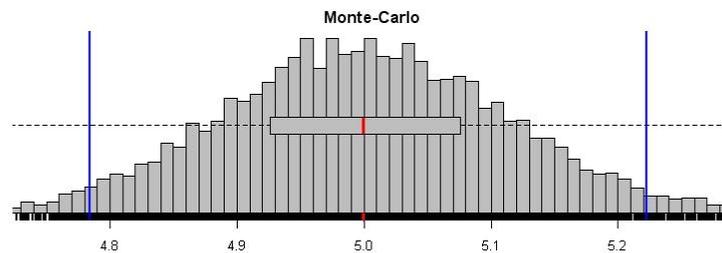


Figure 3.9: Analysis of qPCR data by Monte-Carlo simulation. In the horizontal boxplot, red bar denotes the median value. The box is drawn that contains 50% of the data around the median. Whiskers range from lower to upper limit at 95% confidence interval.

Table3.3: Primers used for RRBS data validation.

Gene (Primer used)	Forward Primer (5'-3')	Reverse Primer (5'-3')	Gene ID (NCBI)	Co-ordinates	size (bp)
LOC100117821			341864960	c7593895-7593681	215
(Pre-amplification)	TTGGTTATGTTATAT GGTTATTTAG	TATCTACTAACA ACTCTCATT CATC			
(Discriminative)	TATGGTTATTTAG/ideoxyI/GGTTT+C	TCTAATCCTAAT TATTATTAACC			
(Non-discriminative)	TATGGTTATTTAG/ideoxyI/GGTTT			c7593883-7593782	102
WDR36			341864959	c 30423392-30423117	276
(Pre-amplification)	TAATTTGGGGATTTT TTTGTGG	TCAATATTCAAC AAATTCAACCA AC			
(Discriminative)	GTTATTTTAGATTTT TTAGATAGTA +C	TAACCAAACA AACATAATAAC			
(Non-discriminative)	GTTATTTTAGATTTT TTAGATAGTA			C30423277-30423148	130
LOC100117390			341864961	14674461-14674739	279
(Pre-amplification)	TTATTTGTAAAAATG TATTTTTTTT	ATTATAACACC AATTAATTCTAC TC			
(Discriminative)	GTTTG/ideoxyI/ GAATTAGTTGGATT+ C	ATAACTTATTAA TAACACTTACCC		14674567-	113
(Non-discriminative)	GTTTG/ideoxyI/GAAT TAGTTGGATT			14674679	

3.3 Results

Photoperiodic response of AsymC strain

The wild-type AsymC strain exhibited a robust photoperiodic response (Fig. 3.10). The progeny of females that were kept in either long (LD) or short day (SD) showed distinct phenotypes. With the increase in the entrainment time (5, 9 and 10 days), the number of diapausing progeny decreased considerably in LD, and increased in SD. A minimal exposure of 5 days was sufficient for inducing substantial difference in the photoperiodic response (Two-tailed Student *t*-test, $t = -5.99$, $df = 82$, $p < 0.001$). After 9 days, all the progeny arising from wasps kept under SD conditions were in diapause.

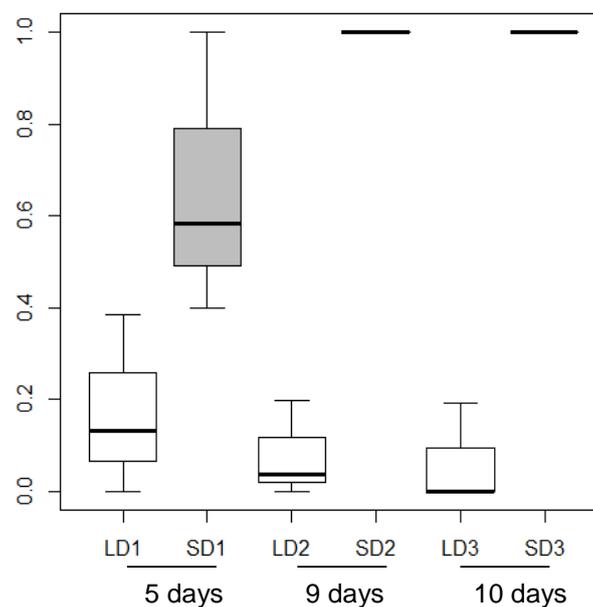


Figure 3.10: **Photoperiodic diapause response in *Nasonia* (AsymC)**. Females kept in either LD (whitebox) or SD (grey box) showed a robust photoperiodic response after 5, 9 and 10 days at 18°C (30-33 females in each group, progeny of each female collected from two hosts, average $n=26$ larvae per female). The line inside each box represents the median, the top and bottom 75 and 25th percentiles respectively, error bars range from minimum to maximum.

The Nasonia methylome

The *Nasonia* methylome was constructed by mapping the sequence reads to the reference genome (Nvit 2.0), accompanied by stepwise quality control, alignment and CpG methylation steps. After removal of low-quality reads, 27,255,357 reads (1.06Gbp) from the long day sample (LD) and 30,818,609 (1.25Gbp) reads from the short day sample (SD) were obtained. The mapping efficiency of these reads to the *Nasonia* reference genome was 84.1% (LD) and 83.2% (SD). The average coverage depth of the reads was 92.6X for the short day sample, and 86.4X for the long day sample. By the

same metric, 4.53% (1,270,716) of CpG sites in the genome were covered by both samples at an average depth of 95.72X for the SD sample, and 88.97X for the LD sample. The LD reads covered 1,304,087 CpG sites (4.65 % of all genomic CpGs) while the SD reads covered 1,385,067 CpG (4.94 %). Based on this, 1618 and 2018 methyl CpGs in the LD and SD samples respectively were detected at a 1% false discovery rate (FDR, see methods above), comprising 0.124% and 0.146% of the cytosines with sequence coverage that were analysed.

While most of the methylated cytosines were in a CpG context (Figure 3.11A), a small fraction (< 6.7%) of these were in non CpG context (mCHG and mCHH, where H = C, T or A), in both the LD and SD samples. CpG methylation showed a large amount of overlap between the two samples (Fig. 3.11B), whereas the other contexts show little (CHH; 1/125 sites), or no overlap (CHG), which suggested that the detected non-CpG methylation largely represented experimental noise. The methylation of CpG sites also showed a bimodal distribution (Fig. 3.11C): CpGs seem to be either heavily methylated or lightly methylated with little in between (the methylation level of a cytosine is here defined as the percentage of reads that are methylated). Whether this bimodality is driven by the same mechanism underlying the bimodality of CpG_{observed/expected} ratio (a proxy measure of CpG methylation) that was previously reported (Park *et al.*, 2011) is yet to be determined. The majority (83.6%) of the methylated CpGs were located in exons, while only 9.79% were located in introns, 1.1% in promoter regions and 5.5% in intergenic regions (Fig. 3.11 D, E).

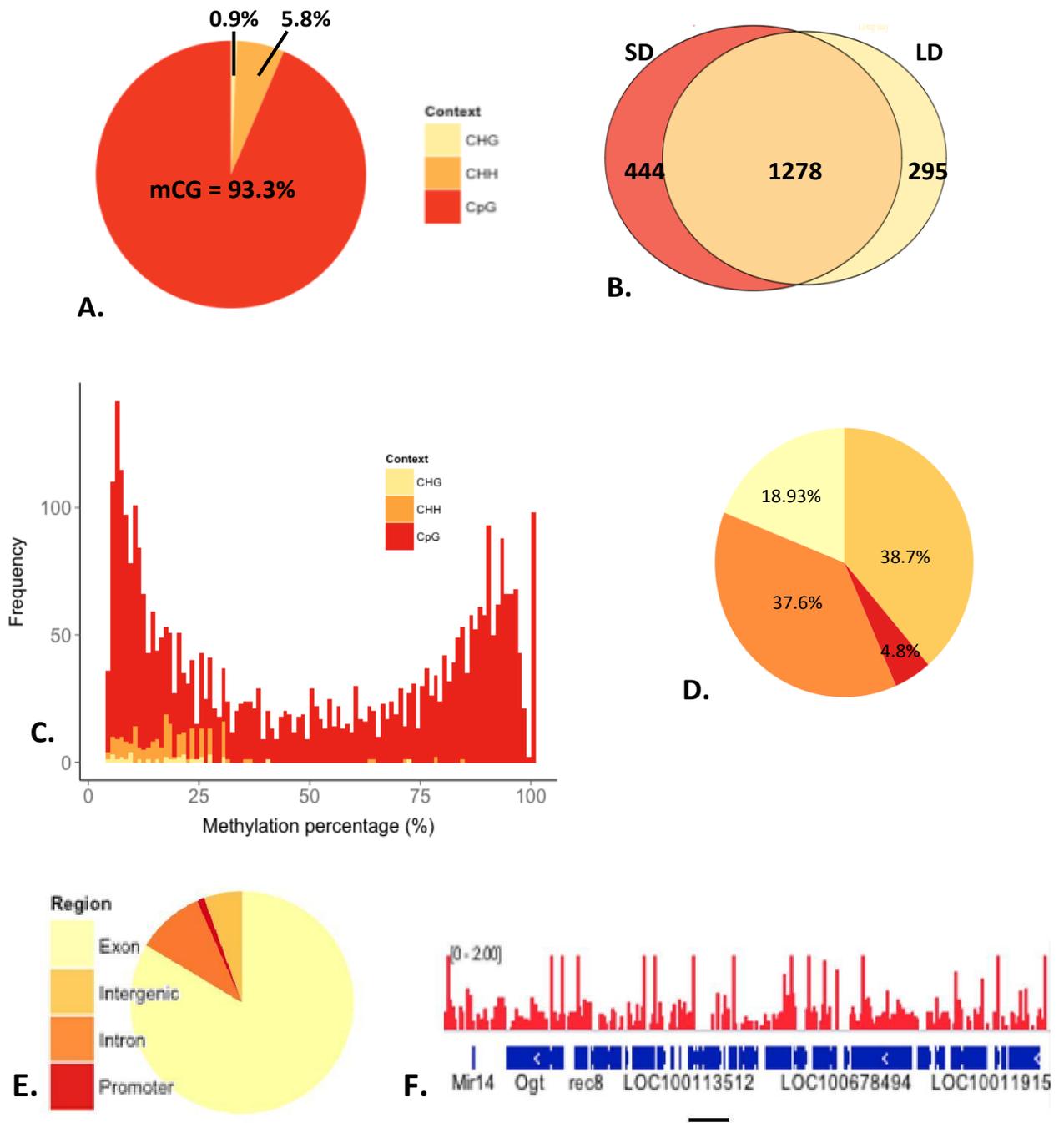


Figure 3.11: **The *Nasonia* methylome.** **A.** Distribution of mCs identified in different sequence contexts (mCG, mCHG and mCHH, where H = non G). **B.** Venn diagrams showing that substantial proportion of the mCGs are methylated under both long- and short-day conditions, in contrast to other sequence contexts that show negligible overlap (see text), suggesting that non-CpG methylation represents largely experimental noise. **C.** The methylation level of CpG sites (percentage of covering reads that are methylated) shows a bimodal distribution. This bimodality is not seen in CHG, or CHH (shaded yellow and orange) methylation contexts. **D.** Fraction of CGs analysed in exon, intron, promoters (2Kb upstream genes) and intergenic regions. **E.** Fraction of mCGs identified in exon, intron, promoters (2Kb upstream genes) and intergenic regions. **F.** A snapshot of the methylome (751kb region) is depicted overlaid with the raw long day methylation data (scaled to 0-2% methylation). Scale bar: 50Kb.

Identification of differentially methylated genes

I used Fisher's exact tests to compare methylation in long and short photoperiods, and identified 51 differentially methylated CpG regions, mostly in exons (DMRs; Benjamini–Hochberg multiple testing, false discovery rate (FDR) <0.05), which were mapped to 39 unique genes. Approximately half of the DMRs (23/48) showed reduced methylation in long day compared to short day, while the other showed the opposite trend (i.e. hypermethylation). The genes represented a wide range of functions, such as GTPase activity, protein binding, and transferase activity (acyl and glycosyl groups). A few of these genes harboured multiple DMRs, including the *UbiA prenyltransferase domain containing protein 1(Ubiad1)* gene. Out of 42 CpG sites in this gene that were covered by our library, 6 were significantly more methylated in the short-day sample (Figure 3.12B). In contrast, the gene *Misshapen (Msn)*, encoding a protein kinase, was represented by seven CpGs sites in our reads, five of which showed increased methylation in the long-day sample (at the FDR-corrected $q < 0.01$; Figure 3.12C).

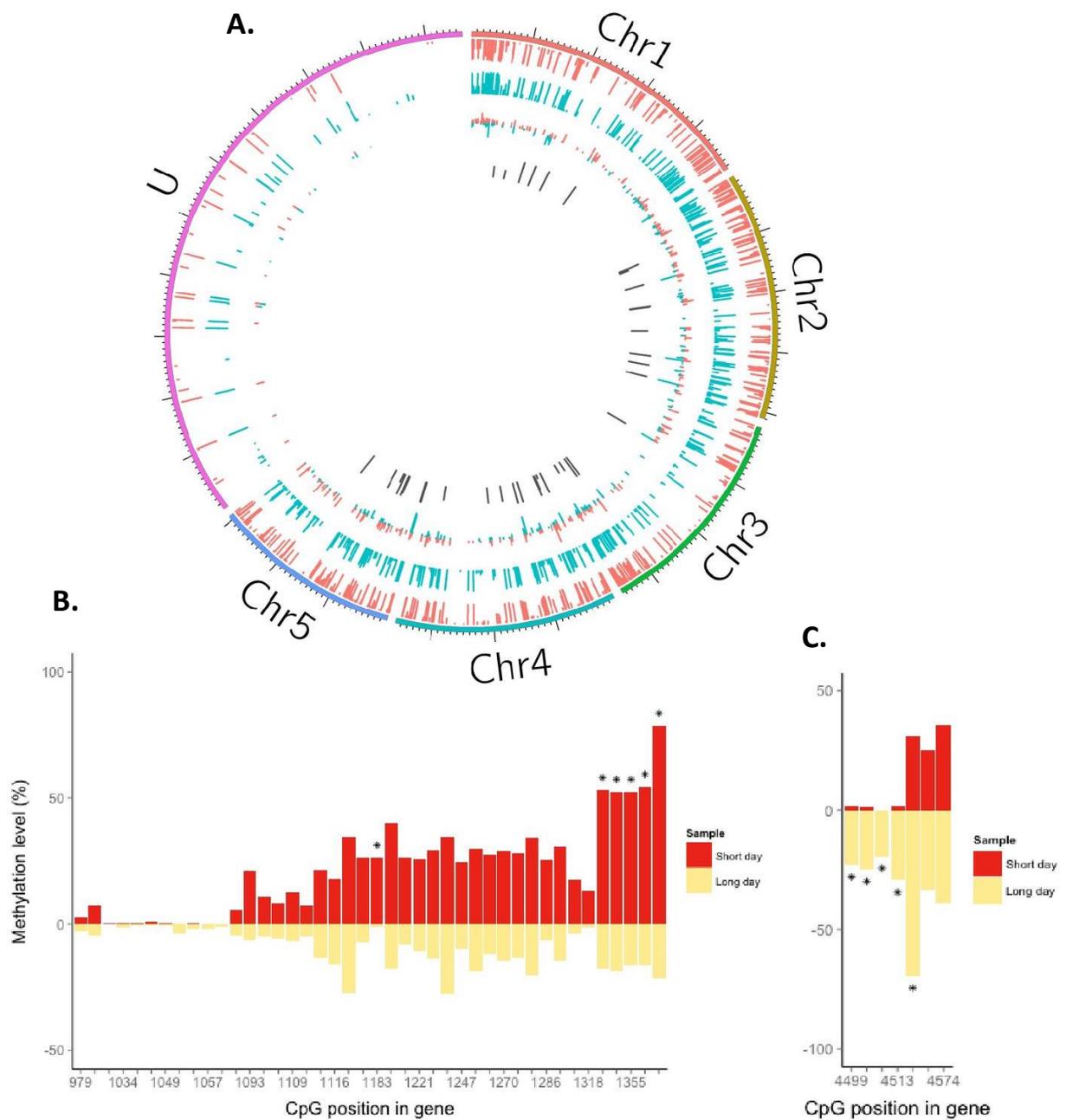


Figure 3.12: **Differential DNA methylation associated with photoperiod in *Nasonia*.** **A.** The methylation is shown across the five chromosomes and the unplaced (U) scaffolds. From outer to inner circle, the data tracks are short day methylated CpGs (red), long day methylated CpGs (blue), methylation difference (red/blue), and significantly differentially methylated sites ($q < 0.05$, dark grey). **B.** Differential methylation in the *Ubiad1* (*LOC100118618*) gene. Out of 42 CpG sites that were covered, six were significantly more methylated in short day ($q < 0.05$). **C.** Differential methylation in *Msn* indicates increased methylation in long day: There are 7 covered CpG sites on this gene, of which 5 are methylated at $q < 0.01$.

RRBS data validation

To validate the RRBS data, three highly differentially methylated genes: *LOC100117390*¹, *LOC100117821*, *WDR36* were tested. Three biological replicates were pooled from LD and SD photoperiods as described above, and used to analyze difference in the level of methylation per gene using MethylQuant. For SYBR Green real time PCR, two technical replicates were included for each biological replicate and assessed individually with the discriminative and non-discriminative primer pair. The methylation in three genes was higher in LD condition than in SD (Table 3.4).

Gene	NCBI gene ID	'C' position	Difference in methylation (%)	q-value
LOC100117390	GeneID:100117390	14674587	100	1.56E-10
LOC100117821	GeneID:100117821	7593794	99.26	3.00E-30
WDR36	GeneID:100679400	30423252	94.7	1.15E-08

Table 3.4: **Locii selected for validation.** For each tested gene (represented in each row) namely, *LOC100117390*, *LOC100117821*, *WDR36*, difference in the level of methylation between LD and SD samples is shown along with the corresponding q-values, exact position of the test cytosines and NCBI gene accessions.

Fig. 3.13 shows a significantly positively fold change ($p < 0.05$ by permutation²), which indicated higher expression in the test (LD) group. High expression of the target gene (D primer set) in the LD group implied more number of molecules with cytosine at the specific test locus at 3' end. This clearly suggested more methylation in the LD group than SD group for the test genes. Thus, using the above method three highly differentially methylated sites were successfully validated.

¹Gene IDs starting with LOCXXX represent genes with unknown function.

²Exchange of group memberships resulted in a smaller ratio occurred at frequency 5%.

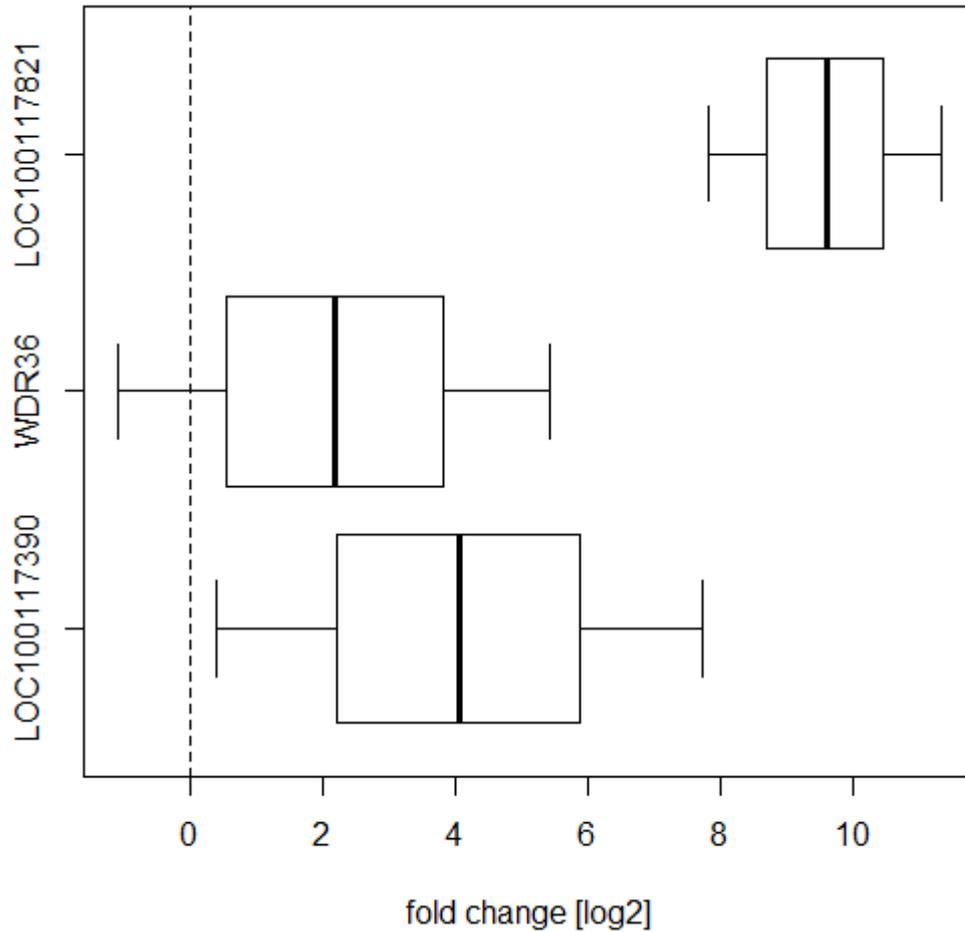


Figure 3.13: **Validation of differential methylation by qPCR.** The [log₂] fold change in the number of copies of specific test genes namely *LOC100117390*, *WDR36*, *LOC10011782* in the LD sample group against the SD, associated with the level of methylation at the specific CpG sites. The positive [log₂] fold change for the specific test genes (using discriminative primer), indicated more number of molecules with 'C' (cytosine) at the particular test CpG site at the 3' position in the test (LD) as compared to the control (SD) group after bisulfite treatment. For each plot the line inside box represents the median [log₂] fold change while the error bars range from minimum to maximum based on Monte-Carlo simulation at 95% confidence intervals.

Functional enrichment of the identified differentially methylated sites

The gene ontology (GO) analysis of the identified differentially methylated genes revealed an enrichment for genes involved in metabolic and cellular processes (figure 3.14). This finding was consistent with the study of Wolschin and Gadau (Wolschin & Gadau, 2009a) that revealed the dominance of metabolic enzymes and maintenance proteins during diapausing stages in *Nasonia*. For the complete list of the GO terms associated with the differentially methylated genes refer to section 8.1 in the appendix.

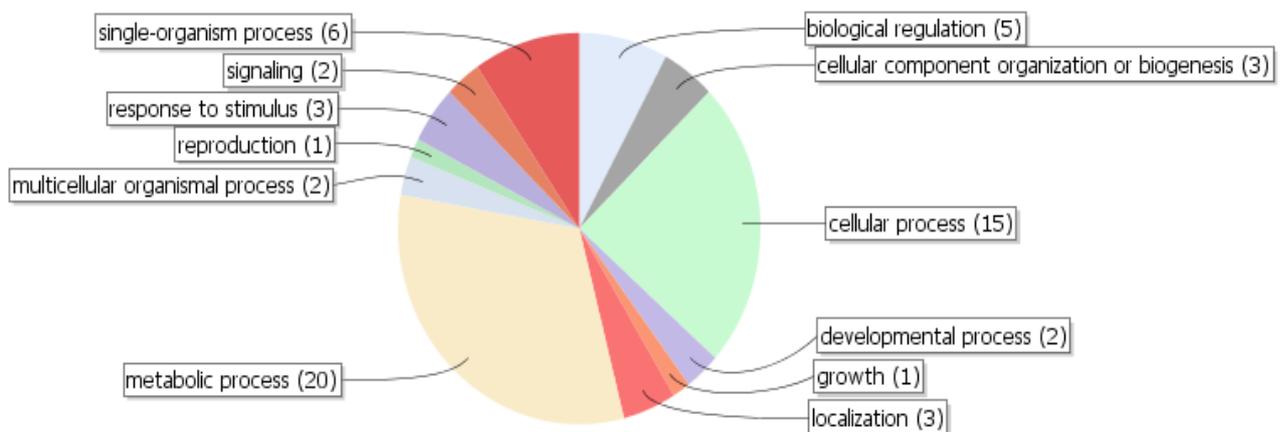


Figure 3.14: **GO analysis of differentially methylated genes in *Nasonia***. The gene ontology analysis (based on 51 differentially methylated sites) revealed an enrichment of genes involved in metabolic and cellular process (light orange and green shaded parts) in *Nasonia* as analysed by Blast2GO tool (Gotz *et al.*, 2008).

3.4 Discussion

The photoperiodic response in *Nasonia* is robust and easy to simulate under artificial laboratory conditions. The female wasp produces entirely diapausing progeny after 10 days of incubation in SD at a moderate temperature (18°C), and very high proportion of non-diapausing developing larvae under LD conditions (Fig. 3.10), and this was the stage that was selected for methylation profiling by RRBS. Although by using RRBS, only small fraction of the genome, comprising only 1%, was analysed, which provided only a snapshot of the methylome, it enabled a reliable detection of methylation of the covered and mapped cytosine molecules. The advantage of using this technique was the resulting high sequence depth that allowed an efficient testing of differential methylation.

Functional significance of Differential methylated regions (DMRs)

In summary, 84.1% and 83.2% of total reads were mapped from the LD and SD samples respectively to the *Nasonia* reference genome. The analysis of mapped reads further revealed 0.13% and 0.15% of covered cytosine molecules to be methylated in LD and SD samples respectively (Figure 3.11, 3.12). The low level of methylation observed in *Nasonia* is consistent with the whole genome methylation profiles found in *Apis mellifera* (0.12%) (Lyko *et al.*, 2010) and *Bombyx mori* (0.11%) (Xiang *et al.*, 2010b). Another shared characteristic with other invertebrate organisms, is the observation that the CpG methylation was predominantly found in the exons in *Nasonia* (see figure 3.11D) (Lyko *et al.*, 2010; Xiang *et al.*, 2010b; Elango *et al.*, 2009). During the course of work, a few other studies have been published on *Nasonia* that showed a similar prevalence of low level of methylation restricted to CpG dinucleotides in the exons (Park *et al.*, 2011; Zwier *et al.*, 2012; Wang *et al.*, 2013b). Moreover, the whole-genome DNA methylation profile generated by Wang *et al.* (2013) also revealed positive correlation between DNA methylation and gene expression based on the 1.6% of the total covered CpG sites.

On close inspection of the fraction of cytosine methylation, 39 individual genes were found to be differentially methylated between LD and SD conditions (see Appendix section 8.1). While conventional wisdom argues that differential methylation regulates transcription via methylation of promoter regions, the differential methylation in *Nasonia* was primarily present in the exons (figure 3.11). This is consistent with several studies that have suggested the potential function of intragenic DNA methylation

is the regulation of alternative splicing (Maunakea *et al.*, 2010; Shukla *et al.*, 2011; Foret *et al.*, 2012b). For example in honey bees, methylated genes, harbouring DNA methylation chiefly in the exons, have shown enrichment for alternative splicing (Flores *et al.*, 2012). The intragenic methyl marks and alternative splicing were also seen to be positively correlated with the gene length. This gives rise to the possibility of generation of varied alternative transcripts from the same gene, presumably with new functions, whilst still retaining the existing isoforms. Moreover, in a separate study, the intragenic methyl marks have also been shown to direct which exons to be included during alternative splicing in the mature transcripts in the honeybees (Foret *et al.*, 2012b). Interestingly, the authors have clearly exemplified that exons which are highly methylated are included in the mature transcripts. However, the biological function of the intragenic methylation is unclear and may also vary depending on the organism and the pattern of DNA methylation. Nonetheless, intragenic methylation has been shown to promote transcription of genes in anemone (*Nematostella vectensis*) and the silkworm (*Bombyx mori*). In *Nasonia* too, differential CpG methylation has been suggested to regulate splicing of a gene (*transformer*) related to sex determination (Park *et al.*, 2011). It was also suggested that sparse level of methylation (as seen in *Nasonia*) may facilitate specific transcriptional avenues including access to alternative transcription start sites, increasing sequence mutations, and exon skipping (Roberts & Gavery, 2012). DNA methylation could also regulate gene expression by modulating the access to alternative promoter sites. A recent study in mammals revealed that intragenic methylation restricts the generation of alternate gene transcripts by masking intragenic promoters (Maunakea *et al.*, 2010).

Exon skipping is another possible mechanism by which a transcriptional variant might be generated. In *Apis mellifera*, the gene *GB18602* has two forms (long and short), which are distinguished by a cassette-exon being skipped in the long form (Lyko & Maleszka, 2011). This exon possesses a stop codon that creates a shorter, alternative transcript. This phenomenon would be consistent with the differential methylation that could lead to alternative transcripts under different environmental conditions.

The work presented here generated the first differential methylation map of *Nasonia* induced by the photoperiod, and highlighted the possible importance of DNA methylation in seasonal timing. In the next chapter, I would describe experiments testing the functional significance of DNA methylation in the photoperiodic response of the wasp.

CHAPTER 4

FUNCTIONAL SIGNIFICANCE OF DNA METHYLATION IN PHOTOPERIODISM IN *NASONIA*

4.1 Introduction

Comparison of the *Nasonia* methylomes in long and short days revealed a substantial differential DNA methylation (Chapter 3). Yet, these methylation differences may either represent various downstream processes that mediate the photoperiodic response, or alternatively, they may constitute the upstream (input) loci that are important for coding the photoperiodic information. Here, I conducted a series of functional assay experiments to test the causal role of DNA methylation in the photoperiodic response of *Nasonia*. RNA interference (*RNAi*) was used to knock-down the target *DNMT's*; *DNMT1a*, *DNMT1b*, *DNMT1c* (*DNMT1*) and *DNMT3* and the effect on the photoperiodic response was analysed. The multiple sequence alignment of *Dnmt1* paralogs revealed conservation of various domains such as DNMT1-RFD, required for cytosine specific DNA methylation (Figure 4.1). Pairwise comparisons of the protein sequences (NCBI-Blast) showed 63 % similarity between *Dnmt1a* and *Dnmt1b*, 44 % between *Dnmt1a* and *Dnmt1c* and 42% between *Dnmt1b* and *Dnmt1c*. Here, in order to determine the specific role of each DNA methyltransferase enzymes, the non-homologous region was selected as the target for *RNAi*.

Additionally, to determine a broader impact of inhibition of DNA methylation on the photoperiodic response, a pharmacological agent, 5-aza-2-deoxycytidine, was used. 5-aza-2-deoxycytidine is an azacytidine derivative that was originally developed and tested as nucleoside antimetabolite with clinical specificity for acute myelogenous leukemia (Cihak, 1974; Sorm *et al.*, 1964). The drug functions as a nucleoside analogue (cytosine analogue), as depicted by Z in the figure 4.2. Soon after administration, the drug gets incorporated as a nucleoside (cytosine) analogue into the replicating DNA strands, and inhibits the activity of DNMTs in an irreversible manner (Christman *et al.*, 1983; Creusot *et al.*, 1982), which leads to compromised maintenance and *de-novo* methylation due to inhibition of *Dnmt1* and *Dnmt3* activity respectively. The mechanism of irreversible inhibition of *Dnmt1a* by 5-aza-2-deoxycytidine is well understood, but the effect of the drug on the activity or conformation of *Dnmt3* is yet unknown. These recently identified *de-novo* methyltransferases are predicted to have

the same or much more sensitive response to the methylation inhibitor (Creusot *et al.*, 1982; Christman, 2002).



Figure 4.1: Multiple Sequence Alignment of *DNMT1* Paralogs. The three *Nasonia* paralogs of *DNMT1*: *Dnmt1a* (NP_001164521.1), *Dnmt1b* (XP_008217946.1) and *Dnmt1c* (XP_001607336) were aligned using CLUSTALW, a multiple sequence alignment tool. The conserved domains (NCBI Conserved Domain Database) present among protein sequences are highlighted and marked.

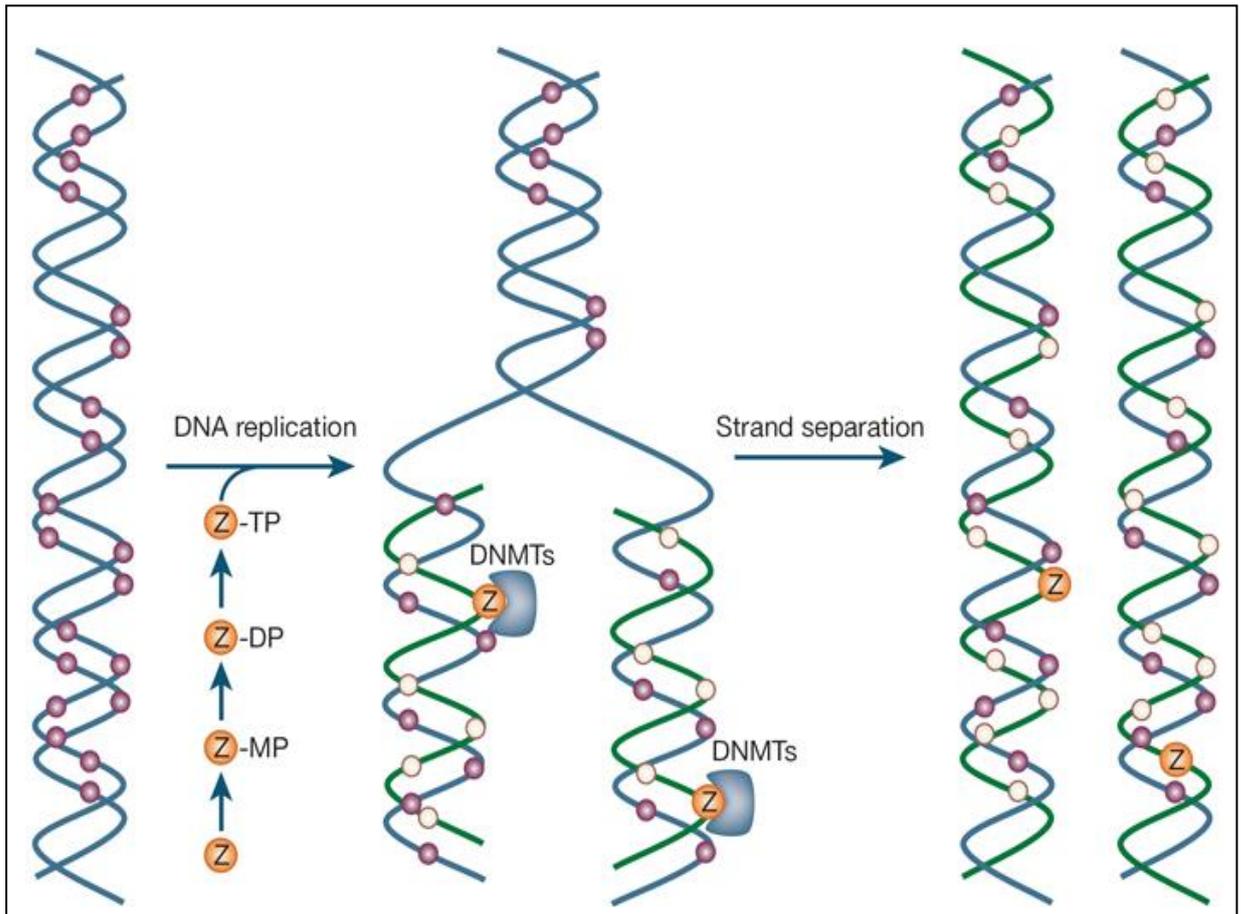


Figure 4.2: **Mechanism of action of cytosine analogue inhibitors (depicted by Z).**The incorporation of cytosine analogue (converted into triphosphate (TP) from diphosphate (DP) and monophosphate (MP)) in the replicating DNA strands leads to subsequent passive loss of methylation in the resulting daughter DNA molecules.(Adapted from Egger *et al.*, 2004)

4.2 Methods

Expression analysis of DNMTs

The level of relative gene expression of *DNMTs* (*DNMT1a*, *DNMT1b*, *DNMT1c* and *DNMT3*) was determined in females maintained in either long (LD) or short day (SD). To assess the level of gene expression, the wasps were collected after 10 days of exposure to differential photoperiods (as per the diapause experiment described in Chapter 2). 10 individual female wasps constituted one biological replicate from each specific photoperiodic condition, and four biological replicates were used from each photoperiod (LD or SD) for the assay. Total RNA (200 ng) was extracted from the whole body (using TRIzol reagent, Invitrogen) and reverse transcribed from each sample by first strand cDNA synthesis kit (Invitrogen) as described in Chapter 2. The synthesized cDNA was further diluted 10 times and used as a template for real-time PCR. The real-time PCR (qPCR) experiment (described in Chapter 2) was carried out with two technical replicates for each biological replicate, -RT (negative reverse transcription) controls and no template controls (NTC). To determine the relative expression level, *Aequorin* mRNA was used as the external reference gene.

Knock-down of DNMT's using RNAi

To sequentially knock-down the *DNMTs*, *RNAi* (RNA interference) approach was followed (as described in Chapter 2). Total RNA was extracted (TRIzol reagent) and reverse transcribed using *SuperScript II Reverse Transcriptase* (Invitrogen). The cDNA was used as a template to amplify the target genes such as *DNMT1a*, *DNMT1b*, *DNMT1c* and *DNMT3*. The PCR product was then used as a template to synthesize double-stranded RNA (dsRNA) by in-vitro transcription. The synthesized dsRNA was injected in the abdomen of yellow-pupae aged female wasps and incubated at 25°C for 4-5 days. In parallel, *GFP* (green fluorescent protein) dsRNA was also injected in separate set of adult female pupae, as a control. After 4-5 days the eclosed females (both from *DNMT* and *GFP-RNAi*) were collected and exposed to either LD or SD conditions at 18°C. After 5 days of incubation in differing photoperiods the surviving females were snap frozen in liquid nitrogen and used for validation of knock-down by real-time PCR. The progeny from the adult female injected wasps was later analyzed (20 days from the start of diapause experiment) for the photoperiodic response; level of diapause (Chapter 2).

To validate the down-regulation of the target genes, SYBR green real-time PCR approach was adopted as described in Chapter 2. Total RNA was extracted from both treated (injected with *Dnmt-dsRNA*) and control (injected with *GFP-dsRNA*) group of wasps, maintained in either LD or SD condition at 18°C for five days (6 independent RNA extractions; 3 independent RNA extractions from LD and SD reared wasps), and used as a template for reverse transcription. The synthesized cDNA pool was then used as a template to quantify the relative expression level of the target gene by standard curve qPCR method (Gilsbach *et al.*, 2006) and analyzed by ANOVA statistical test.

Monitoring changes in CpG methylation by real-time PCR

The effect of *DNMTs-RNAi* on genomic CpG methylation was determined by Methyl Quant, a real-time PCR technique (see Chapter 3 for details). The sodium bisulfite converted DNA was used for PCR testing each of the three highly methylated genes: *LOC100117390*, *LOC100117821*, *WDR36* with primers flanking the specific CpG of interest, designed using MethPrimer software (Li & Dahiya, 2002). The quantification data of the real-time PCR was carried using the R statistical software with the *qpcR* package (Ritz & Spiess, 2008a)(see Chapter 2 for details).

Blocking DNA methylation by 5-aza-2-deoxycytidine

In order to block methylation by 5-aza-2-deoxycytidine (5-aza-2dC), 10 µM non-lethal drug dosage (Amarasinghe *et al.*, 2014) was administrated to females, mixed in 20% sucrose solution. The control group was fed by 20% sucrose solution, without the drug. The wasps were individually placed in glass vials with 200 µl of either drug or sucrose solution and kept at 25°C at 12:12 (light: dark) condition for first two days of the experiment (figure 4.3). Both the drug and sucrose solution were coloured green by the use of a food colourant and few wasps were dissected after 24 and 48 hours to make sure that they were drinking the solution. Traces of green colour were observed in the gut region of the dissected wasps (both from drug treated and control group). Thereafter, on the third day, the surviving wasps were transferred into separate vials containing two host pupae along with the drug or sucrose solution and placed in LD (18:6) or SD (6:18) conditions at 18°C. Fresh solutions (10 µM drug solution or 20% sucrose) were added daily throughout the experiment during the light phase (0800-1400 hrs). After 5 days of LD (long) or SD (short-day), the hosts were recovered and placed in 18°C (DD) and the surviving wasps were snap frozen in liquid nitrogen and stored in -80°C. The level of

diapausing progeny was scored in the hosts after 20 days from the start of the diapause experiment.

