DNA METHYLATION IN *DAPHNIA MAGNA*

by

CAMILA GONÇALVES ATHANÁSIO

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

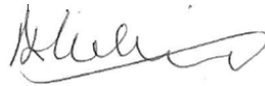University of Birmingham

August 2016

Abstract

Daphnia magna is gaining interest as a model for epigenetic research, especially as it is easy to maintain in large numbers under laboratory conditions and has low genetic diversity due to parthenogenetic reproduction. The *D. magna* genome is responsive to a wide range of stimuli, multigenerational studies can be conducted in short period of time, and a wide range of genomics resources is being developed for this species. Despite these great advantages, information regarding the epigenome of *D. magna* and its regulation is still lacking. Thus, the main aim of this work was to describe the methylome of *D. magna* and investigate its regulation and responsiveness to environmentally relevant exposure conditions. Despite the low levels of global DNA methylation, a defined profile could be identified. DNA methylation in *D. magna* is sporadic and mainly found at coding regions. These data suggest that *D. magna* encodes a complete set of genes for DNA methylation reactions. Evidence of direct effects on the DNA methylation profile were also found when animals were exposed to the DNA Methylation Inhibitor 5-azacytidine and these changes were persistent after the removal of the stressor. Acute and chronic exposures to environmentally relevant concentrations of stressors (arsenic and hypoxia) also induced changes in gene transcription levels and concentrations of one-carbon pathway metabolites. These findings indicate that the epigenome of *D. magna* is responsive to changes in the environment and thus support its use as an environmentally relevant model organism for epigenetics research. Furthermore, the maintenance of some of the epigenetic changes in the absence of the initial stressor provides evidence in support of the concept of 'epigenetic memory' and its potential use in chemical risk assessment.

Some of the methods and data presented in chapter 2 contributed towards a first author publication (Athanasio et al. (2016) Optimisation of DNA extraction from the crustacean Daphnia, Peer J, 4:e2004). The co-authors can confirm that the paper was written by the first author. Thus the sentences that are used from this paper in pages 35-37 of the thesis submitted by Camila Goncalves Athanasio are also entirely written by her.

Camila Goncalves Athanasio (first author)

Leda Mirbahai (senior author)

Acknowledgements

I would like to thank my supervisors Professor Kevin Chipman, Professor Mark Viant and Dr Leda Mirbahai for all their support throughout my PhD. Special thanks to Leda, for staying with me from the beginning to the end, helping me to develop my ideas and keeping me positive when things did not work as planned.

I would like to thank the University of Birmingham and CAPES (Coordenação de Aperfeiçoamento de Pessoal em Nível Superior) for funding this PhD and NBAF for funding the metabolomics study. I am grateful to Dr Ulf Sommer for his expertise and contributions to the metabolomics study. Also, special thanks to Dr Tim Williams for reading through my thesis drafts and the valuable comments.

In addition, I would like to thank all my friends in Birmingham who have contributed to my sanity during this time. Martin for sharing those great playlists and Maria for the great cakes, you made my time in the lab more enjoyable! Arielle, Carol and Lais for the support outside the lab, the great food, great talks and for sharing the coffee, the cakes and chocolate with me. Life will probably take us in different directions, but I am very happy to say that I have made amazing friends, hopefully for life!

Finally, I would like to thank my mom, my sister Carol and all my family for the support and encouragement through this time. Gustavo, I could not have done this without your support and patience. Thanks for being by my side even when we were separated by an ocean!

Thank you!

Table of contents

List of figures

List of Tables

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
a-KG	α-ketoglutarate
ANIMATE	Animal trace elements
ANOVA	Analysis of Variance
BBM	Bolds Basal Medium
BER	Base excision repair
BMHT	Betaine-homocysteine methyltransferase
BPA	Bisphenol A
BSP	Bisulfite sequencing PCR
cDNA	(complementary) Deoxyribonucleic acid
CGIs	CpG islands
CT	Threshold cycle
CTAB	Cetrimonium bromide
DM	Differential methylation
DMA	dimethylarsinic acid
DMAP	DNA methyltransferase-associated protein
DMG	Dimethylglycine
DMR	Differentially methylated region
DNMT	DNA (cytosine-5-)-methyltransferase
E2	17β-Estradiol
EC50	Half maximal effective concentration
FDR	False discovery rate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GNMT	Glycine N-methyltransferase
HDAC	Histone deacetylases

HHCOMBO	High hardness combo media
HIF	Hypoxia-inducible factor
HTS	High throughput sequencing
IAD	Activation-Induced Cytidine Deaminase
IAP	Intracisternal A particle
ICM	Inner cell mass
LC-MS	Liquid Chromatography - Mass Spectrometry
M/Z	Mass to charge ratio
MAT	Methionine adenosyltransferase
MBD	Methyl-CpG binding domain
MePC2	Methyl-CpG-binding protein 2
MMA	monomethylarsonic acid
MOA	Modes of action
MRM	Multiple reaction monitoring
MS	Methionine synthase
MTHFR	Methylenetetrahydrofolate reductase
MTRR	Methionine synthase reductase
OECD	Organisation for Economic Cooperation and Development
PC	Phosphatidylcholine
PCB	Polychlorinated biphenyl
PCR	Polymerase chain reaction
PGCs	Primordial germ cells
piRNA	Piwi-interacting RNA
POPs	Persistent organic pollutants
RT-PCR	Real time-polymerase chain reaction
SAH	S-adenosylhomocysteine
SAHH	S-adenosylhomocysteine hydrolase
SAM	S-adenosylmethionine
TBE buffer	Tris-borate-EDTA buffer
TDG	Thymine-DNA glycosylase

TET	Methylcytosine dioxygenase ten-eleven translocation
TF	Transcription factor
THF	Tetrahydrofolate
tRNAs	Transfer RNA
TSQ	Triple stage quadrupole tandem mass spectrometry
TSS	Transcription start site
WAX	Weak anion-exchange
WGA	Whole genome amplification
WGBS	Whole genome bisulfite sequencing

Chapter 1

General introduction

1.1   Introduction

An introductory overview of epigenetics is provided in section 1.2, including DNA methylation, methyltransferase enzymes and the mechanisms and pathways involved in DNA methylation. Section 1.3 is mainly focused on discussing the evolution of DNA methylation with a specific emphasis on differences in DNA methylation distribution across the genome and its function between vertebrates and invertebrates. Furthermore, the role of DNA methylation as an interface between the environment and the genome is discussed in section 1.4. In this section the concept of "epigenetic memory" is introduced and examples of stressor-induced DNA methylation changes and their potential effects are provided. The model organisms often used for epigenetic studies are described in section 1.5, followed by introducing the model organism *Daphnia* and discussing its potential for use in epigenetic studies. Finally, the aims and objectives of the research presented in this thesis are outlined in section 1.6.

1.2   Epigenetics

Epigenetics refers to mitotically heritable molecular factors and process that regulate genome accessibility and consequently gene expression and potentially organisms' phenotype, without altering the DNA sequence (Skinner *et al.*, 2010). Epigenetic mechanisms can be thought of as a second layer of information on top of the DNA sequence, regulating and interacting with the genome (from the prefix epi-, meaning over, outside of or around genetics). Epigenetic mechanisms include, but are not limited to, post-transcriptional chemical modifications of histones, non-coding RNA, including microRNAs and long non-coding RNAs, and chemical modifications of DNA (Felsenfeld, 2014). It is important to highlight that these

2

different epigenetic mechanisms are not separate entities but rather components of a complex system that work in conjunction to influence and regulate chromatin structure and eventually the function of a cell (Probst *et al.*, 2009). These mechanisms have been identified and investigated in wide range of organisms, from prokaryotes to eukaryotes, including fungi, plants and animals. The focus of this thesis is on one particular epigenetic mechanism referred to as DNA methylation and it is studied in the crustacean *Daphnia magna*.

1.2.1 DNA methylation

Methylation of DNA at cytosine bases is one of the most frequently studied epigenetic markers (Suzuki and Bird, 2008; Tan and Shi, 2012). DNA methylation is defined as post-replication addition of a methyl group to the 5th carbon position of a cytosine base. This process is mediated by a family of enzymes known as DNA methyltransferases (DNMTs), transferring a methyl group from the universal methyl donor, *S*-adenosylmethionine (SAM) to the cytosine base (Figure 1.1). DNA methylation is involved in many biological processes and plays an important role in regulation of gene expression. In specific, it can function as a method for cells to maintain a memory of the genes that require long-term transcriptional inactivation (Jaenisch and Bird, 2003). For example, in vertebrates it is important for X-chromosome inactivation, imprinting and tissue-specific gene expression (Crider *et al.*, 2012).
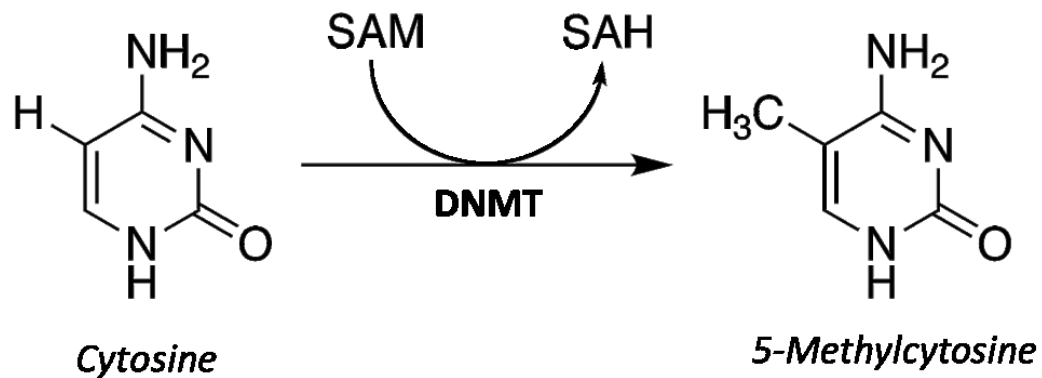
Figure 1.1 DNA methylation reaction. DNMTs mediate the transfer of a methyl group from *S*-adenosylmethionine (SAM) to cytosine, resulting in the formation of *S*-adenosylhomocysteine (SAH) and 5-methylcytosine.

In animals, methylation of cytosine bases predominantly occurs in a CpG context, where a cytosine nucleotide is followed by a guanosine. Despite DNA methylation often being associated with CpG dinucleotides, it has been detected at CHG and CHH sites too (H = A, C or T), however CHG and CHH methylation are rare events in animals.

For many years, the main focus of study has been on the methylation status of CpG nucleotides within transcription initiation regions. It was expected that the relationship between DNA methylation and gene expression would be an inverse correlation, where high methylation levels are associated with repressed gene expression, due to direct or indirect blocking of transcription. However, it is now becoming evident that the relationships between DNA methylation and gene expression are much more complex than originally thought. DNA methylation can be both associated with gene activation and inactivation. The function of DNA methylation is highly dependent on both its location within the genome and the context of the methylated site (Crider *et al.*, 2012; Jones, 2012). A good example is the different impact

of DNA methylation on regulation of gene expression at CpG sites within regions referred to as CpG islands versus CpG sites located along gene bodies. CpG islands (CGIs) are defined as regions with high density of CpGs and are often found in association with promoters and first exon regions. Usual parameters for CGIs prediction include a region with at least 200 bp, containing a CG percentage greater than 50% and an observed/expected ratio of at least 60%. In mammals, CpGs within CGI context are usually unmethylated and associated with transcriptionally active genes. Their methylation usually leads to transcriptional inactivation (follows the general assumption) (Jones, 1999). In contrast, CpG sites that are found along the gene bodies are usually linked to transcriptionally active genes, deviating from the past assumption (Jones and Takai, 2001; Suzuki and Bird, 2008).

1.2.2 One-carbon pathway

The one-carbon pathway functions as a metabolic integrator of nutrient status. It is a bi-cyclic metabolic pathway comprised of the folate and methionine cycles (Locasale, 2013) (Figure 1.2). Several dietary nutrients, such as folate, choline and some amino acids, are required for the maintenance of the pathway, ensuring sufficient products for downstream reactions, including DNA methylation (Crider *et al.*, 2012). The one-carbon pathway is comprised of a series of reactions, shown in Figure 1.2, that lead to the production of SAM, the immediate methyl donor for DNA methylation reaction and several other reactions, involving methylation of proteins, phospholipids and xenobiotic compounds (Lu, 2000). Several enzymes are involved in this pathway. Their levels are highly regulated to ensure appropriate production of SAM and removal of the rate limiting metabolite, S-adenosylhomocysteine (SAH) (Herceg and Vaissière, 2011; Ulrey *et al.*, 2005). SAH is produced after the transfer of the methyl group

from SAM to the cytosine. SAH has a higher binding affinity to methyltransferase than SAM and thus is a potent inhibitor of SAM-dependent methyltransferases. Accumulation of SAH is known to disrupt the one-carbon pathway, leading to decreased functionality of methyltransferases and subsequently altering the levels of DNA methylation (Mirbahai *et al.*, 2013; Tollefsbol, 2012; Yi *et al.*, 2000). Therefore, a series of reactions, as illustrated in Figure 1.2, ensure efficient removal of SAH and continuation of the cycle (Herceg and Vaissière, 2011; Ulrey *et al.*, 2005).
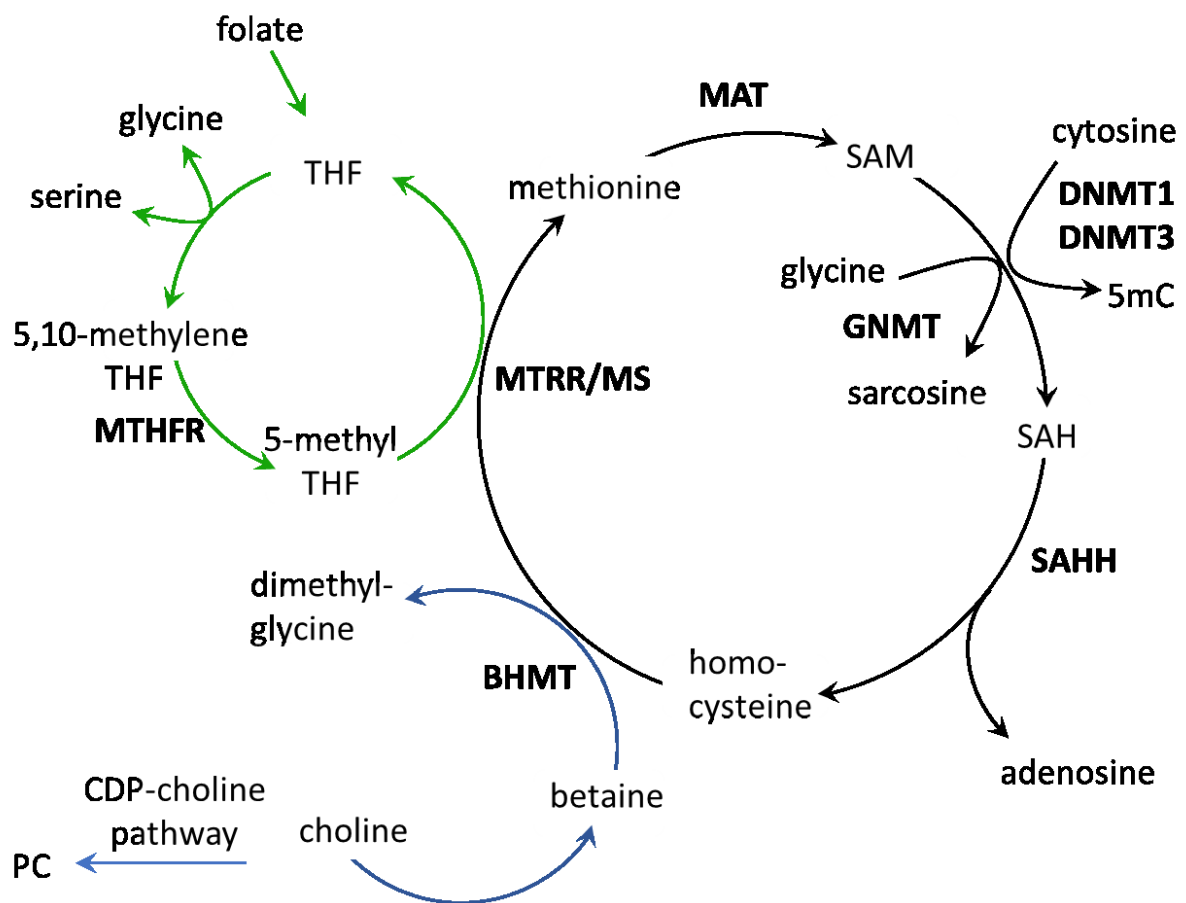
Figure 1.2 The one-carbon pathway. Methionine cycle is presented in black. Folate cycle is represented in green. Choline input in the pathway is shown in blue. Metabolites and enzymes (in bold) are displayed in the pathway.