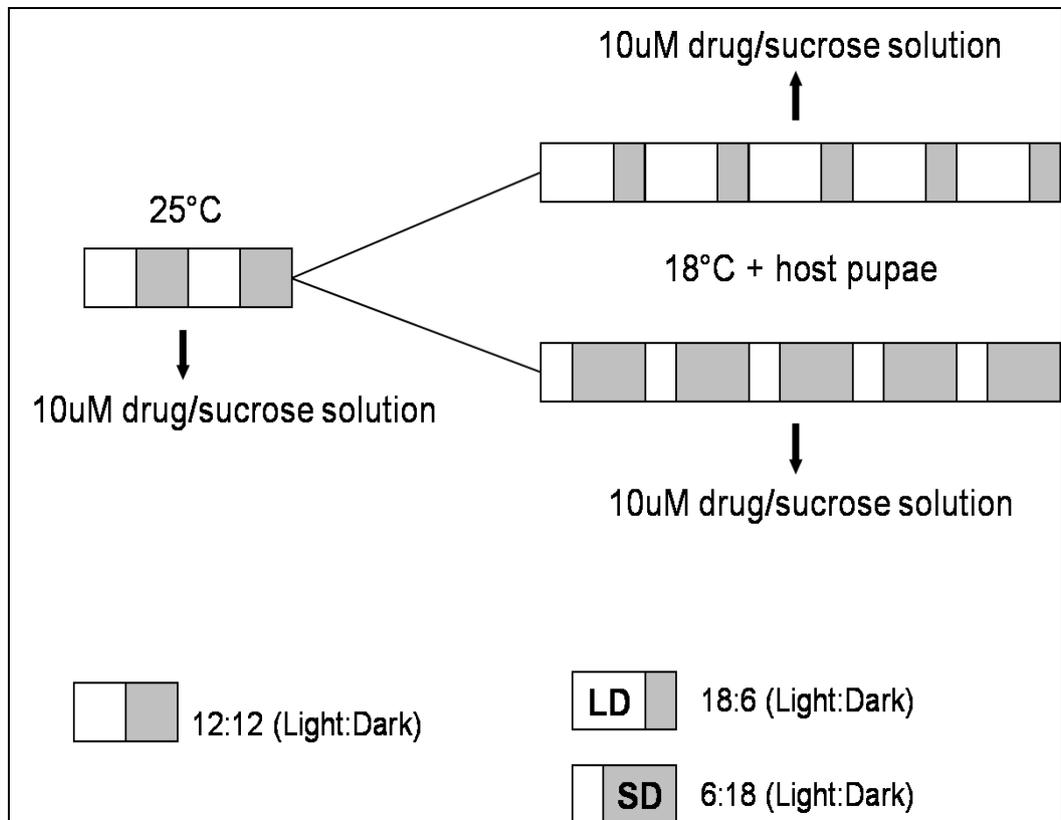


Figure 4.3: **Scheme of the drug experiment.**

To validate the loss of methylation, genomic DNA was extracted from 10 female wasps maintained in either LD or SD, fed with either drug or sucrose solution. Thereafter, the extracted DNA was subjected to bisulfite treatment as described above. The bisulfite converted DNA was pre-amplified for the highly methylated target genes; *LOC100121005*, *LOC100118618*, *LOC100679542* and *LOC100122664* (see chapter 3) with primers designed using MethPrimer (table 4.1). The purified PCR product from drug treated and control (sucrose-fed) conditions were adjusted to a uniform concentration of 10ng/μl for *LOC100118618*, *LOC100679542* and *LOC100122664* and 6ng/μl for *LOC100121005*. 10,000x dilutions were prepared for each test sample across all the four test genes. The diluted (10,000X) pre-amplified product was then assessed using Agilent 2X SYBR green mix by real-time PCR (see Chapter 2).

Table 4.1: List of primers used for analysing CpG methylation after drug treatment.

Gene (Primer used)	Forward primer (5'-3')	Reverse primer (5'-3')	Gene ID (NCBI)	Co-ordinates	size
LOC100121005 (Pre-amplification) (Selective amplification)	TGATTTTAGATTTTATAGTAGATAGTAAAA AGATAGTAAAAGTTTTTCGC	ACCCCTCAAAACATAAACTAATAC	341864961	13336775- 13337025 13336794- 13337025	251 232
LOC100118618 (Pre-amplification) (Selective amplification)	TGGTATTGTTGTTGTAGTATAGTTT TCGGTTTGCATTATGAAGGC	ACCAAAATAAAACACCTTACTCTC	341864959	29619968- 29620208 29620048- 29620208	241 161
LOC100679542 (Pre-amplification) (Selective-amplification)	GGTTATAATATTAGGGGAGTTAATTGG GGTTATAATATTAGGGGAGTTAATTGGC	AACTCCTAAAACCTCTCAAAAAAAA	341864958	c7427666- 7427536	131 132
LOC100122664 (Pre-amplification) (Selective-amplification)	AAGGGTTTAGGTATTTGGAAATAT AAGGGTTTAGGTATTTGGAAATCTC	TAAATCATCAATACTTTCTCTCAATCT AT	341864959	c2578027- 2577894	134 135

4.3 Results

Expression level of DNMTs

There was a substantial difference in the expression levels of *Dnmt1a*, *Dnmt1b*, *Dnmt1c* and *Dnmt3* in females (figure 4.4). Both *Dnmt1a* and *Dnmt3* were expressed at substantially higher levels than *Dnmt1b* and *Dnmt1c* ($Dnmt1b \ll Dnmt1c < Dnmt1a, Dnmt3$). However, between long and short photoperiods the DNMTs expression did not differ significantly (for example, *Dnmt1a* Student *t*-test, $df = 6, t = 0.9, p > 0.05$) between the photoperiods (i.e. LD Vs SD) in *Nasonia*. This clearly suggested that the transcript level of *DNMTs* does not differ per se between the photoperiod.

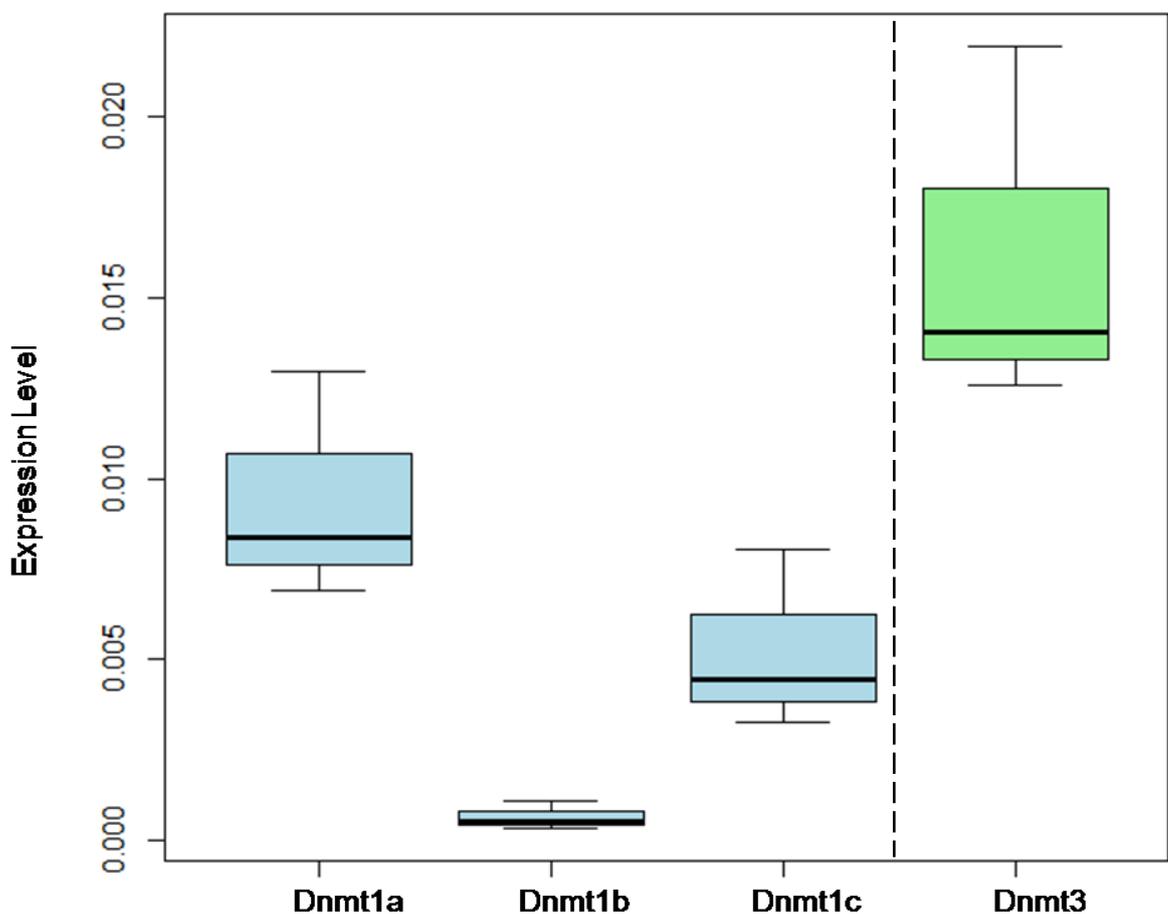


Figure 4.4: **Relative expression level of DNMTs in *Nasonia*.** The expression level of *DNMTs* (*DNMT1* paralogs represented by blue boxplots and *DNMT3* as green boxplots) when compared in the adult females ($n=80$) was found to be statistically different from each other (One-way ANOVA, $F_{3,12} = 3.83, p = 0.04$). The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum.

Effect of DNMTs knock-down on the photoperiodic response

The photoperiodic response (diapause %) arising from long (LD) and short days (SD) reared adult females was tested after each *DNMT5-RNAi* (such as *DNMT1a-RNAi*, treated) and control (*GFP-RNAi*) group using Two-way ANOVA. From the analysis, an overall significant difference for the level of diapause was observed between the two photoperiods (Two-way ANOVA, $F_{1,415} = 57.6$, $p < 0.001$), between the compared groups ; knock-down (*RNAi*) and control groups ($F_{4,415} = 8.3$, $p < 0.001$) with significant interaction between gene knock-down and photoperiod ($F_{4,415} = 2.8$, $p = 0.02$). Furthermore, on close inspection of the photoperiodic response between LD and SD in each group (treated or control), the normal photoperiodic diapause response was found to be disrupted in the progeny arising from the females injected with *DNMT1a* dsRNA (figure 4.5A). While the control females (injected with *GFP*-dsRNA) exhibited the strong photoperiodic response typical of *Nasonia*, with increased diapause in short days ($F_{1,20} = 14.48$, $p = 0.001$), the response of *Dnmt1a*-dsRNA injected females was similar in both long and short days ($F_{1,36} = 2.1$, $p = 0.16$), with a substantial increase in diapausing progeny of females maintained under long day conditions. Similarly, the knock-down of *Dnmt3* also showed aberrant photoperiodic response ($F_{1,40} = 3.9$, $p > 0.05$) in the progeny arising from females kept in either LD or SD conditions (figure 4.5D). This effect of *dsRNAi* was observed within 10 days from injection. For progeny laid 10-15 days after injection, diapause frequencies reverted to normal levels in the *Dnmt1a*-dsRNAi group and *Dnmt3*-dsRNAi group (i.e. low diapause incidence in long days, see Appendix section 8.4).

In contrast, the knockdown of *Dnmt1b* ($F_{1,118} = 6.6$, $p = 0.01$, Figure 4.5B) and *Dnmt1c* ($F_{1,139} = 20.09$, $p < 0.0001$, Figure 4.5C) generated normal photoperiodic response in the progeny. Therefore, these data clearly suggest that optimum transcript levels of *Dnmt1a* and *Dnmt3* are indispensable for seasonal timing in *Nasonia*. Nearly two fold reduction in the transcript levels of key DNA methyltransferases (*Dnmt1* or *Dnmt3*) led to disturbed level of diapause in the progeny arising from both the long and short day group.

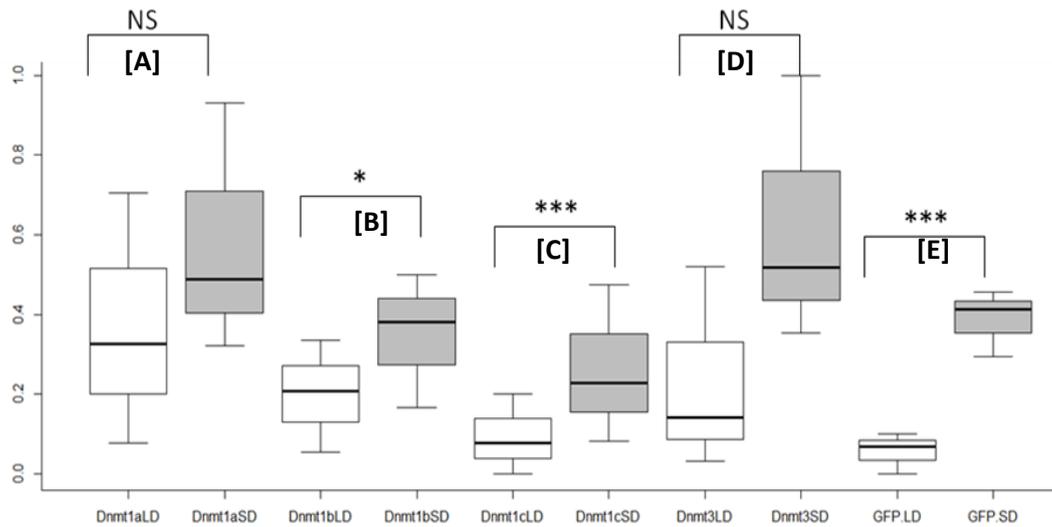


Figure 4.5: **Level of Diapause after DNMTs knock-down.** The progeny of females injected with *DNMT-dsRNA* exhibited variable photoperiodic response, as tested after 5 days at 18°C. In **A.** *Dnmt1a*-RNA and **D.** *Dnmt3*-RNA injected females, the difference in diapause levels between long (18 hr light, white box) and short day (6 hr light, grey box) is abolished (18-20 females in each group, progeny of each female collected from two hosts, average n=24 larvae per female). In **B.** *Dnmt1b*-RNAi, **C.** *Dnmt1c*-RNAi and **E.** *GFP*-RNAi normal photoperiodic response typical of *Nasonia*: low diapause incidence induced in LD compared with SD is seen. The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum. Significance Codes as per TukeyHSD (ANOVA): '***' p < 0.001, '**' p < 0.01, '*' p < 0.05, 'NS' p > 0.05.

The expression level of the knocked-down genes was verified in both the treated (*DNMTs-dsRNAi*) and the control (*GFP-dsRNAi*) group using the endogenous reference gene (*RpL-32*) or the exogenous reference (*Aequorin*). The knock-down of *Dnmt1a* resulted in two-fold reduction in the treated group (One-way ANOVA, $F_{1,10} = 5.0$, $p < 0.05$) against the control group. Similarly, the knock-down of *Dnmt1b* (One-way ANOVA, $F_{1,12} = 15.7$, $p = 0.001$), *Dnmt1c* (One-way ANOVA, $F_{1,12} = 12.8$, $p = 0.003$) and *Dnmt3* (One-way ANOVA, $F_{1,22} = 11.3$, $p = 0.002$) led to successful down-regulation of the respective target gene (figure 4.6).

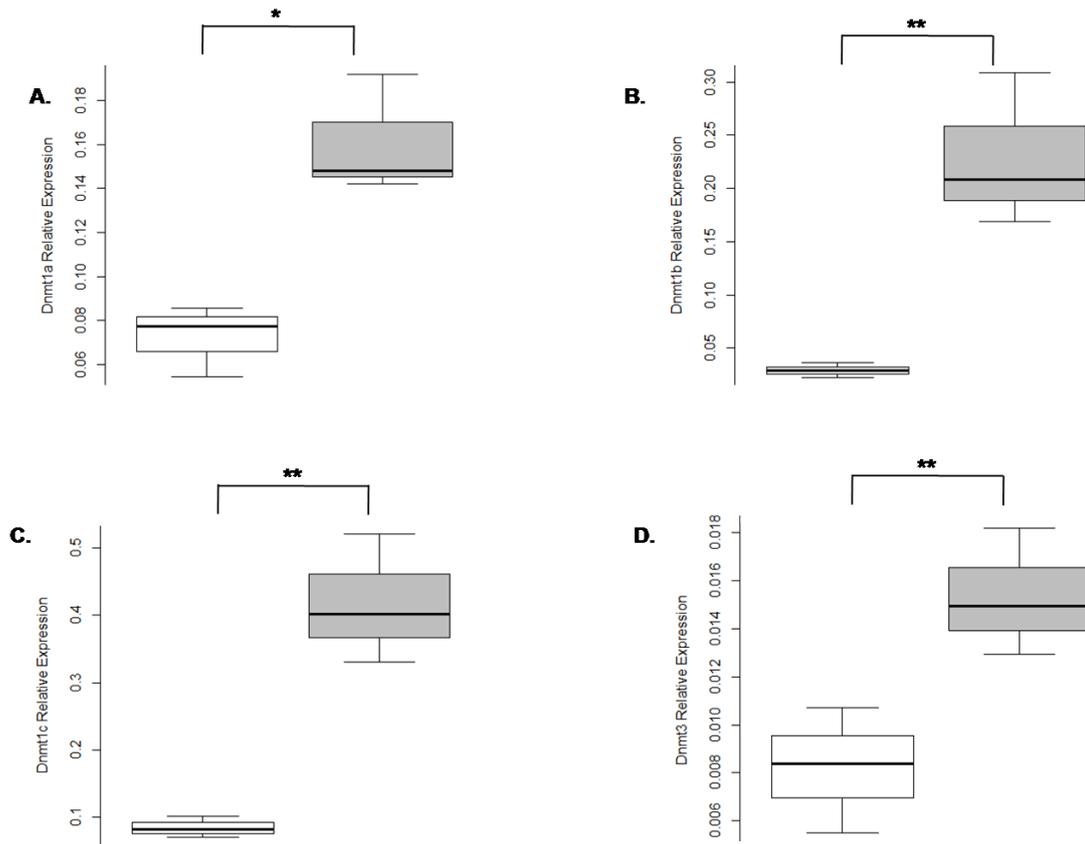


Figure 4.6: **Validation of knock-down of DNMTs.** The normalized expression level of **A. *Dnmt1a*** **B. *Dnmt1b*** **C. *Dnmt1c*** and **D. *Dnmt3*** were compared in the treated group ($n \geq 60$; white boxplots) to the control group ($n \geq 60$, grey box plots). The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum. Significance codes as per TukeyHSD (ANOVA): '****' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$, 'NS' $p > 0.05$.

Effect of DNMTs down-regulation on genomic CpG methylation

To verify the effect of the *RNAi* knockdown on the CpG methylation, qPCR (MethyQuant) was used. Genomic DNA was extracted from long (LD) and short-day (SD) exposed females following *Dnmt3* knock-down and control group. Figure 4.7(A) shows the significant reduction of methylation ($p < 0.05$, calculated by permutation) in each of the tested genes: *LOC100117390*, *WDR36* and *LOC10011782*. In each tested CpG site, the level of methylation was reduced in the treated group (*DNMTs-dsRNAi*) compared to the control group (*GFP-dsRNAi*).

Similarly, two (out of three) genes showed reduction ($p < 0.05$), in the level of CpG methylation after *Dnmt1a* knock-down (figure 4.7B). The CpG site present in the gene *WDR36*, however, showed higher methylation in the treated group after the *Dnmt1a* knock-down.

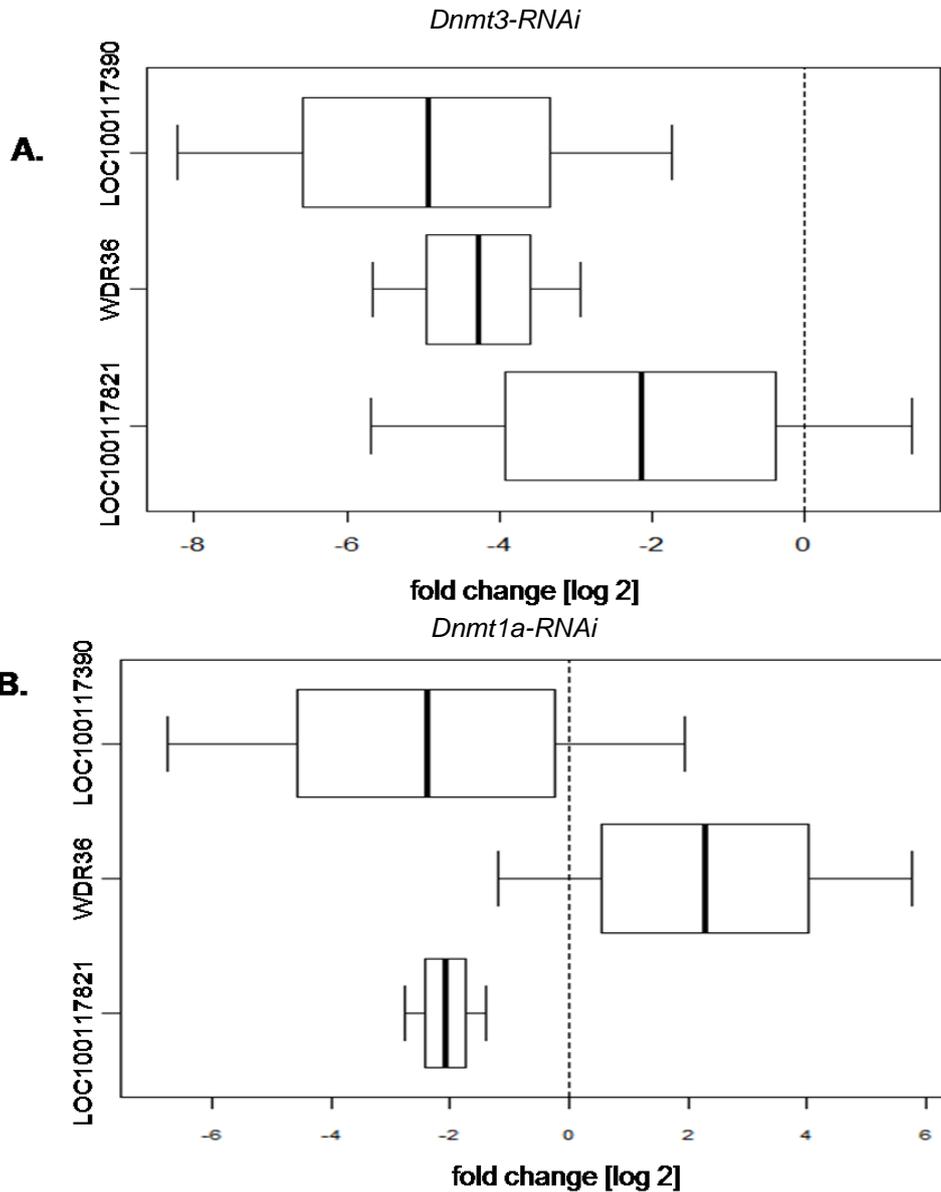


Figure 4.7: **Reduced methylation after knockdown** of **A.***Dnmt3* and **B.***Dnmt1a*. Boxplots depict the distribution of the replicates and the error estimated through permutation. The dotted line describes the value for which the ratio is one (i.e. significant change in methylation).

Effect of pharmacological blocking of methylation by 5-aza-2-deoxycytidine on photoperiodic response

In another set of experiments, 5-aza-2-deoxycytidine was used to inhibit the action of DNA methyltransferases and blocking of methylation. After administration of non-lethal dosage of the drug (10uM) to ovipositing females, the level of diapause was determined in the progeny. Fig. 4.8 shows a strong effect on the photoperiodic response. While control females showed the normal photoperiodic response (i.e. a substantial difference between long and short day females: (One-way ANOVA, $F_{1,79} = 17.4$, $p = 0$), wasps fed with 10 μM 5-aza-dC responded similarly to long and short days (One-way ANOVA, $F_{1,77} = 0.7$, $p = 0.41$). Intriguingly, the diapause in the progeny of wasps maintained under short day decreased (median drop from 35% to 25%), while increasing in long-day progeny (8% to 20%).

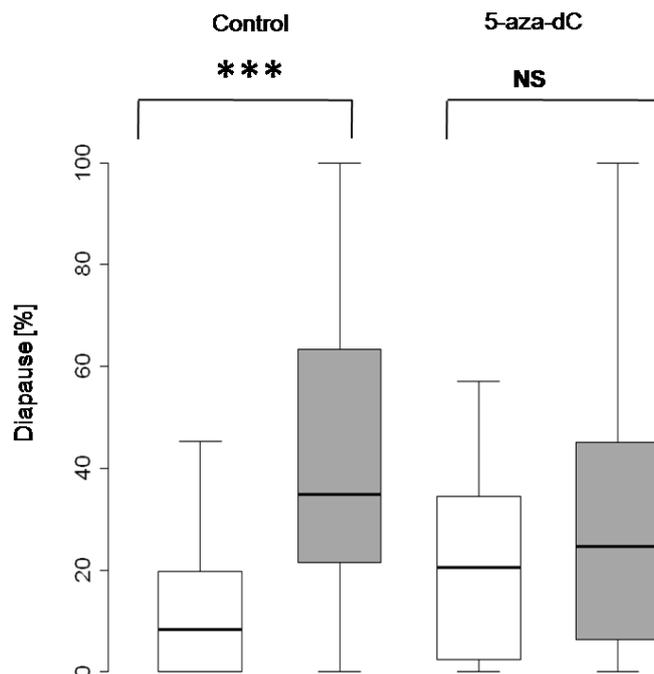


Figure 4.8: **Blocking DNA methylation by 5-aza-dC.** The progeny of control females (fed by sucrose solution) exhibited the normal photoperiodic response typical of *Nasonia*: low diapause incidence induced in long day (18 hr light, white box) compared with short day (6 hr light, grey box), as tested after 5 days at 18°C. In females fed with 10 μM 5-aza-dC, the difference in diapause levels between long and short day was abolished (35-40 females in each group, progeny of each female collected from two hosts, $n = 27-30$ larvae per female). The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum. Significance codes as per TukeyHSD (ANOVA): '****' $p < 0.001$, '***' $p < 0.01$, '**' $p < 0.05$, 'NS' $p > 0.05$.

As expected, the level of CpG methylation declined considerably after the 5-aza-2-deoxycytidine treatment, as tested at specific CpG sites (which normally are highly methylated) in four genes: *LOC100121005*, *LOC100118618*, *LOC100679542* and *LOC100122664*. Here, the critical qPCR cycle (CpD, a measure of template DNA) was used in omnibus ANOVA, that indicated a significant effect of the drug treatment ($F_{1,8} = 9.45$, $p = 0.015$, see figure 4.9). All the tested genes showed high CpD value in comparison to the control group, indicating lower proportion of methylated gene copies in the drug treated group (C/T transition after bisulfite treatment) and therefore confirmed the reduction in the level of methylation.

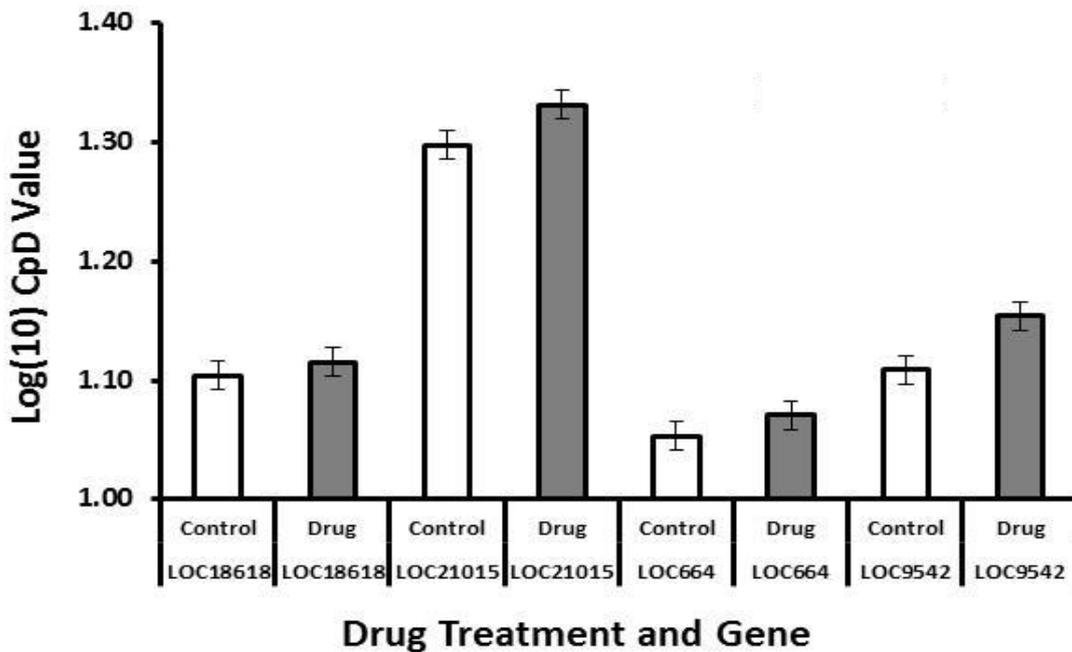


Figure 4.9: **Reduction in the level of CpG methylation after drug treatment.** The log (10) CpD values for the drug treated group (grey vertical plots) for *LOC100121005*, *LOC100118618*, *LOC100679542* and *LOC100122664* showed significant increase compared to the control group (white plots) as analysed by ANOVA mixed effects model. 5-aza-2-deoxycytidine had a significant effect on level of methylation ($F_{1,8} = 9.45$, $p = 0.015$). Gene model also showed a significant effect ($F_{3,8} = 157.98$, $p < 0.001$). There was no interaction effect observed between gene and treatment ($F_{3,8} = 0.74$, $p = 0.56$).

4.4 Discussion

In *Nasonia*, DNA methylation genes, *Dnmt1* (*Dnmt1a*, *Dnmt1b*, *Dnmt1c*), *Dnmt2* and *Dnmt3* DNA methyltransferases are present. The individual knock-down of the highly expressed *Dnmt1a* and *Dnmt3* in females generated an aberrant photoperiodic response underpinning the importance of DNA methylation in photoperiodism (figure 4.5). The observed increase in diapause incidence under long day conditions following the *RNAi* manipulation suggests that under normal conditions, CpG methylation of key photoresponse genes prevents diapause development when days are long. However, the more pervasive pharmacological reduction of methylation by 5-aza-dC also led to disturb photoperiodic response, but not in quite the same way as the *RNAi* did (Fig. 4.8). This may be due to the fact that the drug has a broad impact on DNA methylation (in contrast to *DNMTs*). As such, the differences in response in the two experiments may reflect multiple pathways and targets of DNA methylation, with various effects on diapause, which are likely to be locus-, and tissue-specific.

Although the methylation level was found to differ between the adult female wasps (Chapter 3) exposed to long (LD) and short day (SD), intriguingly the transcript levels of the *DNMTs* remained the same. This indicates that the context dependent differential methylation could result either from the recruitment of the active methyltransferases that are regulated post-transcriptionally or differ in their expression between various tissues that gets masked during the whole body analysis.

Recently, a study in hamsters (Stevenson & Prendergast, 2013) showed that exposure to SD or melatonin lead to lower *DNMTs* expression, causing hypomethylation of the promoter region of *dio3*, an important gene involved in the flow of photoperiodic information to the brain and neuro-endocrine reproductive system. Hypomethylation of *dio3* promoter region caused by *Dnmt3b* reduced activity, leads to increase in the gene transcription which influence the release of gonadotropin- releasing hormone (GnRH). Hence, *Dnmt3b* acts as the key methyltransferase that mediates the photoperiodic timing and seasonal response in reversible fashion.

In vertebrates, *Dnmt1* is considered as a maintenance methyltransferase, and is responsible for placing existing methyl marks on the daughter strands of replicating DNA molecules, while *Dnmt3* is regarded as the *de novo* DNA methyltransferase that confers methyl marks on CpGs in response to physiological, behavioural and environmental influences (Law & Jacobsen, 2010). In invertebrates however, the role of

DNMTs is yet not well defined. Thus, there exists a possibility that Dnmt1 and Dnmt3 have overlapping roles in *Nasonia*; down-regulation of both *Dnmt1a* and *Dnmt3* individually, led to loss of CpG methylation (Figure 4.7) that resulted in disturbed photoperiodic response.

Interestingly, the *Dnmt1a* knock-down, was not accompanied by reduction in CpG methylation in the gene *WDR36*. This particular site possibly could be substrate for other DNMTs, responsible for CpG methylation in *Nasonia*, which may indicate the possibility of the compensatory effect of the *Dnmt1a* knock-down by recruiting other DNMTs such as *Dnmt3* to methylate the CpG site.

Indeed, accumulating evidence (in mammalian systems) suggest that the functional division between the DNMTs may not be that clear. Dnmt1 which was considered as the key maintenance methyltransferase has been also shown to place novel methyl marks under in-vitro conditions (Fatemi *et al.*, 2001). On the other hand, Dnmt3 has been credited as maintenance methylase in ES (embryonic-stem) cells (Liang *et al.*, 2002) due to its ability to restore methylation, missed by Dnmt1 during replication. A parallel study in cancerous cells has even proposed combinatorial action of Dnmt1 and Dnmt3 which leads to establishment and maintenance of DNA methyl marks during cell replication (Robert *et al.*, 2003; Hsieh, 2005). According to one of the proposed models based on this combined action (figure 4.10), Dnmt3 initiates methylation on single DNA strand at specific CpG sites (filled circles) and Dnmt1 (hemi-methylated DNA preferred) methylates the complementary strand (open circles) at these hemimethylated sites soon after DNA replication. The hemimethylated DNA could also trigger Dnmt1 to methylate the former unmethylated CpG sites (open circles indicated with '?'). In addition to this, the pre-existing CpG methyl marks can instigate *de novo* methylation activity of Dnmt1 by cis as well as transmechanism (Christman *et al.*, 1995; Tollefsbol & Hutchison, 1995; Carotti *et al.*, 1998) suggesting that Dnmt1 can be stimulated by both hemimethylated and symmetrically methylated DNA.

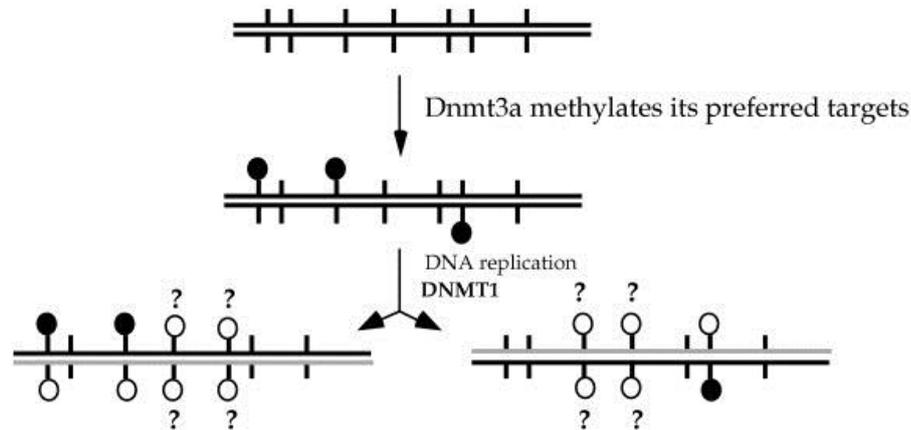


Figure 4.10: **Combined action of Dnmt1a and Dnmt3 in maintenance and establishment of methyl marks in the genome.** Model adapted from Hsieh 2005.

The finding that DNA methylation is functional in *Nasonia*, with a clear and easy to measure phenotypic effect (photoperiodic response), as evident both after the knock-down of key methyltransferases such as *Dnmt1a* and *Dnmt3* and drug manipulation, provide a new insect model system that could be used in the future, for dissecting the individual role of each of the components of the DNA methylation system. As a model system, *Nasonia* offers various benefits, similar to the *Drosophila*. It is therefore likely that further research into epigenetic function in the wasp would provide a substantial contribution, in the same way that research in the fruitfly (which unfortunately lacks CpG methylation) advanced our understanding in broad range of fields.

CHAPTER 5

TESTING FOR THE PRESENCE OF 5-HYDROXYMETHYLCYTOSINE

(5-hmC) IN *NASONIA*

5.1 Introduction

5-hydroxymethylcytosine (5-hmC) is produced by the oxidation of 5-methyl cytosine (5mC), and may potentially have dual roles, by acting both as a stable DNA base, and as an intermediate in the DNA de-methylation pathway (Pfeifer *et al.*, 2013). Recent studies have shown that in mammals, 5hmC levels fluctuate substantially across different cell types and tissues, being highest in the brain, particularly in neurons (Kriaucionis & Heintz, 2009; Globisch *et al.*, 2010; Jin *et al.*, 2011). In these studies, the conversion of 5mC to 5hmC has been shown to lower the levels of 5mC at a specific nucleotide position. More generally, this conversion of 5mC into 5hmC has been shown to propagate DNA demethylation (Zhang *et al.*, 2011; Guo *et al.*, 2011a), either through passive or active erasure of the modified base and/or removal of the methyl group from cytosine in DNA (Song & He, 2013)(figure 5.1). The elevated levels of 5hmC could also lead to binding of specific proteins such as UHRF1 (Frauer *et al.*, 2011), MBD3 (Yildirim *et al.*, 2011), MeCP2 (Mellen *et al.*, 2012). However, the biological significance of this interaction is not clearly understood. As a consequence, 5hmC might act as an epigenetic module per se, with its own unique biochemical coding properties (Jin *et al.*, 2011).

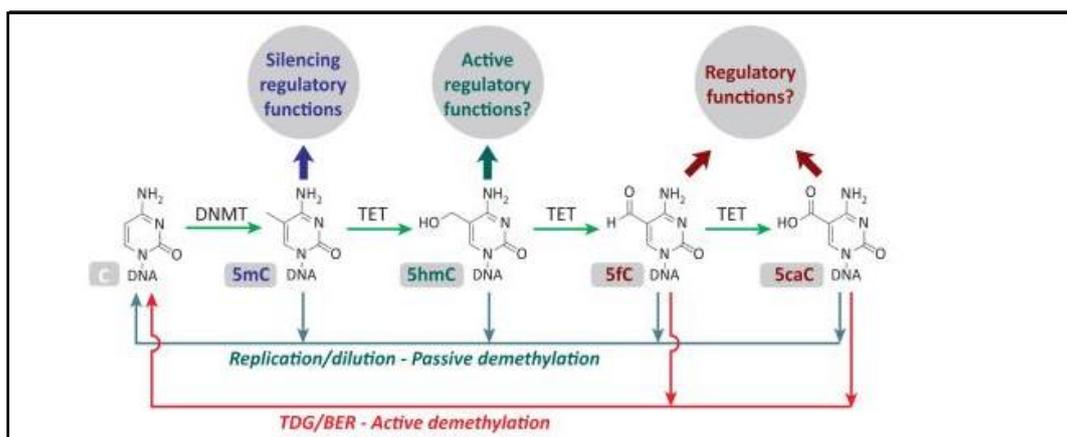


Figure 5.1: **DNA methylation and demethylation in mammals.** DNA methyltransferases (DNMTs) act on cytosine molecules and generate 5mC (5-methyl cytosine), whereas demethylation can be accomplished through either an active or a passive pathway. During the passive mechanism, 5mC gets converted back to cytosine through replication-dependent dilution, due to absence of maintenance methyltransferase such as Dnmt1 (represented by the green arrows). In the active mechanism, 5mC gets actively oxidized by TET proteins to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (red arrows). These intermediates are removed by passive demethylation or active mechanism via thymine DNA glycosylase (TDG) and base excision repair (BER). (Adapted from Song, 2013)

During the passive DNA demethylation process, conversion of 5mC into 5hmC leads to inhibition of DNA maintenance DNMT1 activity that recognises pre-existing methylation patterns from hemimethylated DNA strands and copies them on to the daughter strands during DNA replication (Valinluck & Sowers, 2007; Hashimoto *et al.*, 2012). The active demethylation process includes stepwise oxidation of 5hmC by TET (Ten-eleven translocation elements) proteins, generating 5- formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito *et al.*, 2011; He *et al.*, 2011b), that are removed by different base excision repair enzymes, including Thymine DNA glycosylase (TDG) (He *et al.*, 2011b; Cortellino *et al.*, 2011) and Single- strand selective Monofunctional Uracil DNA Glycosylase (SMUG1) (Boorstein *et al.*, 2001).

In metazoa, the oxidation state of 5mC (5-hydroxymethylcytosine; 5hmC)) was recently found to be catalysed by ten-eleven translocation (TET) enzyme family (Tahiliani *et al.*, 2009; Ito *et al.*, 2010). In vertebrates, 5hmC is produced from 5mC in an enzymatic process involving three 5mC oxidases namely TET1, TET2, and TET3. The metazoan TET proteins possess a unique conserved cysteine-rich region. Vertebrate TET1 and TET3 (and orthologs) also possess a CXXC domain and binuclear Zn-chelating domain that helps to distinguish between methylated and unmethylated DNA (Allen *et al.*, 2006). Based on the homology search the TET orthologs have been identified across various organisms. In *Nasonia vitripennis*, the putative protein, methylcytosinedioxygenase TET2, shares close similarity with mammalian TET3 with conservation of TET domain ([pf:Tet JBP](#)) and CXXC zinc finger motif ([pf:zif-CXXC](#)) (KEGG database ID: nvi100114438) (figure 5.2).