Abbreviations: BMHT: betaine- homocysteine methyltransferase, CDP-choline: cytidine diphosphate-choline, DNMT: DNA methyltransferase, GNMT: glycine N-methyltransferase, MAT: methionine adenosyltransferases, MS: methionine synthase, MTHFR: methylenetetrahydrofolate reductase, MTRR: Methionine synthase reductase, PC: phosphatidylcholine; SAH: S-adenosylhomocysteine, SAHH: S- adenosylhomocysteine hydrolase, SAM dependent MT: S-adenosylmethionine dependent methyltransferase, THF: tetrahydrofolate.

1.2.3 DNA methyltransferases

The DNA methyltransferases (DNMTs) are the primary enzymes involved in the establishment and maintenance of DNA methylation. Three families of DNMTs, DNMT1, DNMT2 and DNMT3, have been identified in vertebrates while in invertebrates the number of detected DNA methyltransferases can vary from organism to organism (Goll and Bestor, 2005).

In vitro experiments suggest that DNMT1 has a preference for hemimethylated CpG, therefore it is mainly involved in maintaining the methylation pattern of a newly synthesised DNA strand after replication. Often, mutation in this enzyme leads to global hypomethylation, confirming the role of DNMT1 in DNA methylation maintenance (Bestor, 2000; Chen *et al.*, 2013; Jaenisch and Bird, 2003). However, it has been demonstrated that under *in vitro* conditions, DNMT1 can cause *de novo* methylation (Pradhan *et al.*, 1999).

The DNA methyltransferase 3 family are classified as *de novo* DNA methyltransferases as they show equal affinity towards both hemi- and un-methylated DNA, adding the methyl group to the DNA regardless of the methylation status of the complementary strand (Arand *et al.*, 2012; Gowher and Jeltsch, 2001; Jones and Liang, 2009). DNMT3 is required for genome-wide *de novo* methylation and is essential during early development and gametogenesis (Klose and Bird, 2006; Okano *et al.*, 1999). In *Mus musculus* (mouse), the knockout of DNMT3s caused increases in hemi-methylated CpG sites in regions with repeats, suggesting that DNMT3s are important in methylation and inactivation of these regions (Jones and Liang, 2009).

Therefore, despite the classic classification of DNMT1 and DNMT3 as maintenance and *de novo* methyltransferases, respectively, as described above there are overlaps between their functions (Elliott *et al.*, 2016). It is thought that this is partly to ensure that DNMT1 and 3 can

compensate for each other to ensure maintenance of normal DNA methylation pattern within a cell even if the function of one of them is compromised (Jin and Robertson, 2013; Rhee *et al.*, 2002).

DNMT2 was originally assigned as a DNA methyltransferase due to its highly conserved DNA methylase domain. Goll *et al.* (2006) has demonstrated that in fact this DNMT is responsible for methylation of small tRNAs rather than DNA, and it is involved in regulation of protein synthesis and tRNA stability (Schaefer *et al.*, 2010; Tuorto *et al.*, 2012). Some of the DNA methyltransferases found in insects, such as *Drosophila melanogaster*, are closely related to the mammalian DNMT2 and it seems to be conserved along evolution (Hendrich and Tweedie, 2003; Okano *et al.*, 1998).

Vertebrate species generally possess genes of all the three DNMT families. Variation is mostly found in the numbers of genes in each family. While mammals have a single copy of DNMT1 and three copies of DNMT3, Zebrafish have six copies of DNMT3 (Capuano *et al.*, 2014; Kamstra *et al.*, 2015b; Li *et al.*, 2010; Robertson *et al.*, 1999; Smith *et al.*, 2011).

Invertebrates present far greater variability in the presence and numbers of the DNMTs. Insects, as an example, present distinct DNMTs profiles (Figure 3.16). The classic genetic model organism, *D. melanogaster*, only has DNMT2 and so far no defined DNA methylation pattern has been identified, suggesting that DNMT2-only genomes are unmethylated (Capuano *et al.*, 2014; Lyko *et al.*, 2000; Raddatz *et al.*, 2013; Rasmussen and Amdam, 2015). *Bombyx mori*, *Nasonia vitripennis* and *Apis mellifera* have similar levels of global DNA methylation, although *B. mori* does not encode the gene for DNMT3 while the others do. Regarding DNMT1, *B. mori* has one copy, while *A. mellifera* has two copies and *N. vitripennis*

9

has three copies (Beeler *et al.*, 2014; Lyko *et al.*, 2010; Pegoraro *et al.*, 2016; Wojciechowski *et al.*, 2014; Xiang *et al.*, 2010). Therefore, a clear relationship between number of DNMTs and DNA methylation levels cannot be directly identified. This is shown in Figure 3.16 and discussed in detail in chapter 3.

Furthermore, DNMTs can be involved in the removal of DNA methylation. Recent studies have demonstrated that *de novo* methyltransferases in mammals can act as DNA dehydroxymethylases, converting 5-hmC to C (Chen *et al.*, 2012). Also, DNMTs were reported to act as Ca^{2+} ion- redox state-dependant demethylases, at least in *in vitro* systems (Chen *et al.*, 2013).

1.2.4 DNA methylation and transcription regulation

DNA methylation, in association with other genetic and epigenetic mechanisms, is involved in regulation of transcription. Although the underlying mechanisms are not fully understood, it is thought that epigenetic regulation of transcription is partly achieved by remodelling the chromatin structure (Spruijt and Vermeulen, 2014; Vaissière *et al.*, 2008). Transcription repression can be achieved by two main mechanisms: (i) occupying transcription factor (TFs) binding sites and (ii) recruitment of methylated-DNA binding domain proteins (MBD) (Figure 1.3).

(i) Occupying binding sites for TFs: It was discovered that methylation of DNA at CpG dinucleotide sites located within the binding site of transcription factors can inhibit their binding and subsequently prevent transcription (Iguchi-Ariga and Schaffner, 1989). However,

it is now known that this is not the main mechanism for transcription repression via DNA methylation (Kass *et al.*, 1997; Vaissière *et al.*, 2008).

(ii) Methylated-DNA binding domain proteins (MBD): The alternative model of transcription repression involves the recruitment of MBD protein. These proteins will recognise and bind to methylated CpG sites. Their binding will block both recruitment and binding of all the required molecular factors for activation of transcription. This is achieved partly by their interactions with several proteins, such as histone deacetylases (HDACs), and induction of local changes to the chromatin structure (Wade, 2001; Vaissière *et al.*, 2008).

Repression of transcription by methylation is not restricted to the promoter regions of genes. In vertebrates, more than 90% the methylated cytosines are located at the transposable repetitive elements, also referred to as transposons (Baccarelli and Bollati, 2009; Yang *et al.*, 2004). Maintenance of methylation levels of repetitive elements is crucial for both chromosomal stability and genome integrity (Putiri and Robertson, 2011).
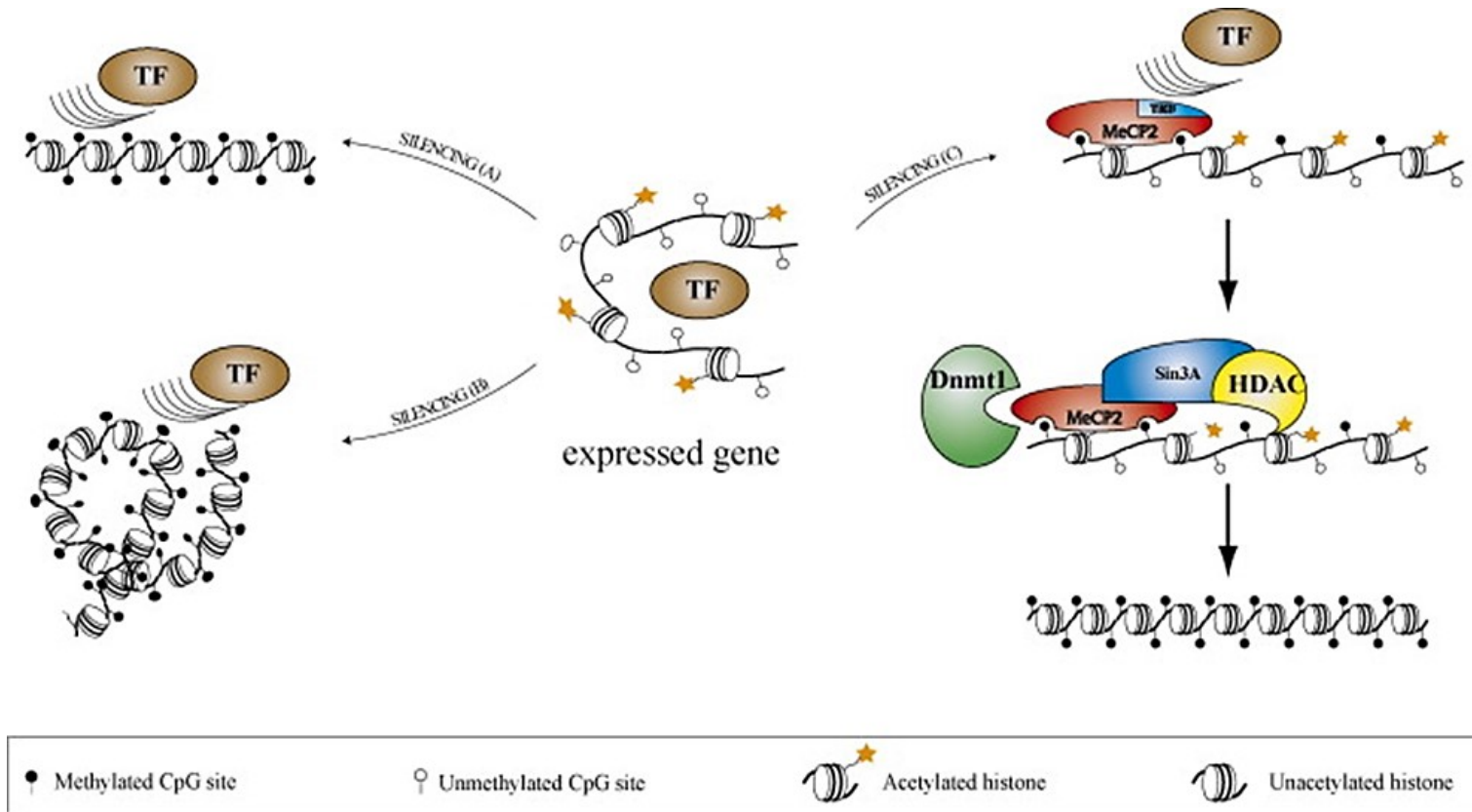
Figure 1.3 Epigenetic mechanisms for gene silencing. Methylated sites in promoter regions can be directly repressed due to blocking of transcription factors biding. Methylated DNA can also form heterochromatin, preventing the access of transcription factors to DNA. Methylation of DNA can also result in binding of methyl-binding proteins (e.g. MePC2) which recruit histone deacetylases (HDAC) leading to a chromatin state that prevents transcription factors binding to the promoter (Reproduced from Vaissière *et al.*, 2008).

Despite gene silencing being recognised as the main function of DNA methylation for many years, several other functions have been identified for DNA methylation. For example, DNA methylation at gene bodies has been associated with transcriptionally active genes (Jones, 2012). As highlighted by Jones (1999), there is an apparent paradox in DNA methylation. DNA methylation in the promoter regions is inversely correlated with gene expression; on the other hand, methylation in the gene body is positively correlated with expression. Therefore, presence of a methylated CpG site alone is not sufficient to determine the outcome on transcription level. Both the location of the methylated CpG site and presence or absence of other interacting molecular factors play a significant role in determining the outcome of DNA methylation on transcription level (Schübeler, 2015).

DNA methylation in genic regions is highly conserved throughout the phylogenetic tree; however, it is still poorly understood at both molecular and functional levels. Studies suggest that is involved in disruption of chromatin, such as nucleosome displacement, which is caused by elongating RNA polymerase (Jones, 2012; Simmen *et al.*, 1999; Suzuki *et al.*, 2007). Genic methylation has also been linked to regulation of alternative splicing (Lev Maor *et al.*, 2015; Maunakea *et al.*, 2013). However, experimental evidence is still lacking and there is no full characterisation of how the splicing machinery can be affected by methylation (Schübeler, 2015; Shukla *et al.*, 2011).

The main point to consider when analysing transcription regulation by DNA methylation is the fact that it is not an isolated event and it works in collaboration with other epigenetic mechanisms. Additionally, it cannot be related to only one function, since it is likely variable according to different genomic contexts.

1.2.5 DNA demethylation pathways

DNA demethylation can either occur through active or passive pathways. Ten-Eleven-Translocation (TET) enzymes are responsible for the active removal of methylation by a multistep reaction, while DNA methylation can be passively lost during replication of DNA, due to malfunction of enzymes from the one-carbon pathway or absence of methyl donors (Piccolo and Fisher, 2014; Song *et al.*, 2013; Tahiliani *et al.*, 2009).

In contrast to the well-studied DNA methylation mechanism, the pathway for active demethylation of DNA has only recently been revealed and thus it is still not completely characterised (Bhutani *et al.*, 2011; Kohli and Zhang, 2013). The intermediate bases of the demethylation pathway were first observed in 1972 in mammalian genomes (Penn *et al.*, 1972), but it was only with the discovery of TET enzymes that these observations received attention and the pathway was proposed. Since then, different demethylation mechanisms have been proposed which are summarised in Figure 1.4 (Kohli and Zhang, 2013).

After the discovery of TET hydroxylases and identification of 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) bases, a multistep reaction was proposed for active removal of methylated cytosine (Ito *et al.*, 2011; Song *et al.*, 2013; Tahiliani *et al.*, 2009). After oxidation or deamination of 5mC, DNA glycosylase TDG removes 5hmU or 5caC and the base-excision-repair pathway adds an unmethylated cytosine to the gap (Gong and Zhu, 2011).
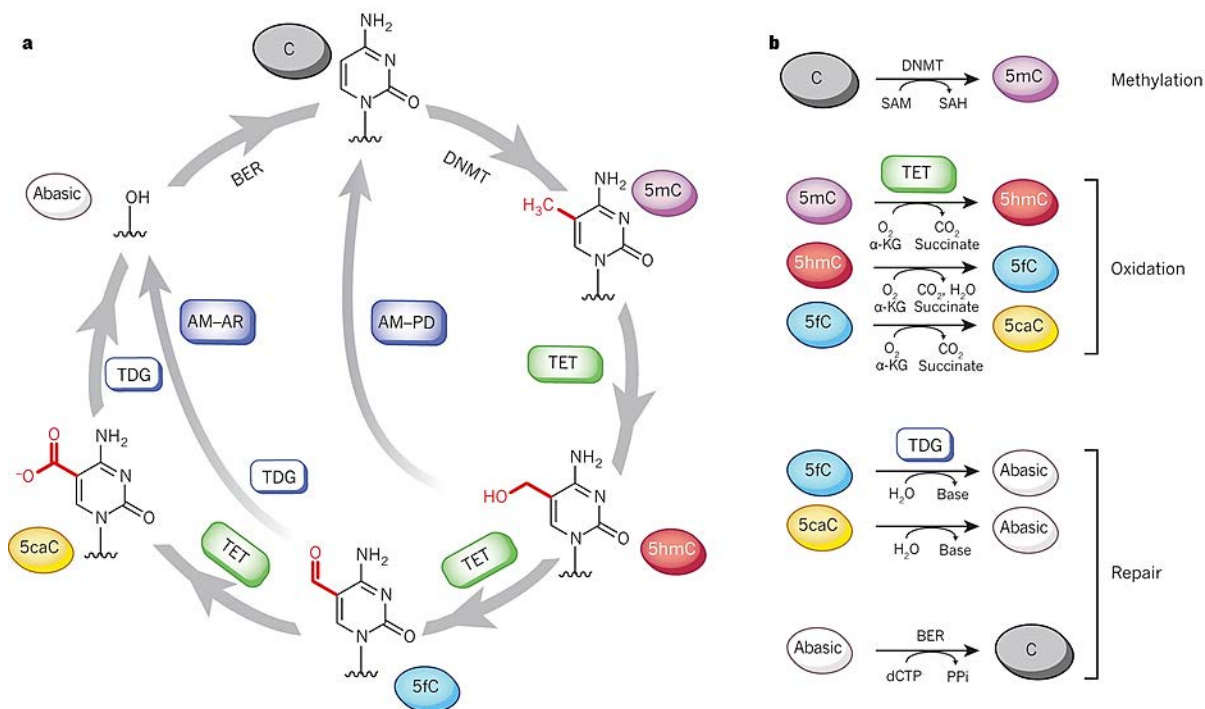
Figure 1.4 Pathways for dynamic demethylation of cytosine. A) DNA demethylation through TET enzymes. 5mC bases, introduced by DNA methyltransferase (DNMT) enzymes, can be oxidized iteratively to 5hmC, 5fC and 5caC. Then, modified bases can either be passively removed or excised by TDG generating an abasic site as part of the base excision repair (BER) process that regenerates unmodified C. B) The individual reactions in the pathway are shown with all reactants depicted. α-KG, α-ketoglutarate; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine (Reproduced from Kohli and Zhang, 2013).

The role of the additional chemical modifications is not completely clear. Besides 5hmC being an intermediate base during DNA methylation by TET enzymes, it is proposed that it can also be a stable base with influences on chromatin structure and transcriptional activity. As well as 5mC, 5hmC can recruit selective hmC-binding proteins or exclude methyl-CpG-binding proteins affecting chromatin modifications. It can also facilitate passive loss of methylation since it is poorly recognized by DNMT1 (Tahiliani *et al.*, 2009). Studies also suggest 5fC may have functional roles as an epigenetic mark of regulatory elements (Song *et al.*, 2013).

The demethylation pathway is completed by the thymine DNA glycosylase (TDG). TDG can directly excise the TET oxidation products, 5fC and 5caC. Then, the excised base is replaced with a cytosine by base excision repair (BER) pathway (Kohli and Zhang, 2013; Robertson *et al.*, 2009).

DNA methylation can also be passively lost during DNA replication. When DNA is replicated the DNMTs recognise the methylated cytosine on the parent strand, adding the methyl group to the new strand, therefore, maintaining the normal levels of DNA methylation. When this mechanism is disrupted, for example either due to malfunction of methylation enzymes or absence of methyl donors, DNA methylation is passively lost. This will result in global loss of DNA methylation and passive demethylation (Piccolo and Fisher, 2014).

1.2.6 DNA methylation reprogramming during development

Epigenetics modifications are usually stable in somatic cells, but in germ cells and during early embryogenesis (Figure 1.5) they undergo an extensive reorganisation involving large scale loss and resetting of the DNA methylation patterns and remodelling of histones, referred to as reprogramming. The first epigenetic reprogramming event occurs in the fertilised eggs. It is necessary that the previous sperm- and oocyte-specific epigenetic patterns are removed and the DNA methylation pattern is reset in a tissue-specific manner as the embryo develops. Furthermore, the resetting of DNA methylation markers will reduce the risk of perpetuation of epimutations through mitotic and meiotic division (Feng *et al.*, 2010b; Lees-Murdock and Walsh, 2008). Likewise, a demethylated state may increase epigenomic plasticity, facilitating the massive transcriptional changes associated with embryogenic development (Seisenberger *et al.*, 2013). As described in Figure 1.5, the demethylation events in maternal and paternal

alleles occur through different mechanisms, with the paternal genome actively demethylated while the maternal genome is passively demethylated (Seisenberger *et al.*, 2013). The only genes that escape this wave of de-methylation event are imprinted genes. Imprinted genes are genes whose expression is determined by the parent that contributed them. Therefore, imprinted genes in the embryo will inherit the DNA methylation pattern of the parent, with one allele reflecting maternal DNA methylation patterns and the other the paternal DNA methylation pattern. Once the DNA is demethylated, the methylation pattern is re-established by the *de novo* DNMT3a and 3b just prior to implantation stage (Kelsey and Feil, 2013; Seisenberger *et al.*, 2013).

The second reprogramming event occurs in primordial germ cells (PGC) once they reach the embryonic gonads (Feng *et al.*, 2010b; Jirtle and Skinner, 2007; Law and Jacobsen, 2010). This reprogramming is potentially important for the development of the germ line as well as being essential for removal of parental imprinting and setting the base for totipotency intrinsic to this cell lineage and establishment of germ-cell specific methylation marks, leading to formation of sperm- and oocyte-specific patterns. It is important to highlight that the DNA methylation patterns of imprinted genes are only reset during this reprogramming event. Thus any mistake during their re-methylation is maintained in the following generations, leading to transgenerational epigenetic inheritance (Seisenberger *et al.*, 2013). In general, epigenetic reprogramming events may occur at an extremely critical stage, with greater vulnerability, where the environment can affect the mechanism of resetting the epigenetic marks and may lead to disruption of epigenetic mechanisms.

The embryonic epigenetic reprogramming event in invertebrates has not been studied to the same extent as it has been in vertebrates, such as mammals. However, theories about DNA

methylation reprogramming in invertebrates are being developed and indicate similarities between the reprogramming events in vertebrates and invertebrates (Head, 2014; Patalano *et al.*, 2012). These associations have been made between the events of reprogramming occuring in mammals and the reprogramming of castes in insects, such as *Apis mellifera*.
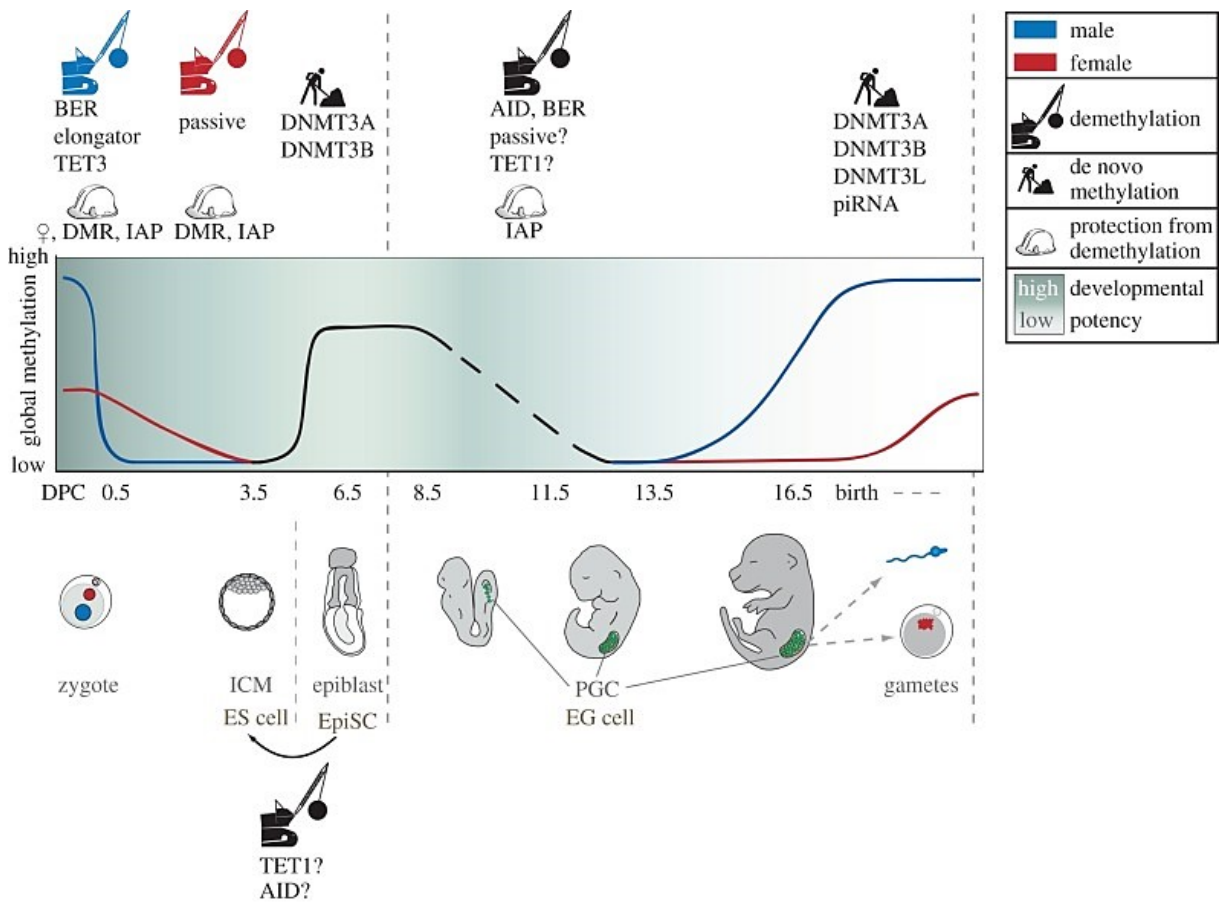
Figure 1.5 DNA methylation reprogramming events in mammalian development. Two waves of demethylation occur in the genome. The first happens following fertilization when the methylation in gametes is erased. Paternal gamete (in blue) undergoes rapid active demethylation, followed by a passive demethylation event in maternal gametes (in red). Methylation pattern is then re-established through *de novo* DNA methylation. Primordial germ cells (PGCs) (in green) suffer a second global demethylation event. DNA methylation pattern is once more re-established during sex determination. BER – Base excision repair; TET – Ten-eleven translocase enzymes; DMR – Differentially methylated region; IAP – intracisternal A particles; ICM – inner cell mass; AID - Activation-Induced Cytidine Deaminase; piRNA - Piwi-interacting RNA (Reproduced from Seisenberger *et al.*, 2013).

1.3   DNA methylation in vertebrates and invertebrates

DNA methylation is an epigenetic mechanism highly conserved throughout evolution (Zemach *et al.*, 2010). It is present in many organisms ranging from prokaryotes to eukaryotes. However, there are some differences between organisms in terms of function, distribution, level and bases that are methylated. For example, methylation in prokaryotes mostly occurs at adenine bases, while in eukaryotes it is largely detected on cytosines (Collier, 2009; Feng *et al.*, 2010a; Rivière, 2014; Suzuki and Bird, 2008; Zemach *et al.*, 2010). However, in recent years other methylated bases, such as methylated adenosine, have been detected in different eukaryote organisms, such as *C. elegans*, *C. reinhardtii*, *D. melanogaster* and *H. sapiens* (Greer *et al.*, 2015; Luo *et al.*, 2015; Wu *et al.*, 2016).

Furthermore, although DNA methylation is detected in majority of the organisms, the level and distribution pattern of DNA methylation can vary dramatically among species (Jiang *et al.*, 2014; Rivière, 2014). For example, vertebrates present higher DNA methylation levels than invertebrates. Table 1.1 summarises the main findings regarding DNA methylation in key species throughout different taxa.

In vertebrates, DNA methylation is distributed throughout the whole genome, with sections, close to the promoter regions and the 5' end of the genes presenting a drop in methylation level (Figure 1.6). Also, genes differ in their CpG content. Usually, promoters with low numbers of CpG sites are often hypermethylated while promoters with high repeats of CpG dinucleotides (CpG islands) are often demethylated, leading to maintaining an active transcriptional status (Rivière, 2014; Zemach *et al.*, 2010).