The discovery of 5hmC

5hmC was first identified in certain bacteriophages (Hershey *et al.*, 1953) and later reported in mammalian tissues (Penn *et al.*, 1972). However, the level of 5hmC reported by Penn *et al.* in mammalian tissues seemed too high and could not be confirmed in subsequent studies (Kothari & Shankar, 1976). Nevertheless, in few recent studies the occurrence of 5hmC in mammalian cells has been shown unambiguously (Kriaucionis & Heintz, 2009; Tahiliani *et al.*, 2009).

In mice, the primordial germ cell (PGC) undergoes epigenetic changes sequentially involving genome-wide DNA demethylation to reset the epigenome for totipotency. It has been successfully demonstrated that removal of CpG methylation

(5mC) in PGCs occurs via conversion to 5-hydroxymethylcytosine (5hmC), driven by high levels of TET1 and TET2 (Hackett *et al.*, 2013). An important immunofluorescent study further showed that TET1 and TET2 were expressed at significantly higher levels in the nucleus of PGCs than neighbouring somatic cells. This clearly suggested that erasure of 5mC (i.e. demethylation) in PGCs takes place through conversion to 5-hydroxymethylcytosine (5hmC) by TET1/TET2 (Ancelin *et al.*, 2006; Hajkova *et al.*, 2008). The global DNA demethylation was found to be imperative for germ cells to establish allele-specific DNA methylation pattern that is required for genomic imprinting (Lee *et al.*, 2002). Compared with the enzymes and mechanisms involved in DNA methylation that have been well studied, the research of DNA demethylation is still in its infancy (Wu & Zhang, 2010).

To date, there is no evidence for 5hmC in insects. However, in *Tribolium castaneum* a wave of demethylation was shown to occur in larvae, pupae and adults (Felicciello *et al.*, 2013). The cycling of DNA methylation ranged from strong overall cytosine methylation in embryos to a weak DNA methylation in the following developmental stages. The authors suggested that the most probable route of DNA demethylation would involve the formation of 5-hydroxymethylcytosine by the action of TET enzyme in other developmental stages (Felicciello *et al.*, 2013), but this has not been demonstrated.

Thus, *Nasonia vitripennis* presented an excellent opportunity to test for the prevalence of 5hmC in the genome and its involvement in DNA demethylation. In the previous chapters, data were presented that show that the difference in the level of CpG methylation plays a crucial role in seasonal timing in *Nasonia*. The differential methylation may represent the net outcome of both methylation and demethylation of DNA. Here, I investigated the prevalence of 5hmC using chemical tagging technology at a single nucleotide resolution (Song *et al.*, 2011). In addition, I have used RNA interference to knock-down TET, and examined the effect on the photoperiodic response.

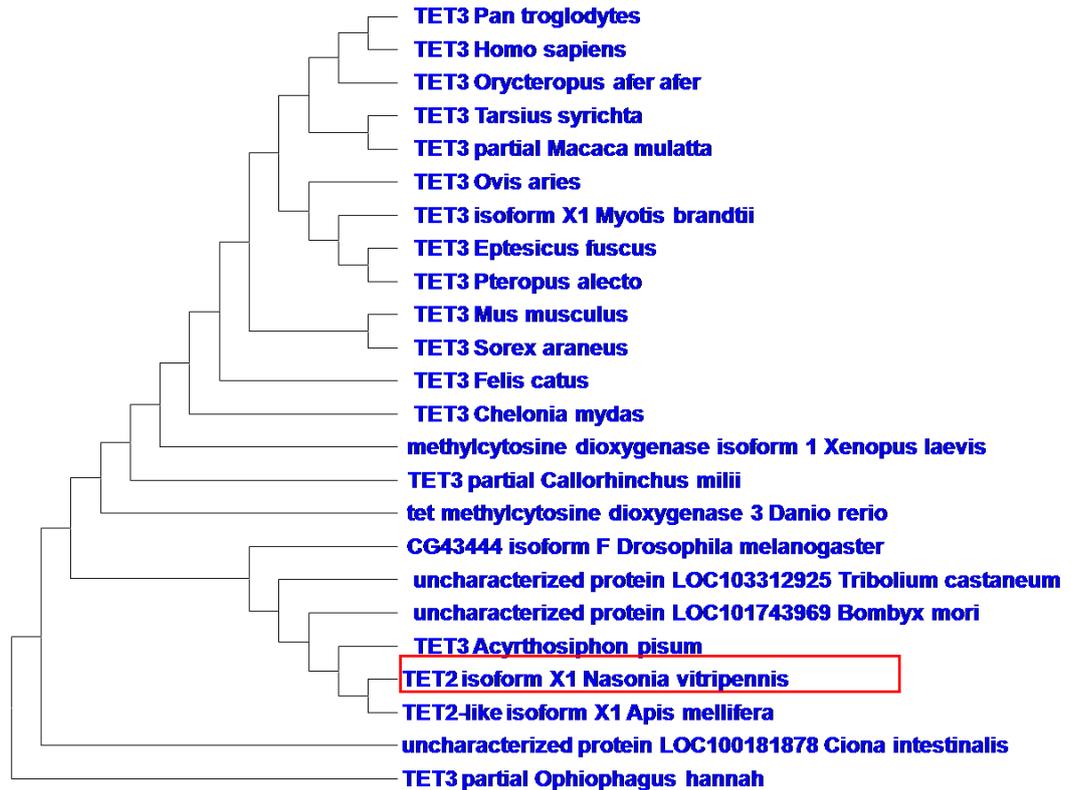


Figure 5.2: **Phylogenetic Tree based on *TET* orthologs.** The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The analysis involved 24 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 235 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.*, 2013).

5.2 Material and methods

Analysis of 5hmC mark by chemical tagging

The analysis of 5hmC across the *Nasonia* genome was done using a chemical tagging technology (Hydroxymethyl Collector kit, Active Motif, cat. no. 55013) that can distinguish between 5mC and 5hmC in complex genomes, and enrich 5hmC containing DNA at single base resolution level (Song *et al.*, 2011). In this procedure, β -glucosyltransferase (β -GT) was employed to transfer a chemically modified glucose moiety (UDP-6-N₃-Glu) onto 5hmC residues, yielding β -glucosyl-5-Hydroxymethyl-cytosine (Figure 5.3A) double stranded DNA. An azide group was then transferred to the 5hmC using the chemically modified glucose moiety (UDP-Glu) (Figure 5.3B). To the azide group, a biotin conjugate was chemically introduced that led to covalent coupling with the streptavidin magnetic beads, extensively used for enrichment of DNA fragments containing the modified 5-hydroxymethyl cytosine fragments. The strong binding properties of the biotin-streptavidin promoted high stringency, leading to accurate enrichment of the modified 5hmC residues.

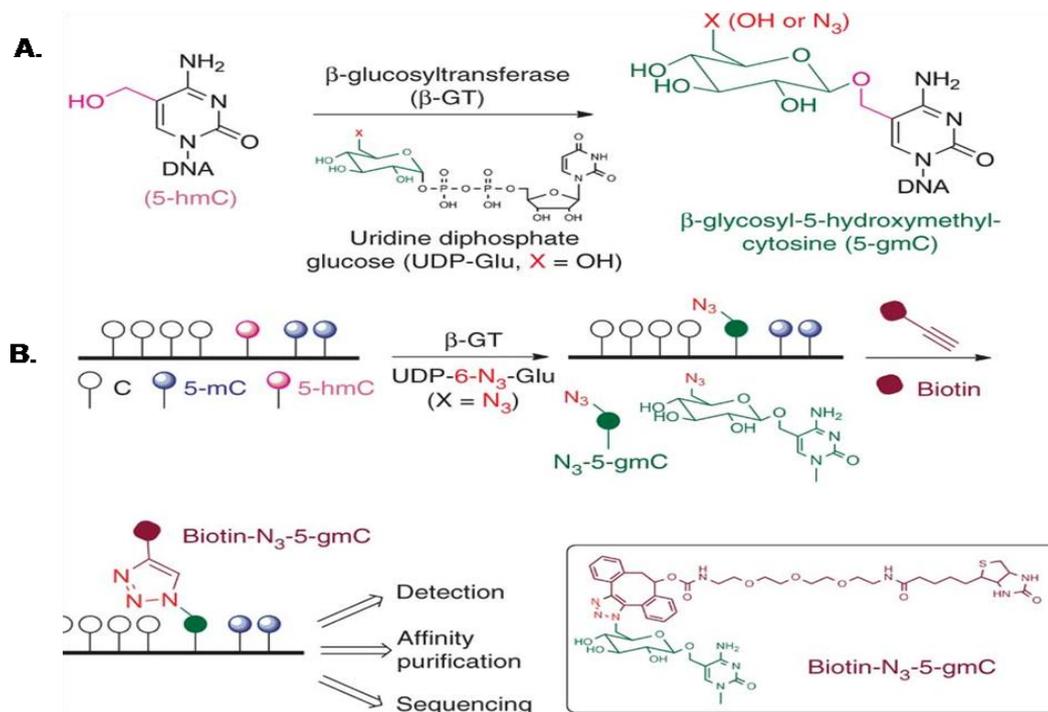


Figure 5.3: **Chemical labeling of 5hmC for subsequent detection.** In the figure A) addition of glucose moiety to the 5hmC molecule results in the formation of β -glucosyl-5hmC (5-gmc) and B)- tagging of 5-gmc molecule by biotin for subsequent enrichment by streptavidin beads. (Adapted from Song *et al.* 2011) (see text for details).

Nasonia vitripennis (Asym C strain) was raised in 25°C and fed with commercially available blowfly pupae as explained in Chapter 2. The newly eclosed female wasps were snap frozen in liquid nitrogen and collected for the assay. Initially, genomic DNA was extracted separately from each biological sample, composed of 10 adult female wasps, using DNeasy Blood and Tissue kit (Qiagen) as described in Chapter 3. The extracted DNA was quantified on Qubit fluorometer using dsDNA BR assay kit. From each sample, 5µg genomic DNA was fragmented to a size range of 200-500 bases by sonicator (10 cycles) and vacuum dried. The fragmented genomic DNA was spiked with mouse liver DNA (300ng) which was used as a positive control in the assay. 1% of total input DNA (50ng) was stored separately at 4°C for future use in the enrichment analysis. The remaining DNA was subjected to chemical tagging assay developed by the He lab at the University of Chicago (known as Chemical capture or hMeSeal) (Song *et al.*, 2011) and marketed as Hydroxymethyl Collector Kit by Active Motif. The first glucosylation step of the assay was done using 30 units of T4 phage β-Glucosyltransferase (New England Biolabs, Cat no. M0357L) followed by 2hr incubation at 37°C (Thomson *et al.*, 2013). After the addition of the glucose moiety, biotinylation reaction was performed as per the kit protocol by 1 hr further incubation at 37°C. Subsequently, the fragments containing chemically tagged 5hmC residues were captured by streptavidin beads and the enriched DNA was finally purified using the PCR purification kit (MinElute PCR purification kit, Qiagen, cat. no. 28004) for downstream analysis.

Downstream analysis of the enriched fraction

The immunoprecipitated fractions were assessed for 5-hydroxymethyl cytosine DNA enrichment using SYBR Green real-time PCR as explained in Chapter 2. In theory, prevalence of 5hmC residues (if present) are highly likely to be concentrated at the differential CpG methylated regions due to dynamic transition (5mC-5hmC) as part of TET assisted DNA demethylation pathway. Based on this, the amount of three methylated *Nasoniagenes* that possess multiple methylated CpG sites, *Misshapen* (*msn*), *UbiA-1* and *Cact-2* were compared between the enriched and the corresponding input fractions by qPCR. For instance, if high amount of test gene (*msn*) copies were found to be present in the 5hmC enriched fraction compared to the input, then this would indicate the presence of 5hmC in that gene. In addition, the mouse *Gapdh* and *tex19.1* were used as negative (due to lack of 5hmC residues) and positive controls respectively. The qPCR

was done using two-step cycling conditions with initial denaturation at 95°C for 15 minutes, followed by 45 cycles of 15 seconds at 94°C and 45 seconds at 60°C were performed. The melting curve analysis and quality checks on the Opticon monitor platform were done as described in Chapter 2. The comparative Ct (Cycle threshold) values indicating the number of test gene copies in the enriched and input fractions were recorded from the qPCR assay. With the Ct values, the gene recovery (%) corresponding to 1% input DNA was calculated in the enriched fraction using the below equation, where Ct_{input} = Ct of the test gene in input fraction and Ct_{enrich} = Ct of the test gene in the enriched fraction (Butcher & Beck, 2010).

$$\text{Recovery (\%)} = 2^{[(Ct_{input} - \log_2(100)) - Ct_{enrich}]} * 100$$

Table 5.1: List of specific primers used for real-time PCR for 5hmC detection.

Gene	Forward (5'-3')	Reverse (5'-3')	Gene ID (NCBI)	Co-ordinates	Size (bp)
<i>nvMsn</i>	CATTTATCCGCGATCAAC CT	AGCTGGCTCATCATCTTCGT	341864961	17786041-17786183	143
<i>nvUbiA-1</i>	GACCGATATCCGTGCTGT TC	TCTCGAGGTCCCTGGTATTG	341864959	c29619956-29620087	132
<i>nvCact2</i>	GGACAAACAGCAGGTGG ATT	GCAAGCCTTAGCAGGTTICAG	341864958	c2065624-2065771	148
<i>Gapdh</i>	CCACTCCCCTTCCCAGTT TC	CCTATAAATACGGACTGCAGC			
<i>Tex 19.1</i>	GGGAGATATGTAAATGA GCTGG	CATCCTTACCTCCCTGACTGA G			

TET knock-down

The knock-down of TET was performed by using *RNAi* as described in Chapter 2. In order to synthesize *TET-dsRNA*, total RNA (1 ug) was used as a template for first strand cDNA synthesis using random primers (Invitrogen). The synthesized cDNA was used to amplify the specific region (510 bp) of the gene of interest (*TET-2*) by PCR approach. Finally, the purified PCR product was used as a template for in-vitro transcription and the synthesized dsRNA was assessed both quantitatively (Nanodrop) and qualitatively (1% agarose gel) before use. The synthesized *dsTET-2-RNA* (and *dsGFP-RNA* as control) was injected in the abdomen of female wasp pupae at appropriate stage and incubated for 4-5 days at 25°C. The surviving eclosed adult female wasps were then exposed to either LD or SD at 18°C and followed as per the diapause experiment regime (Chapter 2 for details). The level of diapausing progeny arising from *dsTET-2-RNA* (treated) and *dsGFP-RNA* (control) adult females were assessed and compared after 5 days of exposure to long and short photoperiods.

5.3 Results:

5hmC enrichment in *Nasonia vitripennis*

Fig. 5.4 shows the results of the 5hmC chemical capture assay. The enrichment of *Gapdh* (negative control) was found to be at minimum across all the three biological replicates indicating a negligible or very low % 5hmC recovery as expected. In contrast, the *tex19.1* gene (positive control) was highly enriched, indicating the robustness and reliability of the assay for 5hmC enrichment. The *Nasonia vitripennis* genes when compared to *Gapdh* using Dunnett's test; *msn* (Dunnett's test, estimate = 0.2, $t = 0.7$, $p > 0.05$) and *UbiA-1* (Dunnett's test, estimate = -0.04, $t = -0.2$, $p > 0.05$), showed no significant difference. But showed a substantially higher recovery for *Cact2* (Dunnett's test, estimate = 1.2, $t = 4.4$, $p < 0.05$) in all the enriched fractions.

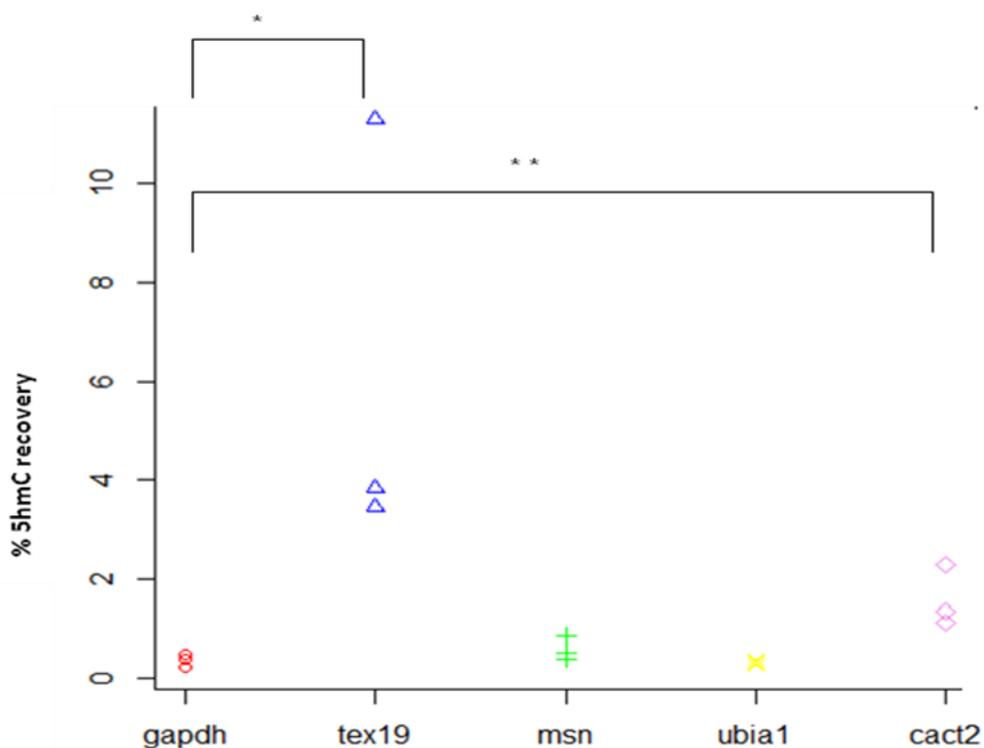


Figure 5.4: **5hmC enrichment of the target genes.** The quantification levels (% 5hmC recovery) of the tested genes *msn*, *ubia1*, *cact2*, along with *Gapdh*, *tex19* (controls), was assessed in the 5hmC enriched fractions and plotted as indicated by different colours (see text for details). The significant difference between the mean quantification levels was computed using Dunnett's test (ANOVA) at 95% confidence intervals. Significance Codes as per Dunnett's test (ANOVA): '****' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$, 'NS' $p > 0.05$.

Effect of *TET-2* knockdown in *Nasonia vitripennis*

Fig 5.5 shows the photoperiodic response of the progeny of the females with knock-down of *TET-2* (treated) and the control (*GFP-RNAi*) groups. The photoperiodic response observed in the progeny arising from the females injected with *ds-TET2-RNA*, resembled closely with the control group as determined using general linear model based logit model for binomial distribution. As a result, significant differences were noticed between the level of diapause in LD and SD conditions for both treated and the control group ($\chi^2= 110.7, p < 0.001$).

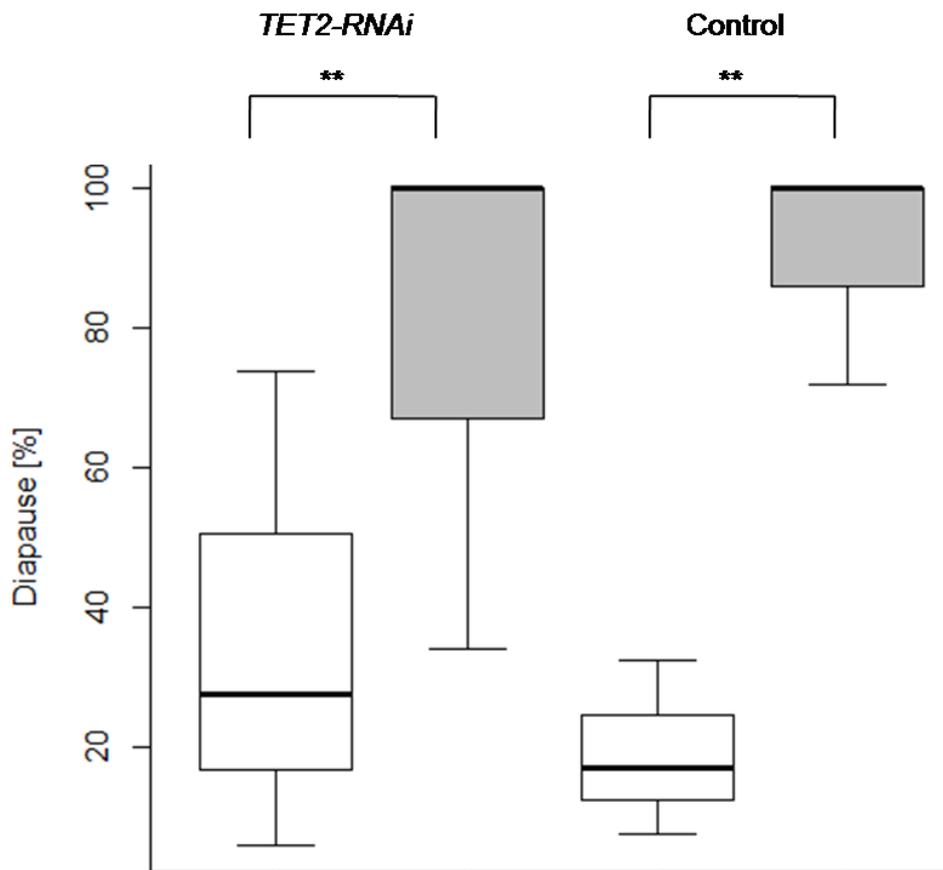


Figure 5.5: **Level of Diapause after *TET-2* knock-down.** The progeny of control females (*GFP*-dsRNA injected) exhibited the normal photoperiodic response typical of *Nasonia*: low diapause incidence induced in long day (18 hr light, white box) compared with short day(6 hr light, grey box), as tested after 5 days at 18°C. The difference in the level of diapause between long and short day in the treated group resembled the control group (20-22 females in each group, progeny of each female collected from two hosts, average n=16 larvae per female). The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum.

However, the magnitude of difference in the level of diapause was reduced after the *TET2-RNAi* in comparison to the control group (figure 5.5). In the treated group, the incidence of diapause considerably increased for the LD and decreased for the SD group in comparison with the control. Moreover, the interaction between the gene and photoperiod for the level of diapause was highly significant as tested by general linear model using analysis of deviance (figure 5.6). Hence, the interaction analysis indicated towards mild effect of *TET2* knockdown on the photoperiodic response.

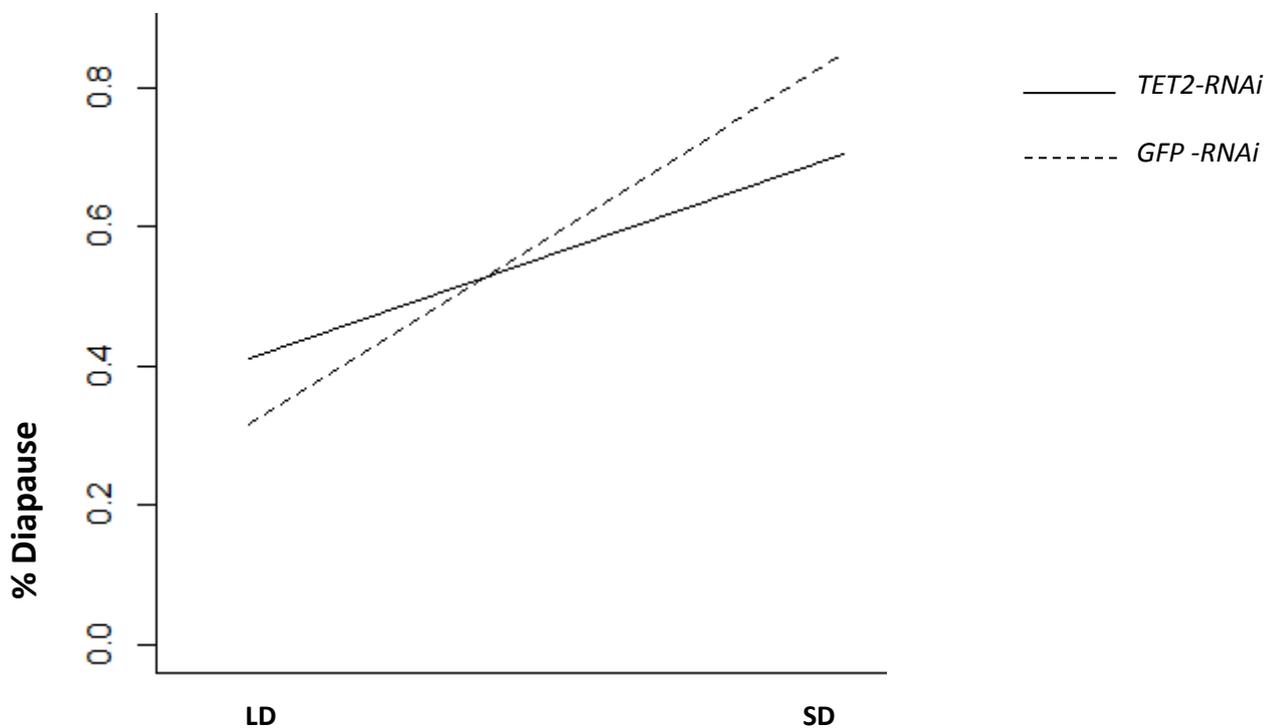


Figure 5.6: **Interaction plot diapause level after *TET-2* knockdown.** The interaction between the gene (*GFP*, *TET*) and photoperiod (LD, SD) was found to be highly significant as analysed by general linear based logit model for binomial distribution using analysis of deviance ($\chi^2=16.6, p<0.001$).

5.4 Discussion

This study suggests that 5hmC is present in *Nasonia*, and therefore could potentially act as an intermediate in DNA demethylation in *Nasonia* through the action of TET enzymes. In addition, the modified base could act as an epigenetic mark per se by dynamically modulating the DNA methylation (Branco *et al.*, 2012). Demethylation in the genome may be driven by either an active or passive mechanisms (see Introduction on page 72), and these possible pathways should be further tested in *Nasonia*.

The finding that 5hmC is present in *Nasonia* constitute a technical challenge due to the fact that bisulfite sequencing based approach was used for mapping methylation (e.g. Chapter 3). The problem is that this approach cannot distinguish between 5hmC and 5mC marks in the genome and could, therefore, potentially generate overlapping signals (Yu *et al.*, 2012) and inaccurate estimation of level of 5mC in the genome. However, the three genes tested here (Fig. 5.4) were found to be hypermethylated under long day (Chapter3), though low level of 5hmC was found only in one of the genes (*cact2*). This strongly suggests that the differential methylation observed between the photoperiods was contributed mostly by the 5mC and not 5hmC. Nevertheless, 5hmC does seem to be present in *Nasonia* and might have functional importance.

One possible role for 5hmC might be in driving demethylation, since knock-down of *TET*, (mediating DNA demethylation via 5hmC in mammals, see Introduction), led a significant effect on the photoperiodic response (Fig. 5.6). Although the level of diapause did differ between the photoperiods, the magnitude of this difference was reduced as compared to the control group. Noticeably, the systemic low levels of TET enzyme could instigate imbalance in the spatial distribution of methylated CpG marks and lead to the observed (weakened) photoperiodic response. To gain a better understanding of the role of TET and DNA demethylation via 5hmC in *Nasonia*, a more detailed and extensive mapping of 5hmC should be carried, including tissue specific analysis. Previous studies have shown that a wave of demethylation, essential for epigenetic reprogramming takes place in a highly orchestrated fashion at the zygotic stage in mammals (Wossidlo *et al.*, 2011; Inoue & Zhang, 2011; Iqbal *et al.*, 2011), through the conversion of 5mC to 5hmC, resulting in subsequent erasure via replication-dependent passive dilution. Thus, inspecting other life stages such as embryonic, or

larval, would enable to gain more insights into the function role of 5hmC. Experiments to that end are currently in progress by other members of Tauber lab.

CHAPTER 6

ROLE OF *microRNA* IN PHOTOPERIODISM

6.1 Introduction

Various classes of non-coding RNA are important regulators that can be classified as an epigenetic mechanism, particularly those that are involved in RNA interference (RNAi), as their signalling can be inherited through cell division (Goldberg *et al.*, 2007). In addition, non-coding RNAs can play a crucial role in other epigenetic processes such as directing cytosine methylation or histone modifications in multicellular organisms and regulating the gene expression (Costa, 2005). Moreover, various complex epigenetic mechanisms such as silencing of transposon activity by piwi interacting RNAs (piRNAs), position effect variegation, X-chromosome inactivation and parental imprinting have direct or indirect association with RNA components (Costa, 2007). Here, I focus on another important class of non-coding RNAs, microRNA (miRNA), which are important regulators of gene expression (Clifford *et al.*, 2013).

miRNA are short non-coding RNA which bind to the target 3' untranslated region (UTR) of the mRNA molecules and regulate expression of the gene at the translational level (figure 6.1). Each miRNA can possibly regulate multiple genes, with predictions that more than one third of all human genes may be regulated by miRNA molecules.

The discovery of miRNA

The first miRNA was discovered in 1993 in *Caenorhabditis elegans* by the identification of two transcripts arising from the *lin-4* locus: the 22 nucleotide long *lin-4s*, and the 61 nucleotide *lin-4L* (Lee *et al.*, 1993; Wightman *et al.*, 1993). The studies showed that *lin-14* translation was regulated by *lin-4* through its 3' untranslated region (UTR) by some anti-sense mechanism. After seven years, Reinhart *et al.* (Reinhart *et al.*, 2000) demonstrated the involvement of 21 nucleotide long *let-7* miRNA in temporal regulation of *lin-41* by binding target sites within its 3' UTR. The discovery of *lin-4* and *let-7* added a new dimension to our understanding of complex gene regulatory networks and since then thousands of putative microRNA have been identified in various organisms using computational and experimental tools.

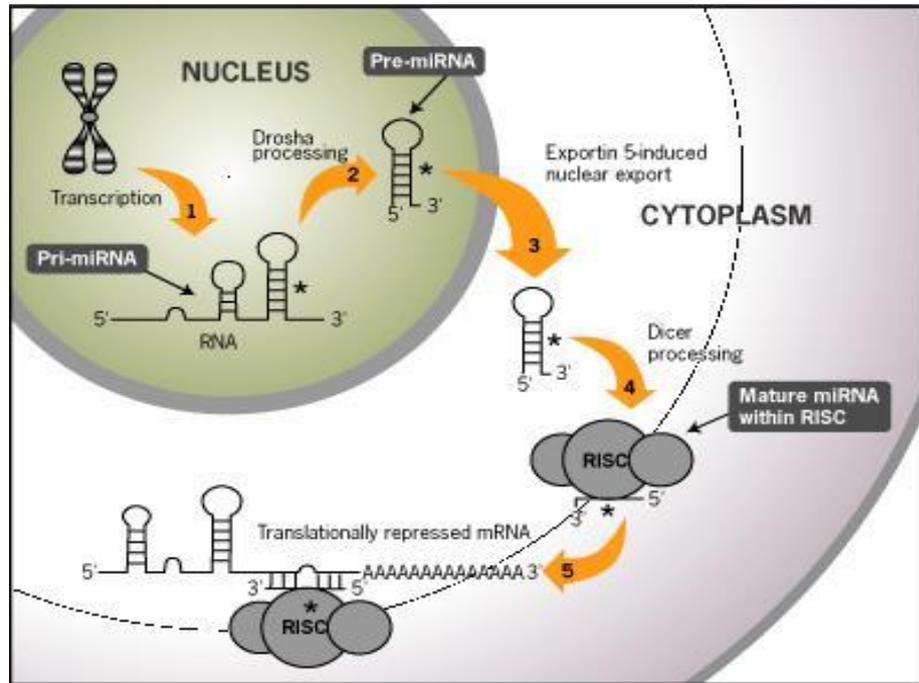


Figure 6.1: **miRNA processing pathway**
(Adapted from Ambion resources)

miRNA in animals

There are three main classes of regulatory small RNA molecules in animals; microRNA (miRNA), small interfering RNAs (siRNA), and piwi-interacting RNAs (piRNA). The difference between these small RNA classes was defined on the basis of their size and interaction with the protein Argonaute (Ago) (Kawaji & Hayashizaki, 2008; Kim *et al.*, 2009; Zhang, 2009). Typically in insects, 22-23 nucleotide (nt) miRNAs interact with Ago-1, 21 nt siRNAs are loaded onto Ago-2, and 24-31 ntpiRNAs are associated with the Piwi-subfamily of Ago proteins. However, a deeper understanding of processing of small RNA molecules has blurred the boundaries between these classes. While the majority of miRNAs are derived from intergenic regions and found as independent transcription units, some miRNA genes are also located in intronic regions and have been shown to be transcribed in parallel with their host transcript by RNA pol II (Rodriguez *et al.*, 2004). Interestingly, several miRNA genes exist as clusters of 2-7 genes and are co-transcribed as one bi-cistronic or polycistronic primary-miRNA transcript (Lee *et al.*, 2002).

Mode of action of miRNA

In eukaryotic cells, miRNA are expressed in the nucleus as parts of long primary miRNA transcripts (pri-miRNA) that have 5' guanine cap and 3' poly(A) tails (Smalheiser, 2003; Cai *et al.*, 2004) (Figure 6.1). The hairpin structures that are likely to form around the miRNA sequence of the pri-miRNA acts as a signal for digestion by a double-stranded (ds) ribonuclease (Drosha) to produce the precursor miRNA (pre-miRNA) (Lee *et al.*, 2002). A cytoplasmic dsRNA nuclease (Dicer) then cleaves the pre-miRNA leaving 3' overhangs (Lee *et al.*, 2002; Yi *et al.*, 2003). Finally, the single-stranded mature miRNA associates with a complex, that is similar if not identical, to the RNA Induced Silencing Complex (RISC) (Hutvagner & Zamore, 2002). Subsequently, the miRNA/RISC complex represses protein translation by binding to sequences in the 3' untranslated region of specific mRNAs. Single miRNA can bind to multiple mRNA (target molecules) and exert gene regulation at post-transcriptional level or multiple miRNAs can bind to single mRNA molecule in an orchestrated fashion and regulate gene expression. The binding of miRNAs to target mRNAs is governed by various factors and results in: (1) degradation of mRNA molecule due to perfect base complementarity between miRNA seed region (nucleotides 2-8) and mRNA 3'UTR (untranslated region), or (2) imperfect complementarity, leading to translation repression. The mismatches and bulges might be present in the miRNA- mRNA duplex but not in the region where the miRNA associate with Ago (Sun *et al.*, 2010). To stabilise the miRNA- mRNA duplex and induce post-transcriptional gene repression, complementary base pairing between the 3' half of the miRNA (generally nucleotides 13-16) and the 3' UTR of the mRNA is crucial. As a consequence, miRNA are mostly known to down-regulate the target mRNA molecules by binding to the 3'UTR (or 5'UTR) target region. However, few recent studies have also shown an up-regulative role of miRNA - mRNA binding (Vasudevan *et al.*, 2007; Buchan & Parker, 2007; Place *et al.*, 2008).

In plants, most miRNAs bind perfectly to the target mRNA molecules resulting in mRNA cleavage and degradation. In contrast, in animal cells this phenomenon is rare, and the duplex (often by imperfect pairing) leads to post- transcriptional repression (Filipowicz *et al.*, 2008). Plants and animals differ not only in miRNA-mRNA duplex

formation but also in the mode of action. In plant cells, the duplex recruit miRNA/RISC complex leading to the degradation of mRNA transcript by *RNase* activity associated with Ago. In animal cells, miRNA regulate protein translation by binding of more than one miRNA molecule to the 3' UTR region in a co-ordinated manner. Thus, miRNA can suppress protein translation at various levels, such as repression at the initiation step (Sun *et al.*, 2010) by blocking the ribosome assembly for the initiation of translation (Chendrimada *et al.*, 2007), repression at the post initiation steps (Petersen *et al.*, 2006) and inhibition of elongation or termination of protein translation. This indirect mode of translation suppression by miRNA occurs at mRNA level mainly through deadenylation.

miRNA in Nasonia vitripennis

To date, genes encoding 53 distinct putative miRNAs have been identified in *Nasonia*, and are listed in miRNA databases such as miRBase (Kozomara & Griffiths-Jones, 2014a) and miROrtho (Gerlach *et al.*, 2009). The list of the *Nasonia* miRNA has been derived from the published *Nasonia* genome sequence (Werren *et al.*, 2010). It is likely however, that the actual number of miRNA is significantly higher, given that the number of miRNA in *Drosophila* exceeds 300 (Lyu *et al.*, 2014).

miRNA as an epigenetic mechanism

Recent studies have shown that miRNA holds the potential to establish and maintain tissue- specific expression profiles (Giraldez *et al.*, 2005; Wienholds & Plasterk, 2005), a phenomenon largely associated with epigenetic mechanisms such as DNA methylation and histone modifications. Being transcribed as non-coding RNA, miRNAs can be considered as an epigenetic mechanism which regulates the gene expression at different levels (Saetrom *et al.*, 2006; Chuang & Jones, 2007).

In a current study, the miRNA induced changes in the gene expression were found to be heritable (Bali *et al.*, 2014) in nature, supporting the role of miRNAs as a stable epigenetic mark. In addition, the stable passage of maternal miRNA per se to the developing embryo influencing the zygotic transcription has been well studied (Choi *et al.*, 2011; Soni *et al.*, 2013) (see below).

The combinatorial action of miRNA, DNA methylation and histone modifications seems to be important in regulating the gene expression. Some recent

studies have suggested that epigenetic mechanisms such as DNA methylation and histone modification, not only regulate the expression of protein coding genes, but can also regulate the transcription of miRNAs (Sato *et al.*, 2011). Conversely, miRNAs can in turn control the expression of key epigenetic regulators, including DNA methyltransferases, histone deacetylases and modulate gene expression. This coordinated network of miRNAs and other epigenetic pathways forms an epigenetics-miRNA regulatory circuit which is important for gene regulation (figure 6.2). In animals, specific subset of miRNAs called epi-miRNA has been reported to indirectly influence the expression of oncogenes and tumor suppressing genes by regulating the translation of important epigenetic regulators such as DNA methyltransferases involved in DNA methylation (Fabbri *et al.*, 2013). Similarly in plants, a small class of miRNA molecules, 24 nucleotide in length, have been shown to regulate the level of DNA methylation at the locus from which it is transcribed (cis-regulation), and also influenced the level of DNA methylation of the target genes in trans via close association with AGO4 protein complex (Liang & Yu, 2010). Therefore, miRNA can mediate the target mRNA expression by interfering with either DNA methylation/histone modifications or through feedback mechanism resulting in fluctuating levels of pri-miRNA transcripts between distinct tissues, conditions or developmental stages.

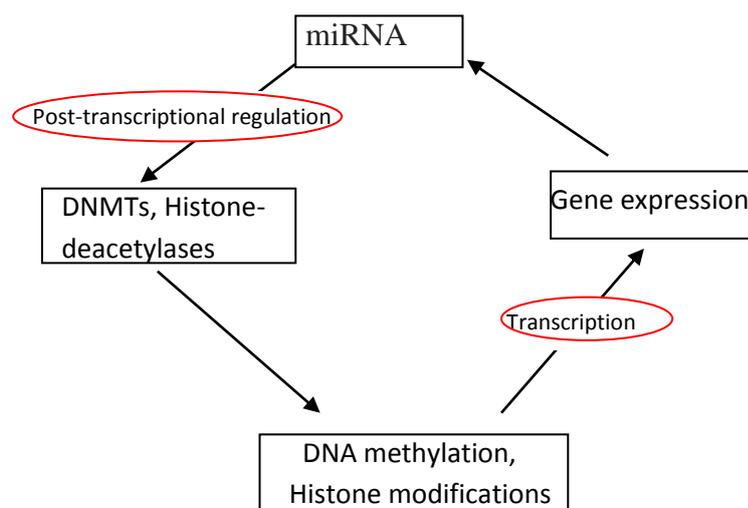


Figure 6.2: **Feedback mechanism of miRNA regulation**(see text for details).

miRNA involved in phenotypic plasticity in insects

The repertoire of miRNA has been profiled and was associated with different phenotypes in insects. In honeybees, differential miRNA expression was found to be associated with different castes viz. worker and queen bees (Guo *et al.*, 2013). According to that study feeding worker and queen bees with worker and royal jelly respectively led to development of distinct phenotypes due to difference in the miRNA content of the feed. The worker jelly was shown to contain 48 known bee miRNA (as annotated from mirBase13.0) in contrast to royal jelly that had only 25 of the 58 known bee miRNA. Thus, miRNAs were found to be more abundant in worker jelly than in royal jelly. In addition, a specific miRNA (miRNA-184) which is found in the worker jelly is important for adult morphology and the transition of destined queen larvae in the direction of worker bee phenotype. This demonstrated that a maternal like effect executed by transfer of miRNAs from adult nurse bees or their hypo pharyngeal secretions to the young developing larvae led to inheritance of developmental programs and societal roles from their foster mother.

The role of miRNA in driving phenotypic plasticity was also observed in aphids (*Acyrtosiphon pisum*). Aphids produce sexual and asexual (parthenogenetic) females in response to seasonal changes (Le Trionnaire *et al.*, 2008). Differential expression of 17 miRNAs were noticed amongst distinct sexual morphs (Legeai *et al.*, 2010), suggesting their involvement in polyphenism and morph-specific gene expression.

Another example of differential miRNA expression was also associated with the distinct social phases in locusts. Locusts are typical insects that modulate the phases in response to alteration in the environment and population density, switching from solitary to gregarious (migratory) phase. In a study, deep sequencing approach was adopted to compare the small RNA profiles of the solitary and gregarious phases of *Locusta migratoria* (Wei *et al.*, 2009). The profiles demonstrated difference in the composition of small RNAs, including miRNAs, between the two phases with nearly twice as much miRNAs in the gregarious phase as compared to the solitary phase.

Therefore, in the above studies condition specific miRNA expression has been noticed between the distinct phenotypes in various insects. The mechanism which controls the context dependent expression level of the miRNA is still not clear. One of

the possibilities leading towards differential expression of miRNA could involve maternal inheritance of adjusted level of the miRNA depending upon the environmental condition as explained below.

Maternal inheritance of miRNA in Drosophila and Zebrafish

A study revealed maternal inheritance of *miRNA-34* in *Drosophila* embryos (Soni *et al.*, 2013). *miRNA-34* was found to be involved in somatic cell reprogramming and early neuronal differentiation (Choi *et al.*, 2011). The maternal transfer of *miRNA-34* in *Drosophila* embryos indicated that localization of miRNA to the oocytes was dependent on the maternal miRNA processing machinery. Thus, along with miRNA, maternal Dicer-1 (involved in miRNA processing in *Drosophila*) was also passed to the oocytes. Similarly, in the same study (Soni *et al.*, 2013) authors have also found that the knock – down of maternal *miR-34* in zebrafish embryos led to developmental defects in the neuronal system.

In *Caenorhabditis elegans*, during embryo development, mRNA molecules are passed into the developing oocytes to translate into the proteins which are imperative for the development of fertilized zygote (Hollander *et al.*, 2010). Hollander and colleagues have showed that the regulation of these maternally transferred mRNA molecules was under the control of maternal miRNA molecules, found to be present at stable concentrations due to their slow turn-over rate. In support of this, another study revealed the importance of maternal miRNA, *let-7*, for the progression of larval stages in *Caenorhabditis elegans* (Grosshans *et al.*, 2005). The maternal transfer of miRNA in the developing zygote was made easy by the stability of miRNA in cytoplasm of the fertilized oocytes.

Overall these examples demonstrate that miRNA can regulate gene expression at the post-transcriptional level, and shape the morphology, development and behavioural pattern of an organism. Here, I investigated whether miRNA has a role in seasonal timing in *Nasonia*. I tested whether different photoperiods experienced by the females would lead to differential miRNA expression that could contribute to the diapause phenotypes in their progeny.

Role of Dicer in photoperiodism

In addition, I have also tested the role of RNase III enzyme, Dicer, an important enzyme in the miRNA processing pathways (Tsutsumi *et al.*, 2011). Dicer activity drives the cleavage of pre-miRNA into ~22-nt RNA duplexes with 2-nt overhangs at both 3' ends and frequent internal mismatches, called miRNA-miRNA* duplexes (Tsutsumi *et al.*, 2011). In *Drosophila*, two paralogs of Dicer: *Dicer-1* (*Dcr-1*) and *Dicer-2* (*Dcr-2*) are found with presumably mutually exclusive functions (Lee *et al.*, 2004). The evolution history of Dicer protein revealed early independent duplication of the gene in animal and plant lineage, followed by functional refinement and subsequent gene-loss modifying RNAi across various lineages (Mukherjee *et al.*, 2013). *Dcr-1*, in *Drosophila*, is chiefly responsible for miRNA maturation by specifically identifying the single-stranded terminal loop structure of pre-miRNA through its N-terminal helicase domain. It then checks the terminal loop size and measures the length from the 3' overhang to the terminal loop. This unique mechanism has been recently shown to allow *Drosophila* *Dcr-1* to verify the authenticity of pre-miRNA structures (Tsutsumi *et al.*, 2011). A parallel small RNA pathway, the small interfering RNA (siRNA) pathway, starts with endogenous or exogenous long dsRNA in the cytoplasm. Long dsRNA are also processed by Dicer into siRNA duplexes (Kim *et al.*, 2009), which are loaded onto Ago (Argonaute) proteins by the chaperone machinery (Iwasaki *et al.*, 2010; Iki *et al.*, 2010; Miyoshi *et al.*, 2010). In *Drosophila*, particularly *Dcr-2* and its partner protein R2D2 are employed for loading of the siRNA duplexes onto Ago2 (Liu *et al.*, 2003; Pham *et al.*, 2004; Tomari *et al.*, 2004). Although the role of *Dcr-1* in *Nasonia* is still unknown, it bears close similarity with the *Drosophila* *Dcr-1* as shown in the figure 6.3.

<i>D.melanogaster</i>	E I E L E I M L D G Y D S L E R S I G Y T F R N R S Y L L Q A F T H A S Y Q P N R L T D C Y Q R L E F L G
<i>Nasonia</i>	T E E L D Q L L S G F E E F E E S L G Y K F R D R S Y L L Q A M T H A S Y T P N R L T D C Y Q R L E F L G
	D A V L D Y L I T R H L Y E D P R Q H S P G A L T D L R S A L V N N T I F A S L A V R C G F H
	D A V L D Y L I T R H L Y E D P R Q H S P G A L T D L R S A L V N N T I F A S L A V R H G F H

Figure 6.3: **Comparison of the RNaseIII domain** between *Drosophila* (top- NCBI accession AAF56056.1) and *Nasonia* (bottom- NCBI accession XP_001605287.1) found in the *Dcr-1*. The protein sequences shared 85% similarity (NCBI Blast analysis) for the RNaseIII domain.

Here, in the second part of this study I have also tested the effect of *Dcr-1* knock-down by *RNAi* on the photoperiodic response.

6.2 Material and Methods

Profiling of miRNA between different photoperiods

Nasonia vitripennis (AsymC strain) was reared in 25°C and fed with commercially available blowfly pupae as explained in Chapter 2. Newly eclosed female wasps were incubated individually in long (LD) or short day (SD) like conditions (see Diapause experiment in Chapter 2) at 18°C for 10 days. After 10 days surviving wasps were snap frozen individually in liquid nitrogen and stored in -80°C. For differential miRNA profiling, 10 individual wasps, collected in microcentrifuge tubes, constituted one biological sample. At least three biological replicates (total of 30 wasps) were collected per photoperiodic condition; long (LD) and short day (SD).

The list of computationally predicted *Nasonia* miRNA was derived from miRBase (Kozomara & Griffiths-Jones, 2014b) and miROrtho (Gerlach *et al.*, 2009) databases, as explained above (miRNA list in the Appendix section 8.2). The multiple sequence alignment profile of the mature miRNA sequences was generated using CLUSTALW (figure 6.4). miRNA molecules bearing close resemblance (16 bases from 5' end and/ or 8 bases from 3' end) were pooled and multiplexed during the reverse transcription step in separate groups (as explained below).

Extraction of total miRNA from the adult female wasp was carried using the All- in- One Purification Kit (Norgen, cat. no. 24200). The kit facilitated sequential isolation and purification of large RNA, genomic DNA, miRNA and proteins. To prepare the animal tissue lysate, 400 µl of Lysis solution (supplied in the kit) was added and homogenized using electric homogenizer. The lysate was then transferred into an *RNase*- free microcentrifuge tube and spun for two minutes to pellet any cell debris. The supernatant was transferred to another tube, and section 3 of the kit protocol marked for Isolation of total large RNA, genomic DNA, microRNA and proteins was followed as per manufacturer instructions. The quantity and quality of the eluted miRNA was assessed on Agilent 2100 Bioanalyzer (figure 6.5) using the Agilent small RNA kit.

```

nvi-miR-71          -----UGAAAGAC--AUGGG--UAG-UGA----- 19
nvi-miR-31a        -----GGCAAGAU--GUCGGCAUAGCUGA----- 22
nvi-miR-7          -----UGGAAGACUAGUGAUUUUG-UUGU----- 23
nvi-miR-190        ---AGAU AUGUUUGAU AUUCUUGGUUGUA---- 26
nvi-miR-34         -----UGGCA-GUGUGUUAGCUGGUUG-- 22
nvi-miR-928        -----GUGGC---UGUGGA-AGCUCGCGAA- 21
nvi-miR-276        -----UAGGAACUUCAUACCGUGCUCU----- 22
nvi-miR-275        ---UCAGGUACCUGAAGUAGCGCGCG----- 23
nvi-miR-279        --UGACUAGA UCCACACUCA----UAAA----- 22
nvi-miR-219        --UGAUUG--UCCAAACGCAAUUCUUG----- 23
nvi-miR-184        -----UGGACGGAGAACUGAUAGGGC----- 22
nvi-miR-1          -----UGGAAUGUAAAAGAAGUAUGGAG----- 22
nvi-miR-263b       -----UUGGCACUGGAAGAAUUCAC----- 20
nvi-miR-2796       -----AGGCCGGCGGAAACUACUUGC----- 21
nvi-miR-124        -----UAAGGCACGCGGUGAAUGCCAAG----- 23
nvi-miR-929        -----AUUGAC---UCUAGUA-GGGAGUCC- 21
nvi-miR-3477       -----CAUAUUACC---UC--GU--GGGAUUUC- 21
nvi-miR-8          -----UAAUACUG--UCAGGUA-AAGAUGUC- 23
nvi-miR-252        -----CUAAGUAC-----UAGUGCCGCAGGAG- 22
nvi-miR-277        -----UAAAUGCACUAUCUGGUACGACA----- 23
nvi-miR-3478       -----AUUGGAUGAAUCCUACCCGGUGAG---- 24
nvi-miR-92a        -----AUUGCACUUGUCCCGCCUAU----- 21
nvi-miR-317        --UGAACA-CAGCU--GGUGGUAUCUCAGU--- 25
nvi-miR-283        ---AAUAUCAGCU--GGUAAU-UCU----- 20
nvi-miR-993        ---GAAGCUCGUCUCUACAGGUAUCU----- 23
nvi-miR-137        -----UUAUUGCUUGAGAAUACACGUA----- 22
nvi-miR-iab-4      ---CGGUAUACCUUCAG--UAUACGUAAC---- 24
nvi-let-7          ---UGAGGUA-GUAGGUUGUAUAGU----- 21
nvi-miR-iab-4as    ---UUACGUAUACUGAAGGUAUACCG----- 23
nvi-miR-10         -----ACCCUGUAGA UCCGAAUUUGU----- 21
nvi-miR-100        -----AACCCGUAGA UCCGAAUUUGU----- 22
nvi-miR-125        -----UCCUG-AGACCCUAACUUGUGA----- 22
nvi-miR-2b         ---UAUCACAGCCAGCUUUGAUGUGC----- 23
nvi-miR-2c         ---UAUCACAGCCAGCUUUGAUGAGC----- 23
nvi-miR-2a         ---UAUCACAGCCAGCUUUGUUGAGC----- 23
nvi-miR-13b        ---UAUCACAGCCAUUUUUGACGAGU----- 23
nvi-miR-13a        ---UAUCACAGCCAUUUU-GAUGAG----- 21
nvi-miR-307        -----UCACAACCUUUUUGAGUGAG----- 20
nvi-miR-29b        -----UAGCA-CCAUUUGAAAUCAGU----- 20
nvi-bantam         ---UGAGAUCAUUGU---GAAAGCUGAUU---- 23
nvi-miR-210        --UUGUG--CGU-GU---GACAGCGGCUA---- 21
nvi-miR-315        --UUUUGAUUGUUGCUCAGAAAGC----- 22
nvi-miR-12         -----UGAGUAUUACAUCAGGUACUGGU----- 23
nvi-miR-133        ---UUGGUCCCUUCAACCA--GCUGU----- 22
nvi-miR-9a         UCUUUGGU-----UAUCUA--GCUGUAUGA-- 23
nvi-miR-14         ---UCA---GUCUUUUUCU---CUCUCCUA-- 21
nvi-miR-281        -UGUCAUG-GACUUGCUCU----CUUUGU---- 23
nvi-miR-282        -GAUUU---AGCCUCUCCUAG-GCUUUGUCUGU 28
nvi-miR-305        --AUUG---UACUUCAUCAGGUGCUCUG----- 23
nvi-miR-927        ---UUUUA-GAAUUC--CUAC-GCUUUACC--- 23
nvi-miR-375        ---UUUGU-UCGUUCGGCU-C-GAGUUA----- 22
nvi-miR-33         -----GU-GCAUU--GUAGUUGCAUUG- 19
---
| nvi-miR-932      -----U-CAAUCCGUAGU-GCAUUGCAG- -22

```

Figure 6.4: Multiple sequence alignment of 53 putative miRNA sequences from *Nasonia*. Sequences sharing close similarity such as clusters highlighted in pink and yellow were multiplexed separately during reverse transcription.

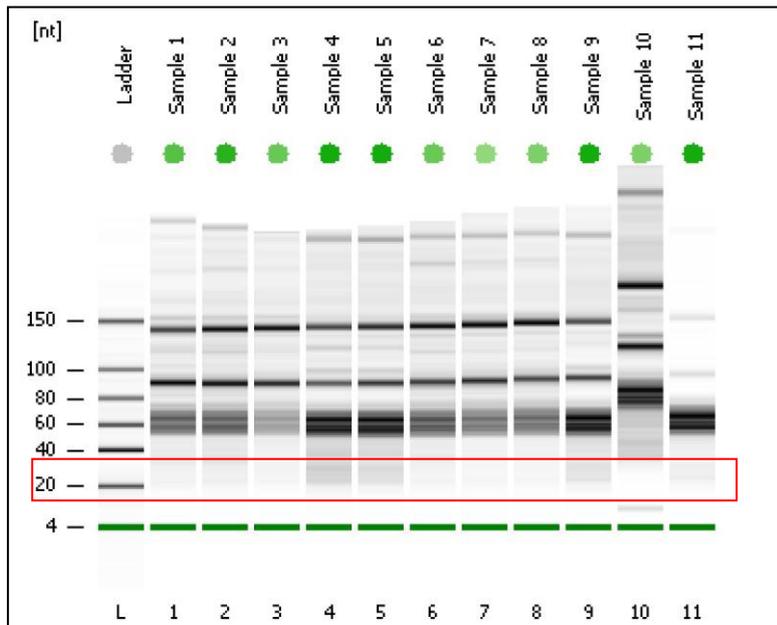


Figure 6.5: **Quantitative and Qualitative analysis of the miRNA extractions using BioAnalyzer.** From the left the lane represented as L denotes the marker lane and lanes 1-11 represents sample lanes. The area marked within the box (red) represents the miRNA proportion (20-30bp) in the tested samples.

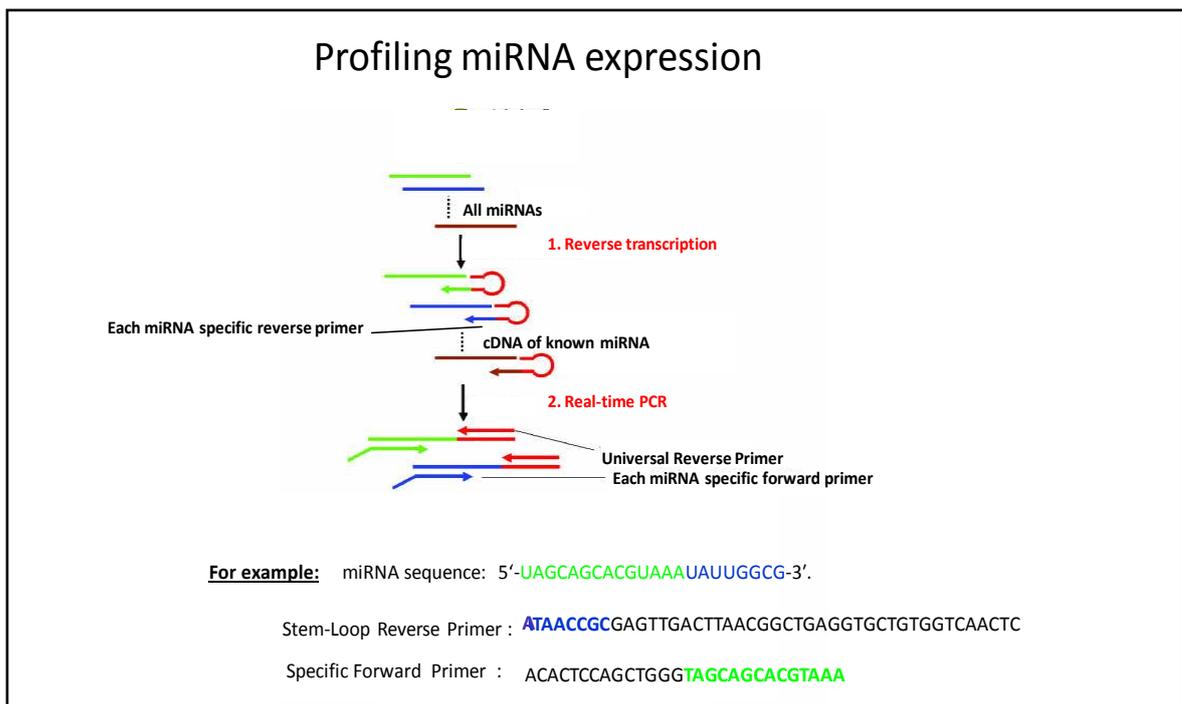


Figure 6.6: **Profiling of miRNA expression between photoperiods** (see text for details). (Adapted from Tang, 2006)

Based on the multiple sequence alignment of the predicted set of *Nasonia* miRNA, the sequences were grouped in sub-sets of 10 miRNA each. The reverse transcription reaction was multiplexed using specific stem-loop reverse primer for each miRNA in the sub-set along with the spiked artificial exogenous reference synthetic miRNA as per the published protocol (Tang *et al.*, 2006). For reverse transcription SuperScript II first strand reverse transcriptase (Invitrogen, cat. no. 18064-014) was used. In a reaction tube, 1µl of 2 picomole specific stem-loop reverse primer mix (for particular sub-set of miRNA), 250 ng total miRNA, 1µl 10mM dNTP mix and 1µl 10 picomoles synthetic reference RNA (*RNase* free double HPLC purified, Integrated DNA Technologies, UK: UCCCUGAGACCUCAAGUGUGA) was added in a 12 µl total reaction volume. The kit protocol specified for gene specific primer (GSP) was followed and the cDNA pool was diluted 10 times with *RNase*- free water.

The expression profiling of the putative set of miRNA was done using Tang *et al.* (2006) published protocol coupled with stem-loop real time PCR technique (Chen *et al.*, 2005). This technique proved to be highly robust and sensitive in detection of specific mature miRNA, allowing monitoring of miRNA from the same family (such as *miRNA-2a*, *miRNA-2b* and *miRNA-2c*, Figure 6.4) or with just one nucleotide difference. The qPCR was performed on DNA Engine II Opticon (BioRad), using 2X SYBR green mix (Agilent technologies), and included three technical replicates, -RT (RNA samples) and no template controls (NTC). Each real-time PCR reaction contained 5µl cDNA (diluted 10X), 0.4µM each of specific tailed forward and universal reverse primer (Box 6.1) in a 25µl total reaction volume. The PCR was done with two-step thermo-cycle conditions of 95°C for 15 seconds and 60°C for 45 seconds for 45 cycles followed by melt curve analysis from 50.0 °C to 95.1°C for every 0.2 °C at 1 sec interval.

A number of quality checks were performed before statistical analysis of the expression data. First, negligible expression profiles from -RT and NTC were confirmed in order to eliminate the chances of sample contamination. Then, the exponential curve (Log fluorescence Vs Cycle threshold) was analyzed for each miRNA along with the melting curve profiles. If the melting curve showed two or more distinct peaks corresponding to specific primer pair, indicative of non-specific amplification, the data was not assessed further for differential expression profiling.

Where the exponential curve and melt profiles passed the quality assessment, the raw fluorescent values were exported and analysed using the 'R' qPCR package (see Chapter 2 for details). In addition, a fraction of PCR products (5µl) were assessed on a

high percent (3%) agarose gel to verify the specific amplification of the desired miRNA and eliminate the risk of quantification bias from the formation of primer dimers. Essentially, after the reverse transcription (specific stem-loop reverse gene specific primer) and amplification using specific tailed forward, the amplicon length was expected to range between 70-76 bases.

BOX 6.1 Designing of primers for stem- loop Real- time PCR

The primers used in the expression profiling of miRNA were designed based on Chen *et al.* (2005).

For example: **miRNA sequence - 5' UAGCAGCACGUAUUUUGGCG 3'.**

Specific forward

specific reverse

The stem-loop reverse primer, specific for each miRNA, was composed of fixed 8 nucleotides stem with 20 nucleotides loop, and complementary miRNA sequence specific 8-mer at 3' end.

3'ATAACCGCGAGTTGACTTAACGGCTGAGGTGCTGTGGTCAACTC 5'

Complementary reverse

Stem

Loop

Stem

The specific tailed forward was composed of a fixed 14 nucleotides tail at 5' end with miRNA sequence specific 16- mer at 3' end.

5' ACACTCCAGCTGGGTAGCAGCACGTAAATA 3' (Specific tailed forward)

Tail

Complementary forward

For real- time PCR: 5'ACACTCCAGCTGGGTAGCAGCACGTAAATA 3': Specific Tailed forward and 3'GGCTGAGGTGCTGTGGTCAACTC5' : Universal reverse (based on stem-loop) was used.

Target prediction of the differential expressed miRNA between photoperiods

The mRNA targets of the differentially expressed miRNA were predicted computationally by inspecting the 3'UTR. The list of 3' UTR *Nasonia* sequences were derived from UTRdb database (Grillo *et al.*, 2010) (<http://UTRdb.ba.itb.cnr.it/advdownload>). UTRdb is a curated database for the untranslated sequences, obtained from various sources of eukaryotic mRNA primary data.

The PITA algorithm was used to predict the targets of the differentially expressed miRNA (Kertesz *et al.*, 2007). The PITA algorithm is based on target-accessibility function. Binding of miRNA at the 3' UTR region of mRNA leads to formation of the miRNA-mRNA duplex. The algorithm calculates the free energy associated with the formation of the miRNA-mRNA duplex, by formulating the energy attained by the base pairing interactions between the miRNA and mRNA (at the 3'UTR) and the energy liberated on unpairing the mRNA target to make it accessible for the miRNA (Kertesz *et al.*, 2007). The energetic score $\Delta\Delta G$ is being computed, with the lower (more negative) $\Delta\Delta G$ values indicating a stronger binding between the miRNA and the mRNA target (Box 6.2). Here, the interactions corresponding to $\Delta\Delta G$ values -10 or lower were used as likely to represent the functional targets. To assign an overall microRNA-gene target score, the below formula was used for negative energy scores:

$$\text{TARGET_SCORE} = -\log(\sum(e^{-\Delta\Delta G_i}))$$

In addition, the program also computes the seed size: mismatch: G-U wobbles for each of the interactions in the form of X:Y:Z notation. For example, the score 8:0:0 representing an 8-mer seed with no mismatch and G:U wobble, was used to filter the possible functional mRNA targets. The predicted target mRNA were further analysed for functional annotation by Blast2GO tool (Gotz *et al.*, 2008) and DAVID algorithm (Huang *et al.*, 2009). The list of gene identifiers were converted to the corresponding Refseq_Protein identifiers using the Gene Accession Conversion Tool (DAVID). The Refseq_Protein identifiers were limited to *Nasonia vitripennis* and further assessed for functional annotation.

BOX 6.2 miRNA Target Prediction

Some of the features considered across various algorithms for miRNA target prediction.

1. Conservation: Both miRNA and the targets are highly conserved across various species (Brennecke *et al.*, 2005). When applied for target prediction, if the same seed matches the 3' UTR of more than one species, the seed match is considered to be conserved.

2. Free energy: Free energy signifies the strength of binding between miRNA and mRNA target. Generally the free energy is a negative real value with units kcal/mol. Lower the free energy, stronger is the binding between the duplex. The free energy of miRNA -mRNA binding is normally assigned by RNA fold program-Vienna RNA package (Schuster *et al.*, 1994).

3. In- site features : Important features can also be deciphered from other parts of 3' UTR along with the seed region based on the free energy, the number of matches, mismatches, G:C matches, A:U matches ,G:U matches, bulges in mRNA etc.

4. Accessibility energy: Accessibility energy ($\Delta\Delta G$ values) (Kertesz *et al.*, 2007) is the energy obtained by miRNA binding to its targets. ΔG open is the energy required to take the target region accessible for miRNA binding and can be calculated as (Yue *et al.*, 2009) given below. The lower the accessibility energy, higher is the chance of the predicted 3' UTR to be a functional target.

$$\Delta\Delta G = \Delta G_{\text{duplex}} - \Delta G_{\text{open}}, \Delta G_{\text{open}} = G_{\text{free}} - G_{\text{unpair}}$$

Expression level of Dicer-1 in different photoperiods

The expression level of *Dicer-1* was determined and compared between LD and SD photoperiods. To assess the transcript level of *Dicer-1*, females were collected after 10 days of exposure to either LD or SD conditions. 10 individual females constituted one biological replicate from specific photoperiodic condition. Likewise, four biological replicates were used from each photoperiod (LD or SD) for the assay. Total RNA (200 ng) was extracted (using TRIzol reagent) from each sample and reverse transcribed using first strand cDNA synthesis kit (Invitrogen) as described in Chapter 2. The cDNA was further diluted 10 times and used as a template for qPCR. The Real-time PCR experiment was run with two technical replicates for each biological replicate, -RT controls and NTC. To determine and compare the relative *Dcr-1* expression level between the photoperiods, *Aequorin* (see Chapter 2) was used as the external reference gene. The data obtained from DNA Engine II Opticon (BioRad) was analysed by the standard curve method.

Knockdown of Dicer-1 in Nasonia

RNAi was employed to knock-down *Dicer-1* (*Dcr-1*) (as described in Chapter 2). To synthesize dsRNA, 1 µg of wasp RNA was used as a template for first strand cDNA synthesis using random primers. The synthesized cDNA was then used to amplify the specific region (450-550 bases) of the *Dcr-1* by nested PCR approach. The final amplicon was used as a template for in-vitro transcription. The synthesized dsRNA was assessed both quantitatively (Nanodrop) and qualitatively (1% agarose gel). The synthesized *dsDcr-1-RNA* (and *dsGFP-RNA* as control) was injected in the abdomen of female wasp pupae at appropriate stage and incubated for 4-5 days at 25°C. The surviving adult females were then exposed to either LD or SD at 18°C and followed as per the diapause experiment regime (Chapter 2 for details). The diapausing progeny from female wasps injected with *dsDcr-1-RNA* (or *dsGFP-RNA*) were assessed periodically after 5, 9 and 10 days of exposure to differential photoperiods.

After 5 days of exposure to different photoperiods, surviving females were collected and used for validation of the knock-down by qPCR (as described in Chapter 2). Three independent RNA extractions (each constituted from 10 females) were used each from LD and SD incubated wasps, injected with either *dsDcr-1-RNA* or *dsGFP-*

RNA. RNA was extracted from the respective groups (target and control) using TRIzol reagent (Invitrogen) as explained in Chapter 2. The extracted RNA was treated with *rDNaseI* to eliminate any genomic DNA residues. Equal amounts of RNA (475 ng from each sample) was used to synthesize cDNA using random primers. cDNA was then used as a template to quantify the transcript levels of *Dcr-1* in the target and the control group by SYBR Green qPCR as explained above.

6.3 Results

Out of the published 53 *Nasonia* putative miRNA, 31 miRNA were successfully amplified and experimentally validated (see Appendix section 8.2). As shown in the figure 6.8, out of the 31 validated miRNA, nine miRNAs were significantly up-regulated and seven were down-regulated in LD (compared to SD). The fold change of the differentially expressed miRNA was calculated based on Monte Carlo simulation distribution and analysed for statistical significance by permutation approach (see chapter 2 for details). The difference in the level of expression ranged from fold change [log2] +4 to -6 with *miRNA-125* most highly expressed and *miRNA-275* least expressed in LD. *miRNA-125* was also reported to be evolutionary conserved from nematodes to humans (Sun *et al.*, 2013)

Gene family	MIPF0001845; mir-125	
Stem-loop	<pre> auaaaagccu c - g ucucca --c c gc gc c cgucgcggu ccgagacuacuuguga gucg g cg cg g guagcgguaa gg cucu gau ggacacu uagc u --aguuuuc u u a u -u ua c cua a </pre> Get sequence	
Genome context	<i>Coordinates (Nvit_2.0)</i> chr1: 23812374-23812483 [+]	
Clustered miRNAs	< 10kb from <i>nvi-mir-125</i> nvi-mir-100 nvi-let-7 nvi-mir-125	chr1: 23809581-23809680 [+] chr1: 23810811-23810909 [+] chr1: 23812374-23812483 [+]

Figure 6.7: ***miRNA-125 (Nasonia vitripennis) sequence***. The mature miRNA sequence is represented by highlighted (pink) nucleotides in the stem-loop sequence. Two clustered miRNA; *miRNA-100* and *let-7* are found <10kb from *miRNA-125*. The data is derived from miRBase.

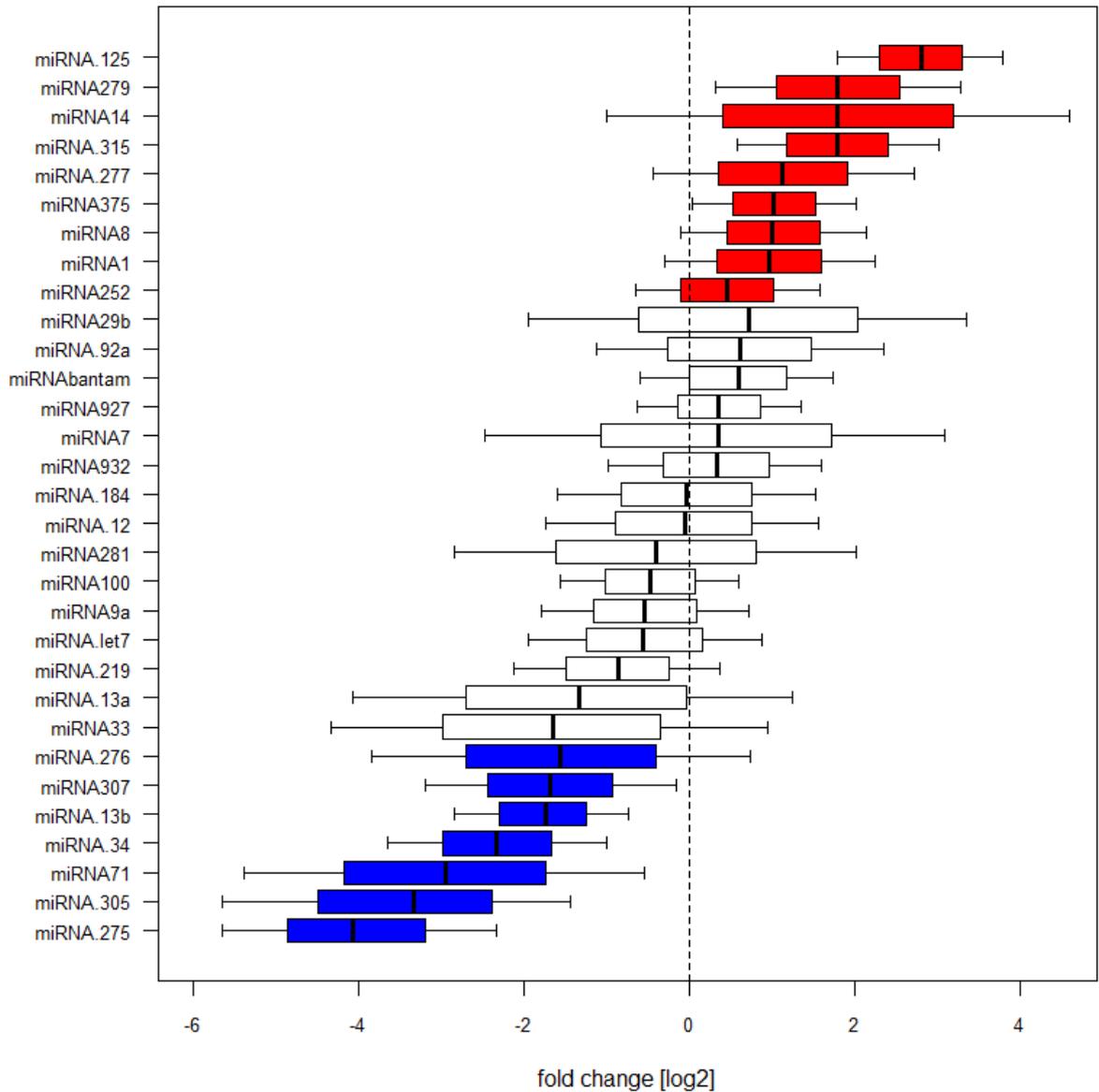


Figure 6.8: **Differentially expressed miRNA induced by photoperiods.** miRNA were profiled in LD and SD conditions and the expression level was compared. The miRNA represented by red boxplots showed up-regulation ($p < 0.05$) in LD, while the miRNA represented by blue boxplots showed down-regulation ($p < 0.05$) in LD. miRNA represented by the white boxplots showed no significant differential expression. The fold change was determined by the permutation approach using qPCR (R-statistical software). The line inside each box represents the median [log₂] fold change, and the error bars represent the minimum to maximum range, based on Monte-Carlo simulation at 95% confidence intervals.

Table 6.1 shows the regulation of miRNA that could be grouped in three clusters based on their genomic location (< 10 kb) and potentially under the same transcriptional regulation. In two of these clusters (Chr1 and Chr4) the miRNA show the same trend of regulation. In the third cluster, opposite trend is seen.

microRNA	Clustered miRNA	Expression level between photoperiods	Genomic location (<10 kb distance)
miRNA-71	miRNA-13b	Downregulated	Chr1: 8525129-8525800
miRNA-275	miRNA-305	Downregulated	Chr4: 1482167-1482444
miRNA-34	miRNA-277	miRNA-277: upregulated, miRNA-34: downregulated	Chr4: 7941792-7946106

Table 6.1: **List of clustered miRNA** (miRBase) showing trends in the expression level between the photoperiods.

Target prediction of differentially expressed miRNA

Using the PITA algorithm, mRNA targets were computationally predicted for each of the differentially expressed miRNA and scores were generated for every potential target site (based on the selected seed criteria as described above).

A total of 62 targets were analysed by DAVID and Blast2Go. These targets showed a broad range of functions including cell division, metabolism and transcription (Table 6.2 and Figure 6.9). With the DAVID algorithm, only 13 gene records were found with the corresponding gene ontology and pathway information. In addition, from the Blast2Go tool, a list of 44 genes were assigned the GO terms (see appendix section 8.2) and categorized based on the biological process (figure 6.9). However, no functional clustering was found between the predicted targets when analysed using DAVID algorithm.

miRNA	Target Gene	Pathway (KEGG record)	Ontology	Expression
miRN A 375	cyclin B	nvi04914:Progesterone-mediated oocyte maturation	Cell division and chromosome partitioning	upregulated in LD
miRN A 375	similar to 6-phosphogluconate dehydrogenase	nvi00030:Penrose phosphate pathway,nvi00480:Glutathione metabolism,		upregulated in LD
miRN A 13b	similar to ENSANGP00000010144	nvi00020:Citrate cycle (TCA cycle),nvi00310:Lysine degradation		downregulated in LD
miRN A 1	similar to GA14484-PA	nvi00190:Oxidative phosphorylation		upregulated in LD
miRN A 1	similar to GA17761-PA		Lipid metabolism	upregulated in LD
miRN A 275	similar to GA20978-PA		General function prediction only,	downregulated in LD
miRN A 305	similar to Transcription initiation factor IIB	nvi03022:Basal transcription factors	Transcription	downregulated in LD
miRN A 375	similar to aminoacylase, putative		Amino acid transport and metabolism	upregulated in LD
miRN A 375	similar to chloride channel protein 3		General function prediction only,	upregulated in LD
miRN A 71	similar to conserved hypothetical protein	nvi00561:Glycerolipid metabolism,nvi00564:Glycerophospholipid		downregulated in LD
miRN A 315	similar to glutathione transferase o1		Posttranslational modification, protein turnover, chaperones	upregulated in LD
miRN A 375	similar to saposin	nvi04142:Lysosome		upregulated in LD
miRN A 277	similar to translational activator gcn1		Energy production and conversion,RN A processing and modification	upregulated in LD

Table 6.2: **Functional annotation of the predicted targets.** The DAVID algorithm listed 13 gene records for the predicted mRNA targets. The corresponding target genes are represented in separate rows followed by the gene ontology /and KEGG pathway.

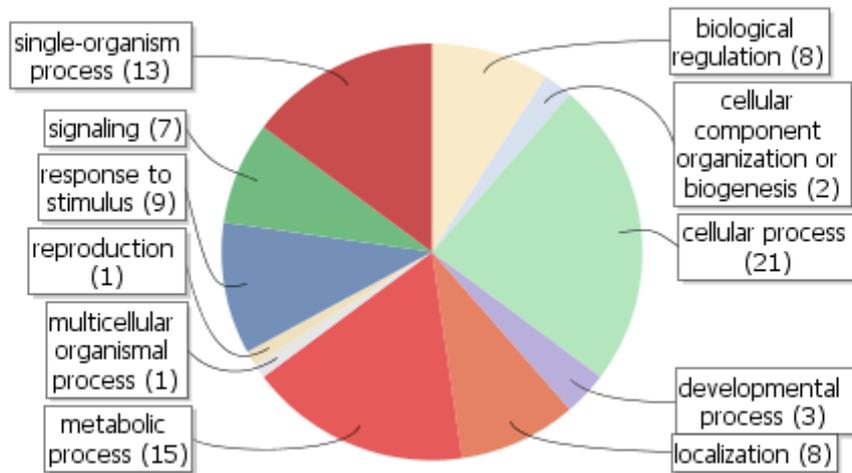


Figure 6.9: **GO analysis of the predicted target genes in *Nasonia*.** The gene ontology analysis based on the 62 predicted targets revealed an enrichment of genes involved in metabolic and cellular process (light red and green shaded parts) in *Nasonia* as analysed by Blast2GO (Gotz *et al.* 2008) tool.

Knock-down of Dicer-1

The phylogenetic analysis of *Dcr* showed prevalence of two paralogs namely, *Dicer-1* (*Dcr-1*) and *Dicer-2* (*Dcr-2*) across animal and plant lineage (Figure 6.10). In insects such as *Nasonia*, *Bombyx*, and *Tribolium*, both paralogs are present, with *Dicer-1* most likely to be involved in miRNA biogenesis. Here, I have silenced *Dcr-1* using *RNAi* and validated the knockdown by qPCR (figure 6.11). The relative expression values of *Dcr-1* was significantly lower in the treated group (*dsDcr-1-RNAi*) compared to the control (*dsGFP-RNAi*), with a fold change of 1.5 (Two-tailed Student *t*-test, *t* stat = -5.26, df = 7, *p* = 0.001). This difference was noticed for the wasps (*dsDcr-1-RNAi*) that survived after 5 days of photoperiodic exposure (in either LD or SD conditions), but didn't produce any viable progeny for phenotypic analysis.