H. sapiens, *M. musculus* and *D. rerio* are the main vertebrate organisms used in epigenetic studies. The overall distribution of DNA methylation is very similar among these organisms and follows what was described above, although the global levels of DNA methylation are usually higher in fish than in mammals (Zhang *et al.*, 2016). Methylation can occur in different features of the genome, including repetitive regions, transposons and gene bodies (Feng *et al.*, 2010a) (Figure 1.7).
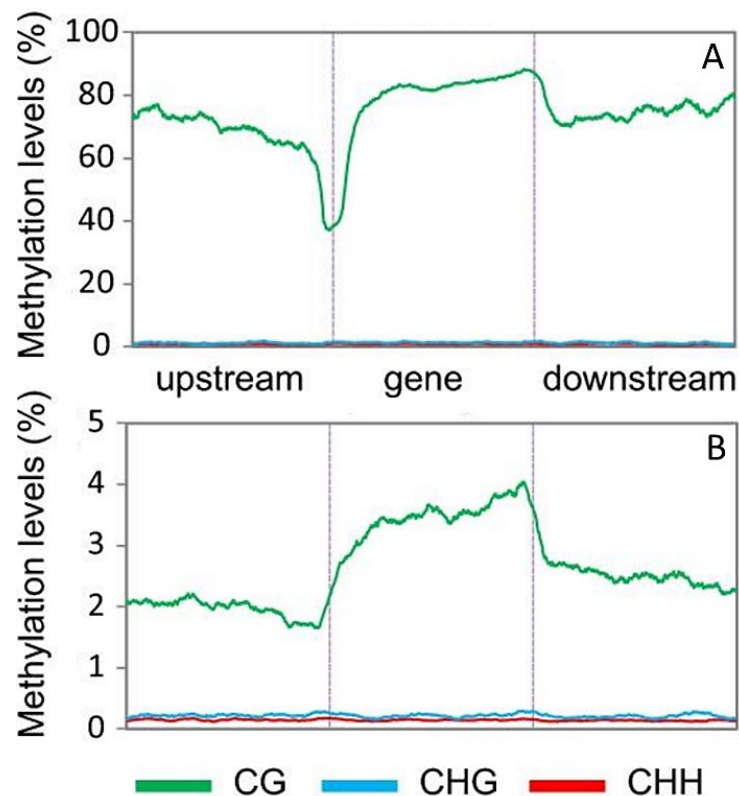
Figure 1.6 General DNA methylation distribution in vertebrate (A) and invertebrate (B) genomes. Methylation level is presented for the different context, CpG, CHG and CHH, showing high levels in CpG context in both. A) a drop in DNA methylation close to the 5' end of the gene and a slight increase in gene bodies is observed for vertebrates; B) overall low level of methylation with enrichment along genes is observed for invertebrates (Reproduced from Feng *et al.*, 2010a).
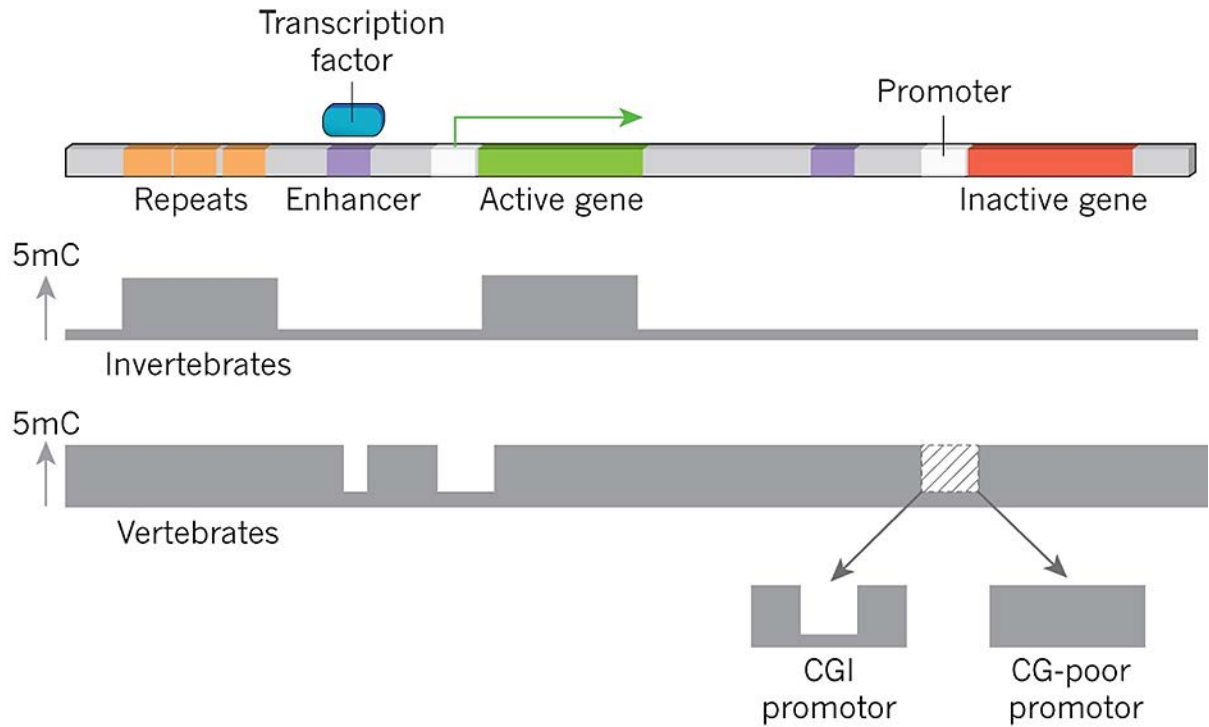
Figure 1.7 Distribution of methylated cytosine in an invertebrate and vertebrate genome. Different genomic regions are represented, including active and inactive genes with proximal promoter, other regulatory region (enhancer) and repetitive elements. The height of grey bars indicates the relative amount of methylation in invertebrates and vertebrates (Reproduced from Schübeler, 2015).

Table 1.1 Main findings regarding DNA methylation occurrence and pattern in different taxa.

Species	Global % 5mC	Distribution pattern	Main findings	Reference
Caenorhabditis elegans	0.0033%	Absent	Very low levels of global DNA methylation. DNMTs were not identified.	Hu *et al.*, 2015
Drosophila melanogaster	0.03%	Absent	Very low levels of global DNA methylation. Only DNMT2 was identified in fruit fly genome. DNA methylation does not follow a defined pattern.	Lyko *et al.*, 2000
Bombyx mori	0.11%	Sporadic	Low levels of DNA methylation, occurring in CpG dinucleotides and targeting gene bodies.	Xiang *et al.*, 2010
Nasonia vitripennis	0.18%	Sporadic	Low levels of DNA methylation, in CpG dinucleotides and occurring in exonic regions.	Beeler *et al.*, 2014
Apis mellifera	0.11%	Sporadic	Low levels of DNA methylation, in CpG dinucleotides and occurring in exonic regions.	Lyko *et al.*, 2010
Crassostrea gigas	1.96%	Sporadic	DNA methylation restricted to CpG sites, enriched in gene bodies and repetitive regions.	Wang *et al.*, 2014
Ciona intestinalis	4.07%	Global	DNA methylation occurring specially in CpG sites along the entire genome, however with the genes being the major target.	Feng *et al.*, 2010a
Danio rerio	~8%	Global	High levels of CpG methylation, globally distributed with enrichment in gene regions and depletion close to TSSs.	Feng *et al.*, 2010a
Homo sapiens	3.93%	Global	High levels of CpG methylation, globally distributed in the genome with depletion close to TSSs.	Li *et al.*, 2010

In contrast, DNA methylation is not distributed evenly throughout the genome of invertebrates. The majority of the genome of invertebrates is unmethylated with limited peaks of DNA methylation (Figure 1.6 B). This pattern of DNA methylation is referred to as a sparse or sporadic pattern (Figure 1.8). The methylation peaks usually correspond to gene bodies (Breiling and Lyko, 2015; Rivière, 2014). However, there are some exceptions where DNA methylation is either extremely low (just above noise) for example the fruit fly *D. melanogaster* or the DNA methylation is restricted to a specific life stage, for example the beetle *Tribolium castaneum* (Breiling and Lyko, 2015; Feng *et al.*, 2010a; Jiang *et al.*, 2014).
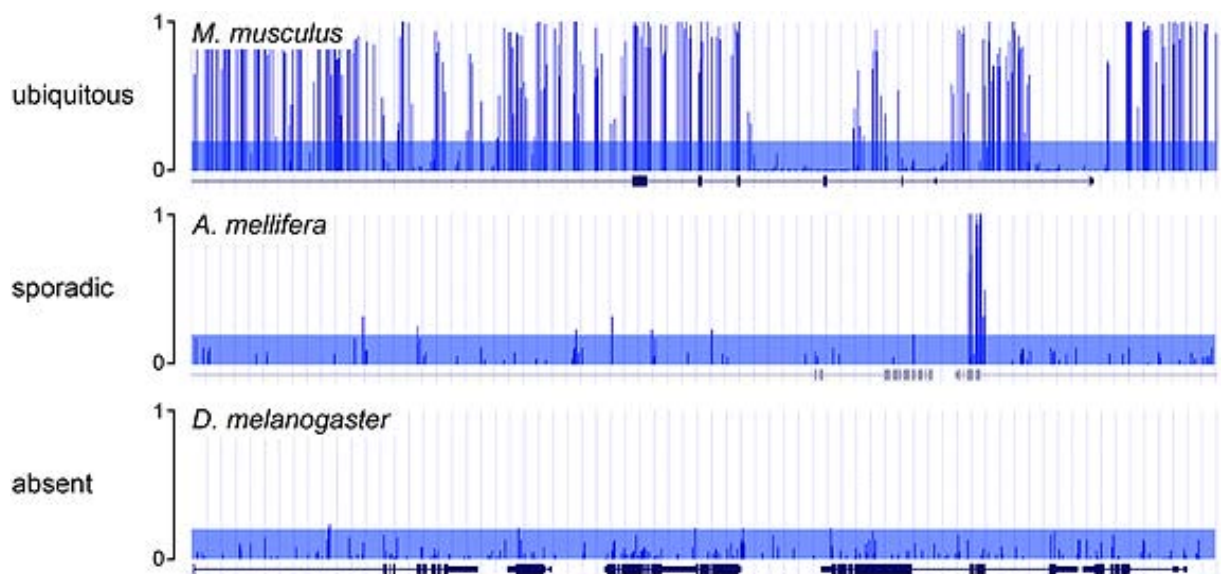
Figure 1.8 Major categories of DNA methylation distribution in animals. Data are presented in (Reproduced from Breiling and Lyko, 2015). Whole genome bisulfite sequencing analyses of mouse (top), honey bee (middle) and *Drosophila* DNA (bottom) are used to exemplify the ubiquitous/global, sporadic/sparse and absent DNA methylation profiles. Methylation ratios for each CpG site are shown in a randomly selected 40 kB window. Transparent blue bars indicate the range of bisulfite conversion artefacts (methylation levels below 20%).

The presence of methylated DNA in *D. melanogaster* has been discussed for many years (Capuano *et al.*, 2014; Lyko and Maleszka, 2011; Lyko *et al.*, 2000; Raddatz *et al.*, 2013). Recently, it has been shown that the global DNA methylation level measured by chromatographic methods is 0.03%, and it is too low to be correctly assessed at single-base resolution by bisulfite sequencing methods (Capuano *et al.*, 2014). As shown in Figure 1.8, it is difficult to distinguish the methylated sites from bisulfite conversion artefacts, therefore, no DNA methylation pattern could be identified in *D. melanogaster* (Breiling and Lyko, 2015; Raddatz *et al.*, 2013).

As cited above, some insects, such as the flour beetle *T. castaneum,* present DNA methylation only during embryonic stages. For many years it was believed that the *T. castaneum* genome lacked methylation even though its genome encodes DNMTs genes. Feliciello *et al.* (2013), was the first researcher to discover a cyclic DNA methylation event, occurring in embryonic stages followed by a loss of DNA methylation in later stages in these species. The evidence suggests that the absence of DNA methylation in some insect groups is probably due to lineage-specific loss events (Glastad *et al.*, 2011).

High-throughput sequencing technologies have rapidly increased the knowledge regarding DNA methylation in invertebrates. With the understanding of this mechanism in various groups it is now possible to infer about the evolution of DNA methylation marks, together with its conservation and divergence, and its significance for the different species (Standage *et al.*, 2016).

Assumptions regarding the evolutionary conservation of DNA methylation function have been made for several years (Regev *et al.*, 1998; Tweedie *et al.*, 1997). The divergence in the DNA

methylation distribution patterns between vertebrates and invertebrates indicates that it can have distinct functions for the two groups, and in relation to its context (Head, 2014).

Analysing the relationship between DNA methylation distribution and evolution, it was possible to identify that gene methylation was already present in invertebrates, in the ancestral organism of vertebrata (the chordate *C. intestinalis*), and are still identified in higher taxa, such as fish and mammals (*D. rerio* and *H. sapiens*) (Table 1.1). On the other hand, global DNA methylation arose only in the Chordata group along with the appearance of vertebrates (Regev *et al.*, 1998; Tweedie *et al.*, 1997). Therefore, in terms of function of DNA methylation, the usage for global genome silencing (e.g. silencing of repeats and transposons) was likely gained in vertebrate groups. Zemach *et al.* (2010) proposed that the evolution of global methylation profile was driven by sexual reproduction and the need to silence the transferred transposable elements.

Gene methylation has been identified in several vertebrate and invertebrate organisms. Interestingly, it is not related exclusively to silencing of the genes. In fact, DNA methylation within genes is related to transcription activation, in both vertebrates and invertebrates (Feng *et al.*, 2010a; Sarda *et al.*, 2012; Spruijt and Vermeulen, 2014; Suzuki *et al.*, 2007; Wang *et al.*, 2013).

Gene body methylation has also been suggested to increase mutation rates leading to reduced numbers of CpG dinucleotides in DNA sequences (Goll and Bestor, 2005; Zemach *et al.*, 2010). This observation led to the idea that methylated genes should show reduced sequence conservation among different taxa. However, studies conducted in insects and plants have

contradicted this hypothesis. In fact, highly methylated genes have shown greater sequence conservation than low methylated genes in invertebrates (Sarda *et al.*, 2012).

Additionally, the new insights presented for gene body methylation, suggest that the function is not only related to quantity of transcripts, but rather the composition, since some studies have already linked it to alternative splicing (Head, 2014; Suzuki and Bird, 2008; Wang *et al.*, 2013).

Although some hypotheses have arisen, DNA methylation function and impact on vertebrate and invertebrate genomes has not been comparatively studied. This is interesting, since the conservation of gene body methylation from invertebrates to vertebrates has been demonstrated many years ago (Regev *et al.*, 1998; Tweedie *et al.*, 1997), and has been cited in many studies and reviews throughout this time (Feng *et al.*, 2010a; Sarda *et al.*, 2012; Zemach *et al.*, 2010). In vertebrates it could be due to the lack of clear relationship between DNA methylation distribution and the different contribution to gene expression, especially due to the constant focus on DNA methylation in promoter regions. In invertebrates it can be attributed to the absence of an ideal invertebrate model organism for epigenetic studies. The near absence of DNA methylation in *D. melanogaster* and *C. elegans*, traditional model organisms for invertebrate genetics, has contributed to this lack of interest in investigating the functional differences between DNA methylation in vertebrates and invertebrates (Capuano *et al.*, 2014; Park *et al.*, 2011; Simpson *et al.*, 1986).

Therefore, it is important to avoid assumptions based on previous observations, because often they will lead to the simplistic idea of DNA methylation as a repressor. Instead, the highly conserved DNA methylation in gene bodies can be hypothesised as the key feature, often

associated with transcription activation, while the repression of unwanted regions should be seen as an added function of DNA methylation, observed in higher chordate taxa (vertebrates) (Tweedie *et al.*, 1997).