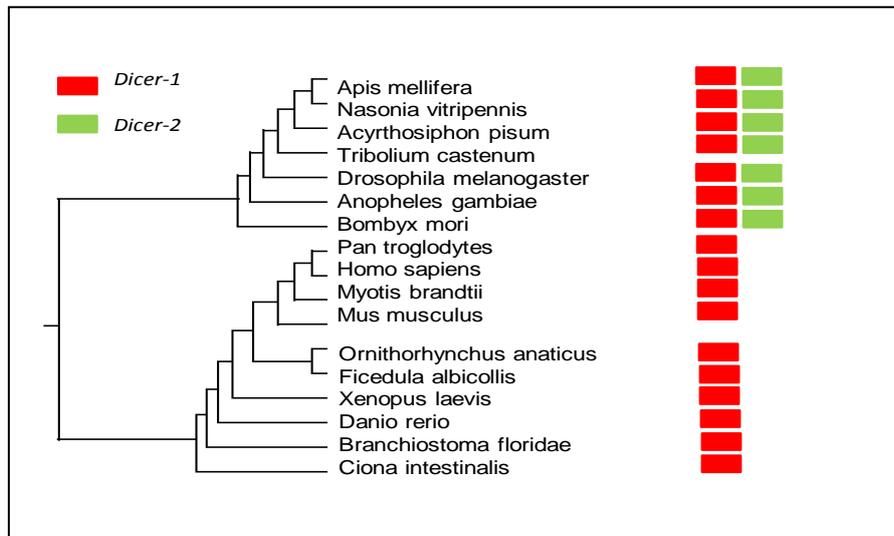


Figure 6.10: **Phylogenetic tree showing the prevalence of Dicer protein family across the taxa.** The rooted tree (UPGMA) was constructed using a CLUSTALW- Multiple sequence alignment of the protein Dcr-1 sequences from various species.

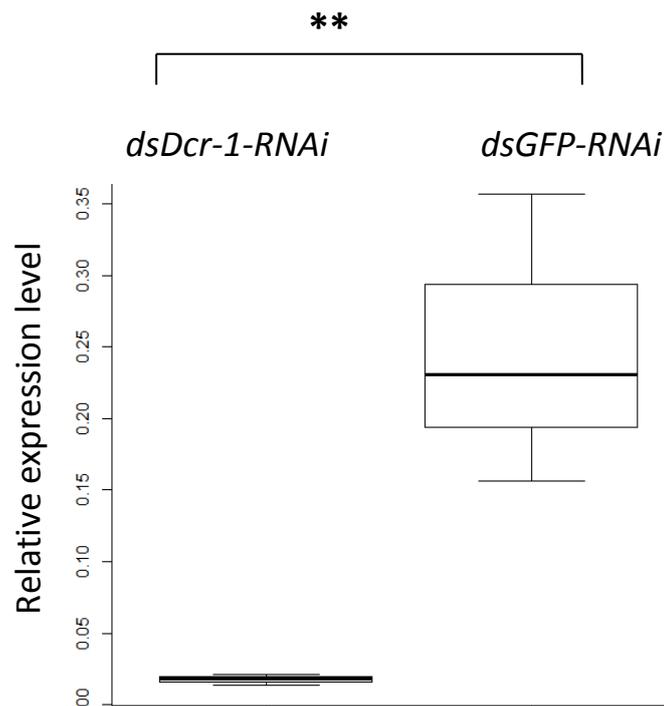


Figure 6.11: **Relative expression level of *Dcr-1* after the knock-down.** The relative expression level of *Dcr-1* was substantially down-regulated in the treated (*dsDcr-1RNAi*) group ($n \geq 60$) represented by grey boxplot as compared to the control (*dsGFP RNAi*, $n \geq 60$) represented by white boxplot. The line inside each box represents the median, the top and the bottom the 75 and 25th percentiles respectively, error bars range from minimum to maximum.

Consequently, the knockdown of *Dcr-1* seems to be lethal, as substantial reduction in the number of progeny was observed in *dsDcr-1-RNAi* group compared to the control (Figure 6.12). A considerable difference in the number of progeny was seen between *dsDcr-1-RNAi* and *dsGFP RNAi*, after 5 and 9 days, but not after 10 days. Thus, this clearly pointed out that the lethal effect of the *Dcr-1* knock-down seemed to fade with time, leading to restoration in the number of progeny.

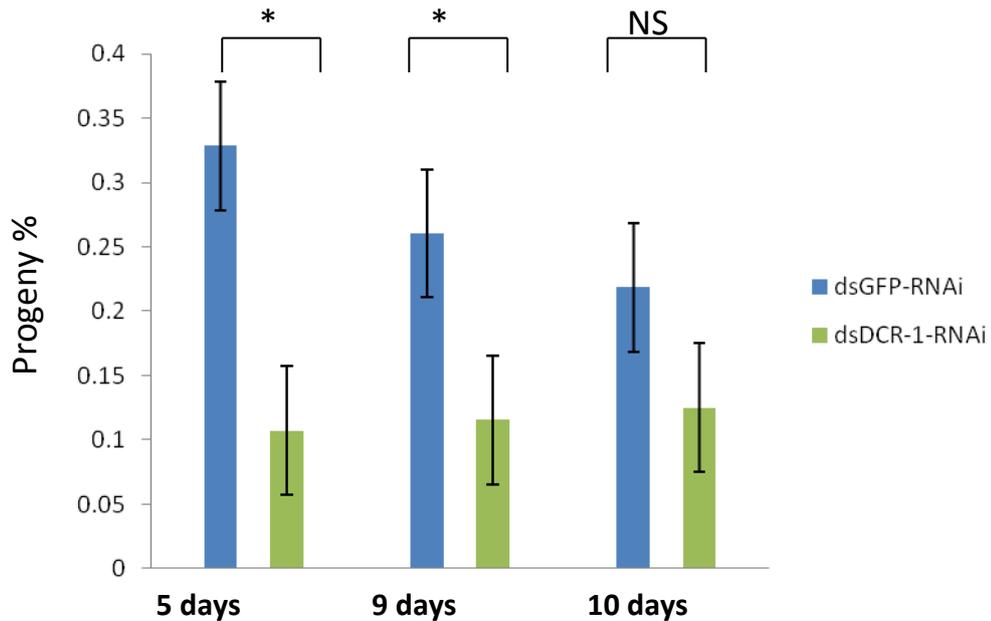


Figure 6.12: **Number of progeny after *Dcr-1* knockdown.** The number of females producing viable progeny (represented as progeny %) declined in *ds-Dcr1RNAi* group (treated, green bars), as tested after exposing for 5 (Fisher's Exact test, $p = 0.0001$) and 9 days (Fisher's Exact test, $p = 0.01$) in the differential photoperiods, compared to the control group (blue bars). No difference was, however, observed after 10 days of incubation between the treated and control groups.

Moreover, the level of diapause analysed in progeny from the females injected with *dsDcr-1-RNA* did not show any abnormal response between LD and SD, and resembled the typical photoperiodic response (One-way ANOVA, $F_{1,19} = 5.36$, $p = 0.03$). Presumably, these treated females did not have a considerable reduction in the *Dcr-1* transcript level, and as expected showed classical difference in the level of diapause between the photoperiods.

6.4 Discussion

This study has experimentally validated 31 *Nasonia* miRNA, and assessed the effect of the photoperiod on their expression. Nine miRNAs showed higher expression in LD compared to SD, whilst seven showed lower expression in LD. Obviously the genes targeted by these miRNA are expected to show the reverse trend of regulation, driven by the photoperiod.

One of the differentially expressed miRNA, miRNA-125, showed significantly a higher expression in LD. In *Drosophila*, expression of miRNA-125 was found to be responsible for maturation from larval-to-reproductive adult forms (Chawla & Sokol, 2012). Presumably, high expression of miRNA-125 in LD condition in *Nasonia* could also inhibit diapause and promote development of the progeny. In addition, miRNA-100 and let-7 were found to be present in the close genomic proximity of miRNA-125 (Table 6.1). A set of miRNA that are distributed closely in the genome are termed a miRNA cluster (Lagos-Quintana *et al.*, 2003; Lai *et al.*, 2003). During the miRNA biogenesis, the splicing of one pri-miRNA transcript can result into multiple pre-miRNA due to the presence of the clustered miRNA on the specific pri-miRNA (Kim & Nam, 2006). Recent expression studies have shown that the clustered miRNA are often co-expressed to regulate various biological processes through functional co-ordination (Cullen, 2004; Baskerville & Bartel, 2005; Ambros, 2004). However, unlike miRNA-125, the clustered miRNA (miRNA-100 and let-7) did not show up-regulation in LD, raising the doubt over the co-expression of the clustered miRNA. On the contrary, the differentially expressed and clustered miRNA (table 6.1), miRNA-275 and miRNA-305 (see appendix section 8.2) showed uniform down-regulation in LD.

In theory, different miRNA molecules hold the potential to regulate distinct mRNAs whose expression might vary in environment, tissue or development specific manner. In *Nasonia* none of the differentially expressed miRNAs (Fig. 6.8) were found to regulate the same target gene. Moreover, the analysis using the DAVID algorithm did not identify any functional cluster among the predicted mRNA targets. However, it is likely that given that the actual number of miRNA in *Nasonia* is much higher, it is still possible that some of the miRNA may share the same targets.

The differential expression of miRNA could arise from the condition specific difference in methylation at the gene locus, from which the miRNA is transcribed (cis-

regulation), as seen in human cancer cells (Saito, Jones 2006). This difference in methylation state could exist either on the same gene at different positions resulting in differential miRNA expression (as seen for clustered miRNA) or at different genes. The difference in methylation regions (Chapter 3) driven by the photoperiod could thus be involved in the transcription of the differentially expressed miRNA molecules. But as per the current data set, the gene regions involved in the transcription of the differentially expressed miRNA did not overlap with the identified DMRs (Chapter 3).

Alternatively, the differentially expressed miRNA may regulate the expression of epigenetic modifiers such as DNA methyltransferases at post-transcriptional level (as explained above), resulting in differences in methyl marks on target genes (trans-regulation) responsible for the phenotypic plasticity (Wu *et al.*, 2010). This could lead to a light dependent target gene expression. Currently, an RNA-seq study is being carried at the Tauber lab. This would determine the difference in the transcriptome between the photoperiods, which could further be analysed for the presence of any overlap with the list of differentially methylated candidate genes (Chapter 3), as well as computationally predicted mRNA targets.

A more direct epigenetic route could be a maternal transfer of differentially expressed miRNA (encoding the photoperiod), or specific maternal transcripts that were regulated by the miRNA from the female and were transferred to the developing oocytes. Maternal transfer of the miRNA can potentially elicit the regulation of target mRNA expression (zygotic and maternal) imperative for phenotypic plasticity in the progeny (non-diapausing or diapausing larvae). In a separate study, the ortholog of one of the identified differentially expressed miRNAs (*miRNA-34*) in *Nasonia* (fig. 6.8), was shown to be maternally inherited in *Drosophila* and *Danio rerio* (Soni *et al.* 2013). In that study, knocking down maternal *miRNA-34* resulted in neuronal developmental defects in the growing embryo. Thus, the profiling of miRNA in the progeny (4-5 hours-old-embryos) in the future would further elucidate the inheritance of differentially expressed miRNA crucial for phenotypic plasticity in *Nasonia*.

Functional characterization of the predicted targets

The predicted targets that were generated by the PITA algorithm were analysed and characterized based on their functional annotation (Table 6.2). The highly expressed *miRNA-375* (in LD) was found to regulate *cyclin B*, most likely reducing the expression of *cyclin B* under LD. This protein has been shown to be involved in diapause in *Bombyx mori* (Takahashi *et al.*, 1996), which is consistent with the data here. The *miRNA-1* (higher expression in LD), was predicted to regulate the level of the protein GA14484 involved in oxidative phosphorylation, and the protein GA17761 involved in lipid metabolism. Low level of *miRNA-1* in SD (winter-like) would lead to up-regulation of these target proteins, which fits well with a recent proteomic study in *Nasonia* (Wolschin & Gadau, 2009a), showing that these two proteins are abundant during diapause. Similarly, a putative aminoacylase, found in the early stages of diapausing larvae in *Nasonia* (Wolschin & Gadau, 2009b), and Saposin, are targets of *miRNA-375*, which was found here to be upregulated in LD. (Fig. 6.8). Another protein similar to glutathionetransferase-I (GST) was found to be present in the diapause larvae (Wolschin & Gadau, 2009b) and is predicted to be targeted by the highly expressed *miRNA-315* in LD (Fig 6.8).

Knock-down of *Dicer-1* induced lethality

Another part of the study focused on the role of *Dicer-1* in the photoperiodic response. The successful knockdown of *Dicer-1* in adult females induced lethality in the progeny at the larval stage (figure 6.12) underpinning the important role of this protein during development. Interestingly, in two recent studies, the knock-down of *Dicer-1* in the phylogenetically basal cockroach *Blattella germanica* has resulted in disrupted oocyte development (Tanaka & Piulachs, 2012), and impaired metamorphic process particularly due to disturbed miRNA maturation pathway (Gomez-Orte & Belles, 2009). In addition, *Dicer-1* has been shown to affect the expression of maternally inherited miRNA in *Drosophila* (Soni *et al.* 2013). Therefore, potentially down-regulation of *Dicer-1* could lead to inhibition of the miRNA maturation machinery and adversely affect the transfer of miRNA to the developing embryo resulting in the developmental arrest and lethality, as seen in *Nasonia*.

Overall, the experiments described here demonstrate the potential importance of miRNA in photoperiodism, to my knowledge, for the first time. The current finding underpins the importance of miRNA as an epigenetic mechanism that can modulate target gene expression, in response to environmental cue.

CHAPTER 7

GENERAL DISCUSSION

The study of *Nasonia vitripennis* offers a great opportunity to test the role of epigenetic mechanisms in seasonal timing. The use of Reduced Representation Bisulfite Sequencing (RRBS) allowed a glimpse into the changes of DNA methylation associated with seasonal timing in *Nasonia*. The knockdown of the various *DNMTs* (*Dnmt1a*, *Dnmt3*; Chapter 4) has indicated a causal role in the photoperiodic response, although the specific mechanisms that translate the epigenetic information are yet to be identified (see below). In plants, similarly previous studies have shown that abiotic stress such as salinity or exposure to salicylic acid resulted in differential methylation at organism-or-tissue level (Downen *et al.*, 2012; Karan *et al.*, 2012). This differential methylation was also shown to affect the level of gene expression between the distinct tissues or/and genotypes

The differential methylation that I identified in this study was profiled using the whole body of the wasp. It is likely that various tissues, possibly even specific cells, undergo different methylation changes. In the honeybee for example, high proportion of differentially methylated genes were found in the larval tissues (2,399) compared to the adult brain (561) in queen and worker bees. In addition, the vast majority of differentially methylated genes were reported to be more methylated in workers than the queen bees (Foret *et al.*, 2012a).

Nevertheless, the results here suggested that the switch from long to short day may induce either increase or decrease in methylation. These may reflect specific loci (in specific cells) that carry basal level of methylation (Table 7.1), and activation of *DNMTs* in specific tissues at specific photoperiod may result in overall increase or decrease in methylation. It may also suggest the presence of de-methylation system (discussed below, table 7.1).

Condition	Gene A	Gene B	Gene C	Gene D	Gene E	Gene F
Normal	●●○○	●○●○	○○○○	●●●●	●●○●	●●○○●
LD	●●●●	●○●○	○○○○	●●●●	●●○●	●○○○
SD	●●○○	●●●○	○○○○	●●●●	●○○○	●●○○●
Result	Increase in LD	Increase in SD	Unmethylated	Uniform methylation	Decrease in SD	Decrease in LD

Table 7.1: **Fate of CpG sites under different photoperiods.** Each column represents different type of possible gene locus based on the level of CpG methylation present across 4 distinct CpG sites. The methylated CpG site is denoted by filled circle and the unmethylated CpG is denoted by open circle. For Gene A and Gene B, increase in the level of methylation in LD and SD conditions respectively whilst for Gene E and Gene F decrease in the level of methylation, contributes towards the establishment of differential methylation. Gene C and Gene D represent no difference with complete lack and uniform methylation present respectively throughout the CpG sites between the contrasting photoperiodic conditions.

In LD for example, *Msn* (*Misshapen*) showed increased level of methylation whereas decreased level of methylation was seen across *LOC100123268*. This clearly pointed out that differential CpG methylation was not driven by overall increase/decrease in a particular photoperiod, but was rather dependent on the candidate genes. Thus, the existing data supported the viewpoint that discrete set of genes resembling Gene A (table 7.1) and Gene B would result in increased level of CpG methylation in LD and SD respectively and lead to the observed phenotypic plasticity. On the contrary, removal of specific methyl marks (Gene E and Gene F) could also result in the photoperiod dependent differential methylation across the *Nasonia* genome.

How does DNA methylation gets translated to phenotypic plasticity is not yet understood. In mammals, regulation of gene expression is thought to be mediated by methylation of promoter regions (Jones & Takai, 2001). However, the recent mapped methylomes of silkworm (Xiang *et al.*, 2010a), honeybees (Flores & Amdam, 2011) and *Nasonia* (Chapter 3) indicated that most of the methylation occurs in gene body, rather than in promoter regions, pointing to other possible mechanisms. Recently, various studies have suggested the importance of differential methylation in alternative splicing and exon skipping leading to altered transcript level. In *Nasonia*, the alternatively spliced exon present in the sex-determining locus *transformer* was also found to be heavily methylated (Park *et al.*, 2011) along with sparsely methylated introns. It is

therefore possible that alternative splicing (or exon skipping) is the mechanism that regulates the photoperiodic response, which is worthy of further investigation.

As mentioned earlier, DNA de-methylation could also be involved in the photoperiodic differential methylation (Figure 7.2). Generally, DNA demethylation can occur via active or passive mechanisms (Chapter 5). In brief, the passive mechanism involves subsequent loss of methylation due to absence of maintenance methyltransferase (Dnmt1), after several rounds of DNA replication. On the other hand, the action of enzymes such as Thymine DNA glycosylase (TDG), Base excision repair (BER) leads to direct erasure of the methyl marks and promotes active DNA demethylation. Another important route of DNA demethylation is facilitated by spontaneous deamination which results in replacement of methylated cytosine to thymine that gets further removed due to T/G mismatch. In addition, few recent studies have unequivocally pointed out towards DNA demethylation mechanism through the action of TET enzymes on 5hmC intermediate, particularly in mammals (Zhang *et al.*, 2011; Guo *et al.*, 2011b). The current study in *Nasonia* suggest that 5hmC is present (probably at low level), and the knock-down of *TET-2* in the adult females did show an aberrant photoperiodic response that invites a more thorough analysis of the 5hmC marks in the *Nasonia* genome (e.g. using next-generation sequencing based approach). Interestingly, the analysis of the *Nasonia* genome reveals the presence of the G/T mismatch-specific Thymine DNA Glycosylase-like protein (TDG), what was shown to be involved in active DNA demethylation (Cortellino *et al.*, 2011; He *et al.*, 2011a). The gene is computationally predicted to be present on the chromosome 2 in *Nasonia* (Fig. 7.1 NCBI ID: XP_001603860.2). Thus, other active DNA demethylation agents are potentially present in *Nasonia*, and their study, as well as the genome-wide mapping of 5hmC would allow better understanding of DNA demethylation and its role in photoperiodism. On a more general note, the work here establishes *Nasonia* as an efficient insect model system for studying DNA methylation, and therefore may serve as an alternative model to *Drosophila*, in which CpG methylation does not exist (Kunert *et al.*, 2003).

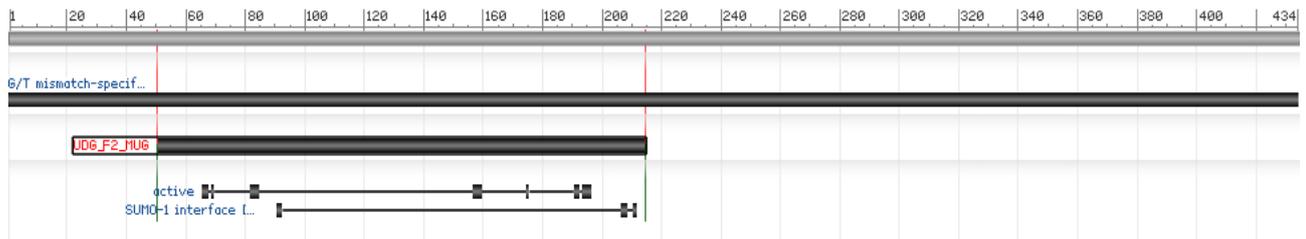


Figure 7.1: **Graphical representation of the predicted Thymine DNA glycosylase in *Nasonia*.** Conservation of the protein domain UDG_F2_MUG (highlighted in red) confers the removal of the thymine/uracil bases mispaired with guanine due to cytosine deamination. Source (NCBI database)

Certain miRNA were also found to be differentially expressed across the photoperiods in *Nasonia*. The differentially expressed miRNA could either regulate the expression of the target genes indispensable for seasonal timing (as shown in Chapter 6), or influence the magnitude of its inheritance in the developing progeny. For example, increased level of *miRNA-125* in the ovipositing adult females, reared under LD conditions (figure 6.8, Chapter 6), could potentially result in increased (maternal) *miRNA-125* levels in the developing embryo. To explore this, I attempted to profile the miRNAs in early embryos (4-5 hours old; no endogenous miRNA at this stage yet; (Verhulst *et al.*, 2010)) generated by females in either LD or SD conditions, but these experiments are still in progress. To test the functional role of each of the differentially expressed miRNA, loss-of-function studies would be carried out in the developing embryo. Injection of 2'-O-methyl antisense oligoribonucleotides in the embryo could be used for the specific depletion of the target miRNA in vivo, as demonstrated in *Drosophila* (Leaman *et al.*, 2005). The systematic loss-of-function analysis would also enable to study the effect on the expression level of the putative mRNA targets and their role in the photoperiodic clock. Alternatively, *pRNAi* as previously described in Chapter 2, can be adopted to knock-down the predicted target mRNA molecules and determine the effect on the photoperiodic response.

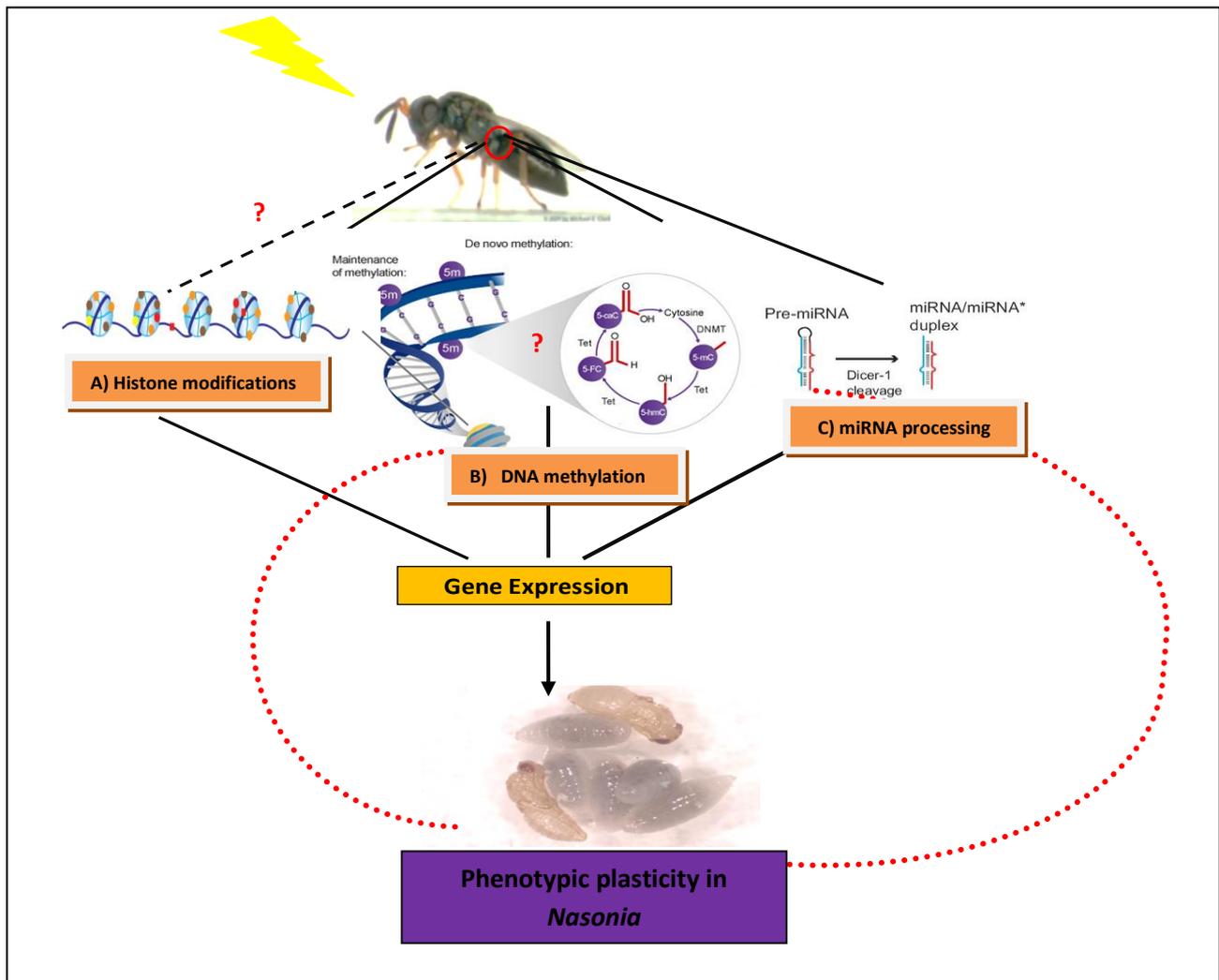


Figure 7.2: **Model of epigenetic mechanisms underlying seasonal timing in *Nasonia***. The adult female wasps perceives the day-length and generates diapausing or developing progeny via mechanisms of A. Possible post-translational histone modifications B. change in DNA methylation (or demethylation) and C. miRNA. These epigenetic mechanisms could either modulate candidate gene expression imperative for seasonal timing (black arrow) in the mother or lead to altered maternal inheritance (represented by red dashed curves) in the progeny to generate the phenotypic plasticity.

Fig 7.2 provides a summary of various epigenetic pathways which could lead to phenotypic plasticity as seen in response to seasonal cues in *Nasonia*. DNA methylation may encode the photoperiodic information via two plausible routes. The differential photoperiodic-driven methyl marks could be directly transferred from the mother to the developing progeny (figure 7.2). These inherited methyl marks could in turn lead to differences in zygotic transcription and generation of phenotypic plasticity. Alternative pathway could involve differential gene expression in the mother, due to differential methylation and then the transfer of the differentially expressed transcripts by the

mother to the embryo. The maternal contribution of specific transcripts for example in SD will trigger the diapause cascade in the embryo.

Similarly, the differential expression of miRNA that was identified in this study (Chapter 6) may relay the photoperiodic information in the embryo, where the photoperiod-driven miRNA could be transferred from the mother to the developing oocytes (Soni *et al.*, 2013). Alternatively, these miRNAs could modulate the target gene expression at the post-transcriptional level in the mother and lead to transfer of photoperiod-specific profile of maternal transcripts to the progeny. Apart from DNA methylation, various other possible epigenetic mechanisms could induce phenotypic plasticity (Fig. 7.3), and be important for the photoperiodic response.

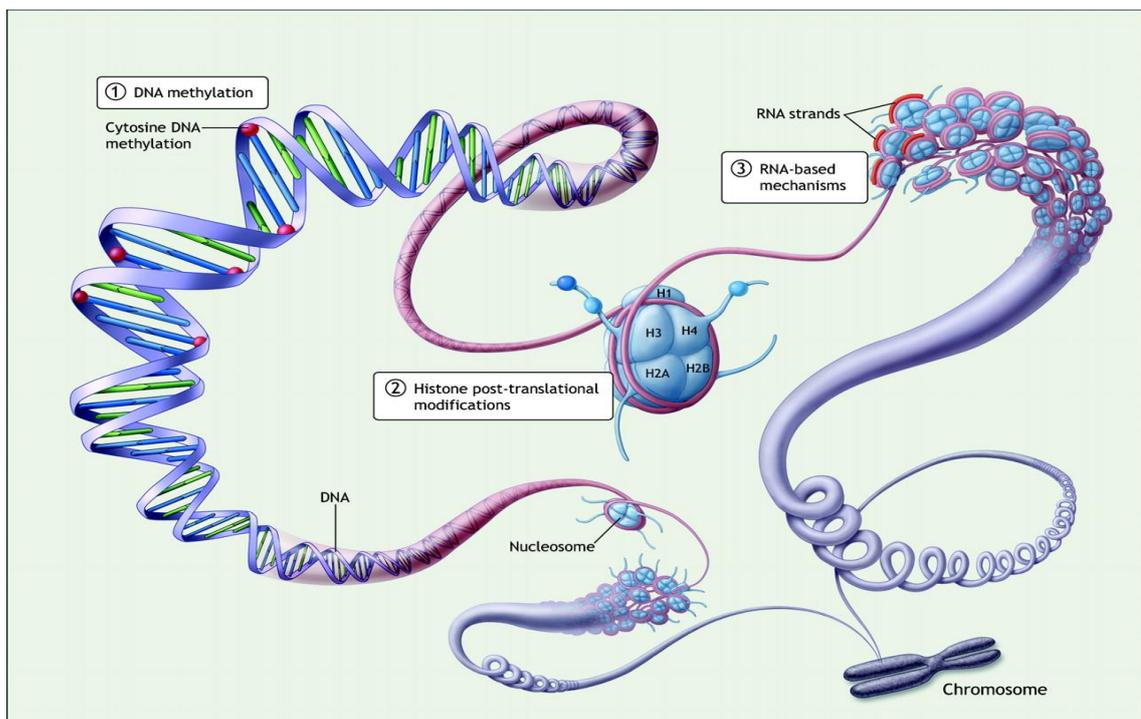


Figure 7.3: **Plethora of Epigenetic mechanisms** 1) DNA methylation involves addition of the methyl group to the cytosine molecule (represented by red balls). 2) Post-translational histone modifications (represented by light and dark blue balls) and recently discovered 3) RNA based mechanisms are shown to influence the chromatin structure and gene expression (represented by red strands coating the chromatin). (Adapted from Charles, 2010)

The nucleosome for example, (figure 7.3) is the basic building block of the chromatin which is composed of the octamer of histone proteins (H2A, H2B, H3, H4) and connector H1 wrapped around the DNA strand. Post-translational modifications at the N-terminal tail of the histone proteins by methylation (Jenuwein & Allis, 2001), phosphorylation, acetylation (Wade *et al.*, 1997), and ubiquitination (Shilatifard, 2006) have been shown to regulate the level of gene expression (Jenuwein & Allis, 2001). Importantly, these modifications have also been shown to play an important role in modulating the phenotype based on the environmental cue by reversible chromatin modeling (Feil & Fraga, 2012).

Among the various modifications, histone methylation has been shown to be involved in fundamental processes such as gene transcription and DNA repair (Bannister & Kouzarides, 2005). In vertebrates and plants, histone methylation has also been shown to interact closely with DNA methylation and lead to gene silencing (Fuks *et al.*, 2003). Moreover, some studies have also suggested an integrated role of histone modifications, DNA methylation and MBD (methyl-CpG-binding domain) in gene regulation (Brooks *et al.*, 2010). In a recent study, close association has been observed between the genes marked for DNA methylation and histone modifications in insects (Hunt *et al.*, 2013). This study proposed the conservation of the genomic targets for histone modifications to a greater extent between the taxa. As a consequence, by comparing the patterns of conserved DNA methylation in fire-ant and honeybees to histone modifications in *Drosophila melanogaster*, the authors found a strong correlation signaling towards their integrated relationship in governing processes such as gene transcription. Furthermore, in a separate study Dickman *et al.* (Dickman *et al.*, 2013) has identified 23 distinct post-translational modifications in the histone proteins H3 and H4 in honeybees and suggested towards the comprehensible possibility of the involvement of these modifications in nutritional-driven phenotypic plasticity evident in this eusocial insect. Likewise, post-translational histone modifications could potentially also exist in *Nasonia* and consequently influence the phenotypic plasticity arising due to photoperiodism.

In summary, the studies presented in this thesis are the first step towards understanding the role of epigenetic mechanisms such as DNA methylation and non-coding RNA in seasonal timing, and more broadly establish *Nasonia* as a powerful insect model system for studying various epigenetic pathways and the interaction between these pathways.

CHAPTER 8

APPENDIX

8.1 *N.vitripennis* methylome

List of differentially methylated regions using RRBS

Gene information	Chromosome	Type	Gene strand	Methylation position	Methylation strand	Methylation difference (respect to LD)	Qvalue
ID=gene907;Name=LOC100116941;Dbxref=GeneID:100116941,NASONIABASE:NV13337;gbkey=Gene;gene=LOC100116941	NC_015867.1	exon	-	9893080	-	27.520	0.042245
ID=gene1087;Name=LOC100121005;Dbxref=GeneID:100121005,NASONIABASE:NV16413;gbkey=Gene;gene=LOC100121005;psudo=true	NC_015867.1	exon	-	13336813	+	85.714	0.071698
ID=gene1167;Name=LOC100117390;Dbxref=GeneID:100117390,NASONIABASE:NV18148;gbkey=Gene;gene=LOC100117390	NC_015867.1	exon	+	14674587	+	100.000	1.56E-10
ID=gene1277;Name=msn;Dbxref=GeneID:100114663,NASONIABASE:NV18625;gbkey=Gene;gene=msn	NC_015867.1	exon	+	17786144	+	20.867	0.005668
ID=gene1277;Name=msn;Dbxref=GeneID:100114663,NASONIABASE:NV18625;gbkey=Gene;gene=msn	NC_015867.1	exon	+	17786152	+	23.391	0.000148
ID=gene1277;Name=msn;Dbxref=GeneID:100114663,NASONIABASE:NV18625;gbkey=Gene;gene=msn	NC_015867.1	exon	+	17786158	+	19.266	0.000292
ID=gene1277;Name=msn;Dbxref=GeneID:100114663,NASONIABASE:NV18625;gbkey=Gene;gene=msn	NC_015867.1	exon	+	17786164	+	27.289	1.28E-05
ID=gene1277;Name=msn;Dbxref=GeneID:100114663,NASONIABASE:NV18625;gbkey=Gene;gene=msn	NC_015867.1	exon	+	17786188	-	38.634	1.02E-05
ID=gene1485;Name=LOC100114164;Dbxref=GeneID:100114164,NASONIABASE:NV16579;gbkey=Gene;gene=LOC100114164	NC_015867.1	exon	+	21597600	-	-38.042	4.66E-07
ID=gene1485;Name=LOC100114164;Dbxref=GeneID:100114164,NASONIABASE:NV16579;gbkey=Gene;gene=LOC100114164	NC_015867.1	exon	+	21597609	-	-27.284	0.00016
ID=gene1692;Name=LOC100114300;Dbxref=GeneID:100114300,NASONIABASE:NV18104;gbkey=Gene;gene=LOC100114300	NC_015867.1	intron	-	29964721	+	-22.444	5.62E-05
ID=gene1692;Name=LOC100114300;Dbxref=GeneID:100114300,NASONIABASE:NV18104;gbkey=Gene;gene=LOC100114300	NC_015867.1	intron	-	29964755	+	-12.048	0.010563
ID=gene3184;Name=LOC100117821;Dbxref=GeneID:100117821,NASONIABASE:NV18104;gbkey=Gene;gene=LOC100117821	NC_015868.1	exon	-	7593864	-	99.265	3.00E-30

ASE:NV50484;gbkey=Gene;gene=LOC100117821;psuedo=true							
ID=gene3210;Name=LOC100115546;Dbxref=GeneID:100115546,NASONIABASE:NV15853;gbkey=Gene;gene=LOC100115546	NC_015868.1	exon	-	8113335	+	-16.667	0.042245
ID=gene3542;Name=Brm;Dbxref=GeneID:100123524,NASONIABASE:NV11766;gbkey=Gene;gene=Brm	NC_015868.1	exon	+	13285953	+	72.941	0.006188
ID=gene3872;Name=LOC100122036;Dbxref=GeneID:100122036,NASONIABASE:NV13848;gbkey=Gene;gene=LOC100122036	NC_015868.1	exon	-	19562661	-	-42.877	0.000123
ID=gene3876;Name=LOC100122200;Dbxref=GeneID:100122200,NASONIABASE:NV13851;gbkey=Gene;gene=LOC100122200	NC_015868.1	exon	+	19682752	-	-31.290	0.004612
ID=gene4038;Name=LOC100113804;Dbxref=GeneID:100113804,NASONIABASE:NV13108;gbkey=Gene;gene=LOC100113804	NC_015868.1	exon	+	26207451	+	16.881	0.016862
ID=gene4429;Name=LOC100118687;Dbxref=GeneID:100118687,NASONIABASE:NV14557;gbkey=Gene;gene=LOC100118687	NC_015868.1	exon	-	33255224	-	75.714	3.35E-07
ID=gene4430;Name=LOC100118729;Dbxref=GeneID:100118729,NASONIABASE:NV14558;gbkey=Gene;gene=LOC100118729	NC_015868.1	exon	+	33255224	-	75.714	3.35E-07
ID=gene4576;Name=LOC100122518;Dbxref=GeneID:100122518,NASONIABASE:NV14674;gbkey=Gene;gene=LOC100122518	NC_015868.1	exon	+	35786771	+	28.051	0.0019
ID=gene4576;Name=LOC100122518;Dbxref=GeneID:100122518,NASONIABASE:NV14674;gbkey=Gene;gene=LOC100122518	NC_015868.1	exon	+	35786855	-	24.239	0.059644
ID=gene4876;Name=LOC100117025;Dbxref=GeneID:100117025,NASONIABASE:NV15224;gbkey=Gene;gene=LOC100117025	NC_015868.1	exon	+	39097293	+	89.349	4.60E-11
ID=gene4877;Name=LOC100679730;Dbxref=GeneID:100679730;gbkey=Gene;gene=LOC100679730	NC_015868.1	exon	-	39097293	+	89.349	4.60E-11
ID=gene5764;Name=LOC100114102;Dbxref=GeneID:100114102,NASONIABASE:NV17553;gbkey=Gene;gene=LOC100114102	NC_015869.1	exon	-	10728639	+	61.143	0.003677
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620049	+	-56.874	5.09E-13
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620056	+	-37.868	8.71E-05
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620067	+	-35.854	0.000506

ASE:NV15033;gbkey=Gene;gene=LOC100118618							
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620076	+	-33.792	0.003108
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620083	+	-35.494	0.000664
ID=gene6433;Name=LOC100118618;Dbxref=GeneID:100118618,NASONIABASE:NV15033;gbkey=Gene;gene=LOC100118618	NC_015869.1	Exon	-	29620248	+	-25.316	0.000202
ID=gene6503;Name=WDR36;Dbxref=GeneID:100679400;gbkey=Gene;gene=WDR36	NC_015869.1	Exon	-	30423252	-	94.737	1.15E-08
ID=gene6618;Name=LOC100117948;Dbxref=GeneID:100117948,NASONIABASE:NV14276;gbkey=Gene;gene=LOC100117948	NC_015869.1	Exon	+	32292989	+	-29.450	0.03872
ID=gene6766;Name=LOC100119027;Dbxref=GeneID:100119027,NASONIABASE:NV15549;gbkey=Gene;gene=LOC100119027	NC_015869.1	Exon	-	33979466	-	20.801	0.042245
ID=gene7112;Name=Cact2;Dbxref=GeneID:100119253,NASONIABASE:NV11366;gbkey=Gene;gene=Cact2	NC_015870.1	Exon	-	2065751	-	39.241	1.09E-09
ID=gene7223;Name=LOC100121964;Dbxref=GeneID:100121964,NASONIABASE:NV11453;gbkey=Gene;gene=LOC100121964	NC_015870.1	Intro	-	3375369	+	-31.019	0.059644
ID=gene7311;Name=LOC100123019;Dbxref=GeneID:100123019,NASONIABASE:NV11527;gbkey=Gene;gene=LOC100123019	NC_015870.1	Exon	+	3987274	-	-23.342	0.022743
ID=gene7335;Name=LOC100123268;Dbxref=GeneID:100123268,NASONIABASE:NV11549;gbkey=Gene;gene=LOC100123268	NC_015870.1	Exon	+	4264085	-	-77.586	5.30E-07
ID=gene7640;Name=LOC100680213;Dbxref=GeneID:100680213;gbkey=Gene;gene=LOC100680213	NC_015870.1	Exon	+	9501397	+	97.727	8.74E-08
ID=gene7642;Name=LOC100119692;Dbxref=GeneID:100119692,NASONIABASE:NV18435;gbkey=Gene;gene=LOC100119692	NC_015870.1	Exon	-	9501397	+	97.727	8.74E-08
ID=gene7694;Name=LOC100115227;Dbxref=GeneID:100115227,NASONIABASE:NV16159;gbkey=Gene;gene=LOC100115227	NC_015870.1	Exon	+	10879153	+	79.817	1.01E-08
ID=gene7983;Name=LOC100115582;Dbxref=GeneID:100115582,NASONIABASE:NV15403;gbkey=Gene;gene=LOC100115582	NC_015870.1	Exon	-	14994681	-	-12.755	0.018959
ID=gene8136;Name=LOC100120295;Dbxref=GeneID:100120295,NASONIABASE:NV16720;gbkey=Gene	NC_015870.1	Exon	+	18985232	-	-31.325	0.018959

e;gene=LOC100120295							
ID=gene9138;Name=LOC100121209;Dbxref=GeneID:100121209,NASONIABASE:NV13544;gbkey=Gene;gene=LOC100121209	NC_015871.1	Exon	+	200405	-	84.406	1.96E-06
ID=gene9182;Name=mor;Dbxref=GeneID:100121971,NASONIABASE:NV13586;gbkey=Gene;gene=mor	NC_015871.1	Exon	-	425616	+	89.873	0.000115
ID=gene9505;Name=Atp5b;Dbxref=GeneID:100118124,NASONIABASE:NV12146;gbkey=Gene;gene=Atp5b;gene_synonym=Atp5n-beta	NC_015871.1	Exon	-	5609405	+	-33.121	0.010122
ID=gene9539;Name=LOC100119184;Dbxref=GeneID:100119184,NASONIABASE:NV12166;gbkey=Gene;gene=LOC100119184	NC_015871.1	Exon	+	5884954	-	31.707	1.43E-06
ID=gene9651;Name=LOC100122046;Dbxref=GeneID:100122046,NASONIABASE:NV12262;gbkey=Gene;gene=LOC100122046	NC_015871.1	Exon	+	7043621	-	-35.637	6.36E-05
ID=gene9723;Name=LOC100122869;Dbxref=GeneID:100122869,NASONIABASE:NV12317;gbkey=Gene;gene=LOC100122869	NC_015871.1	Exon	-	7582575	+	-30.354	0.042245
ID=gene9920;Name=Lin;Dbxref=GeneID:100187720;gbkey=Gene;gene=Lin	NC_015871.1	Exon	-	10082196	-	-37.209	0.006197
ID=gene10305;Name=LOC100114611;Dbxref=GeneID:100114611,NASONIABASE:NV17739;gbkey=Gene;gene=LOC100114611	NC_015871.1	Exon	-	18278577	-	-43.885	1.15E-08

Gene Ontologies analysis of the predicted DMRs

Seq. Name	Seq. Description	Seq. Length	#Hits	min. eValue	mean Similarity	#GOs	GOs	Enzyme Codes
gi 156549378 ref XP001602055.1	zinc finger protein	451	20	0	72.65%	2	F:binding; F:DNA binding	-
gi 345489116 ref XP003426057.1	wd repeat-containing protein 36	830	20	0	75.80%	6	F:hydrolase activity; P:nucleobase-containing compound metabolic process; F:DNA binding; F:nucleotide binding; C:intracellular; F:protein binding	-
gi 345483345 ref XP001599568.2	uncharacterized glycosyltransferase aer61-like	537	20	0	85.85%	1	F:transferase activity	-
gi 345485740 ref XP003425328.1	ubiquitin carboxyl-terminal hydrolase 36-like	895	20	0	74.70%	3	P:protein metabolic process; P:catabolic process; F:peptidase activity; F:hydrolase activity	EC:3.1.2.15 ; EC:3.4
gi 345485737 ref XP003425327.1	ubiquitin carboxyl-terminal hydrolase 36-like	888	20	0	74.80%	3	P:protein metabolic process; P:catabolic process; F:peptidase activity; F:hydrolase activity	EC:3.1.2.15 ; EC:3.4
gi 156547946 ref XP001604800.1	ubiquitin carboxyl-terminal hydrolase 36-like	881	20	0	75.50%	3	P:protein metabolic process; P:catabolic process; F:peptidase activity; F:hydrolase activity	EC:3.1.2.15 ; EC:3.4
gi 156550899 ref XP001602544.1	ubia prenyltransferase domain-containing protein 1 homolog	331	20	0	84.00%	3	C:cell; F:transferase activity; P:biosynthetic process	EC:2.5.1
gi 345490491 ref XP003426390.1	t-complex protein 1 subunit beta-like	534	20	0	93.75%	4	C:cytoplasm; P:protein metabolic process; F:nucleotide binding; F:protein binding	-
gi 345490493 ref XP001602878.2	t-complex protein 1 subunit beta-like	537	20	0	93.80%	4	C:cytoplasm; P:protein metabolic process; F:nucleotide binding; F:protein binding	-
gi 345485679 ref XP001605573.2	swi snf complex subunit smarcc2	976	20	0	81.95%	11	P:transcription, DNA-templated; P:regulation of biological process; C:nucleoplasm; C:protein complex; F:DNA binding; F:protein binding; C:nucleus; F:chromatin binding; P:organelle organization	-
gi 156543868 ref XP001608158.1	probable trna threonylcarbamoyladenosine	335	20	0	87.85%	6	C:cytoplasm; F:binding; F:transferase activity; F:hydrolase activity; C:nucleus; P:nucleobase-containing compound	-

	biosynthesis protein osgep-like						metabolic process	
gi 229577230 ref NP001153331.1 	polyubiquitin-a-like isoform x2	540	20	0	99.50%	8	P:cell death; P:embryo development; C:cytoplasm; P:reproduction; P:growth; C:nucleus; P:multicellular organismal development; F:protein binding	-
gi 156553659 ref XP001601204.1 	nuclear receptor 2c2-associated	67	20	2.93E-38	67.40%	1	P:biological_process	-
gi 156549935 ref XP001602626.1 	nodal modulator 2	1210	20	0	68.20%	1	F:carbohydrate binding	-
gi 156550151 ref XP001606126.1 	nad-dependent deacetylase sirtuin-1	871	20	0	78.60%	1	F:nucleotide binding	-
gi 156543828 ref XP001606625.1 	n-alpha-acetyltransferase 60-like	242	20	1.46E-175	82.65%	7	C:cytoplasm; C:protein complex; C:nuclear envelope; P:organelle organization; P:cell cycle; P:biological_process; P:cellular protein modification process; P:transport; F:transferase activity	EC:2.3.1.88
gi 156548502 ref XP001605802.1 	low quality protein: frizzled-like	580	20	0	80.45%	11	C:cell; F:receptor activity; P:multicellular organismal development; P:signal transduction; P:cell differentiation; F:protein binding; P:anatomical structure morphogenesis; P:embryo development	-
gi 345492232 ref XP001600241.2 	inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase-like	2221	20	0	87.35%	4	F:kinase activity; F:protein binding; P:metabolic process; F:hydrolase activity	EC:3.1.3.2
gi 345484640 ref XP001606343.2 	hbs1-like protein	1048	20	0	85.50%	5	P:nucleobase-containing compound metabolic process; P:catabolic process; F:hydrolase activity; C:cytoplasm; C:nucleus; F:nucleotide binding	-
gi 345484780 ref XP001599359.2 	eukaryotic translation initiation factor 4h	277	20	6.69E-107	83.50%	9	P:translation; F:translation factor activity, nucleic acid binding; P:cellular protein modification process; F:protein kinase activity; C:cell; F:hydrolase activity; F:transporter activity; P:transport; P:nucleobase-containing compound metabolic process; P:catabolic process; F:nucleotide binding	EC:2.7.11.12
gi 156551295 ref XP001601373.1 	dolichyl-diphosphooligosaccharide--protein glycosyltra	808	20	0	90.85%	3	P:biosynthetic process; P:cellular protein modification process; P:carbohydrate metabolic process; F:transferase activity; C:cell	EC:2.4.1

	nsferase subunit stt3b-like							
gi 345488022 ref XP003425816.1 	dna repair protein xrcc4-like	311	10	0	58.80%	4	F:DNA binding; P:response to stress; P:DNA metabolic process; C:nucleus	-
gi 345481883 ref XP001605650.2 	chromodomain-helicase-dna-binding protein mi-2 homolog	2009	20	0	87.35%	8	P:transcription, DNA-templated; P:regulation of biological process; F:hydrolase activity; F:DNA binding; F:nucleotide binding; F:transferase activity; C:nucleus; F:binding; F:protein binding	-
gi 156545406 ref XP001606477.1 	carnitine o-palmitoyltransferase mitochondrial	663	20	0	74.30%	3	P:lipid metabolic process; C:organelle; C:intracellular; F:transferase activity	-
gi 345486020 ref XP001605639.2 	carboxy-terminal domain rna polymerase ii polypeptide a small phosphatase 1-like	294	20	5.18E-177	90.50%	9	P:lipid metabolic process; F:lipid binding; P:metabolic process; F:phosphoprotein phosphatase activity; F:transferase activity; F:kinase activity; F:nucleotide binding; P:signal transduction; F:protein binding	EC:3.1.3.16
gi 156544311 ref XP001607169.1 	atp-dependent helicase brm	1587	20	0	87.40%	6	P:transcription, DNA-templated; P:regulation of biological process; F:hydrolase activity; F:DNA binding; F:nucleotide binding; C:nucleus; F:protein binding	-
gi 229577383 ref NP001153366.1 	atp synthase subunit mitochondrial-like	447	20	0	97.60%	5	F:transporter activity; P:ion transport; C:protein complex; C:intracellular; P:generation of precursor metabolites and energy; P:nucleobase-containing compound metabolic process; P:biosynthetic process; F:nucleotide binding	-
gi 156546920 ref XP001601662.1 	60s ribosome subunit biogenesis protein nip7 homolog	180	20	3.33E-131	87.35%	3	F:RNA binding; P:organelle organization; C:nucleus	-

8.2 miRNA profiling by Real-time PCR

List of putative miRNA sequences from MiROrtho

Name	Family	scaffold	start	end	Strand	mature_start	mature_end	sequence_mature
nvit-mir-14	mir-14	SCAFFOLD1	384344	384446	+	67	87	UCAGUCUUUUUCUCUCUCCUA
nvit-mir-29b	mir-29	SCAFFOLD1	561541 4	561551 2	-	66	85	UAGCACCAUUUGAAAU CAGU
nvit-mir-33	mir-33	SCAFFOLD1	903796 0	903806 0	-	31	49	GUGCAUUGUAGUUGCA UUG
nvit-mir-281	mir-46	SCAFFOLD10	692904	693006	-	61	83	UGUCAUGGACUUGCUC UCUUUGU
nvit-mir-7	mir-7	SCAFFOLD10	210857 7	210866 3	+	15	37	UGGAAGACUAGUGAUU UUGUUGU
nvit-mir-137	mir-137	SCAFFOLD10	255853 3	255863 2	+	60	81	UUAUUGCUUGAGAAUA CACGUA
nvit-mir-31a	mir-31	SCAFFOLD10	318192 7	318202 5	+	28	49	GGCAAGAUGUCGGCAU AGCUGA
nvit-mir-71	mir-71	SCAFFOLD12	145019 6	145029 7	+	26	44	UGAAAGACAUGGGUAG UGA
nvit-mir-2	mir-2	SCAFFOLD12	145036 7	145046 8	+	63	85	UAUCACAGCCAGCUUUG AUGAGC
nvit-mir-13a	mir-2	SCAFFOLD12	145051 6	145061 8	+	67	87	UAUCACAGCCAUUUUG AUGAG
nvit-mir-13b	mir-2	SCAFFOLD12	145077 6	145086 7	+	52	74	UAUCACAGCCAUUUUU GACGAGU
nvit-mir-2	mir-2	SCAFFOLD12	145097 0	145106 2	+	56	78	UAUCACAGCCAGCUUUG UUGAGC
nvit-mir-2	mir-2	SCAFFOLD12	145199 3	145207 7	+	51	73	UAUCACAGCCAGCUUUG AUGUGC
nvit-mir-92a	mir-25	SCAFFOLD14	410729	410826	+	63	83	AUUGCACUUGUCCCGGC CUAU
nvit-mir-283	mir-283	SCAFFOLD15	829662	829763	+	21	40	AAUAUCAGCUGGUAA UUCU
nvit-mir-12	mir-12	SCAFFOLD15	830969	831052	+	13	35	UGAGUAUUACAUCAGG UACUGGU
nvit-mir-184	mir-184	SCAFFOLD153	173577	173668	-	56	77	UGGACGGAGAACUGAU AAGGGC
nvit-mir-929	mir-929	SCAFFOLD2	378391 3	378401 2	+	17	37	AUUGACUCUAGUAGGG AGUCC
nvit-mir-190	mir-190	SCAFFOLD2	629259 1	629268 6	-	24	49	AGAUUGUUUGAUAAU CUUGGUUGUA
nvit-mir-iab-4as	mir-iab-4as	SCAFFOLD23	101630 2	101637 0	+	6	28	UUACGUAUACUGAAGG UAUACCG
nvit-mir-iab-4	mir-iab-4	SCAFFOLD23	101630 2	101637 0	-	42	65	CGUAUACCUUCAGUA UACGUAAC
nvit-mir-10	mir-10	SCAFFOLD23	161584 5	161594 4	+	22	42	ACCCUGUAGAUCCGAAU UUGU
nvit-mir-100	mir-99	SCAFFOLD251	29839	29938	+	19	40	AACCCGUAGAUCCGAAC UUGUG
nvit-let-7	let-7	SCAFFOLD251	31069	31167	+	13	33	UGAGGUAGUAGGUUGU AUAGU
nvit-mir-125	mir-125	SCAFFOLD251	32632	32741	+	27	48	UCCUGAGACCCUAACU UGUGA

nvit_223800	NULL	SCAFFO LD26	605703	605787	+	53	73	AGGCCGGCGGAAACUAC UUGC
nvit-mir-279	mir-279	SCAFFO LD26	143203 5	143213 4	+	67	88	UGACUAGA UCCACACUC AUUAA
nvit-mir-34	mir-34	SCAFFO LD29	257419	257517	-	21	42	UGGCAGUGUGGUUAGC UGGUUG
nvit-mir-277	mir-277	SCAFFO LD29	261831	261917	-	54	76	UAAAUGCACUAUCUGG UACGACA
nvit-mir-317	mir-317	SCAFFO LD29	275742	275831	-	55	79	UGAACACAGCUGGUGG UAUCUCAGU
nvit-mir-925	mir-925	SCAFFO LD329	99366	99465	+	16	40	AGGGAUUCGGUUUUGU AACAUUCGC
nvit-mir-275	mir-275	SCAFFO LD4	148216 7	148226 0	+	61	83	UCAGGUACCUGAAGUA GCGCGCG
nvit-mir-305	mir-305	SCAFFO LD4	148235 5	148244 4	+	16	38	AUUGUACUUCAUCAGG UGCUCUG
nvit-mir-276	mir-276	SCAFFO LD4	267660 8	267669 1	-	52	73	UAGGAACUUCAUACCG UGCUCU
nvit-mir-124	mir-124	SCAFFO LD4	298409 1	298416 7	+	51	73	UAAGGCACGCGGUGAA UGCCAAG
nvit-mir-315	mir-315	SCAFFO LD44	722281	722362	-	12	33	UUUUGAUUGUUGCUCU GAAAGC
nvit-mir-219	mir-219	SCAFFO LD47	29864	29965	+	20	42	UGAUUGUCCAAACGCA AUUCUUG
nvit-mir-133	mir-133	SCAFFO LD5	133836 5	133845 3	-	57	78	UUGGUCCCUUCAACCA GCUGU
nvit-mir-1	mir-1	SCAFFO LD5	135615 3	135623 3	-	50	71	UGGAAUGUAAAGAAGU AUGGAG
nvit-mir-282	mir-282	SCAFFO LD5	405499 4	405508 5	+	15	42	GAUUUAGCCUCUCCUAG GCUUUGUCUGU
nvit-mir-307	mir-67	SCAFFO LD5	443172 2	443181 9	-	78	97	UCACAACCUUUUUGAG UGAG
nvit-mir-210	mir-210	SCAFFO LD6	105752 3	105760 7	-	51	71	UUGUGCGUGUGACAGC GGCUA
nvit-bantam	bantam	SCAFFO LD6	170407 5	170416 1	+	50	72	UGAGAUCAUUGUGAAA GCUGAUU
nvit-mir-9a	mir-9	SCAFFO LD6	179037 2	179045 8	+	15	37	UCUUUGGUUAUCUAGC UGUAUGA
nvit-mir-8	mir-8	SCAFFO LD7	230096 0	230104 5	-	53	75	UAAUACUGUCAGGUAA AGAUGUC
nvit-mir-375	mir-375	SCAFFO LD7	252713 4	252723 5	+	65	86	UUUGUUCGUUCGGCUC GAGUUA
nvit-mir-927	mir-927	SCAFFO LD70	57154	57255	-	21	43	UUUUAGAAUCCUACG CUUUACC
nvit-mir-252	mir-252	SCAFFO LD2	383975 1	383985 1	+	16	37	CUAAGUACUAGUGCCGC AGGAG
nvit-mir-932	mir-932	SCAFFO LD2	542843 1	542853 0	-	16	37	UCAAUUCCGUAGUGCA UUGCAG
nvit-mir-263b	mir-263	SCAFFO LD22	177954 3	177962 3	-	11	30	UUGGCACUGGAAGAAU UCAC
nvit-mir-993	mir-993	SCAFFO LD23	165787 4	165799 2	-	85	107	GAAGCUCGUCUCUACAG GUAUCU
nvit-mir-928	mir-928	SCAFFO LD9	223053 4	223063 3	-	12	32	GUGGCUGUGGAAGCUC GCGAA
NULL	NULL	SCAFFO LD15	830357	830448	+	60	80	CAUAUUACCUCGUGGG AUUUC
NULL	NULL	SCAFFO LD40	180821 3	180829 2	+	10	33	AUUGGAUGAAUCCUAC CCGGUGAG

Primers used for the stem-loop qPCR

miRNA	Stem-loop - Reverse primer (5'-3')	Specific Tailed forward (5'-3')	Assessed
nvit-mir-14	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTAGGAGAG	ACACTCCAGCTGGGTGAGTC TTTTTCTC	✓
nvit-mir-29b	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACTGATTT	ACACTCCAGCTGGGTAGCAC CATTGAA	✓
nvit-mir-33	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGCAATGCAA	ACACTCCAGCTGGGGTGCAT TGTAGTTG	✓
nvit-mir-281	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACAAAGAG	ACACTCCAGCTGGGTGTCAT GGACTGC	✓
nvit-mir-7	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACAACAAA	ACACTCCAGCTGGGTGGAAG ACAGTGA	✓
nvi-mir-3a	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTCAGCTAT	ACACTCCAGCTGGGGGCAA GATGTCGGC	✗
nvit-mir-71	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTCACTACC	ACACTCCAGCTGGGTGAAAG ACATGGGT	✓
nvit-mir-13a	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTCACTCAA	ACACTCCAGCTGGGTATCAC AGCCATT	✓
nvit-mir-2	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGGCTTACCA	ACACTCCAGCTGGGTATCAC AGCCAGT	✗
nvit-mir-13b	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACTCGTCA	ACACTCCAGCTGGGTATCAC AGCCATT	✓
nvit-mir-2a	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGGCTCAACA	ACACTCCAGCTGGGTATCAC AGCCAGT	✗
nvit-mir-92a	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGATAGGCCG	ACACTCCAGCTGGGATTGCA CTTGCCC	✓
nvit-mir-12	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACCAGTAC	ACACTCCAGCTGGGTGAGTA TTACATCA	✓
nvit-mir-283	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGAGAATTAC	ACACTCCAGCTGGGAAATAT CAGCTGGT	✗
nvit-mir-184	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGGCCCTTAT	ACACTCCAGCTGGGTGGACG GAGAATG	✓
nvit-let-7	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGACTATAACA	ACACTCCAGCTGGGTGAGGT AGTAGGT	✓
nvit-mir-100	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGCACAAGT	ACACTCCAGCTGGGAACCCG TAGATCCG	✓
nvit-mir-125	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTCAACAAGT	ACACTCCAGCTGGGTCCCTG AGACCCTA	✓
nvit-mir-929	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGGGACTCCC	ACACTCCAGCTGGGATTGAC TCTAGTAG	✗
nvit-mir-190	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTACAACCA	ACACTCCAGCTGGGAGATAT GTTTGATA	✗
nvit-mir-iab-4as	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGCGGTATAC	ACACTCCAGCTGGGTTACGT ATACTGAA	✗
nvit-mir-279	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTTAATGAG	ACACTCCAGCTGGGTGACTA GATCCACA	✓
nvit-mir-34	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGCAACCAGC	ACACTCCAGCTGGGTGGCAG TGTGGTTA	✓
nvit-mir-277	CTCAACTGGTGTCTGGAGTCGGCA ATTCAGTTGAGTGTCTGAC	ACACTCCAGCTGGGTAAATG CACTATCT	✓

nvit-mir-275	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCGCGCGCT	ACACTCCAGCTGGGTCAGGT ACCTGAAG	✓
nvit-mir-iab-4	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGTTACGTA	ACACTCCAGCTGGGCGGTAT ACCTTCAG	✗
nvit-mir-305	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCAGAGCAC	ACACTCCAGCTGGGATTGTA CTTCATCA	✓
nvit-mir-276	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGAGAGCACG	ACACTCCAGCTGGGTAGGAA CTTCATAC	✓
nvit-mir-10	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGACAAATTC	ACACTCCAGCTGGGACCCTG TAGATCCG	✗
nvit-mir-315	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGCTTTCTG	ACACTCCAGCTGGG TTTTGATTGTTGCT	✓
nvit-mir-219	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCAAGAATT	ACACTCCAGCTGGG TGATTGTCCAAACG	✓
nvit-mir-1	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCTCCATAC	ACACTCCAGCTGGG TGGAATGTAAAGAA	✓
nvit-mir-307	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCTCACTCA	ACACTCCAGCTGGG TCACAACCTTTTTG	✓
nvit-mir-993	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGAGATACTT	ACACTCCAGCTGGG GAAGCTCGTCTCTA	✗
nvit-bantam	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGAATCAGCT	ACACTCCAGCTGGG TGAGATCATTGTGA	✓
nvit-mir-9a	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGTCATACAG	ACACTCCAGCTGGG TCTTTGGTTATCTA	✓
nvit-mir-8	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGACATCTT	ACACTCCAGCTGGG TAATACTGTCAGGT	✓
nvit-mir-375	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGTAACTCGA	ACACTCCAGCTGGG TTTGTTTCGTTCCGC	✓
nvit-mir-927	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGGTAAAGC	ACACTCCAGCTGGG TTTTAGAATTCTTA	✓
nvit-mir-252	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCTCCTGCG	ACACTCCAGCTGGG CTAAGTACTAGTGC	✓
nvit-mir-317	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGACTGAGAT	ACACTCCAGCTGGGTGAACA CAGCTGGT	✗
nvit-mir-925	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGCGAATGT	ACACTCCAGCTGGGAGGGAT TCGGTTTT	✗
nvit-mir-932	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCTGCAATG	ACACTCCAGCTGGG TCAATTCCGTAGTG	✓
nvit-mir-137	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGTACGTGTA	ACACTCCAGCTGGGTTATTG CTTGAGAA	✗
Reference RNA (UACGUCGA UUCUAACG CAGGC)	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGGCTGCGT	ACACTCCAGCTGGGTACGTC GATTCTAA	✓
nvit-mir-282	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGACAGACAA	ACACTCCAGCTGGG GATTTAGCCTCTCC	✗
Universal Reverse	CTCAACTGGTGTCGTGGAGTCGG		
Universal	CTCAACTGGTGTCGTGGAGTCGGC		

Reverse +C			
------------	--	--	--

List of predicted mRNA targets based on PITA algorithm

UTR	miRNA	Start	End	Seed	Loop	duple x	dG 5	dG3	dG0	dG1	dG open	ddG
utr 3NVIR000988 LOC100113774	mirna34	102	94	8:00:00	0	-26.2	-15	-	-	-	-5.88	-
utr 3NVIR002189 LOC100115197	mirna375	833	825	8:00:00	0	-17.36	-9.3	-	-	-	-0.7	-
utr 3NVIR000869 LOC100118083	mirna252	265	257	8:00:00	0	-20.14	-8.4	-	-32	-	-4.02	-
utr 3NVIR003469 GstO1	mirna315	324	316	8:00:00	0	-17.4	-6.8	-	-9.26	-7.25	-2	-
utr 3NVIR001141 LOC100118717	mirna71	56	48	8:00:00	0	-16	-9.7	-6.3	-	-	-0.81	-
utr 3NVIR001840 LOC100118323	mirna375	49	41	8:00:00	0	-19.9	-9.5	-	-	-12.2	-5.05	-
utr 3NVIR000763 LOC100118966	mirna275	197	189	8:00:00	0	-19.91	-	-	-	-	-5.12	-
utr 3NVIR000271 LOC100116428	mirna34	1571	1563	8:00:00	0	-20.7	-15	-5.7	-	-	-6.14	-
utr 3NVIR000684 LOC100120186	mirna375	447	439	8:00:00	0	-20.75	-9.5	-	-	-	-6.27	-
utr 3NVIR001752 LOC100121922	mirna375	594	586	8:00:00	0	-20	-9.5	-	-	-	-5.7	-
utr 3NVIR001423 LOC100123339	mirna275	184	176	8:00:00	0	-19.4	-	-6.5	-	-	-5.21	-
utr 3NVIR003118 Emc	mirna375	656	648	8:00:00	0	-15.79	-9.5	-	-	-27.3	-1.63	-
utr 3NVIR002285 LOC100114283	mirna375	106	98	8:00:00	0	-19.2	-9.2	-10	-26.3	-	-5.11	-
utr 3NVIR001094 LOC100123437	mirna275	119	111	8:00:00	0	-22.31	-	-	-45.9	-	-8.34	-
utr 3NVIR002713 LOC100116888	mirna375	205	197	8:00:00	0	-21	-9.5	-	-24.6	-	-7.22	-
utr 3NVIR000870 LOC100118454	mirna277	499	491	8:00:00	0	-20.8	-8.4	-	-	-	-7.05	-
utr 3NVIR001678 LOC100118803	mirna275	166	158	8:00:00	0	-22	-	-8.2	-	-	-8.45	-
utr 3NVIR001464 LOC100120778	mirna305	53	45	8:00:00	0	-19.7	-8.2	-	-	-18.8	-6.26	-
utr 3NVIR002908 LOC100115609	mirna375	101	93	8:00:00	0	-15.94	-9.3	-	-22.4	-	-2.52	-
utr 3NVIR000979 LOC100113505	mirna375	75	67	8:00:00	0	-15.6	-9.5	-6.1	-	-	-2.19	-13.4
utr 3NVIR000568	mirna375	42	34	8:00:00	0	-21.5	-9.5	-12	-	-	-8.44	-

LOC100120510										25.2 3	16.7 8		13.0 5				
utr 3NVIR003443 Dsx	mirna276	541	533	8:00:00	0	-21.11	-	-	-	10.9 2	10.9 1	37.5 2	29.4 6	-8.05	-	13.0 5	
utr 3NVIR003530 Dsx	mirna276	685	677	8:00:00	0	-21.11	-	-	-	10.9 2	10.9 1	37.5 2	29.4 6	-8.05	-	13.0 5	
utr 3NVIR000892 LOC100114880	mirna375	269	261	8:00:00	0	-17	-9.5	-7.5	-	-	-	32.3 1	28.3 6	-3.95	-	13.0 4	
utr 3NVIR000915 LOC100118125	mirna305	268	260	8:00:00	0	-16.8	-8.2	-8.6	-	-	-	26.8 6	23.0 8	-3.77	-	13.0 2	
utr 3NVIR001012 LOC100114836	mirna1	36	28	8:00:00	0	-14.8	-	-4.4	-	10.4 4	-	14.0 4	12.1 5	-1.88	-	12.9 1	
utr 3NVIR001105 LOC100123624	mirna1	224	216	8:00:00	0	-13.73	-	-	-	10.4 4	3.33	22.4 3	21.6 2	-0.81	-	12.9 1	
utr 3NVIR000536 LOC100119520	mirna305	500	492	8:00:00	0	-17	-8.2	-8.8	-	-	-	37.6 4	33.3 7	-4.26	-	12.7 3	
utr 3NVIR002000 LOC100121412	mirna34	37	29	8:00:00	0	-18.1	-15	-3.1	-	-	-	10.4 7	-4.82	-5.65	-	12.4 4	
utr 3NVIR000311 LOC100123516	mirna277	290	282	8:00:00	0	-21.6	-9	-	-	12.6	-	35.5 2	26.3 2	-9.2	-	12.3 9	
utr 3NVIR002017 LOC100117219	mirna14	43	35	8:00:00	0	-19.6	-9.3	-	-	10.3	-	15.3 1	-7.91	-7.4	-	12.1 9	
utr 3NVIR003629 LOC100122234	mirna71	248	240	8:00:00	0	-18.4	-	-8.3	-	10.1	-	40.3 9	34.0 9	-6.29	-12.1		
utr 3NVIR003449 Cpap3-d1	mirna275	114	106	8:00:00	0	-19.9	-	-7	-	12.9	-	42.2 7	34.3 5	-7.92	-	11.9 7	
utr 3NVIR001608 LOC100113585	mirna71	1843	1835	8:00:00	0	-14.7	-	-4.1	-36.4	10.6	-	-	33.6 1	-2.79	-11.9		
utr 3NVIR002182 LOC100114421	mirna275	889	881	8:00:00	0	-17.4	-	-4.3	-	13.1	-	23.3 4	17.6 9	-5.65	-	11.7 4	
utr 3NVIR002340 LOC100115094	mirna305	234	226	8:00:00	0	-19.2	-7.9	-	-	11.3	-	27.3 8	19.8 6	-7.52	-	11.6 7	
utr 3NVIR002743 LOC100121836	mirna71	483	475	8:00:00	0	-14.7	-	-4.1	-	10.6	-	28.6 7	25.5 6	-3.1	-	11.5 9	
utr 3NVIR003615 cycB	mirna375	105	97	8:00:00	0	-19.52	-9.5	-	-	10.2	-	41.2 8	33.3 1	-7.97	-	11.5 4	
utr 3NVIR000276 LOC100117442	mirna375	473	465	8:00:00	0	-17.3	-9.3	-8	-	-	-	33.2 4	-27.3	-5.93	-	11.3 6	
utr 3NVIR003323 Pdf	mirna315	190	182	8:00:00	0	-12.3	-6.8	-5.5	-	-	-	30.5 4	-29.6	-0.93	-	11.3 6	
utr 3NVIR000445 LOC100118814	mirna34	149	141	8:00:00	0	-23	-15	-8	-41.5	-	-	-	29.8 5	11.64	-	11.3 5	
utr 3NVIR001984 LOC100118733	mirna252	152	144	8:00:00	0	-15.8	-8.2	-7.6	-	-	-	40.6 7	36.0 6	-4.61	-	11.1 8	
utr 3NVIR002550 LOC100122759	mirna305	207	199	8:00:00	0	-12.8	-8.2	-4.6	-	-	-	34.4 5	32.7 6	-1.69	-	11.1 1	
utr 3NVIR001946 LOC100120735	mirna13b	160	152	8:00:00	0	-14.51	-	-	-	11.1	3.41	25.3 4	21.6 2	-3.71	-	10.7 9	
utr 3NVIR003229 Ttc39b	mirna277	331	323	8:00:00	0	-15.05	-9	-	-	-	-	6.05	41.8	37.6	-4.27	-	10.7

									8	1		7
utr 3NVIR001966 LOC100115911	mirna315	74	66	8:00:00	0	-15.61	-6.8	-	-22.3	-	-4.86	-
								8.81		17.4		10.7
										4		4
utr 3NVIR001456 LOC100119748	mirna71	180	172	8:00:00	0	-17.34	-	-	-	-	-6.61	-
								10.6	6.74	29.9	23.3	10.7
								6	6	6	4	2
utr 3NVIR002468 LOC100122076	mirna375	86	78	8:00:00	0	-17	-9.3	-7.7	-	-8.77	-6.28	-
									15.0			10.7
									5			1
utr 3NVIR001829 LOC100116811	mirna375	476	468	8:00:00	0	-20.2	-9.5	-	-	-	-9.58	-
								10.7	39.9	30.3		10.6
									1	2		1
utr 3NVIR000920 LOC100118379	mirna13b	247	239	8:00:00	0	-23.33	-	-	-	-	-	-
								11.7	11.6	51.2	38.4	10.5
									3	5	8	5
utr 3NVIR001400 LOC100121904	mirna375	245	237	8:00:00	0	-13.34	-9.5	-	-	-	-2.96	-
								3.84	29.0	26.1		10.3
									9	2		7
utr 3NVIR002287 LOC100120328	mirna252	160	152	8:00:00	0	-14.81	-8.2	-	-	-	-4.46	-
								6.61	38.2	33.7		10.3
									1	4		4
utr 3NVIR000766 LOC100119640	mirna305	148	140	8:00:00	0	-18.48	-8	-	-	-	-8.22	-
								10.4	43.8	35.6		10.2
								8	6	4		5
utr 3NVIR000528 LOC100119245	mirna277	109	101	8:00:00	0	-16.41	-8.4	-	-	-	-6.18	-
								8.01	28.6	22.4		10.2
									3	4		2
utr 3NVIR000611 LOC100122597	mirna305	267	259	8:00:00	0	-17.9	-8	-9.9	-	-	-7.68	-
									40.7	33.0		10.2
									3	5		1
utr 3NVIR001634 LOC100116455	mirna13b	1261	1253	8:00:00	0	-15.59	-	-	-	-	-5.43	-
								11.1	4.49	22.5	17.1	10.1
									6	3		5
utr 3NVIR001413 LOC100122851	mirna13b	504	496	8:00:00	0	-19.7	-8	-	-	-	-9.55	-
								11.7	32.4	22.8		10.1
									1	5		4
utr 3NVIR001797 LOC100119722	mirna375	319	311	8:00:00	0	-16.21	-9.5	-	-	-24.5	-6.07	-
								6.71	30.5			10.1
									7			3
utr 3NVIR000256 LOC100115149	mirna375	83	75	8:00:00	0	-18.02	-9.5	-	-	-22.9	-7.91	-10.1
								8.52	30.8			
									1			
utr 3NVIR002843 LOC100115309	mirna305	190	182	8:00:00	0	-15.94	-8.2	-	-	-	-5.84	-
								7.74	39.1	33.3		10.0
									6	2		9
utr 3NVIR001037 LOC100116861	mirna1	720	712	8:00:00	0	-15.2	-	-4.3	-	-	-5.13	-
								10.9	25.8	20.7		10.0
									9	5		6
utr 3NVIR000850 LOC100120366	mirna1	126	118	8:00:00	0	-18.9	-	-8	-	-	-8.89	-10
								10.9	38.4	29.5		
									6	7		

Annotation list of putative targets based on DAVID algorithm

Putative targets	Protein ID (DAVID)	Species	Gene Name
CYCB	XP_001605826	Nasonia vitripennis	cyclin B
CYCB	NP_001154932	Nasonia vitripennis	cyclin B
EMC	NP_001135435	Nasonia vitripennis	extra macrochaetae
	XP_001600762	Nasonia vitripennis	similar to glutathione transferase o1
LOC100113505	XP_001599003	Nasonia vitripennis	similar to tropomyosin 1
LOC100113585	XP_001600552	Nasonia vitripennis	hypothetical protein LOC100113585
LOC100113774	XP_001602713	Nasonia vitripennis	similar to GA11059-PA
LOC100114283	XP_001599353	Nasonia vitripennis	similar to aminoacylase, putative
LOC100114421	XP_001599450	Nasonia vitripennis	similar to GA20978-PA
LOC100114880	XP_001599758	Nasonia vitripennis	similar to cathepsin o
LOC100115094	XP_001599173	Nasonia vitripennis	similar to GH07239p
LOC100115149	XP_001608244	Nasonia vitripennis	hypothetical protein LOC100115149
LOC100115197	XP_001599993	Nasonia vitripennis	similar to fk506-binding protein
LOC100115309	XP_001600193	Nasonia vitripennis	hypothetical protein LOC100115309
LOC100115609	XP_001600290	Nasonia vitripennis	similar to conserved hypothetical protein
LOC100115911	XP_001606776	Nasonia vitripennis	similar to uncharacterized conerved protein
LOC100116428	XP_001600930	Nasonia vitripennis	similar to CG16771-PA
LOC100116455	XP_001600947	Nasonia vitripennis	similar to Chm protein
LOC100116811	XP_001601211	Nasonia vitripennis	similar to CG11081-PB
LOC100116861	XP_001601249	Nasonia vitripennis	similar to GA14484-PA
LOC100116888	XP_001601269	Nasonia vitripennis	similar to leucine-rich transmembrane protein
LOC100117219	XP_001601523	Nasonia vitripennis	similar to enterophilin-2S
LOC100117442	XP_001601680	Nasonia vitripennis	hypothetical protein LOC100117442
LOC100118083	XP_001602148	Nasonia vitripennis	similar to sodium-dependent phosphate transporter
LOC100118125	XP_001604644	Nasonia vitripennis	similar to conserved hypothetical protein
LOC100118323	XP_001602320	Nasonia vitripennis	similar to abc transporter
LOC100118	XP_001606463	Nasonia	similar to ENSANGP00000012893

379		vitripennis	
LOC100118 454	XP_001602416	Nasonia vitripennis	similar to translational activator gcn1
LOC100118 717	XP_001602615	Nasonia vitripennis	similar to Ppib protein
LOC100118 803	XP_001607565	Nasonia vitripennis	similar to Wdr451 protein
LOC100118 814	XP_001602701	Nasonia vitripennis	similar to tnf receptor associated factor
LOC100118 966	XP_001599313	Nasonia vitripennis	similar to myosin 2 light chain
LOC100119 245	XP_001600330	Nasonia vitripennis	similar to kazal-type proteinase inhibitor
LOC100119 520	XP_001604593	Nasonia vitripennis	similar to DEAD box polypeptide 5
LOC100119 640	XP_001600757	Nasonia vitripennis	hypothetical protein LOC100119640
LOC100119 722	XP_001603446	Nasonia vitripennis	similar to saposin
LOC100119 748	XP_001603473	Nasonia vitripennis	similar to ENSANGP00000012286
LOC100120 186	XP_001603848	Nasonia vitripennis	similar to chloride channel protein 3
LOC100120 328	XP_001603979	Nasonia vitripennis	similar to GA20103-PA
LOC100120 366	XP_001604014	Nasonia vitripennis	similar to GA17761-PA
LOC100120 510	XP_001600933	Nasonia vitripennis	similar to 6-phosphogluconate dehydrogenase
LOC100120 735	XP_001605023	Nasonia vitripennis	similar to ENSANGP00000010144
LOC100120 778	XP_001604382	Nasonia vitripennis	hypothetical protein LOC100120778
LOC100121 412	XP_001605025	Nasonia vitripennis	similar to trypsin
LOC100121 836	XP_001605445	Nasonia vitripennis	similar to conserved hypothetical protein
LOC100121 904	XP_001605510	Nasonia vitripennis	similar to GA19110-PA
LOC100121 922	XP_001605531	Nasonia vitripennis	similar to GM06507p
LOC100122 076	XP_001607755	Nasonia vitripennis	similar to RE67445p
LOC100122 234	XP_001607833	Nasonia vitripennis	similar to ENSANGP00000012279
LOC100122 234	NP_001162016	Nasonia vitripennis	similar to ENSANGP00000012279
LOC100122	XP_001606208	Nasonia vitripennis	similar to CG5973-PA
597			
LOC100122 759	XP_001606358	Nasonia vitripennis	similar to conserved hypothetical protein
LOC100122 851	XP_001606456	Nasonia vitripennis	similar to ENSANGP00000026211
LOC100123	XP_001606968	Nasonia	similar to receptor expression enhancing protein

339		vitripennis	
LOC100123 437	XP_001607080	Nasonia vitripennis	similar to transcriptional activator
LOC100123 516	XP_001608013	Nasonia vitripennis	similar to conserved hypothetical protein
LOC100123 624	XP_001607291	Nasonia vitripennis	similar to Transcription initiation factor IIB (General transcription factor TFIIB)
PDF	NP_001155850	Nasonia vitripennis	pigment-dispersing factor
TTC39B	NP_001153334	Nasonia vitripennis	tetratricopeptide repeat domain 39B