1.4   The interactions between the environment and the epigenome

External stressors, such as chemical pollutants, dietary components, predators and temperature changes can disturb an organism's development, metabolism and health. In part, organisms respond to external cues by specifically altering their DNA methylation patterns (Feil and Fraga, 2012). Alternatively, changes in DNA methylation profiles can be due to malfunction of the mechanisms of DNA methylation maintenance. Both responses can lead to changes in the phenotype of the organism, either having a negative effect or as an adaptive response (Jaenisch and Bird, 2003; Vandegehuchte and Janssen, 2013).

In animals, heritable epigenetic adaptation in response to environmental changes is often related to quantitative epigenetic traits rather than individual genes. In contrast, plants present several examples of single or multiple loci where methylation was altered in response to cold, osmotic and salt stress, and was maintained through subsequent generations (Chinnusamy and Zhu, 2009; Heard and Martienssen, 2014; Whittle *et al.*, 2009).

DNA methylation differences can be associated to mechanisms that contribute to biological diversity, such as the mechanisms of cell differentiation, or with mechanisms that allow adaptation or organisms to changing environments. On the other hand, abnormal DNA methylation alterations can be associated with the development and/or progression of several

diseases, such as cancer and neurological disorders, impairing the organisms' health (Bird, 2002; Jones and Baylin, 2002; Jones, 2012).

The effects of several different chemicals and stressors on DNA methylation have already been investigated. Carcinogenic metals, including nickel, cadmium, lead, and particularly arsenic, show a weak mutagenicity capacity, therefore, it is suggested that epigenetic mechanisms underline the carcinogenicity of these compounds (Baccarelli and Bollati, 2009). In general, metals affect DNMTs expression and activity leading to global DNA hypomethylation (Takiguchi *et al.*, 2003; Zhao *et al.*, 1997). However, some compounds can have specific effects on chromatin structure.

Arsenic detoxification occurs by the methylation of inorganic arsenic, using SAM as the methyl donor. Therefore, the metabolism of arsenic can have an effect on the availability of methyl groups for DNA methylation (Reichard and Puga, 2010; Reichard *et al.*, 2007; Zhao *et al.*, 1997). Arsenic is a major concern as an environmental contaminant, naturally present in freshwater and groundwater, affecting populations in different places in the world, including Chile, Argentina, India and the United States of America (WHO, 2011). Even low concentrations of arsenic in the water represent a great hazard to human health due to continuous exposure through drinking water and food. Populations in Bangladesh have been diagnosed with skin lesions and increased risk of developing cancer due to arsenic exposure (Argos *et al.*, 2015).

It is not only metals that can affect the methylome of organisms. Different endocrine disruptors, such as polychlorinated biphenyl (PCBs), 17β-estradiol (E2) and Bisphenol A (BPA), have demonstrated effects on global methylation and in specific regions of the DNA.

PCB exposure causes alteration on the DNA methylation levels when exposed *in vitro* and in analysed populations that were exposed to environmental contamination of different persistent organic pollutants (POPs). For example, PCB-153 caused DNA hypomethylation in murine N2A cell line (Bastos Sales *et al.*, 2013). Hypomethylation was also observed in different human populations exposed to PCBs, including healthy Japanese women, a population of healthy Koreans and in a population of Inuits (Itoh *et al.*, 2014; Kim *et al.*, 2010; Rusiecki *et al.*, 2008). For all those groups the concentration of PCBs in the serum was inversely correlated with global DNA methylation. On the other hand, a study analysing methylation levels in an elderly Sweden population identified hypermethylation of DNA (Lind *et al.*, 2013). Differences were attributed to divergence in age, geographical location, and lifetime exposure levels to different PCBs (Keil and Lein, 2016).

BPA has also been related to effects on DNA methylation. DNMT1 and DNMT3 expression was upregulated in ovaries of zebrafish exposed to 5µg L^{-1} of BPA for 3 weeks, while DNMT4, DNMT6 and DNMT7 showed reduction in relative abundance (Santangeli *et al.*, 2016). In another study, also analysing the effects of BPA on zebrafish (1 mg L^{-1} for 15 days), the expression of DNMT1 was reduced, along with significant global hypomethylation of ovaries and testis DNA (Laing *et al.*, 2016).

Mirbahai *et al.* (2011) demonstrated the effects of environmental exposures in contaminated marine sites off UK shores on the flatfish dab (*Limanda limanda*). High occurrence of liver tumours and altered DNA methylation profiles were related to the increased contamination in the analysed sites, especially endocrine disruptors, such as PCBs.

Besides the contamination from chemical compounds, adverse environmental conditions can also be harmful to aquatic organisms. Hypoxia is an important stressor, since the depletion of oxygen can impair growth, disturb the reproduction and even cause death of aquatic populations (Long *et al.*, 2015). An increase in anthropogenic input of organic matter and nutrients can increase algal growth leading to reduction of dissolved oxygen concentrations in water bodies (Wu, 2002).

Some studies have linked the organisms' responses to hypoxic conditions to regulation by epigenetic mechanisms, however the investigations often use a medical perspective since hypoxia is often linked to tumour progression or ischemic events (Brown and Rupert, 2014; Hattori *et al.*, 2015; Lachance *et al.*, 2014; Tsai and Wu, 2014; Tudisco *et al.*, 2014).

As demonstrated above, several stressors have the ability to modify epigenetic marks, altering DNA methylation profiling. The changes in epigenetic marks can also be maintained during lifetime, and potentially beyond into subsequent generations, emphasizing the concept of "epigenetic memory" (Bird, 2002; Mirbahai and Chipman, 2014; Skinner, 2008).

The "epigenetic memory" also can play a role in the transmission of disrupted epigenetic information to following generations. The possibility of transgenerational epigenetic inheritance has been suggested for multiple generations for several organisms, including mammals, plants and invertebrates (Hauser *et al.*, 2011; Skinner, 2014; Skinner *et al.*, 2010).

The term "memory" in epigenetic studies can have different interpretations. In this study, this refers to the persistence of abnormal epigenetic marks without the presence of the stressor, within the same generation. This can represent a potential initiation factor for negative health outcomes later in life (Head *et al.*, 2012).

Studies analysing the DNA methylation profiles of monozygotic twins provide strong evidence of memory of epigenetic alterations. One of the most important studies was conducted by Fraga *et al.*, (2005), and demonstrated that identical twins can be distinguished later in life when comparing their epigenetic profiles and suggest that differences were caused by the accumulation of epigenetic alterations during their life-times based on their different environmental exposures. These findings help to explain how different phenotypes can originate from the same genotype and how the environment can be the key to modulate such changes.

Therefore, it is proposed that different classes of stressors can induce class-specific alteration in the normal DNA methylation profile. These altered epigenetic marks can potentially have effects on the organism, either immediately or at a later stage of life.

Additionally, the stressor-specific epigenetic profile offers a unique opportunity for environmental monitoring, since it could provide a lifetime history of past exposures and be used to infer to which class of pollutants the organisms were exposed during their lifetime (Mirbahai and Chipman, 2014). Therefore, stressor-specific changes in DNA methylation could be used as biomarkers of exposure to certain stressors as well as for early detection of adverse effects and monitoring of the progression of diseases (Mikeska and Craig, 2014).

The addition of an epigenetic perspective to toxicology and ecotoxicology studies could improve the understanding of modes of action of several compounds that affect the epigenome and leading to changes in gene expression and later on the phenotype (Head *et al.*, 2012). Several endpoints are important in ecotoxicological studies and many of them, such

as growth, development and reproduction, are known to be regulated by epigenetic mechanisms (Head *et al.*, 2012).

Some studies have analysed the effects of the environment on the epigenome. However, most of the studies have been conducted in mammalian model organisms. These are not directly relevant for ecotoxicology but are an important tool to understand the different epigenetic mechanisms (Vandegehuchte and Janssen, 2013).

One of the major challenges for future research is linking epigenetics and ecotoxicology to evaluate the possible effects on populations, since protection of communities and populations is the main goal for ecological risk assessment (Vandegehuchte and Janssen, 2013).

1.5   Model organisms for epigenetic studies

The current knowledge on epigenetics is the result of many studies conducted in a wide variety of organisms, from fungi to mammals (Table 1.1). Each can play a significant role in advancing our understanding of different processes and mechanisms in the epigenetic field.

The model organism *Saccharomyces cerevisiae* (yeast) has been crucial in genetic and epigenetic discoveries. Several genetic and epigenetic regulatory mechanisms, such as histone post-translational modifications, were first discovered in yeast. However, DNA methylation was not observed in this organism (Fuchs and Quasem, 2014; Grunstein and Gasser, 2013).

While *Drosophila melanogaster* (fruit fly) and *Caenorhabditis elegans* (roundworm) are undeniably two of the most powerful model organisms for genetic research, their uses in epigenetic research are limited to markers, such as histone modifications and microRNAs. They are not suitable for DNA methylation studies as either DNA methylation levels are very

low, without a defined pattern and, as yet, no indication of function (*D. melanogaster*) or DNA methylation has not been detected (*C. elegans*) (Capuano *et al.*, 2014; Hu *et al.*, 2015; Park *et al.*, 2011; Simpson *et al.*, 1986; Takayama *et al.*, 2014).

Originally, efforts were focused on the use of vertebrate species such as *H. sapiens* (human), *M. musculus* (mouse) and *D. rerio* (zebrafish). However, the use of mouse requires a large number of animals, requiring large efforts to maintain them in the lab, increasing the costs of research (Mukherjee *et al.*, 2015). Human and mouse cell lines can be applied for different studies in epigenetics mechanisms. They are of great value to understand the molecular bases of epigenetic and genetic mechanisms, especially for medical research. Nevertheless, their limiting factors include the difficulty of inferring the effects on organisms' phenotype, interaction among different cell types and transgenerational effects.

Zebrafish (*D. rerio*) is a suitable model organism for DNA methylation studies in a non-mammalian species, and a promising system to study the epigenetic effects of environmental stressors (Kamstra *et al.*, 2015a). Effects of several chemicals on DNA methylation have been identified, such as global hypomethylation caused by exposure to benzo[a]pyrene, 5-azacytidine and sodium arsenite, and gene specific hypomethylation when exposed to 17α ethinylestradiol (Fang *et al.*, 2013; Li *et al.*, 2009; Olsvik *et al.*, 2014; Strömqvist *et al.*, 2010). However, despite the advantages of using zebrafish for epigenetic studies it does not overcome the ethical constrains of using vertebrates for research.

One of the future steps of epigenetics research is also to investigate the effects of and interactions with different environmental conditions. In light of this, epigenetic mechanisms have been investigated across a wide range of environmentally relevant invertebrate species,

such as the honey bee *Apis mellifera* (Lyko *et al.*, 2010; Rasmussen and Amdam, 2015), the wasp *Nasonia vitripennnis* (Beeler *et al.*, 2014; Wang *et al.*, 2013; Zwier *et al.*, 2012), the ant *Camponotus floridanus* (Glastad *et al.*, 2015), and the oyster *Crassostrea gigas* (Gavery and Roberts, 2010; Rivière, 2014; Wang *et al.*, 2014).

The advantages of using invertebrates are partly linked to their short generation time and small size as well as their ease of culturing and maintenance for multiple generations in the laboratory. Studies in invertebrates in general, are also considered ethically acceptable, and regulations are less strict for these organisms (Mukherjee *et al.*, 2015).

Insects have been highlighted as promising model organisms for epigenetic studies (Mukherjee *et al.*, 2015), and can be valuable for studies trying to comprehend the effects of environmental factors and can help to infer about evolutionary adaptation to stressors and effects on the ecosystem. However, it can be challenging to induce phenotypic plasticity in these species in response to environmental cues under laboratory conditions. The epigenetic studies in these species have, so far, been restricted to analyses of the mechanisms and, in some cases, the relationship with the social behaviour in hymenoptera. No studies analysing the effects of chemicals or stressors on the DNA methylation profile of these organisms have been reported.

Besides all the discussed advantages of using invertebrates for epigenetic studies, the use of *Daphnia* species adds the advantage of using an organism already standardised for ecotoxicological testing and with an extensive database of responses to several chemicals. Phenotypic responses to environmental cues are also described for *Daphnia*, including

morphological modifications to avoid predation and increase in haemoglobin concentration to survive low oxygen environments (Stollewerk, 2010; Zeis *et al.*, 2013).

Therefore, *Daphnia* is a highly suitable organism model for toxicological studies, including the ones investigating epigenetic mechanisms, due the particular features as morphological responses and clonal reproduction, and due to an extensive knowledge base constructed for many years on this recognised model organism.

1.5.1 *Daphnia magna* as a test organism

Daphnia spp*.* are considered keystone species in both lakes and ponds and are well-studied in terms of their ecology and response to stressors, both under laboratory conditions and in the field (Lampert and Kinne, 2011). *Daphnia magna* are freshwater microcrustaceans, with a short life span and short time until reproduction (Figure 1.9). Contrary to other model organisms, *Daphnia* play important roles in ecosystems worldwide and interact with different trophic levels (Lampert and Kinne, 2011). These species are already well-studied in the context of their ecological role and show great potential to be used for epigenetic studies (Harris *et al.*, 2012).
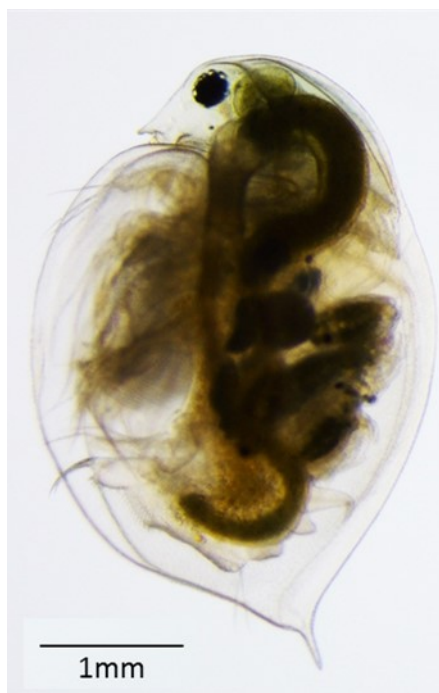
Figure 1.9 Adult *D. magna* showing parthenogenetic offspring in the brood pouch.

Daphnia magna has been used extensively for ecotoxicological assays for many years (OECD, 2004, 2012) while *Daphnia pulex* has been listed as a model system for biomedical research by the National Institutes of Health, USA (Colbourne *et al.*, 2011). Furthermore, both species have been proposed as model organisms for environmental genomics, toxicogenomics and epigenetics studies (Eads *et al.*, 2008; Colbourne *et al.*, 2011; Harris *et al.*, 2012; Miner *et al.*, 2012).

Daphnia are key model organisms used for research into the molecular mechanisms of phenotypic plasticity, adaptation and microevolution (Giessler *et al.*, 1999; Van Doorslaer *et al.*, 2009; Messiaen *et al.*, 2010; Messiaen *et al.*, 2013; Geerts *et al.*, 2015). The extensive use of *Daphnia* spp. in a wide range of research fields has motivated the development and

optimisation of several 'omics technologies to probe the molecular machinery within these species (Taylor *et al.*, 2008; Dircksen *et al.*, 2011; Colbourne *et al.*, 2011).