Gene Ontologies analysis of the predicted target genes

Seq. Name	Seq. Description	Seq. Length	#Hits	min. eValue	mean Similarity	#GOs	GOs	Enzyme Codes
gi 156542554 ref XP001599353.1	aminoacylase-1-like	401	20	0	79.75 %	5	C:cytoplasm; P:cellular amino acid metabolic process; F:peptidase activity; P:biological_process; F:hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	EC:3.5.1.14
gi 156542845 ref XP001599173.1	ankyrin repeat domain-containing protein 29-like	303	20	0	86.60 %	1	F:molecular_function	-
gi 156545299 ref XP001605445.1	glycerol-3-phosphate acyltransferase 3-like isoform 2	590	20	0	77.00 %	2	P:biological_process; F:transferase activity, transferring acyl groups	-
gi 156545557 ref XP001606463.1	heat shock 70 kda protein cognate 3 isoform x1	659	20	0	96.45 %	2	F:ion binding; P:response to stress	-
gi 156545922 ref XP001599003.1	tropomyosin-1	284	20	0	92.70 %	4	F:cytoskeletal protein binding; C:cytoskeleton; C:cytoplasm; P:anatomical structure development	-

gi 156546002 ref XP001607765.1	spry domain-containing protein 7-like	194	20	2.87 E-142	74.45 %	1	F:molecular_function	-
gi 156546454 ref XP001607291.1	transcription initiation factor iib	315	20	0	94.95 %	10	F:translation factor activity, nucleic acid binding; P:translation; F:transcription factor binding; P:cellular nitrogen compound metabolic process; P:biosynthetic process; F:ion binding; C:protein complex; C:intracellular; P:biological_process; C:nucleus	-
gi 156548007 ref XP001605510.1	major facilitator superfamily domain-containing protein 10	451	20	0	79.90 %	2	P:transmembrane transport; C:cellular_component	-
gi 156548400 ref XP001604382.1	eukaryotic translation initiation factor 4e type 3-like	178	20	1.77 E-131	79.70 %	3	C:cytoplasm; P:translation; F:translation factor activity, nucleic acid binding	-
gi 156549567 ref XP001607565.1	wd repeat domain phosphoinositide-interacting protein 3	342	20	0	90.40 %	7	C:cytoplasm; P:response to stress; P:catabolic process; P:cellular component assembly; F:lipid binding; F:ion	-

							binding; F:molecular_f unction	
gi 156550309 ref XP 001603446.1	proactivator polypeptide-like	111 3	20	0	62.95 %	2	C:lysosome; P:lipid metabolic process	-
gi 156551161 ref XP 001605023.1	dihydrolipoylly sine-residue succinyltransfer ase component of 2- oxoglutarate dehydrogenase mitochondrial- like	483	20	0	92.75 %	4	C:protein complex; C:cytoplasm; F:transferase activity, transferring acyl groups; P:generation of precursor metabolites and energy	EC:2.3. 1.61; EC:2.3. 1.12
gi 156554399 ref XP 001604593.1	atp-dependent rna helicase p62-like	551	20	0	87.50 %	6	F:helicase activity; F:ATPase activity; F:signal transducer activity; F:molecular_f unction; P:signal transduction; F:ion binding	-
gi 213972574 ref NP 001135435.1	extra macrochaetae	92	20	5.46 E- 58	74.50 %	5	C:cytoplasm; P:cellular nitrogen compound metabolic process; P:biosynthetic process; F:molecular_f unction; C:nucleus	-
gi 229577253 ref NP 001153334.1	tetratricopeptide repeat protein 39b-like	616	20	0	77.85 %	1	F:molecular_f unction	-
gi 238814342 ref NP 001154932.1	cyclin b	433	20	0	66.55 %	2	P:cell cycle; C:nucleus	-

gi 240849424 ref NP001155850.1	pigment-dispersing hormone peptides-like	81	7	4.44 E-52	68.71 %	3	C:extracellular region; P:biological_process; F:molecular_function	-
gi 270483886 ref NP001162016.1	c-type lectin 5 precursor	217	20	2.89 E-150	96.40 %	1	F:molecular_function	-
gi 288869494 ref NP001165854.1	cuticular protein analogous to peritrophins 3-d1 precursor	231	20	5.93 E-150	84.40 %	5	C:extracellular space; F:molecular_function; P:reproduction; F:structural molecule activity; P:biological_process	-
gi 345483318 ref XP0011602615.2	peptidyl-prolyl cis-trans rhodopsin-specific isozyme-like	234	20	2.14 E-152	78.70 %	15	P:signal transduction; P:transport; C:cell; C:nuclear envelope; C:cellular_component; F:isomerase activity; C:cytoplasmic membrane-bounded vesicle; P:pigmentation; P:biosynthetic process; C:endoplasmic reticulum; P:protein folding; F:molecular_function; P:response to stress; P:cell death; P:cellular protein modification	EC:5.2.1.8

							process	
gi 345484317 ref XP001604014.2	short branched chain specific acyl-mitochondrial-like	416	20	0	86.45 %	4	F:ion binding; F:oxidoreductase activity; C:mitochondrion; P:biological_process	EC:1.3.99
gi 345485535 ref XP001606456.2	mitochondrial uncoupling protein 2-like	320	20	0	75.75 %	3	P:transport; C:cellular_component; C:mitochondrion	-
gi 345487200 ref XP001600947.2	rab proteins geranylgeranyltransferase component a 1	597	20	0	72.90 %	5	P:biological_process; C:protein complex; C:cytoplasm; P:transport; F:transferase activity, transferring alkyl or aryl (other than methyl) groups	EC:2.5.1.60
gi 345490979 ref XP001600930.2	alkaline tissue-nonspecific isozyme-like	523	20	0	78.55 %	2	P:biological_process; F:phosphatase activity	-
gi 345494675 ref XP001603979.2	kelch-like protein 17	497	20	0	84.65 %	1	F:molecular_function	-
gi 345495146 ref XP001603848.2	h(+) cl(-) exchange transporter 5 isoform x1	820	20	0	87.50 %	5	F:molecular_function; F:transmembrane transporter activity; P:transmembrane transport; C:cellular_component; P:transport	-
gi 345497242 ref XP001599993.2	peptidyl-prolyl cis-trans isomerase	217	20	5.07 E-159	86.65 %	12	P:anatomical structure development;	EC:5.2.1.8

	fkbp14-like							P:homeostatic process; P:cellular protein modification process; F:ion binding; P:protein folding; P:signal transduction; C:extracellular space; F:isomerase activity; F:molecular function; P:biological process; C:cellular component; C:endoplasmic reticulum	
gi 345497741 ref XP001602148.2	inorganic phosphate cotransporter	496	20	0	79.85 %	2	P:transmembrane transport; C:cellular component	-	
gi 645022010 ref XP001599758.2	cathepsin o-like	334	20	0	67.50 %	2	P:biological process; F:peptidase activity	EC:3.4	
gi 645026108 ref XP001603473.3	cytoplasmic protein nck1 isoform x1	270	20	4.91 E-174	85.70 %	1	F:molecular function	-	
gi 645035764 ref XP001601269.3	leucine-rich repeat neuronal protein 3-like isoform x3	527	20	0	67.90 %	1	F:molecular function	-	
gi 645007639 ref XP008216351.1	serine protease 100 precursor	297	20	0	50.05 %	2	F:peptidase activity; P:biological process	EC:3.4.21	
gi 645006573 ref XP008204843.1	atp-binding cassette sub-family g member 4	696	20	0	85.90 %	8	F:ATPase activity; F:transmembrane transporter	EC:3.6.3.28	

	isoform x2							activity; C:cellular_co mponent; P:nucleobase- containing compound catabolic process; P:small molecule metabolic process; C:protein complex; C:plasma membrane; F:ion binding	
gi 645033417 ref XP 008214973.1	v-type atpase subunit b-like	340	20	0	97.25 %	13	F:ATPase activity; F:transmembr ane transporter activity; P:cellular nitrogen compound metabolic process; P:small molecule metabolic process; P:transmembr ane transport; P:homeostatic process; C:endosome; C:cell; C:protein complex; C:vacuole; C:cytoplasm; F:ion binding; C:plasma membrane	-	
gi 645003497 ref XP 008212581.1	spry domain- containing socs box protein 1- like	490	20	0	90.30 %	2	P:signal transduction; F:molecular_f unction	-	

gi 645003050 ref XP008208112.1	tnf receptor-associated factor 4-like	427	20	0	79.60 %	15	P:vesicle-mediated transport; F:molecular_function; F:ion binding; P:cell junction organization; P:response to stress; P:cell morphogenesis; P:cell differentiation; P:embryo development; P:biological_process; P:anatomical structure development; P:cellular protein modification process; C:cytoplasm; P:cell death; P:signal transduction; P:anatomical structure formation involved in morphogenesis	EC:6.3.2.19
gi 645002542 ref XP008214431.1	translational activator gcn1	2594	20	0	79.30 %	1	F:molecular_function	-
gi 645006456 ref XP008204796.1	plexin-a4 isoform x3	1443	20	0	87.20 %	5	C:cellular_component; P:signal transduction; P:biological_process; F:molecular_function; C:intracellular	-
gi 645035478 ref XP008216726.1	glutathione s-transferase omega-1-like	241	20	2.07 E-176	78.85 %	9	P:cellular nitrogen compound metabolic process;	EC:1.5.4.1; EC:2.5.1.18; EC:1.8.

							P:biosynthetic process; F:transferase activity, transferring alkyl or aryl (other than methyl) groups; P:sulfur compound metabolic process; P:cellular amino acid metabolic process; F:oxidoreductase activity; F:molecular_function; C:cytoplasm; P:pigmentation	5.1
gi 345498319 ref XP003428202.1	transcription factor e2f3	447	20	0	67.85 %	5	F:nucleic acid binding transcription factor activity; P:cellular nitrogen compound metabolic process; P:biosynthetic process; C:protein complex; C:intracellular	-
gi 645024451 ref XP008213863.1	palmitoyltransferase zdhhc3-like	273	20	0	90.65 %	2	F:ion binding; F:molecular_function	-

gi 645006092 ref XP008204321.1	tubulointerstitial nephritis antigen-like	401	20	0	70.10 %	6	C:extracellular space; F:molecular_function; P:signal transduction; C:external encapsulating structure; F:peptidase activity; P:biological_process	EC:3.4
gi 645024986 ref XP008212869.1	myosin light chain	150	20	1.03 E-103	86.50 %	1	F:ion binding	-

	<pre> a cgcuggc guau cgucaca c aucg ccagcacg g u --cuuag a uaa g u - aaa a </pre> <p>Get sequence</p>
Genome context	<i>Coordinates (Nvit_2.0)</i> chr4: 7941694-7941792 [-]
Clustered miRNAs	< 10kb from <i>nvi-mir-34</i> nvi-mir-277 chr4: 7946106-7946192 [-] nvi-mir-34 chr4: 7941694-7941792 [-]
Database links	ENTREZGENE: 100526334; Mir34

miRNA-275 and *miRNA-305*

Stem-loop sequence <i>nvi-mir-275</i>	
Accession	MI0014746
Description	Nasonia vitripennis miR-275 stem-loop
Gene family	MIPF0000187; mir-275
Stem-loop	<pre> uuacguacag u a u - c ac gc ccaguu gca c cgcgcuacu c gguacuu gacugu a u ggucag cgu ggcggaugagccauggacugaua u -----a - c cau -- gc </pre> <p>Get sequence</p>
Confidence	Feedback: Do you believe this miRNA is real?
Genome context	<i>Coordinates (Nvit_2.0)</i> chr4: 1482167-1482260 [+]
Clustered miRNAs	< 10kb from <i>nvi-mir-275</i> nvi-mir-275 chr4: 1482167-1482260 [+] nvi-mir-305 chr4: 1482355-1482444 [+]
Database links	ENTREZGENE: 100526336; Mir275

Melting curve profiles of miRNA qPCR

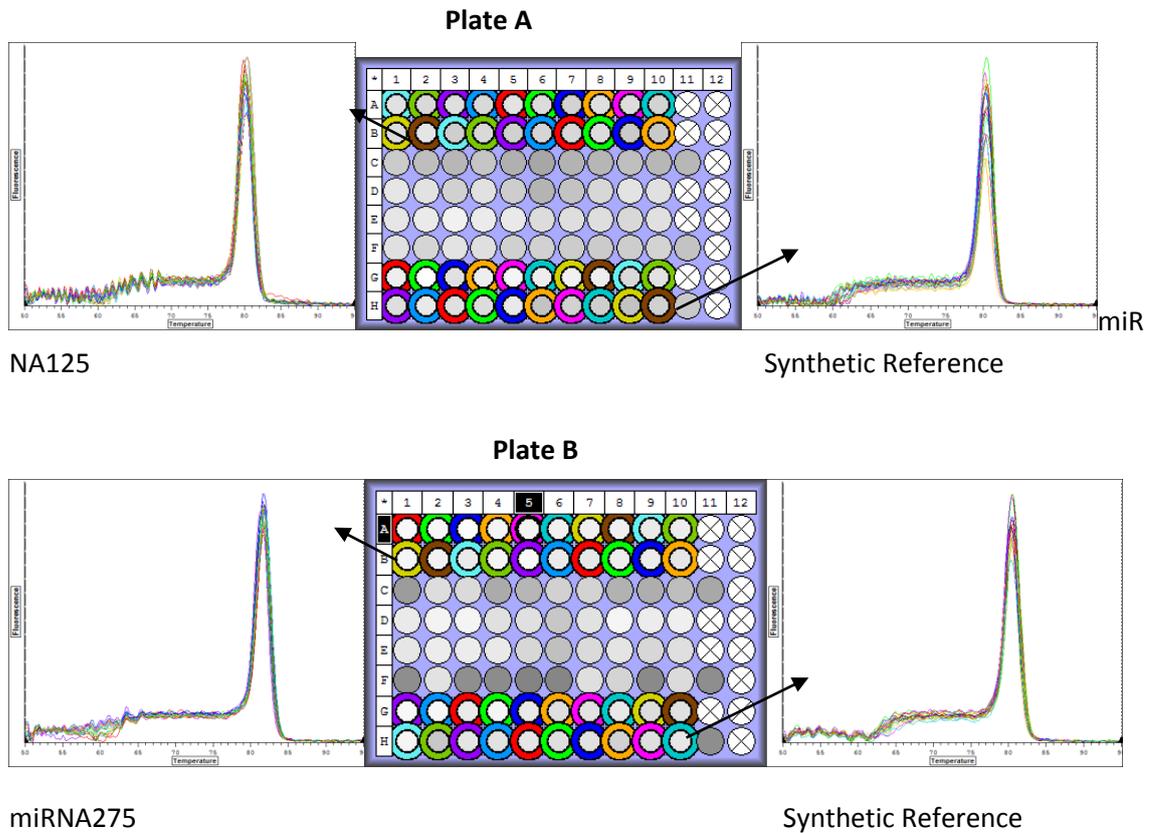


Figure 8.1: **Melting curve profiles generated by miRNA qPCR.** Specific miRNA was profiled by qPCR in separate 96-well plates. Row A and B indicate target miRNA reaction wells (1-10) and Row C,D indicate reference (exogenous) reaction wells in plate A and B. The melt curve profiles were identical for the reference in both the plates but differed for the target miRNA (miRNA125 Vs miRNA275) in respective plates.

8.3 pRNAi in Nasonia

List of primers used to amplify target gene for pRNAi in Nasonia

Primer	Sequence (5'-3') with T7 site at 5' end
<i>nvDnmt1a-forward</i>	TAATACGACTCACTATAGGGAGACCACGCTGTATTGGAGCGAGGAA
<i>nvDnmt1a-reverse</i>	TAATACGACTCACTATAGGGAGACCACGCATTGCAGTCTCCAGTGAA
<i>nvDnmt1b-forward</i>	TAATACGACTCACTATAGGGAGACCACGATCCAGGTCTTGCAGCACT
<i>NvDnmt1b-reverse</i>	TAATACGACTCACTATAGGGAGACCCTCTTGTATTATGGCCCTCCAG
<i>nvDnmt1c-forward</i>	TAATACGACTCACTATAGGGAGACCACGAGACAGTCTGCGTTGGTGA
<i>nvDnmt1c-reverse</i>	TAATACGACTCACTATAGGGAGACCACCATCAACAGTTCGAGGCTCA
<i>nvDnmt3-forward</i>	TAATACGACTCACTATAGGGAGACCCTGGCAAATAGTGGGTCACAA
<i>nvDnmt3-reverse</i>	TAATACGACTCACTATAGGGAGACCCTACCGGCATCTTCAGCTCTT
<i>nvDcr-1 forward</i>	TAATACGACTCACTATAGGGAGACCACCCTCTGGGCTGCAGATAAAG
<i>nvDcr-1 reverse</i>	TAATACGACTCACTATAGGGAGACCCTTCGGAATGTCGATTCCCTC
<i>GFP-forward</i>	TAATACGACTCACTATAGGGAGACCACCTGAAGTTCATCTGCACCA
<i>GFP-reverse</i>	TAATACGACTCACTATAGGGAGACCCTGCTCAGGTAGTGGTTGTCTG
<i>nvTET-2 forward</i>	TAATACGACTCACTATAGGGAGACCACGAAAAGCTAGCAACGGTTCG
<i>nvTET-2 reverse</i>	TAATACGACTCACTATAGGGAGACCACCACACGTTCTTGGCTCGTTA

List of primers used for validation of the knock-down by real-time PCR

Primer used	Sequences (5'-3')	Amplicon (bp)
<i>nvDnmt1a-</i>	GGGACGAGTTCCTTCATCCTG	130
<i>nvDnmt1a-reverse</i>	ACTGCATTGCCAACCTGTCT	
<i>nvDnmt1b-</i>	TGCTGTCATTGACGAAAAGC	125
<i>NvDnmt1b-reverse</i>	GGCATAGGTGTCCATTCCCTG	
<i>nvDnmt1c-</i>	TACTGTTTTGGTGGGGCATC	100
<i>nvDnmt1c-reverse</i>	AACATCATCTTTGCGGGGTA	
<i>nvDnmt3-forward</i>	GGGTTTCTTGGATTGGTGAA	130
<i>nvDnmt3-reverse</i>	TTTTTGCAAGTCACCTTGGTC	
<i>nvDcr-1 forward</i>	TTGATGAGGAAGATGCTCCA	112
<i>nvDcr-1 reverse</i>	TTCCATTCAGGAGTTAGTTTGG	
<i>AQR-forward</i>	TTGACGAGATGGTCTACAAGGCATC	140
<i>AQR-reverse</i>	GAAGGCTTCTACAGCATCTTTGTGTCGT	
<i>nvRpl32-forward</i>	GCCCAACATCGGTTATGGTA	145
<i>nvRpl32-reverse</i>	AACTCCATGGGCAATTTCTG	

8.4 Level of diapause after 9-10 days in *Dnmt1a* and *Dnmt3* knockdown progeny

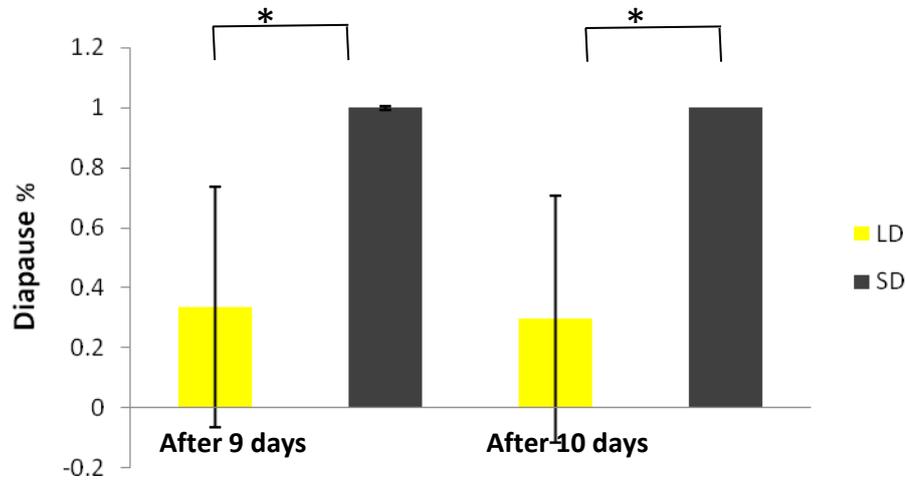


Figure 8.2: **Level of diapause after 9-10 days in *Dnmt1a* knockdown.** The level of diapause was significantly different between LD (18 hr light, yellow box) and SD (6 hr light, grey box) after 9 (Student-*t*-test, $p = 5.12697E-06$) and 10 days (Student-*t*-test, $p = 0.000204$) as analysed for 10-12 females in each group, progeny of each female collected from two hosts, average $n=26$ larvae per female). The error bars range from minimum to maximum (excluding outliers) at 95% confidence interval.

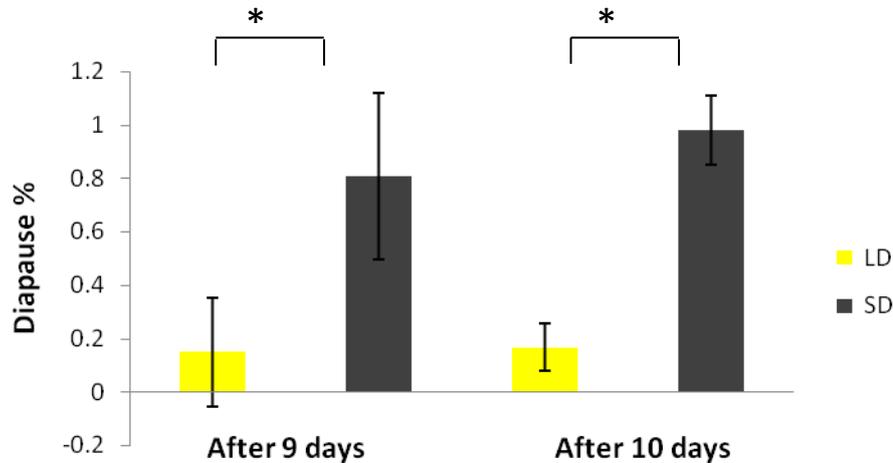
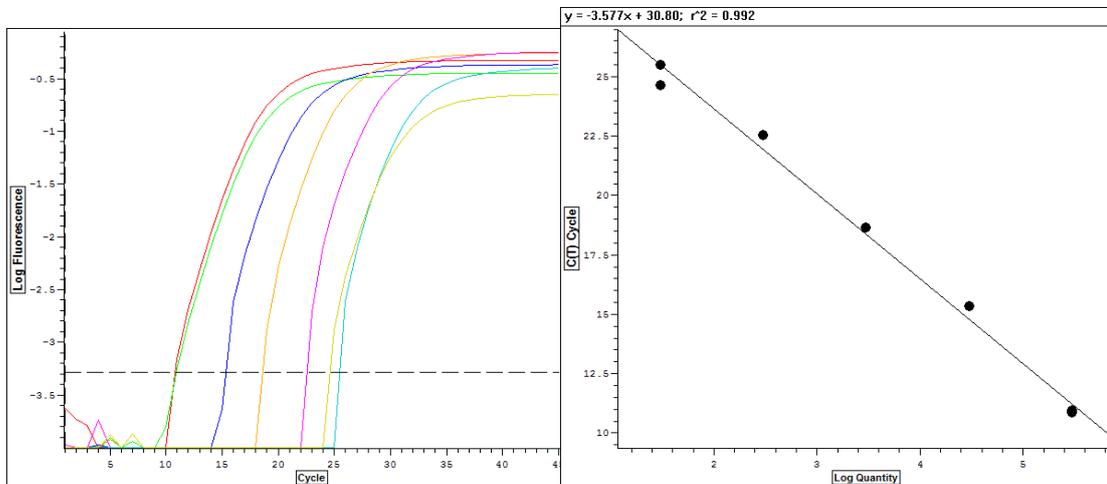


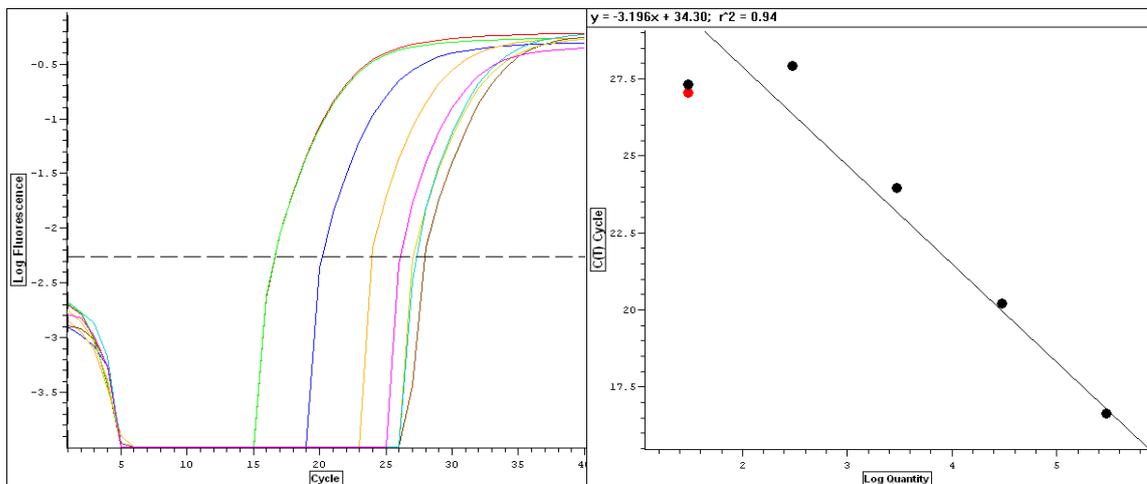
Figure 8.3: **Level of diapause after 9-10 days in *Dnmt3* knockdown.** The level of diapause was significantly different between LD (18 hr light, yellow box) and SD (6 hr light, grey box) after 9 (Student-*t*-test, $p = 6.57064E-22$) and 10 days (Student-*t*-test, $p = 2.36839E-61$) as analysed for 48-50 females in each group, progeny of each female collected from two hosts, average $n= 40$ larvae per female). The error bars range from minimum to maximum (excluding outliers) at 95% confidence interval.

8.5 Standard Curve Quantification graphs

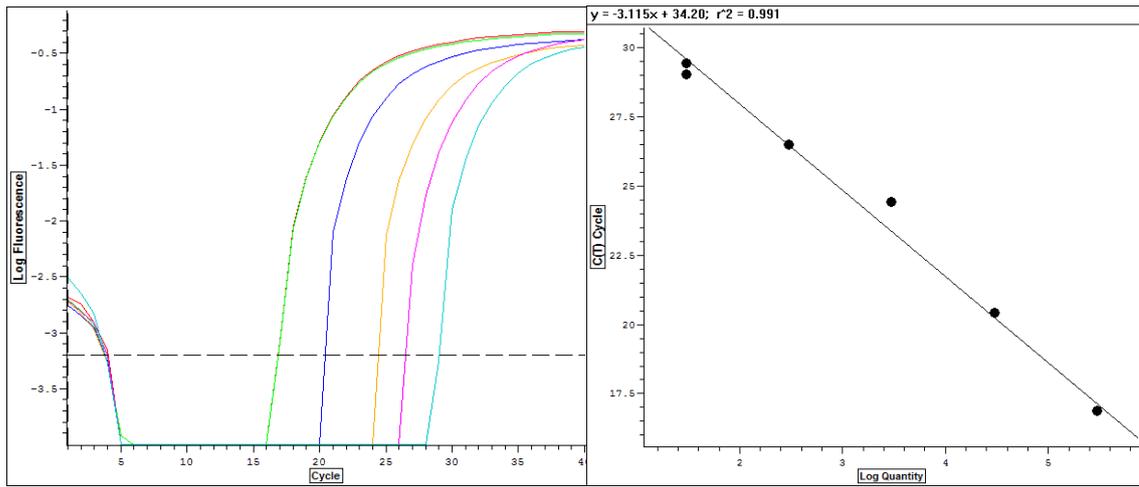
A) Dnmt1a standard graph



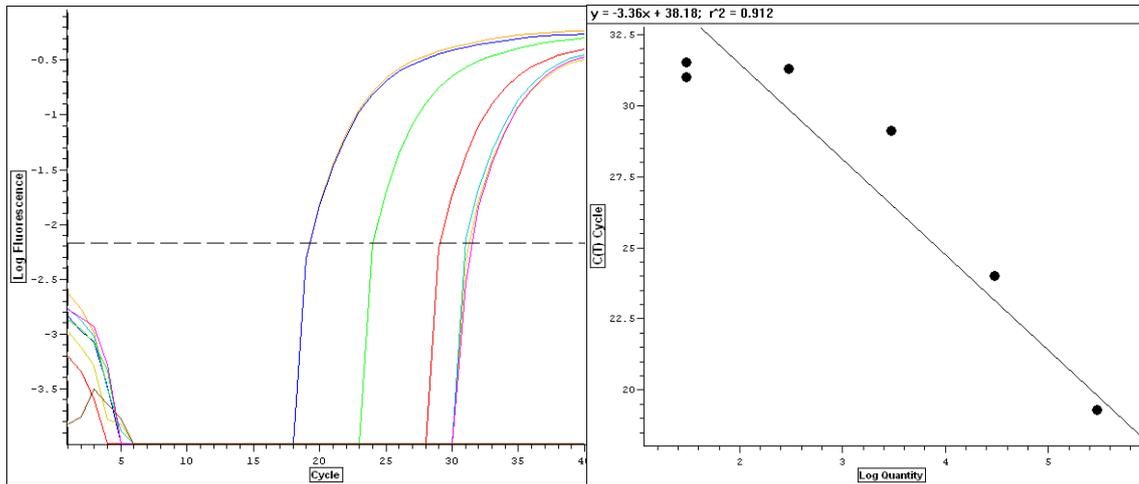
B) Dnmt1b standard graph



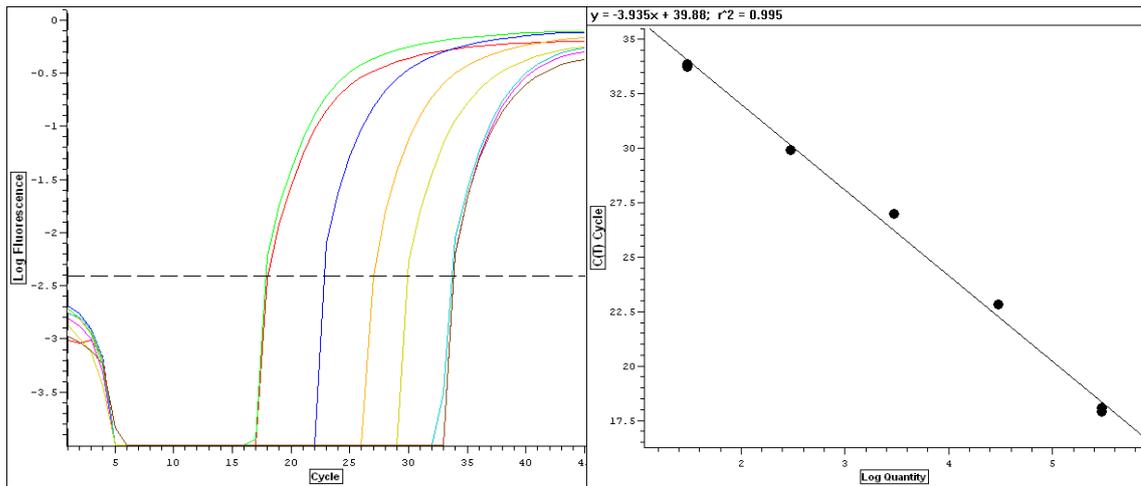
C) Dnmt1c standard graph



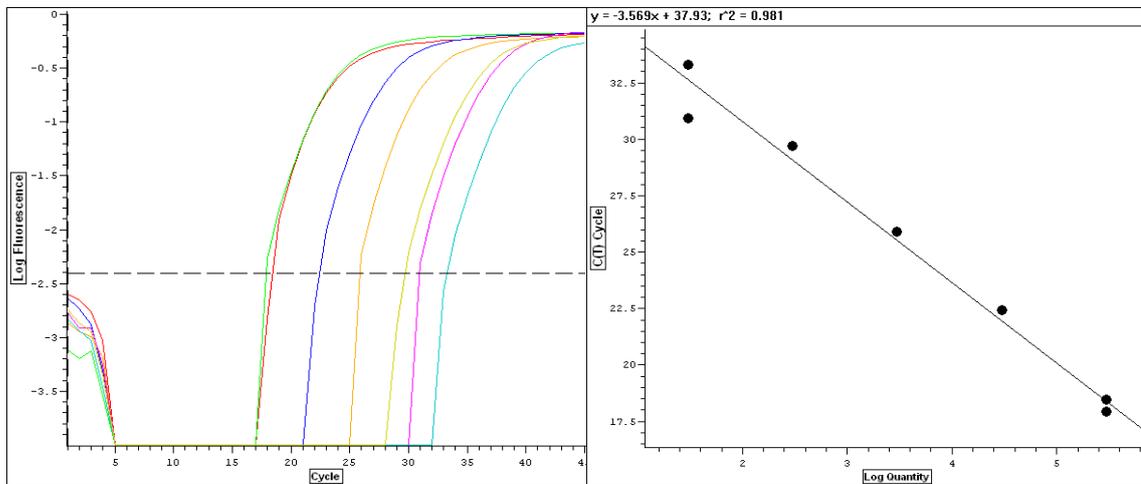
D) Dnmt3 standard graph



E) Dicer-1 standard graph



F) External Reference (Aequorin) standard graph



G) Internal Reference (nvit-RpL32) standard graph

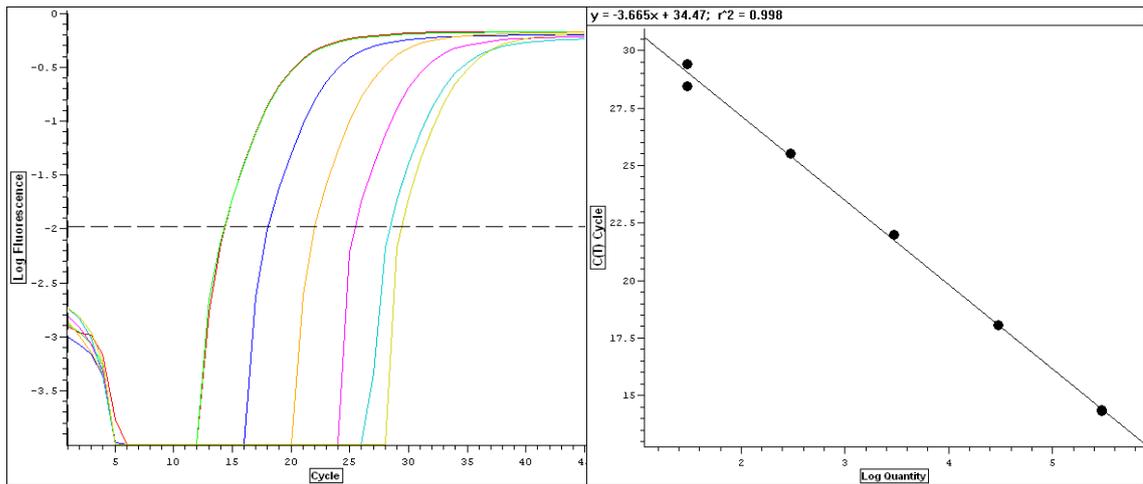


Figure 8.4: **Standard graphs for target and reference genes used in quantification analysis by qPCR.** On the left, amplification profile of the specific gene (*Dnmt1a*, *Dnmt1b*, *Dnmt1c*, *Dnmt3*, *Dicer-1*, *Aequorin*, *nvit RpL-32*) using standard five-fold dilutions (300000-30 copies) by real-time PCR. On the right, corresponding standard curves for the specific genes wherein Ct values were plotted against log₁₀ concentration of template. The equation above the curve was used to determine the correlation co-efficient (r^2) and slope of the curve.

CHAPTER 9

BIBLIOGRAPHY

- Aapola, U., Liiv, I., Peterson, P., 2002. Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. *Nucleic Acids Research*. **30**, 3602-3608.
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*. **13**, R87.
- Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., Warren, A.J., 2006. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *The EMBO Journal*. **25**, 4503-4512.
- Amarasinghe, H., Clayton, C., Mallon, E., 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc. R. Soc.* **281**, .
- Ambros, V., 2004. The functions of animal microRNAs. *Nature*. **431**, 350-355.
- Ancelin, K., Lange, U.C., Hajkova, P., Schneider, R., Bannister, A.J., Kouzarides, T., Surani, M.A., 2006. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biology*. **8**, 623-630.
- Augui, S., Nora, E.P., Heard, E., 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews.Genetics*. **12**, 429-442.
- Bali, K.K., Hackenberg, M., Lubin, A., Kuner, R., Devor, M., 2014. Sources of individual variability: miRNAs that predispose to neuropathic pain identified using genome-wide sequencing. *Molecular Pain*. **10**, 22-8069-10-22.
- Bannister, A.J. & Kouzarides, T., 2005. Reversing histone methylation. *Nature*. **436**, 1103-1106.
- Barrett, D.M., Duggavathi, R., Davies, K.L., Bartlewski, P.M., Bagu, E.T., Rawlings, N.C., 2007. Differential effects of various estradiol-17beta treatments on follicle-stimulating hormone peaks, luteinizing hormone pulses, basal gonadotropin concentrations, and antral follicle and luteal development in cyclic ewes. *Biology of Reproduction*. **77**, 252-262.
- Baskerville, S. & Bartel, D.P., 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA (New York, N.Y.)*. **11**, 241-247.
- Beck, S.D., 1980. *Insect photoperiodism*. 2nd ed. New York: Academic Press.
- Bestor, T.H., 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics*. **9**, 2395-2402.
- Bird, A., 2007. Perceptions of epigenetics. *Nature*. **447**, 396-398.

- Bird, A.P., Taggart, M.H., Smith, B.A., 1979a. Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell*. **17**, 889-901.
- Bird, A.P., Taggart, M.H., Smith, B.A., 1979b. Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell*. **17**, 889-901.
- Blaney, L.T. & Hamner, K.C., 1957. Interrelations among effects of temperature, photoperiod and dark periods on floral initiation of Biloxi soybean. *Botan. Gaz.* **119**, 10-24.
- Boorstein, R.J., Cummings, A., Jr, Marenstein, D.R., Chan, M.K., Ma, Y., Neubert, T.A., Brown, S.M., Teebor, G.W., 2001. Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *The Journal of Biological Chemistry*. **276**, 41991-41997.
- Bradshaw, W.E. & Holzapfel, C.M., 2010. Light, time, and the physiology of biotic response to rapid climate change in animals. *Annual Review of Physiology*. **72**, 147-166.
- Bradshaw, W.E. & Holzapfel, C.M., 2008. Genetic response to rapid climate change: it's seasonal timing that matters. *Molecular Ecology*. **17**, 157-166.
- Bradshaw, W.E., Holzapfel, C.M., Mathias, D., 2006. Circadian rhythmicity and photoperiodism in the pitcher-plant mosquito: can the seasonal timer evolve independently of the circadian clock? *The American Naturalist*. **167**, 601-605.
- Bradshaw, W.E., Quebodeaux, M.C., Holzapfel, C.M., 2003. Circadian rhythmicity and photoperiodism in the pitcher-plant mosquito: adaptive response to the photic environment or correlated response to the seasonal environment? *The American Naturalist*. **161**, 735-748.
- Bradshaw, W.E. & Holzapfel, C.M., 2001. Phenotypic evolution and the genetic architecture underlying photoperiodic time measurement. *Journal of Insect Physiology*. **47**, 809-820.
- Bradshaw, W.E. & Holzapfel, C.M., 2001. Genetic shift in photoperiodic response correlated with global warming. *Proceedings of the National Academy of Sciences of the United States of America*. **98**, 14509-14511.
- Bradshaw, W.E. & Holzapfel, C.M., 2010. What season is it anyway? Circadian tracking vs. photoperiodic anticipation in insects. *Journal of Biological Rhythms*. **25**, 155-165.
- Branco, R., Gabriella, F., Wolf R, 2012. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *13*, 7-13.
- Brennecke, J., Stark, A., Russell, R.B., Cohen, S.M., 2005. Principles of microRNA-target recognition. *PLoS Biology*. **3**, e85.
- Brooks, W.H., Le Dantec, C., Pers, J.O., Youinou, P., Renaudineau, Y., 2010. Epigenetics and autoimmunity. *Journal of Autoimmunity*. **34**, J207-19.
- Buchan, J.R. & Parker, R., 2007. Molecular biology. The two faces of miRNA. *Science (New York, N.Y.)*. **318**, 1877-1878.