The resulting growth of genomic resources for *Daphnia* spp., coupled with the dramatic reduction in costs and accessibility of sequencing technologies and other genomic tools, has fuelled their increasing use in environmental genomics, toxicogenomics and evolutionary biology (Pfrender *et al.*, 2000; Omilian and Lynch, 2009; Orsini *et al.*, 2012; Hochmuth *et al.*, 2015).

Daphnia species are ideal as environmentally-relevant invertebrate model organisms for epigenetic research. They can easily respond to environmental cues, altering some characteristics of their life-cycle, creating phenotypic variability in organisms, allowing them to face environmental changes (Colbourne *et al.*, 2011). The cyclic parthenogenetic reproduction of *Daphnia* creates a stable genetic background very useful for epigenetic studies (Figure 1.10) (Castonguay and Angers, 2012; Robichaud *et al.*, 2012). Therefore, *Daphnia* can be applied to investigate the role of DNA methylation in multiple areas, such as response to stressors, adaptation, phenotypic plasticity and maternal transfer of information, without the variation in genetic background found for other species (Harris *et al.*, 2012; Vandegehuchte and Janssen, 2013).

Daphnia genome sequencing has been pursued by the *Daphnia* Genomics Consortium. The genome of *D. pulex* was published in 2011. The draft genome is an extremely densely packed sequence, it contains more than 30,000 genes, a higher number of genes than predicted for human, and is only 200 megabases (Colbourne *et al.*, 2011; Ezkurdia *et al.*, 2014; Human Genome Sequencing Consortium, 2004). Recently, the transcriptome generated after the

exposure to 12 stressors was published for *D. magna* (Orsini *et al.*, 2016). However, *D. magna* still lacks a complete annotated genome. The predicted size of the *D. magna* genome is 238 Mb (Routtu *et al.*, 2014). The available genome sequence (NCBI BioProject PRJNA298946) is shorter than predicted (122 Mb) and only encodes 65% of the predicted gene transcripts (Orsini *et al.*, 2012; Routtu *et al.*, 2014).
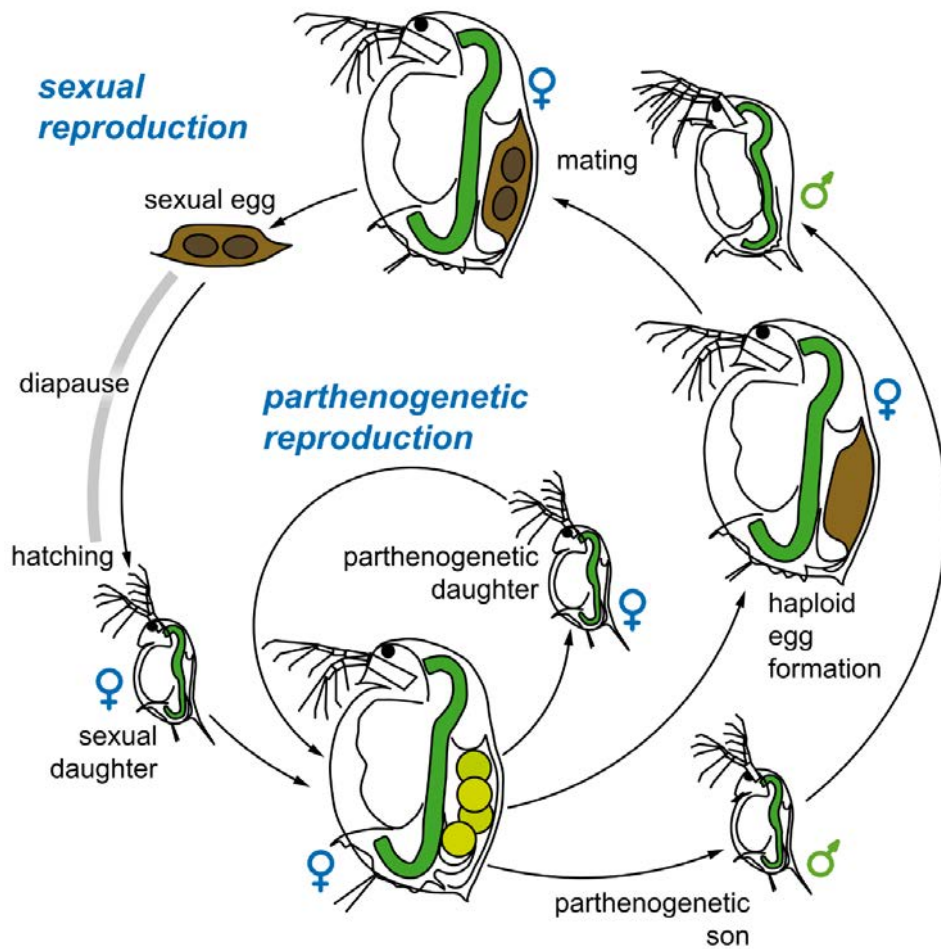
Figure 1.10 Life cycle of *Daphnia* species. Parthenogenetic reproduction occurs under favourable conditions while stress conditions lead to switch to sexual reproduction (Reproduced from Ebert D., 2005).

1.5.2 Epigenetic studies using *Daphnia magna*

As discussed above, *Daphnia magna* shows great potential as an environmentally relevant model organism to investigate the role of DNA methylation in multiple areas, such as response to stressors, adaptation, phenotypic plasticity and maternal transfer of information (Harris *et al.*, 2012).

Despite the great potential of *Daphnia* for epigenetic studies, its DNA methylation toolkit (i.e. genes involved in DNA methylation and demethylation and one-carbon pathway) has not been profiled comprehensively.

The occurrence of DNA methylation in *D. magna* was firstly described in Vandegehuchte *et al.*, (2009a). The homologous genes for the vertebrate DNA methyltransferases were also described for *D. magna*. However, the additional enzymes involved in the DNA methylation pathway have not been described in *D. magna*. Histone modifications are also present in *Daphnia* (Robichaud *et al.*, 2012).

Global levels of DNA methylation, measured by LC-MS, are described for two different inbred strains. For the Iinb1 strain, the global DNA methylation was 0.49 ± 0.19%. The Xinb3 strain presented a global methylation level of 0.52 ± 0.16% (Asselman *et al.*, 2015).

Global DNA methylation changes were observed in *D. magna* in response to several chemicals, including 5-azacytidine, genistein and vinclozolin (Vandegehuchte *et al.*, 2009a, 2009b, 2010a, 2010b) and environmental stressors, such as dissolved humic substances, predation cues, low-quality food and salinity (Asselman *et al.*, 2015; Menzel *et al.*, 2011). Furthermore, recent studies have analysed the distribution of DNA methylation across various regions of genome in *Daphnia* species (Asselman *et al.*, 2016; Strepetkaitė *et al.*, 2015). Gene specific methylation

was also analysed in *D. magna* and showed relationship with gene family size and diversification of genes (Asselman *et al.*, 2016).

Although *D. magna* present a great potential for epigenetic studies and have shown altered global levels of DNA methylation when exposed to chemicals, no studies have been published analysing gene-specific hyper/hypomethylation in response to stressors. Therefore, the major goal of this thesis was to describe the DNA methylation mechanisms in *D. magna*, and to identify the altered methylation profiles caused in response to different stressors.

1.6   Aims

Overall, the aim of this thesis was to describe the genome-wide distribution of DNA methylation as well as the DNA methylation changes induced in response to environmentally relevant exposure conditions in *Daphnia magna*, a potential invertebrate model organism for epigenetic studies.

The specific aims of this study are:

i)   To describe the overall pattern of DNA methylation across the genome of *D. magna*;

ii)   To present the DNA methylation machinery of *D.* magna and the dynamic changes in DNA methylation machinery that occur in response to age. Accomplishing these aims was necessary to achieve a basic knowledge of *Daphnia*'s methylome, enabling specific investigations regarding the role of DNA methylation in *Daphnia*.

iii)   To test different methods of differentially methylated regions (DMRs) identification (biased vs. unbiased methods) using the visualisation software

SeqMonk and whole genome bisulfite sequencing datasets from *Daphnia* exposed to 5-azacytidine and respective control

iv) To identify DMRs induced in *Daphnia* as a result of exposure to 5-azacytidine (3.7 mg L^{-1}, 5 days exposure), arsenic (100 µg L^{-1}, 14 days exposure) and hypoxia (<2 mg L^{-1} of dissolved oxygen, 14 days exposure) using the selected method of analysis in aim (iii).

v) To investigate the sensitivity of the *Daphnia's* epigenome to three stressors: 5-azacytidine, arsenic and hypoxia, using acute and chronic exposures. The epigenome was investigated under three perspectives: (i) methylation of regulatory regions and gene bodies using Whole Genome Bisulfite Sequencing (WGBS) and direct bisulfite sequencing, (ii) metabolites quantitation from the one-carbon pathway, and (iii) expression levels of selected genes.

vi) To test the concept of epigenetic memory and recovery. Therefore, the aim was to assess the presence of stressor-specific alterations on the DNA methylation, metabolites concentration and gene expression, and the maintenance of those alterations once the stressor is removed.

These aims are addressed in chapters 3, 4 and 5.

Chapter 2

Material and methods

2.1   Chemicals

All chemicals were obtained from Sigma-Aldrich Co. (UK) unless otherwise stated.

2.2   Culturing of *Daphnia magna*

Daphnia magna Bham2 strain was used for all the exposures in this study (original animals were obtained from ▨▨▨▨▨▨▨ where they were categorised as IRCHA Clone Type 5). The animals were maintained in a 16:8 hrs light:dark photoperiod and temperature of 20 ± 2°C. The organisms were maintained in density of 20 *Daphnia* in 1200 mL of media. The media was renewed once a week. All cultures were initiated using third brood neonates aged <24 h. Animals were acclimated for a minimum of 3 generations prior to use in any experiments.

2.2.1 Modified high hardness COMBO media preparation

The modified high hardness COMBO medium (mHHCOMBO) was prepared according to the protocol adapted from Baer and Goulden (1998) and Kilham *et al.* (1998). The mHHCOMBO was prepared by the addition of the stocks to distilled water to the final concentrations described in Table 2.1 (Keating and Dagbusan, 1984). The media was aerated for 24h and the pH was adjusted to the range between 7.6 to 7.8 using hydrochloric acid (HCl). Animal trace element (ANIMATE), a micronutrient solution, was also added to the media.

Table 2.1 Modified High Hardness COMBO (mHHCOMBO) for culturing of *Daphnia magna.*

Compound	Stock	Final medium
	(g L^{-1})	(mg L^{-1})
Major Stocks		
CaCl$_2$.2H$_2$O	110.28	110.28
MgSO$_4$.7H$_2$O	55.45	55.45
K$_2$HPO$_4$	1.742	1.742
NaNO$_3$	17	17
NaHCO$_3$	63	126
Na$_2$SiO$_3$.9H2O	28.42	28.42
H$_3$BO$_3$	24	24
KCl	5.96	5.96
Na$_2$SeO$_3$	0.04	0.002
Animal Trace Elements (ANIMATE)		
LiCl	310	0.31
RbCl	70	0.07
SrCl$_2$.6H$_2$O	150	0.15
NaBr	16	0.016
KI	3.3	0.0033

2.2.2 Culturing of Chlorella vulgaris

Chlorella vulgaris was cultured in Bold's basal media (BBM) in a closed aerated system, under constant light. The protocol for preparation of BBM is detailed in Table 2.2.

For algae suspension preparation, a known volume was taken from the culture flasks and the optical density (1:10) was measured at 440 nm. Algae were centrifuged at 2,250 x g for 30 minutes at room temperature. Algae were re-suspended in deionised water to obtain the required optical density of 0.800. The final volume was calculated as follow:

$$Final\ volume\ =\ (measured\ OD\ (1:10)\ \times\ volume\ of\ algae)\ /\ 0.800$$

Cultures were fed daily on suspensions of algae. The volumes varied according to age: <2 days old: 1 mL; 3-7 days old: 1.5 mL; >7 days old: 2 mL per culture.

Table 2.2 Bolds Basal Medium (BBM) for culturing of *Chlorella vulgaris.*

Compound	Stock (g L^{-1})	Final medium (g L^{-1})
KH_2PO_4	17.5	0.175
K_2HPO_4	7.5	0.075
$MgSO_4.7H_2O$	7.5	0.075
$NaNO_3$	25	0.25
$CaCl_2.2H_2O$	2.5	0.025
NaCl	2.5	0.025
EDTA Na_4	50	0.05
KOH	31	0.031
$FeSO_4.7H_2O$	4.98	0.00498
H_2SO_4	10 mL/L	
H_3BO_3	11.42	0.01142
$ZnSO_4.7H_2O$	14.12	0.001412
$MnCl_2.4H_2O$	2.32	0.000232
$CuSO_4.5H_2O$	2.52	0.000252
$Co(NO_3)_2.6H_2O$	0.8	0.00008
$Na_2MoO_4.2H_2O$	1.92	0.000192

2.3   Treatments and exposure design

2.3.1 Exposure design

The exposure design followed the OECD guidelines for assessment of chronic toxicity with some modifications (OECD, 2012). As described in section 1.2.6, in mammals, DNA methylation undergoes two cycles of demethylation and remethylation. The first one occurs immediately after fertilization and the second one occurs in primordial germ cells. The function of these events is related to reprogramming and gain of cell type specific DNA

methylation profiles. These are also the most critical and sensitive periods for disruption of DNA methylation (Feng *et al.*, 2010b).

The reprogramming events are not yet described for *Daphnia*, however the exposures were designed to be compatible with such events.

The exposure design is represented in Figure 2.1. Neonates <24h of age were kept in clean media until the eggs in the brood pouch were visible and detectable (around 8 days old). The female *Daphnia* carrying eggs were randomly divided into two groups; exposure and control. The exposure group were exposed while the control group was maintained in clean media. The first and second broods in both groups were discarded. The third brood was maintained either in clean media (control group) or continuously exposed (treatment group). In contrast to the OECD exposure procedures, this exposure design ensures that the *Daphnia* are exposed throughout embryogenesis.