- Bunning, E., 1936. Die endogene Tagesrhythmik als Grundlage der Photoperiodischen Reaktion. *Berichte Der Deutschen Botanischen Gesellschaft*. **54**, 590-607.
- Burn, J.E., Bagnall, D.J., Metzger, J.D., Dennis, E.S., Peacock, W.J., 1993. DNA methylation, vernalization, and the initiation of flowering. *Proceedings of the National Academy of Sciences of the United States of America*. **90**, 287-291.
- Butcher, L.M. & Beck, S., 2010. AutoMedIP-seq: a high-throughput, whole genome, DNA methylation assay. *Methods (San Diego, Calif.)*. **52**, 223-231.
- Cai, X., Hagedorn, C.H., Cullen, B.R., 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA (New York, N.Y.)*. **10**, 1957-1966.
- Carotti, D., Funiciello, S., Palitti, F., Strom, R., 1998. Influence of pre-existing methylation on the de novo activity of eukaryotic DNA methyltransferase. *Biochemistry*. **37**, 1101-1108.
- Cavalli, G. & Paro, R., 1998. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell*. **93**, 505-518.
- Chawla, G. & Sokol, N.S., 2012. Hormonal activation of let-7-C microRNAs via EcR is required for adult Drosophila melanogaster morphology and function. *Development (Cambridge, England)*. **139**, 1788-1797.
- Chen, T. P., Ueda, Y., Dodge, J. E., Wang, Z. J., & Li, E., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol*. **23**, 5594-5605.
- Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., Guegler, K.J., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*. **33**, e179.
- Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E., Shiekhattar, R., 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature*. **447**, 823-828.
- Choi, Y.J., Lin, C.P., Ho, J.J., He, X., Okada, N., Bu, P., Zhong, Y., Kim, S.Y., Bennett, M.J., Chen, C., Ozturk, A., Hicks, G.G., Hannon, G.J., He, L., 2011. miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nature Cell Biology*. **13**, 1353-1360.
- Chomczynski, P. & Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*. **162**, 156-159.
- Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*. **21**, 5483-5495.
- Christman, J.K., Mendelsohn, N., Herzog, D., Schneiderman, N., 1983. Effect of 5-azacytidine on differentiation and DNA methylation in human promyelocytic leukemia cells (HL-60). *Cancer Research*. **43**, 763-769.

- Christman, J.K., Sheikhnejad, G., Marasco, C.J., Sufrin, J.R., 1995. 5-Methyl-2'-deoxycytidine in single-stranded DNA can act in cis to signal de novo DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*. **92**, 7347-7351.
- Chuang, J.C. & Jones, P.A., 2007. Epigenetics and microRNAs. *Pediatric Research*. **61**, 24R-29R.
- Cihak, A., 1974. Biological effects of 5-azacytidine in eukaryotes. *Oncology*. **30**, 405-422.
- Clifford, R.L., Singer, C.A., John, A.E., 2013. Epigenetics and miRNA emerge as key regulators of smooth muscle cell phenotype and function. *Pulmonary Pharmacology & Therapeutics*. **26**, 75-85.
- Clouaire, T., de Las Heras, J.I., Merusi, C., Stancheva, I., 2010. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Research*. **38**, 4620-4634.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L.K., Bartolomei, M.S., Rambow, F., Bassi, M.R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A.J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., Bellacosa, A., 2011. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*. **146**, 67-79.
- Costa, F.F., 2007. Non-coding RNAs: lost in translation? *Gene*. **386**, 1-10.
- Costa, F.F., 2005. Non-coding RNAs: new players in eukaryotic biology. *Gene*. **357**, 83-94.
- Creusot, F., Acs, G., Christman, J.K., 1982. Inhibition of DNA methyltransferase and induction of Friend erythroleukemia cell differentiation by 5-azacytidine and 5-aza-2'-deoxycytidine. *The Journal of Biological Chemistry*. **257**, 2041-2048.
- Cullen, B.R., 2004. Transcription and processing of human microRNA precursors. *Molecular Cell*. **16**, 861-865.
- Danks, H.V., 2005. How similar are daily and seasonal biological clocks? *Journal of Insect Physiology*. **51**, 609-619.
- Darling, D. & Werren, J., 1990. **Biosystematics of Nasonia (Hymenoptera: Pteromalidae): two new species reared from birds' nests in North America**. *Annals of the Entomological Society of America*. **83**, 352-370.
- Dennis, E.S., Bilodeau, P., Burn, J., Finnegan, E.J., Genger, R., Helliwell, C., Kang, B.J., Sheldon, C.C., Peacock, W.J., 1998. Methylation controls the low temperature induction of flowering in Arabidopsis. *Symposia of the Society for Experimental Biology*. **51**, 97-103.
- De Neve, J., Thas, O., Ottoy, J.P., Clement, L., 2013. An extension of the Wilcoxon-Mann-Whitney test for analyzing RT-qPCR data. *Statistical Applications in Genetics and Molecular Biology*. **12**, 333-346.

- Dickman, M.J., Kucharski, R., Maleszka, R., Hurd, P.J., 2013. Extensive histone post-translational modification in honey bees. *Insect Biochemistry and Molecular Biology*. **43**, 125-137.
- Dolezel, D., Vaneckova, H., Sauman, I., Hodkova, M., 2005. Is period gene causally involved in the photoperiodic regulation of reproductive diapause in the linden bug, *Pyrrhocoris apterus*? *Journal of Insect Physiology*. **51**, 655-659.
- Dombrowsky, A., Arthaud, L., Ledger, T.N., Tares, S., Robichon, A., 2009. Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction. *Genome Research*. **19**, 2052-2063.
- Downen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Downen, J.M., Nery, J.R., Dixon, J.E., Ecker, J.R., 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, E2183-91.
- Dugast-Darzacq, C. & Grange, T., 2009. MethylQuant: a real-time PCR-based method to quantify DNA methylation at single specific cytosines. *Methods in Molecular Biology (Clifton, N.J.)*. **507**, 281-303.
- Egger, G., Liang, G., Aparicio, A., Jones, P.A., 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. **429**, 457-463.
- Elango, N., Hunt, B.G., Goodisman, M.A., Yi, S.V., 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 11206-11211.
- Emerson, K.J., Bradshaw, W.E., Holzapfel, C.M., 2009. Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause. *Trends in Genetics : TIG*. **25**, 217-225.
- Emerson, K., Letaw, A., Bradshaw, W., Holzapfel, C., 2008. Extrinsic light:dark cycles, rather than endogenous circadian cycles, affect the photoperiodic counter in the pitcher-plant mosquito, *Wyeomyia smithii*. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*.
- Fabbri, M., Calore, F., Paone, A., Galli, R., Calin, G.A., 2013. Epigenetic regulation of miRNAs in cancer. *Advances in Experimental Medicine and Biology*. **754**, 137-148.
- Fatemi, M., Hermann, A., Pradhan, S., Jeltsch, A., 2001. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *Journal of Molecular Biology*. **309**, 1189-1199.
- Feil, R. & Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics*. **13**, 97-109.
- Feliciello, I., Parazajder, J., Akrap, I., Ugarkovic, D., 2013. First evidence of DNA methylation in insect *Tribolium castaneum*: environmental regulation of DNA methylation within heterochromatin. *Epigenetics : Official Journal of the DNA Methylation Society*. **8**, 534-541.

- Feng, S., Jacobsen, S.E., Reik, W., 2010. Epigenetic reprogramming in plant and animal development. *Science (New York, N.Y.)*. **330**, 622-627.
- Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews.Genetics*. **9**, 102-114.
- Flores, K.B. & Amdam, G.V., 2011. Deciphering a methylome: what can we read into patterns of DNA methylation? *The Journal of Experimental Biology*. **214**, 3155-3163.
- Flores, K., Wolschin, F., Corneveaux, J.J., Allen, A.N., Huentelman, M.J., Amdam, G.V., 2012. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*. **13**, 480-2164-13-480.
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., Maleszka, R., 2012a. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, 4968-4973.
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., Maleszka, R., 2012b. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, 4968-4973.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., Leonhardt, H., 2011. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One*. **6**, e21306.
- Fuks, F., Hurd, P.J., Deplus, R., Kouzarides, T., 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research*. **31**, 2305-2312.
- Gavery, M.R. & Roberts, S.B., 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*. **11**, 483-2164-11-483.
- Gerlach, D., Kriventseva, E.V., Rahman, N., Vejnar, C.E., Zdobnov, E.M., 2009. miROrtho: computational survey of microRNA genes. *Nucleic Acids Research*. **37**, D111-7.
- Giltsbach, R., Kouta, M., Bonisch, H., Bruss, M., 2006. Comparison of in vitro and in vivo reference genes for internal standardization of real-time PCR data. *Biotechniques*. **40**, 173-177.
- Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., Schier, A.F., 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science (New York, N.Y.)*. **308**, 833-838.
- Glastad, K.M., Hunt, B.G., Yi, S.V., Goodisman, M.A., 2011. DNA methylation in insects: on the brink of the epigenomic era. *Insect Molecular Biology*. **20**, 553-565.

- Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M., Carell, T., 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One*. **5**, e15367.
- Goldberg, A.D., Allis, C.D., Bernstein, E., 2007. Epigenetics: a landscape takes shape. *Cell*. **128**, 635-638.
- Goldman, B.D., 2001. Mammalian photoperiodic system: formal properties and neuroendocrine mechanisms of photoperiodic time measurement. *Journal of Biological Rhythms*. **16**, 283-301.
- Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., Bestor, T.H., 2006. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science (New York, N.Y.)*. **311**, 395-398.
- Gomez-Orte, E. & Belles, X., 2009. MicroRNA-dependent metamorphosis in hemimetabolan insects. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 21678-21682.
- Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*. **36**, 3420-3435.
- Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E., Pesole, G., 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Research*. **38**, D75-80.
- Grosshans, H., Johnson, T., Reinert, K.L., Gerstein, M., Slack, F.J., 2005. The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in *C. elegans*. *Developmental Cell*. **8**, 321-330.
- Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., Meissner, A., 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*. **6**, 468-481.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., Song, H., 2011a. Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle (Georgetown, Tex.)*. **10**, 2662-2668.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., Song, H., 2011b. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. **145**, 423-434.
- Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G.L., Song, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*. **17**, 215-222.

- Guo, X., Su, S., Skogerboe, G., Dai, S., Li, W., Li, Z., Liu, F., Ni, R., Guo, Y., Chen, S., Zhang, S., Chen, R., 2013. Recipe for a busy bee: microRNAs in Honey Bee caste determination. *PLoS One*. **8**, e81661.
- Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., Surani, M.A., 2013. Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. *Stem Cell Reports*. **1**, 518-531.
- Hahn, J.S., Hu, Z., Thiele, D.J., Iyer, V.R., 2004. Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Molecular and Cellular Biology*. **24**, 5249-5256.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., Surani, M.A., 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature*. **452**, 877-881.
- Hall, J.C., 2003. Genetics and molecular biology of rhythms in Drosophila and other insects. *Adv Genet*. **48**, 1-280.
- Hanon, E.A., Lincoln, G.A., Fustin, J.M., Dardente, H., Masson-Pevet, M., Morgan, P.J., Hazlerigg, D.G., 2008. Ancestral TSH mechanism signals summer in a photoperiodic mammal. *Current Biology : CB*. **18**, 1147-1152.
- Hardin, P.E., 2005. The circadian timekeeping system of Drosophila. *Current Biology : CB*. **15**, R714-22.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., Cheng, X., 2012. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*. **40**, 4841-4849.
- Hata, K., Okano, M., Lei, H., Li, E., 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development (Cambridge, England)*. **129**, 1983-1993.
- Hayatsu, H., Wataya, Y., Kai, K., Iida, S., 1970. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*. **9**, 2858-2865.
- Hazlerigg, D., 2012. The evolutionary physiology of photoperiodism in vertebrates. *Progress in Brain Research*. **199**, 413-422.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C., Xu, G.L., 2011a. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)*. **333**, 1303-1307.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C., Xu, G.L., 2011b. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)*. **333**, 1303-1307.
- HERSHEY, A.D., DIXON, J., CHASE, M., 1953. Nucleic acid economy in bacteria infected with bacteriophage T2. I. Purine and pyrimidine composition. *The Journal of General Physiology*. **36**, 777-789.

- Hodkova, 2002. Photoperiodic regulation of the phospholipid molecular species composition in thoracic muscles and fat body of *Pyrrhocoris apterus* (Heteroptera) via an endocrine gland, corpus allatum.
- Hollander, J.A., Im, H.I., Amelio, A.L., Kocerha, J., Bali, P., Lu, Q., Willoughby, D., Wahlestedt, C., Conkright, M.D., Kenny, P.J., 2010. Striatal microRNA controls cocaine intake through CREB signalling. *Nature*. **466**, 197-202.
- Hsieh, C.L., 2005. The de novo methylation activity of Dnmt3a is distinctly different than that of Dnmt1. *BMC Biochemistry*. **6**, 6.
- Huang da, W., Sherman, B.T., Zheng, X., Yang, J., Imamichi, T., Stephens, R., Lempicki, R.A., 2009. Extracting biological meaning from large gene lists with DAVID. *Current Protocols in Bioinformatics / Editorial Board, Andreas D.Baxevanis ...[Et Al.]*. **Chapter 13**, Unit 13.11.
- Hunt, B.G., Brisson, J.A., Yi, S.V., Goodisman, M.A., 2010. Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biology and Evolution*. **2**, 719-728.
- Hunt, B.G., Glastad, K.M., Yi, S.V., Goodisman, M.A., 2013. The function of intragenic DNA methylation: insights from insect epigenomes. *Integrative and Comparative Biology*. **53**, 319-328.
- Hutvagner, G. & Zamore, P.D., 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science (New York, N.Y.)*. **297**, 2056-2060.
- Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhashi, I., Meshi, T., Ishikawa, M., 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Molecular Cell*. **39**, 282-291.
- Inoue, A. & Zhang, Y., 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science (New York, N.Y.)*. **334**, 194.
- Iqbal, K., Jin, S.G., Pfeifer, G.P., Szabo, P.E., 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*. **108**, 3642-3647.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. **466**, 1129-1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*. **333**, 1300-1303.
- Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., Tomari, Y., 2010. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Molecular Cell*. **39**, 292-299.

- Jackson, A.L. & Linsley, P.S., 2010. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery*. **9**, 57-67.
- Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*. **293**, 1074-1080.
- Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., Pfeifer, G.P., 2011. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Research*. **71**, 7360-7365.
- Jones, P.A. & Takai, D., 2001. The role of DNA methylation in mammalian epigenetics. *Science (New York, N.Y.)*. **293**, 1068-1070.
- Karan, R., DeLeon, T., Biradar, H., Subudhi, P.K., 2012. Salt stress induced variation in DNA methylation pattern and its influence on gene expression in contrasting rice genotypes. *PloS One*. **7**, e40203.
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., Sasaki, H., 2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human Molecular Genetics*. **16**, 2272-2280.
- Kawaji, H. & Hayashizaki, Y., 2008. Exploration of small RNAs. *PLoS Genetics*. **4**, e22.
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E., 2007. The role of site accessibility in microRNA target recognition. *Nature Genetics*. **39**, 1278-1284.
- Kim, J., Lee, S.H., Choe, J., Park, T.G., 2009. Intracellular small interfering RNA delivery using genetically engineered double-stranded RNA binding protein domain. *The Journal of Gene Medicine*. **11**, 804-812.
- Kim, V.N. & Nam, J.W., 2006. Genomics of microRNA. *Trends in Genetics : TIG*. **22**, 165-173.
- Klose, R.J. & Bird, A.P., 2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*. **31**, 89-97.
- Kothari, R.M. & Shankar, V., 1976. 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. *Journal of Molecular Evolution*. **7**, 325-329.
- Koveos, D.S. & Tzanakakis M.E, 1989. **Influence of photoperiod, temperature and host plant on the production of diapause eggs in *Petrobia (Tetranychina) harti* (Acari: Tetranychidae)**. *Experimental & Applied Acarology*. **6**, 327-342.
- Kozomara, A. & Griffiths-Jones, S., 2014a. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*. **42**, D68-73.
- Kozomara, A. & Griffiths-Jones, S., 2014b. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*. **42**, D68-73.
- Kriaucionis, S. & Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)*. **324**, 929-930.

- Kristensen, L.S. & Hansen, L.L., 2009. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clinical Chemistry*. **55**, 1471-1483.
- Krueger, F. & Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)*. **27**, 1571-1572.
- Kucharski, R., Maleszka, J., Foret, S., Maleszka, R., 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science (New York, N.Y.)*. **319**, 1827-1830.
- Kunert, N., Marhold, J., Stanke, J., Stach, D., Lyko, F., 2003. A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development (Cambridge, England)*. **130**, 5083-5090.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., Tuschl, T., 2003. New microRNAs from mouse and human. *RNA (New York, N.Y.)*. **9**, 175-179.
- Lai, E.C., Tomancak, P., Williams, R.W., Rubin, G.M., 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biology*. **4**, R42.
- Latorra, D., Arar, K., Hurley, J.M., 2003. Design considerations and effects of LNA in PCR primers. *Molecular and Cellular Probes*. **17**, 253-259.
- Law, J.A. & Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*. **11**, 204-220.
- Le Trionnaire, G., Hardie, J., Jaubert-Possamai, S., Simon, J.C., Tagu, D., 2008. Shifting from clonal to sexual reproduction in aphids: physiological and developmental aspects. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization*. **100**, 441-451.
- Leaman, D., Chen, P.Y., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., Marks, D.S., Sander, C., Tuschl, T., Gaul, U., 2005. Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell*. **121**, 1097-1108.
- Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., Ishino, F., 2002. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development (Cambridge, England)*. **129**, 1807-1817.
- Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. **75**, 843-854.
- Lee, Y., Jeon, K., Lee, J.T., Kim, S., Kim, V.N., 2002. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal*. **21**, 4663-4670.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W., 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*. **117**, 69-81.
- Legeai, F., Rizk, G., Walsh, T., Edwards, O., Gordon, K., Lavenier, D., Leterme, N., Mereau, A., Nicolas, J., Tagu, D., Jaubert-Possamai, S., 2010. Bioinformatic prediction,

- deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrtosiphon pisum*. *BMC Genomics*. **11**, 281-2164-11-281.
- Li, C., Mizutani, E., Ono, T., Wakayama, T., 2009a. Production of normal mice from spermatozoa denatured with high alkali treatment before ICSI. *Reproduction (Cambridge, England)*. **137**, 779-792.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. **25**, 2078-2079.
- Li, L.C. & Dahiya, R., 2002. MethPrimer: designing primers for methylation PCRs. *Bioinformatics (Oxford, England)*. **18**, 1427-1431.
- Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., Jones, P.A., 2002. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and Cellular Biology*. **22**, 480-491.
- Liang, G. & Yu, D., 2010. Reciprocal regulation among miR395, APS and SULTR2;1 in *Arabidopsis thaliana*. *Plant Signaling & Behavior*. **5**, 1257-1259.
- Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. **462**, 315-322.
- Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., Wang, X., 2003. R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science (New York, N.Y.)*. **301**, 1921-1925.
- Lockett, G.A., Kucharski, R., Maleszka, R., 2012. DNA methylation changes elicited by social stimuli in the brains of worker honey bees. *Genes, Brain, and Behavior*. **11**, 235-242.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., Maleszka, R., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology*. **8**, e1000506.
- Lyko, F. & Maleszka, R., 2011. Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics : TIG*. **27**, 127-131.
- Lynch, J.A. & Desplan, C., 2006. A method for parental RNA interference in the wasp *Nasonia vitripennis*. *Nature Protocols*. **1**, 486-494.
- Lyu, Y., Shen, Y., Li, H., Chen, Y., Guo, L., Zhao, Y., Hungate, E., Shi, S., Wu, C.I., Tang, T., 2014. New microRNAs in *Drosophila*--birth, death and cycles of adaptive evolution. *PLoS Genetics*. **10**, e1004096.
- Martin, F.H., Castro, M.M., Aboul-ela, F., Tinoco, I., Jr, 1985. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Research*. **13**, 8927-8938.
- Masaki, S., 1961. Geographical variation of diapause in insects.

- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. **466**, 253-257.
- Melchior, W.B., Jr & Von Hippel, P.H., 1973. Alteration of the relative stability of dA-dT and dG-dC base pairs in DNA. *Proceedings of the National Academy of Sciences of the United States of America*. **70**, 298-302.
- Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., Heintz, N., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*. **151**, 1417-1430.
- Miyoshi, T., Takeuchi, A., Siomi, H., Siomi, M.C., 2010. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nature Structural & Molecular Biology*. **17**, 1024-1026.
- Mocellin, S. & Provenzano, M., 2004. RNA interference: learning gene knock-down from cell physiology. *Journal of Translational Medicine*. **2**, 39.
- Morgan, H.D., Sutherland, H.G., Martin, D.I., Whitelaw, E., 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*. **23**, 314-318.
- Mukherjee, K., Campos, H., Kolaczowski, B., 2013. Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Molecular Biology and Evolution*. **30**, 627-641.
- Nanda, K.K. & Hamner, K.C., 1958. Studies on the nature of the endogenous rhythm affecting photoperiodic responses of Biloxi soy bean. *Botanical Gazette*. **120**, 14-25.
- Nelson, R., Denlinger, D., Somers, D., 2010. Photoperiodism: The biological calendar. *Oxford University Press*. 218-257.
- Nishizuka, M., Azuma A & Masaki, S., 1998. Diapause response to photoperiod and temperature in *Lepisma saccharina* Linnaeus (Thysanura: Lepismatidae). *Entomological Science*,. **1**, 7-14.
- Okamura, K., Matsumoto, K.A., Nakai, K., 2010. Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution. *BMC Bioinformatics*. **11 Suppl 7**, S2-2105-11-S7-S2.
- Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. **99**, 247-257.
- Park, J., Peng, Z., Zeng, J., Elango, N., Park, T., Wheeler, D., Werren, J.H., Yi, S.V., 2011. Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes. *Molecular Biology and Evolution*. **28**, 3345-3354.
- Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., Yura, R., 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *The Biochemical Journal*. **126**, 781-790.

- Petersen, C.P., Bordeleau, M.E., Pelletier, J., Sharp, P.A., 2006. Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell*. **21**, 533-542.
- Pfeifer, G.P., Kadam, S., Jin, S.G., 2013. 5-Hydroxymethylcytosine and its Potential Roles in Development and Cancer. *Epigenetics & Chromatin*. **6**, 10-8935-6-10.
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., Sontheimer, E.J., 2004. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell*. **117**, 83-94.
- Pittendrigh, C.S., Kyner, W.T., Takamura, T., 1991. The amplitude of circadian oscillations: temperature dependence, latitudinal clines, and the photoperiodic time measurement. *J Biol Rhythms*. **6**, 299-313.
- Pittendrigh, C.S. & Minis, D.H., 1964. The entrainment of circadian oscillations by light and their role as photoperiodic clocks. *American Naturalist*. **98**, 261-294.
- Place, R.F., Li, L.C., Pookot, D., Noonan, E.J., Dahiya, R., 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America*. **105**, 1608-1613.
- R Development Core Team, 2010. R: A Language and Environment for Statistical Computing, 2.10.1. Vienna, Austria.: .
- Raddatz, G., Guzzardo, P.M., Olova, N., Fantappie, M.R., Rampp, M., Schaefer, M., Reik, W., Hannon, G.J., Lyko, F., 2013. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, 8627-8631.
- Rae, P.M. & Steele, R.E., 1979. Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other insects. *Nucleic Acids Research*. **6**, 2987-2995.
- Regev, A., Lamb, M. J., Jablonka, E, 1998. The role of DNA methylation in invertebrates: Developmental regulation or genome defense? . *Molecular Biology and Evolution*. **15**, 880-891.
- Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*. **403**, 901-906.
- Ritz, C. & Spiess, A.N., 2008a. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics (Oxford, England)*. **24**, 1549-1551.
- Robert, M.F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A., MacLeod, A.R., 2003. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nature Genetics*. **33**, 61-65.
- Roberts, S.B. & Gavery, M.R., 2012. Is There a Relationship between DNA Methylation and Phenotypic Plasticity in Invertebrates? *Frontiers in Physiology*. **2**, 116.

- Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., Bradley, A., 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Research*. **14**, 1902-1910.
- Saenz de Miera, C., Monecke, S., Bartzen-Sprauer, J., Laran-Chich, M.P., Pevet, P., Hazlerigg, D.G., Simonneaux, V., 2014. A circannual clock drives expression of genes central for seasonal reproduction. *Current Biology : CB*. **24**, 1500-1506.
- Saetrom, P., Snove, O., Nedland, M., Grunfeld, T.B., Lin, Y., Bass, M.B., Canon, J.R., 2006. Conserved microRNA characteristics in mammals. *Oligonucleotides*. **16**, 115-144.
- Sakamoto, T., Uryu, O., Tomioka, K., 2009. The clock gene period plays an essential role in photoperiodic control of nymphal development in the cricket *Modicogryllus siamensis*. *Journal of Biological Rhythms*. **24**, 379-390.
- Sandrelli, F., Tauber, E., Pegoraro, M., Mazzotta, G., Cisotto, P., Landskron, J., Stanewsky, R., Piccin, A., Rosato, E., Zordan, M., Costa, R., Kyriacou, C.P., 2007. A molecular basis for natural selection at the timeless locus in *Drosophila melanogaster*. *Science (New York, N.Y.)*. **316**, 1898-1900.
- Sato, F., Tsuchiya, S., Meltzer, S.J., Shimizu, K., 2011. MicroRNAs and epigenetics. *The FEBS Journal*. **278**, 1598-1609.
- Saunders, D.S., 1982. *Insect clocks*. 2nd ed. Oxford; New York: Pergamon Press.
- Saunders, D.S., Steel, C.G.H. and Lewis, R.D., 2002. *Insect Clocks*. Third Edition ed. Elsevier.
- Saunders, D.S., 1977. *An introduction to biological rhythms*. New York: Wiley.
- Saunders, D.S., 1965. Larval diapause of maternal origin: induction of diapause in *Nasonia vitripennis* (Walk.) (Hymenoptera: Pteromalidae). *J Exp Biol*. **42**, 495-508.
- Saunders, D.S. & Gilbert, L.I., 1990. Regulation of ovarian diapause in *Drosophila melanogaster* by photoperiod and moderately low-temperature. *Journal of Insect Physiology*. **36**, 195-200.
- Saunders, D.S., Henrich, V.C., Gilbert, L.I., 1989. Induction of diapause in *Drosophilamelanogaster* - photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement. *Proceedings of the National Academy of Sciences of the United States of America*. **86**, 3748-3752.
- Saunders, D.S., Steel, C.G.H. and Lewis, R.D., 2002. *Insect Clocks*. Third Edition ed. Elsevier.
- Saunders, D.S. & Sutton, D., 1969. Circadian rhythms in the insect photoperiodic clock. *Nature*. **221**, 559-561.
- Schaefer, M. & Lyko, F., 2007. DNA methylation with a sting: an active DNA methylation system in the honeybee. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*. **29**, 208-211.

- Schiesari, L., Kyriacou, C.P., Costa, R., 2011. The hormonal and circadian basis for insect photoperiodic timing. *FEBS Letters*. **585**, 1450-1460.
- Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I.L., 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proceedings Biological Sciences / the Royal Society*. **255**, 279-284.
- Schmidt, P.S. & Paaby, A.B., 2008. Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*. *Evolution; International Journal of Organic Evolution*. **62**, 1204-1215.
- Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J., Dennis, E.S., 1999. The FLF MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *The Plant Cell*. **11**, 445-458.
- Shilatifard, A., 2006. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual Review of Biochemistry*. **75**, 243-269.
- Shukla, J.N., Jadhav, S., Nagaraju, J., 2011. Novel female-specific splice form of *dsx* in the silkworm, *Bombyx mori*. *Genetica*. **139**, 23-31.
- Simmen, M.W. & Bird, A., 2000. Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. *Molecular Biology and Evolution*. **17**, 1685-1694.
- Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., Bird, A., 1999. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science (New York, N.Y.)*. **283**, 1164-1167.
- Simpson, V.J., Johnson, T.E., Hammen, R.F., 1986. *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Research*. **14**, 6711-6719.
- Siomi, H. & Siomi, M.C., 2011. Stress signaling etches heritable marks on chromatin. *Cell*. **145**, 1005-1007.
- Smalheiser, N.R., 2003. EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biology*. **4**, 403.
- Smith, C.R., Mutti, N.S., Jasper, W.C., Naidu, A., Smith, C.D., Gadau, J., 2012. Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. *PloS One*. **7**, e42433.
- Smith, Z.D. & Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*. **14**, 204-220.
- Song, C.X. & He, C., 2013. Potential functional roles of DNA demethylation intermediates. *Trends in Biochemical Sciences*. **38**, 480-484.
- Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T.J., Zhang, B., Godley, L.A., Hicks, L.M.,

- Lahn, B.T., Jin, P., He, C., 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*. **29**, 68-72.
- Soni, K., Choudhary, A., Patowary, A., Singh, A.R., Bhatia, S., Sivasubbu, S., Chandrasekaran, S., Pillai, B., 2013. miR-34 is maternally inherited in *Drosophila melanogaster* and *Danio rerio*. *Nucleic Acids Research*. **41**, 4470-4480.
- Sorm, F., Piskala, A., Cihak, A., Vesely, J., 1964. 5-Azacytidine, a new, highly effective cancerostatic. *Experientia*. **20**, 202-203.
- Stevenson, T.J. & Prendergast, B.J., 2013. Reversible DNA methylation regulates seasonal photoperiodic time measurement. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, 16651-16656.
- Storey, J.M. & Storey, K.B., 2000. *Environmental Stressors and Gene Responses*.
- Sun, J.G., Liao, R.X., Qiu, J., Jin, J.Y., Wang, X.X., Duan, Y.Z., Chen, F.L., Hao, P., Xie, Q.C., Wang, Z.X., Li, D.Z., Chen, Z.T., Zhang, S.X., 2010. Microarray-based analysis of microRNA expression in breast cancer stem cells. *Journal of Experimental & Clinical Cancer Research : CR*. **29**, 174-9966-29-174.
- Sun, Y.M., Lin, K.Y., Chen, Y.Q., 2013. Diverse functions of miR-125 family in different cell contexts. *Journal of Hematology & Oncology*. **6**, 6-8722-6-6.
- Sutherland, H.G., Kearns, M., Morgan, H.D., Headley, A.P., Morris, C., Martin, D.I., Whitelaw, E., 2000. Reactivation of heritably silenced gene expression in mice. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*. **11**, 347-355.
- Suzuki, M.M. & Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews.Genetics*. **9**, 465-476.
- Suzuki, M.M., Kerr, A.R., De Sousa, D., Bird, A., 2007a. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Research*. **17**, 625-631.
- Suzuki, M.M., Kerr, A.R., De Sousa, D., Bird, A., 2007b. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Research*. **17**, 625-631.
- Syrova, Z., Dolezel, D., Saumann, I., Hodkova, M., 2003. Photoperiodic regulation of diapause in linden bugs: are period and Clock genes involved? *Cell Mol Life Sci*. **60**, 2510-5.
- Szyf, M., 2006. Targeting DNA methylation in cancer. *Bulletin Du Cancer*. **93**, 961-972.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*. **324**, 930-935.
- Takahashi, M., Niimi, T., Ichimura, H., Sasaki, T., Yamashita, O., Yaginuma, T., 1996. Cloning of a B-type cyclin homolog from *Bombyx mori* and the profiles of its mRNA level in non-diapause and diapause eggs. *Development Genes and Evolution*. **206**, 288-291.

- Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S.J., Pachter, L., Martin, D.I., Boffelli, D., 2014. Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Research*. **24**, 821-830.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*. **30**, 2725-2729.
- Tanaka, E.D. & Piulachs, M.D., 2012. Dicer-1 is a key enzyme in the regulation of oogenesis in panoistic ovaries. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization*. **104**, 452-461.
- Tang, F., Hajkova, P., Barton, S.C., Lao, K., Surani, M.A., 2006. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Research*. **34**, e9.
- Tauber, M., Tauber, C. and Masaki, S., 1986. *Seasonal adaptations of insects*. Oxford: Oxford University Press.
- Thomassin, H., Kress, C., Grange, T., 2004. MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. *Nucleic Acids Research*. **32**, e168.
- Thomassin, H., Oakeley, E.J., Grange, T., 1999. Identification of 5-methylcytosine in complex genomes. *Methods (San Diego, Calif.)*. **19**, 465-475.
- Thomson, J.P., Hunter, J.M., Nestor, C.E., Dunican, D.S., Terranova, R., Moggs, J.G., Meehan, R.R., 2013. Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Research*. **41**, e206.
- Tollefsbol, T.O. & Hutchison, C.A., 3rd, 1995. Mammalian DNA (cytosine-5-)-methyltransferase expressed in *Escherichia coli*, purified and characterized. *The Journal of Biological Chemistry*. **270**, 18543-18550.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., Zamore, P.D., 2004. A protein sensor for siRNA asymmetry. *Science (New York, N.Y.)*. **306**, 1377-1380.
- Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., Tomari, Y., 2011. Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nature Structural & Molecular Biology*. **18**, 1153-1158.
- Valinluck, V. & Sowers, L.C., 2007. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Research*. **67**, 946-950.
- Vasudevan, S., Tong, Y., Steitz, J.A., 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science (New York, N.Y.)*. **318**, 1931-1934.
- Veerman, A., 2001. Photoperiodic time measurement in insects and mites: a critical evaluation of the oscillator-clock hypothesis. *Journal of Insect Physiology*. **47**, 1097-1109.
- Verhulst, E.C., Beukeboom, L.W., van de Zande, L., 2010. Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science (New York, N.Y.)*. **328**, 620-623.

W. W. GARNER, H. A. ALLARD, 1920. EFFECT OF THE RELATIVE LENGTH OF DAY AND NIGHT AND OTHER FACTORS OF THE ENVIRONMENT ON GROWTH AND REPRODUCTION IN PLANTS. 415.

Wade, P.A., Pruss, D., Wolffe, A.P., 1997. Histone acetylation: chromatin in action. *Trends in Biochemical Sciences*. **22**, 128-132.

Wagner, G.C., Johnston, J.D., Clarke, I.J., Lincoln, G.A., Hazlerigg, D.G., 2008. Redefining the limits of day length responsiveness in a seasonal mammal. *Endocrinology*. **149**, 32-39.

Walsh, C.P. & Xu, G.L., 2006. Cytosine methylation and DNA repair. *Current Topics in Microbiology and Immunology*. **301**, 283-315.

Wang, H.Q., Tuominen, L.K., Tsai, C.J., 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics (Oxford, England)*. **27**, 225-231.

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.H., Colbourne, J.K., Clark, A.G., Werren, J.H., 2013a. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genetics*. **9**, e1003872.

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.H., Colbourne, J.K., Clark, A.G., Werren, J.H., 2013b. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genetics*. **9**, e1003872.

Wei, Y., Chen, S., Yang, P., Ma, Z., Kang, L., 2009. Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust. *Genome Biology*. **10**, R6-2009-10-1-r6. Epub 2009 Jan 16.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Nasonia Genome Working Group, Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C., Smith, C.D., van de Zande, L., Worley, K.C., Zdobnov, E.M., Aerts, M., Albert, S., Anaya, V.H., Anzola, J.M., Barchuk, A.R., Behura, S.K., Bera, A.N., Berenbaum, M.R., Bertossa, R.C., Bitondi, M.M., Bordenstein, S.R., Bork, P., Bornberg-Bauer, E., Brunain, M., Cazzamali, G., Chaboub, L., Chacko, J., Chavez, D., Childers, C.P., Choi, J.H., Clark, M.E., Claudianos, C., Clinton, R.A., Cree, A.G., Cristino, A.S., Dang, P.M., Darby, A.C., de Graaf, D.C., Devreese, B., Dinh, H.H., Edwards, R., Elango, N., Elhaik, E., Ermolaeva, O., Evans, J.D., Foret, S., Fowler, G.R., Gerlach, D., Gibson, J.D., Gilbert, D.G., Graur, D., Grunder, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C., 4th, Hurst, G.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J., Kim, J., Kiryutin, B., Koevoets, T., Kovar, C.L., Kriventseva, E.V., Kucharski, R., Lee, H., Lee, S.L., Lees, K., Lewis, L.R., Loehlin, D.W., Logsdon, J.M., Jr, Lopez, J.A., Lozado, R.J., Maglott, D., Maleszka, R., Mayampurath, A., Mazur, D.J., McClure, M.A., Moore, A.D., Morgan, M.B., Muller, J., Munoz-Torres, M.C., Muzny, D.M., Nazareth, L.V., Neupert, S., Nguyen, N.B., Nunes, F.M., Oakeshott, J.G., Okwuonu, G.O., Pannebakker, B.A., Pejaver, V.R., Peng, Z., Pratt, S.C., Predel, R., Pu, L.L., Ranson, H., Raychoudhury, R., Rechtsteiner, A., Reese, J.T., Reid, J.G., Riddle, M., Robertson, H.M., Romero-Severson, J., Rosenberg, M., Sackton, T.B., Sattelle, D.B., Schluns, H.,

- Schmitt, T., Schneider, M., Schuler, A., Schurko, A.M., Shuker, D.M., Simoes, Z.L., Sinha, S., Smith, Z., Solovyev, V., Souvorov, A., Springauf, A., Stafflinger, E., Stage, D.E., Stanke, M., Tanaka, Y., Telschow, A., Trent, C., Vattathil, S., Verhulst, E.C., Viljakainen, L., Wanner, K.W., Waterhouse, R.M., Whitfield, J.B., Wilkes, T.E., Williamson, M., Willis, J.H., Wolschin, F., Wyder, S., Yamada, T., Yi, S.V., Zecher, C.N., Zhang, L., Gibbs, R.A., 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science (New York, N.Y.)*. **327**, 343-348.
- Wienholds, E. & Plasterk, R.H., 2005. MicroRNA function in animal development. *FEBS Letters*. **579**, 5911-5922.
- Wightman, B., Ha, I., Ruvkun, G., 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*. **75**, 855-862.
- Wolschin, F. & Gadau, J., 2009a. Deciphering proteomic signatures of early diapause in *Nasonia*. *PLoS One*. **4**, e6394.
- Wolschin, F. & Gadau, J., 2009b. Deciphering proteomic signatures of early diapause in *Nasonia*. *PLoS One*. **4**, e6394.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., Walter, J., 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Communications*. **2**, 241.
- Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., Qi, Y., 2010. DNA methylation mediated by a microRNA pathway. *Molecular Cell*. **38**, 465-475.
- Wu, S.C. & Zhang, Y., 2010. Active DNA demethylation: many roads lead to Rome. *Nature Reviews.Molecular Cell Biology*. **11**, 607-620.
- Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., Dong, Y., Zhao, L., Lin, Y., Cheng, D., Yu, J., Sun, J., Zhou, X., Ma, K., He, Y., Zhao, Y., Guo, S., Ye, M., Guo, G., Li, Y., Li, R., Zhang, X., Ma, L., Kristiansen, K., Guo, Q., Jiang, J., Beck, S., Xia, Q., Wang, W., Wang, J., 2010a. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*. **28**, 516-520.
- Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., Dong, Y., Zhao, L., Lin, Y., Cheng, D., Yu, J., Sun, J., Zhou, X., Ma, K., He, Y., Zhao, Y., Guo, S., Ye, M., Guo, G., Li, Y., Li, R., Zhang, X., Ma, L., Kristiansen, K., Guo, Q., Jiang, J., Beck, S., Xia, Q., Wang, W., Wang, J., 2010b. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*. **28**, 516-520.
- Yaish, M.W., Colasanti, J., Rothstein, S.J., 2011. The role of epigenetic processes in controlling flowering time in plants exposed to stress. *Journal of Experimental Botany*. **62**, 3727-3735.
- Yamada, H. & Yamamoto, M.T., 2011. Association between circadian clock genes and diapause incidence in *Drosophila triauraria*. *PLoS One*. **6**, e27493.

- Yanovsky, M.J.&.S.A.K., 2002. Molecular basis of seasonal time measurement in Arabidopsis. *Nature*.
- Yi, R., Qin, Y., Macara, I.G., Cullen, B.R., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development*. **17**, 3011-3016.
- Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., Fazio, T.G., 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*. **147**, 1498-1510.
- Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics : TIG*. **13**, 335-340.
- Yoshimura, T., Yasuo, S., Watanabe, M., Iigo, M., Yamamura, T., Hirunagi, K., Ebihara, S., 2003. Light-induced hormone conversion of T4 to T3 regulates photoperiodic response of gonads in birds. *Nature*. **426**, 178-181.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Jin, P., Ren, B., He, C., 2012. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature Protocols*. **7**, 2159-2170.
- Yue, D., Liu, H., Huang, Y., 2009. Survey of Computational Algorithms for MicroRNA Target Prediction. *Current Genomics*. **10**, 478-492.
- Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science (New York, N.Y.)*. **328**, 916-919.
- Zhang, C., 2009. Novel functions for small RNA molecules. *Current Opinion in Molecular Therapeutics*. **11**, 641-651.
- Zhang, L., Zhang, L., Zhang, J., Liu, C., Cai, C., 2011. Determination of global DNA methylation in tissues by hydrophilic interaction liquid chromatography. *Se Pu = Chinese Journal of Chromatography / Zhongguo Hua Xue Hui*. **29**, 342-345.
- Zhang, T.Y. & Meaney, M.J., 2010. Epigenetics and the environmental regulation of the genome and its function. *Annual Review of Psychology*. **61**, 439-66, C1-3.
- Zwier, M.V., Verhulst, E.C., Zwahlen, R.D., Beukeboom, L.W., van de Zande, L., 2012. DNA methylation plays a crucial role during early Nasonia development. *Insect Molecular Biology*. **21**, 129-138.