The duration of each exposure differed according to each experiment. For the whole genome bisulfite sequencing (WGBS) experiment the exposure to arsenic and hypoxia lasted for 14 days while animals were exposed to 5-azacytidine for 5 days. For the experiment evaluating accumulation and persistence of the effects on the epigenome the animals were exposed for 21 days to arsenic and hypoxia and 5 days to 5-azacytidine. Each group then was kept in clean media for an additional 7 day recovery period. In both experiments the animals were exposed during embryonic stage, while in the brood pouch, as detailed before.
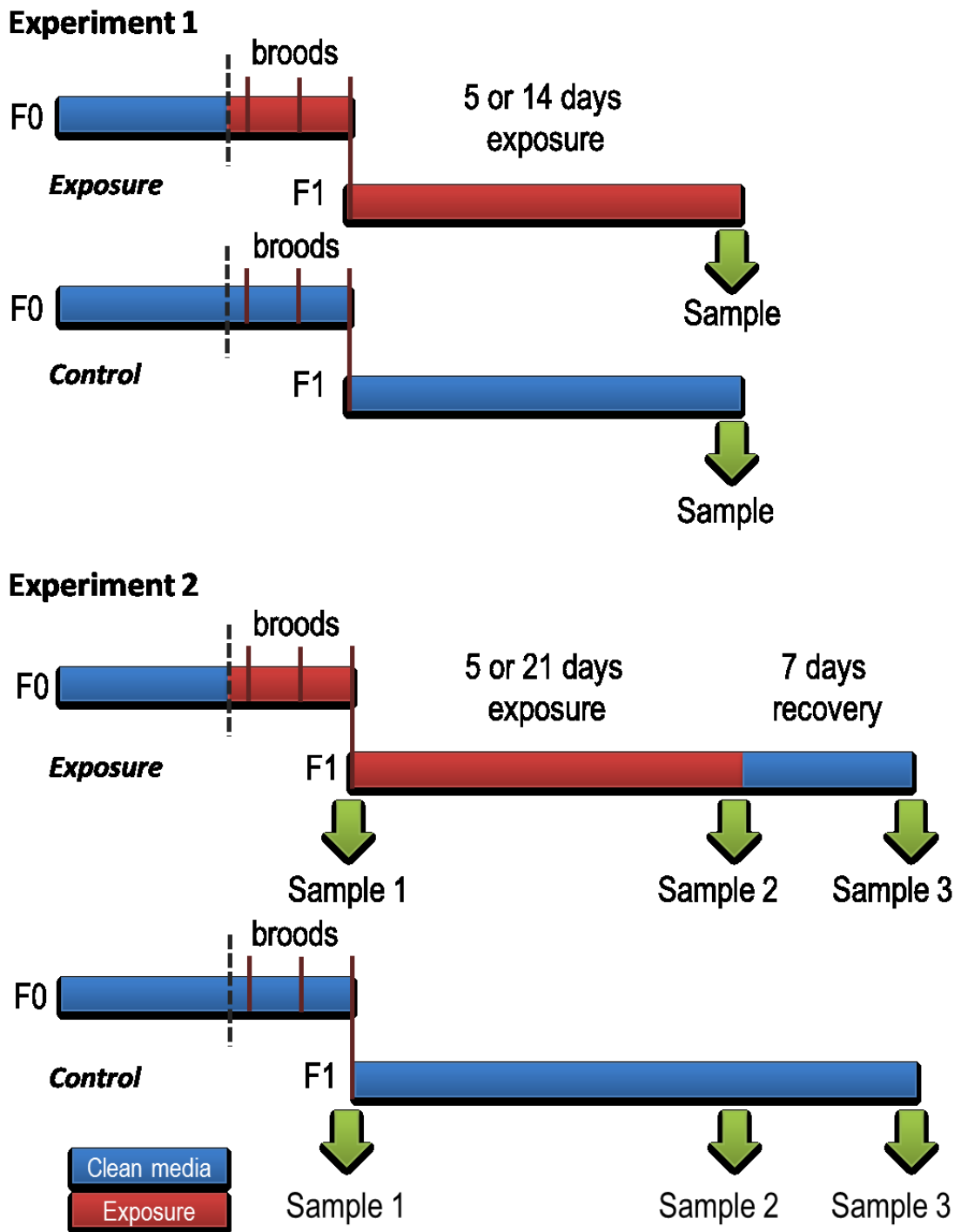
Figure 2.1 Schematic representation of exposure design used for experiments 1 and 2. F0: Animals exposed only after maturity until the release of third brood (broods represented by dark red lines). F1: Animals used for experiment after exposure during developmental stages. Blue: animals maintained in clean media. Red: animals exposed to stressor. Green arrows represent the moment where samples were obtained.

2.3.2 Treatments

The exposures consisted of three different treatments: 5-azacytidine, arsenic and hypoxia. 5-azacytidine was used as a chemical with known effects on DNA methylation (positive control). 5-azacytidine is a cytosine nucleoside analogue that is incorporated into DNA synthesised during replication. DNA methyltransferases (DNMTs) recognise 5-azacytidine as a substrate. During this process the enzyme becomes inactivated, due to a covalent and irreversible biding to the 5-azacytidine, leading to DNA hypomethylation (Santi *et al.*, 1984; Stresemann and Lyko, 2008).

5-Azacytidine (7.4 mg L^{-1}) is known to reduce the global DNA methylation levels in *D. magna* (Vandegehuchte *et al.*, 2010b). A pilot study was conducted using the same concentration reported in this paper, however this concentration caused high levels of toxicity to the *Daphnia,* therefore, half of this concentration (3.7 mg L^{-1}) was used in our study. In addition, a shorter exposure duration of 5 days was used for DNA methylation profiling and evaluation of the concept of "epigenetic memory".

Arsenic is a non-genotoxic carcinogenic metal known to induce changes in DNA methylation and is an important environmental pollutant. There are two pathways described for the metabolism of inorganic arsenic (Hayakawa *et al.*, 2005; Vahter, 2002) and both include biomethylation of arsenic by S-adenosylmethionine (SAM) reducing the amount of SAM available for DNA methylation (Lindberg *et al.*, 2007).

The arsenic concentration used in this study (100 µg L^{-1}) is based on environmentally relevant concentrations. Surface and groundwater values stated in the literature show a wide range from <0.5–5000 µg L^{-1}, but high values have only been related to mining activities.

The third selected stressor is hypoxia, an environmental stressor potentially caused by organic contaminants and eutrophication. The discharge of organic compounds to surface water leads to increased levels of primary production in the ecosystems. Once the organic matter starts to decompose, oxygen is depleted. Dissolved oxygen levels below 2 mg L^{-1} can be harmful to aquatic organisms. The effects of hypoxia are potentially related to changes in DNA methylation. Previous research has demonstrated that in response to hypoxia the hypoxia-inducible transcription factors (HIFs) are expressed and activated in human cell lines and rats. This transcription factor can recognise specific binding sites within the genome and activate and regulate the expression of many downstream genes. However, it has been shown that the recognition site for HIF contains CpG sites which are required to be unmethylated for the TF to be able to access and bind to its binding site (Rössler *et al.*, 2004; Wagner *et al.*, 2003).

To produce the hypoxic condition, air with low oxygen content (4% O2, balanced with nitrogen, BOC, UK) was bubbled into the media. A continuous flow of oxygen and nitrogen gases ensured that the media contained 2 mg L^{-1} of dissolved oxygen at 20°C throughout the experiment. The amount of oxygen was monitored during the exposure using a Unisense microrespiration system (Unisense S/A, Denmark).

2.4   Sample preparation

Fifty neonates (<24h old), 30 juveniles (5 days old), 10 adults (12 and 14 days old) or 5 adults (21 and 28 days old) were used per biological replicate. When necessary embryos were dissected and removed from the brood pouch and the samples were snap frozen in liquid nitrogen and stored at -80°C until processed.

Samples from the first experiment were ground with a plastic pellet pestle, homogenised in methanol:water and DNA was extracted for genome wide DNA methylation analyses. Samples from second experiment were homogenised in 320 µL of methanol and 128 µL of water (both HPLC grade) using a ceramic bead-based system (Precellys 24, Stretton Scientific Ltd, UK). Samples were aliquoted for RNA extraction, metabolites extraction and DNA extraction.

The different methods of DNA extraction from *Daphnia*, including sample storage, homogenisation and extraction were assessed and have been published in Athanasio *et al.* (2016).

2.4.1 DNA extraction

2.4.1.1   CTAB method for DNA extraction

DNA samples from the first experiment were obtained from 10 animals, dissected and frozen, and extracted using a modified CTAB protocol (Doyle *et al.*, 1987). The frozen organisms were ground using a plastic pestle and homogenised in 300 µL of methanol:water (214:86 µL) solution. CTAB buffer (2% hexadecyltrimethyl ammonium bromide CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8, 0.2% β-mercaptoethanol) was added to the homogenised sample (500 µL). After 60 minutes of incubation period at 50°C, the extraction was performed using 500 µL of chloroform:isoamyl alcohol (24:1). The samples were centrifuged at maximum speed (13,000 x g) and the top aqueous layer containing the DNA was transferred to a clean sterile 1.5 mL eppendorf tube. RNAse A (4 µL from 1 µg/µL stock) was added to the sample, and incubated for 30 minutes at 37°C. Another extraction was performed using chloroform:isoamyl alcohol. In a clean tube, 1 volume of isopropanol was added to the aqueous layer to precipitate the DNA. Sample was incubated for at least 1h at -80°C. Then, the

sample was washed using cold 100% (v/v) ethanol followed by 70% ethanol. The pellet was air-dried and resuspended in sterile water. Following extractions, all samples were stored at -80°C.

The extracted DNA was quantified with: (1) 8000 UV-Vis spectrophotometer (NanoDrop, Wilmington, DE, USA) and (2) SYBR Green DNA I dye (Thermo Fisher Scientific, Paisley, UK) using an Infinite® 200 PRO microplate reader (Tecan, Männedorf, Switzerland). These represent an absorbance- and a fluorescence-based methods of DNA quantification, respectively. The quality and integrity of DNA samples and potential RNA contamination were also assessed using a 1% agarose gel in TBE buffer containing Midori Green Advance DNA Stain (Nippon Genetics, Dueren, Germany). Same amount of DNA for each sample was loaded onto the gel and electrophoresis was performed at 80 V for 50 minutes.

2.4.1.2   DNA extraction using protein precipitation method.

DNA samples from the second exposure were extracted using a MasterPure DNA purification kit (Epicentre, USA). The protocol was modified from the manufacturer's instructions. Briefly, samples were preserved and homogenised using the same procedure as described for the CTAB method. After homogenisation, samples were centrifuged at maximum speed (13,000 x g) for 6 minutes to remove methanol:water supernatant on a benchtop centrifuge (Eppendorf, UK). Tissue and Cell Lysis solution (300 µL) and Proteinase K (1 µL at 50 µg/µL) were added to the pellet. The samples were homogenised and incubated at 65°C for 15 minutes, vortexing briefly every 5 minutes. After incubation samples were cooled to 37°C and RNase A (1 µL at 5 µg/µL) was added to the sample. Followed by 30 minutes of incubation at 37°C, samples were placed on ice for 5 minutes.

Precipitation of DNA was achieved by adding 175 µL of MCP Protein Precipitation Reagent. Samples were centrifuged at 4°C for 10 minutes at 13,000 x g. The supernatant was transferred to a new tube. Isopropanol (600 µL) was added to the recovered supernatant. Samples were spun at 4°C for 10 minutes at 13,000 x g. Samples were washed with ethanol as described for CTAB method. Samples were re-suspended in sterile water and stored at -80°C. Quality and yield were assessed using NanoDrop, agarose gel and fluorescence quantitation with Sybr green I.

2.4.2 RNA extraction

RNase free eppendorfs (Axygen, USA), and barrier tips were utilised at all times to prevent degradation of the RNA samples.

The samples homogenised in methanol:water were aliquoted for RNA extraction corresponding to 1/5 of total volume. RNA extraction was performed using RNeasy Micro kit (Qiagen Ltd., UK) according to the manufacturer's instructions with some minor modifications. Samples were centrifuged for 6 minutes at ≥8000 x g to remove methanol:water solution. Buffer RTL (300 µL containing 3 µL of β-Mercaptoethanol) was added directly to the pellet. After mixing, 1 volume of 70% ethanol was added to the lysate. Then, the lysate was transferred into the column and centrifuged for 15 seconds at ≥8000 x g. Column was washed once with 350 µL of RW1 buffer. DNase treatment was performed by addition of 10 µL of DNase I stock and 70 µL of Buffer RDD directly to the column membrane. After 15 minutes of incubation at room temperature, 350 µL of Buffer RW1 was added to the column and centrifuged. Buffer RPE (500 µL) was added to the column and centrifuged, followed by 500 µL of 80% ethanol. Spin column was dried by centrifuging the columns at full speed (13,000 x

g) for 5 minutes. RNase-free water (14 µL) was added to the centre of the column and centrifuged for 1 minute at full speed to elute the RNA.

Samples were quantified with NanoDrop 8000 UV-Vis Spectrophotometer (NanoDrop, USA) and were stored in -80°C until future use.

2.4.3 Extraction of metabolites

Metabolites were extracted from one aliquot (89 µL) of the homogenised samples. The final volumes of the samples were adjusted by addition of 358 µL of methanol:water and transferred from plastic tubes to a 1.8 mL glass vials. Then, 320 µL of chloroform and 160 µL of water were added (final solvent ratio of 2:2:1.8) and samples were vortexed for 30 seconds. Samples were left on ice for 10 minutes and were centrifuged at 1,500 x g at 4°C for 10 minutes. Samples were left at room temperature for 5 minutes to achieve biphasic separation. From the upper layer of the samples containing the polar metabolites, 300 µL were removed and aliquoted (150 µL) into two 1.5 mL microtubes using a glass Hamilton syringe. Polar samples were then dried in a centrifugal concentrator (Thermo Savant, USA) and stored at -80°C until analysed.

2.5   Phenotypic measurements

2.5.1 Body length

Pictures were taken from the exposed and control animals using a stereomicroscope SMZ800 (Nikon, Japan) coupled to a digital camera DS-Fi2 (Nikon, Japan).

The measurement of body length was made from the base of the spine to the top of the head using the Software Image Measurement (KLONK, Denmark) (Figure 2.2). The animals exposed to hypoxia and arsenic were measured at day 1, day 21 and day 28. The group exposed to 5-Azacytidine was measured at day 1, day 5 and day 12.
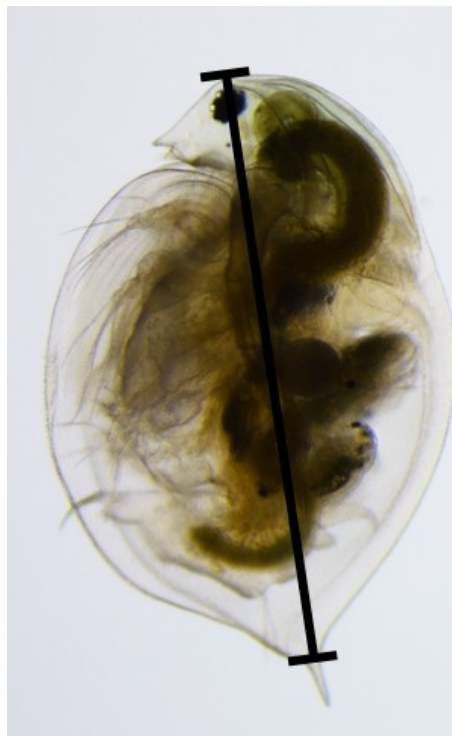
Figure 2.2 Body length measurement using Image measurement software (KLONK, Denmark). Black line indicates the method of measurement, taken from the top of the head/eye until the base of the spine.

2.5.2 Haemoglobin quantification

Haemoglobin (Hb) concentrations of samples exposed to hypoxia and respective controls were quantified according to Yampolsky *et al.* (2014) with modifications.

A single adult *Daphnia* was placed in a 1.5 mL microtube and was frozen in liquid nitrogen. Samples were ground using a plastic pestle and homogenised in 25 µL of Tris-HCl buffer, 0.05M, pH 7.2, and centrifuged at 13,000 x g for 6 minutes.

Supernatant was transferred to a new tube and kept on ice. Absorbance at 414nm, 560nm, 576nm and 600nm was measured using NanoDrop 8000 UV-Vis Spectrophotometer (Nanodrop, USA). Each sample was measured three times for technical replicates. Each group of samples consisted of six biological replicates.

Haemoglobin content was compared between group using Δ576nm values calculated as follow:

$$\Delta 576nm \ = \ Abs \ 576nm \ - \ ((Abs \ 560nm \ + \ Abs \ 600nm)/2)$$

Values were normalized by dividing the Δ576nm values by the total protein concentration measured by the Bradford method (Bradford, 1976). Briefly, Protein Assay Dye Reagent Concentrate (Bio-Rad, USA) was diluted in sterile water to final concentration 1:5. Diluted reagent was filtered using 0.45 µm filter and each cuvette received 1 mL of reagent. Standard curve was constructed using 1, 2, 4, 6, 8 and 10 µg/mL of bovine serum albumin (BSA). Each sample was measured using 2 µL in 1 mL of reagent. Absorbance at 595 nm was measured using a spectrophotometer. Linear regression was calculated based on standards absorbance measurements for known concentrations and use to calculate total protein content of samples.

2.6   Global methylation

2.6.1 DNA hydrolysis

DNA hydrolysis was performed according to Quinlivan and Gregory (2008). Genomic DNA (1 µg) was added to 50 µL of digestion buffer (20mM Tris-HCl (pH 7.9), 100 mM NaCl and 20 mM MgCl$_2$) containing 2.5 U of Benzonase (Sigma Aldrich, UK), 3 mU of phosphodiesterase I (Sigma Aldrich, UK) and 2 U of alkaline phosphate (Sigma Aldrich, UK). Samples were incubated overnight at 37°C. After incubation, samples were filtered using a Ultrafree-MC GV Centrifugal Filter (pore size 0.22 µm). Samples were dried using a centrifugal concentrator (Thermo Savant, USA) and stored at -20°C until further analysis.

2.6.2 LC-MS/MS

The prepared hydrolysed DNA samples were analysed in the Department of Toxicology, University of Wurzburg, Germany for measurement of the percentage of methylated DNA via LC-MS/MS.

DNA hydrolysate was dissolved with 100 µL double-distilled H2O in a vial. LC-MS/MS analysis was performed using an Agilent 1100 series LC coupled to an API 3000 triple quadrupole mass spectrometer equipped with a turbo ion spray source (Applied Biosystems, Germany). Separation was performed by a Reprosil Pur ODS 3 column (150 × 2 mm, 5 µm) with a gradient elution with 0.1% formic acid (solvent A) and methanol (solvent B) using the following conditions: 90% A and 10% B (starting conditions) followed by an increase to 40% in the first 3 minutes and a linear increase to 100% B in 2.5 minutes, at a flow rate of 300 µL/minute. Positive ion mode was used for the detection of the nucleosides at a vaporizer temperature

of 400°C. Data acquisition was performed by multiple reaction monitoring (MRM) of mass transitions of 268.2 mass to charge ratio (m/z) (parent compound) to 152.1 m/z (product) for 2-deoxyguanosine and mass transitions of 242.17 m/z (parent compound) to 108.95 m/z (product) for 5-methyldeoxycytidine. Quantitation of a serial dilution of known amounts of 2-deoxyguanosine and 5-methyldeoxycytidine was used to generate a standard curve for the compounds of interest. Global methylation level was expressed as percentage of methylated cytosines in the total amount of cytosines (measured by the amount of guanosine nucleotide).

2.7   Whole genome bisulfite sequencing (WGBS)

Sodium bisulphite treatment is the standard method for detection of DNA methylation at single nucleotide resolution. When DNA is treated with sodium bisulphite, the unmethylated cytosines are deaminated and converted to uracils while the methylated cytosines are not converted. After PCR amplification, the methylated cytosines remain as cytosines and the unmethylated ones are consequently read as thymines (Frommer *et al.*, 1992). Sodium bisulphite treatment can be combined with whole genome high throughput sequencing allowing the analysis of the methylation status across the entire genome at single nucleotide-resolution (Cokus *et al.*, 2008). Despite being the standard method for DNA methylation analyses, the bisulfite treatment is not able to differentiate 5mC from 5hmC. Due to the general low occurrence of 5hmC, the results obtained with bisulfite treatment are still valid. Few alternatives have been proposed, however, improvements on these techniques are required to be used for genome-wide methylation mapping (Booth *et al.*, 2012).

2.7.1 Sodium bisulfite treatment

DNA samples were treated with EZ DNA Methylation-gold kit (Zymo Research Corporation, USA) for bisulfite conversion following the manufacturer's instructions. Firstly, CT Conversion reagent was dissolved in 900 µL of water, 300 µL of M-Dilution Buffer, and 50 µL M-Dissolving Buffer and vortexed for 10 minutes at room temperature. CT Conversion reagent is light sensitive; therefore, it was handled avoiding exposure to light.

Then, CT conversion reagent (130 µL) was added to the DNA sample (1 µg of DNA in 20 µL). Samples were incubated at 98°C for 10 minutes followed by 64°C for 2.5 hours and chilled at 4°C. M-Binding buffer (600 µL) was added to the Zymo-Spin IC column followed by the addition of sample. Samples were mixed by inverting the column several times and centrifuged at full speed (13,000 x g) for 30 seconds. M-Wash buffer (100 µL) was added to the column and it was centrifuged again. M-desulphonation buffer (200 µL) was added to the columns and samples were incubated for 20 minutes at room temperature. After incubation, they were centrifuged at full speed (13,000 x g) for 30 seconds. Columns were washed twice with M-wash buffer (200 µL) and centrifuged for 30 seconds. Following the final wash, columns were placed in clean eppendorfs and nuclease free water (10 µL) was added directly to the membranes, and then columns were centrifuged (13,000 x g) for 30 seconds to elute the bisulfite treated DNA samples. Treated samples were stored at -80°C until analysed.

2.7.2 Library construction

The libraries used for sequencing were constructed using EpiGnome Methyl-Seq kit (Epicentre, USA). The library construction kit was used for sodium bisulfite treated samples, as well as a

non-converted DNA sample. The latter was used for the *de novo* assembly of *D. magna* genome Bham2 strain. It followed the same procedure for library preparation, with exception of the sodium bisulfite treatment step and adjusting the starting amount of DNA (20 ng).

The first step of the protocol was the annealing of the DNA synthesis primers followed by the synthesis of DNA. The DNA samples (50 ng for sodium bisulfite converted and 20 ng for non-converted samples) were diluted in 9 µL of nuclease-free water, mixed with 2 µL of DNA synthesis primer and incubated at 95°C for 5 minutes. Samples were placed on ice and 5 µL of the mastermix (containing 4 µL of EpiGnome DNA synthesis premix, 0.5 µL of 100 mM DTT and 0.5 µL of EpiGnome polymerase) was added to each sample. The reactions were incubated as 25°C for 5 minutes, 42°C for 30 minutes and 37°C for 2 minutes. Then, 1 µL of exonuclease I was added to each sample and incubated for 10 minutes at 37°C, 3 minutes at 95°C followed by 2 minutes at 25°C.

Next step was the tagging of the DNA. TT master mix was prepared on ice as follows: 7.5 µL of EpiGnome terminal tagging premix and 0.5 µL of DNA polymerase. TT master mix was mixed by pipetting and 8 µL was added to each reaction. Samples were incubated at 25°C for 30 minutes followed by 3 minutes at 95°C. Reactions were cooled to 4°C and purified using AMPure XP system (1.6x beads) (Beckman Coulter Inc., USA) as recommended.

The final step of library construction comprised the amplification of the libraries and addition of barcodes. Each reaction contained 22.5 µL of the purified tagged DNA, 25 µL of FailSafe PCR premix E, 1 µL EpiGnome forward primer, 1 µL of EpiGnome index PCR primer and 0.5 µL of FailSafe PCR enzyme (1.25 U). PCR was performed according with the suggested steps: initial denaturation at 95°C for 1 minute, 10 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 68°C for 3 minutes. Final extension was performed at 68°C for 7 minutes.

Samples were purified once more using AMPure XP beads (1x beads). Samples were re-suspended in 20 µL of nuclease-free water and quantity and quality of the sequencing libraries were assessed.

2.7.3 Library quantitation and quality control

Library quality and quantity was accessed using 2100 Bioanalyzer (Agilent Technologies, USA) and qPCR.

A high sensitivity DNA chip (Agilent Technologies, USA) was used in order to identify library insert size and to access quality. High sensitive dye and gel matrix mixture was prepared by the addition of 15 µL of concentrated high sensitive DNA dye to the gel matrix vial. The mixture was transferred to a spin column and centrifuged at 2240 x g for 10 minutes.

A high sensitive DNA chip was placed in the chip priming station and 9 µL of the gel mix was pipetted into the well of the chip marked as G in black. The plunger was set to 1 mL mark followed by closing of the chip priming station. The plunger was pressed down until it reached the chip surface and was held in this position for 60 seconds and then released. After the chip was removed from the chip priming station, an additional gel-dye mix (9 µL) was pipetted into the two wells of the chip marked as G in grey. Then, 5 µL of the marker was pipette to each well. High sensitive DNA ladder (1 µL) was added to the marked well. The samples (1 µL) were pipetted to the other wells. The chip was vortexed using the Agilent Chip Vortexer for 1 minute at maximum speed (2400 rpm) and analysed in the 2100 Bioanalyzer.

The quantification of the purified libraries was performed using KAPA Library Quantification kit (Kapa Biosystems, Inc., USA). An initial dilution of 1:100 of the libraries was prepared using the library dilution buffer (10 mM Tris-HCl, pH 8.0, 0.05% Tween 20). qPCR reactions were set

up with 12 µL of 2X KAPA SYBR FAST qPCR master mix, 4 µL of PCR-grade water and 4 µL of diluted library or DNA standards (six DNA standards with serial 10-fold dilution). Amplification was performed on a Mx3005P PCR System (Agilent Technologies, USA) with the following program: Initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 95°C for 30 seconds and annealing/extension at 60°C for 45 seconds. Samples and DNA standards were run on triplicate. Libraries concentration was determined by the qPCR relation to the concentration of the annotated NDA standards. Size adjustment was performed to account for the differences between the average fragments size of the library (obtained for Bioanalyzer results) and the DNA standards (452 bp). Concentrations of the undiluted libraries were calculated using the relevant dilution factor.

2.7.4 High throughput sequencing (HTS)

The high-throughput sequencing was performed at The University of Birmingham on an Illumina HiSeq 2500 Platform. The sequencing run was performed using a rapid run flow cell with paired-end and read length of 150bp.

Libraries (11 bisulfite samples and 1 non-bisulfite converted) were combined based on the index sequence of each library to generate two pools for each of the two lanes of the flow cell. A non-converted sample, in addition to PhiX, was run in duplicate in both lanes to account for the over simplification of the base composition of bisulfite treated samples. Both samples accounted for approximately 25% of the library content in each library pool. After being mixed, the samples were denatured for 5 minutes at room temperature with the addition of 10 µL of 0.1M NaOH to 10 µL of pooled libraries. Then, 980 µL of HI1 (hybridization buffer) was added

to the samples. PhiX was spiked in the samples at a final concentration of 1%. Pooled libraries (12 pM) were loaded to cBot (Illumina, USA) for cluster generation.

After template hybridization on the cBot, the flow cell was transferred to the HiSeq 2500 system (Illumina, USA) and run on rapid mode setting.

2.8   Bioinformatics analysis

The data analyses, including quality control, genome assembly and DNA methylation calls, were performed in collaboration with Genotypic Technologies, India.

First, sequencing reads were de-multiplexed to fastq files using Illumina bcl2fastq Conversion Software. The raw reads were quality checked using Genotypic Pvt. Ltd., proprietary tool SeqQC_v2.2. Then, the reads were processed using TrimGalore to remove adapters towards 3'-end, low quality bases (Phred <20), and sequences shorter than 50 bases.

2.8.1 Draft genome assembly

De novo assembly of Illumina HiSeq data was performed using ABySS 3.8 assembler (Simpson *et al.*, 2009). ABySS is a *de novo*, parallel, paired-end sequence assembler that is designed for short reads and that is able to assemble large genomes. ABySS *de novo* assembly was followed by scaffolding using paired-end data. Scaffolding was carried out using SSPACE scaffolder (Boetzer *et al.*, 2011). SSPACE scaffolds pre-assembled contigs by using the distance information of paired-end data, SSPACE is able to assess the order, distance and orientation of contigs and combine them into larger scaffolds. Due to scaffolding using unknown insert between two pairs of reads, the assembly is introduced with distance-estimated numbers of Ns in-between scaffolds. These intra-scaffold gaps (represented by Ns) were closed using

GapCloser tool (Luo *et al.*, 2012), to the maximum level possible. GapCloser takes into account

unaligned reads and performs local assemblies, where one of the read pairs thought to be

originated from the gaps.

2.8.1.1 Draft genome annotation

The draft *de novo* assembled genome was used for prediction of the GpG islands (CGI),

annotation of transcription start sites (TSS) and gene annotation.

 CGI were identified using EMBOSS newcpgreport software

(http://emboss.bioinformatics.nl/cgibin/emboss/newcpgreport) with default parameters

(window size = 100 bp; minimum length = 200 bp; Minimum observed/expected = 0.6;

minimum percentage = 50).

The draft *de novo* assembled genome was also annotated to the available gene sets. First, it

was blasted against the *D. magna* genome v2.4 gene set, generated by gene prediction, for

TSS annotation. Then, due to the new release of *D. magna* gene set (finloc9b) in April of 2016,

the draft genome for Bham2 was annotated again. The transcript sequences were blasted to

the *de novo* assembled Bham2 genome. Then, the annotation of the mRNA sequences was

transferred to the Bham2 assembled genome and used for further analysis on methylation

profiling.

2.8.2 Whole genome bisulfite sequencing analyses

Sequencing quality control for bisulfite treated samples used the same methods as described

in section 2.8. Mapping of the reads and methylation call were performed using Bismark

software (version 0.12.2, with the parameters: –bowtie2 –score_min L, 0, −0.4) using the *de novo* assembled draft genome for *Daphnia magna* Bham2 as reference.

After DNA methylation mapping, the output files from Bismark were visualised using SeqMonk software (www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Methylation levels were quantitated using the 'bisulfite methylation over features' pipeline available in the software. Differential methylation was investigated using different approaches, as described in chapter 4, section 4.3.3.

2.9   Direct bisulfite sequencing PCR (BSP)

Different regions identified with WGBS were also confirmed with BSP. For this, firstly the DNA samples were treated with sodium bisulfite as described in section 2.7.1. Then, using specific primers, the regions were amplified, purified and sequenced for quantification of DNA methylation at site specific resolution.

2.9.1 Design of BSP primers

The primers for BSP were designed with MethPrimer software (Li and Dahiya, 2002). MethPrimer can identify CpG islands in a given sequence using the following parameters; percentage of CG >50, observed/expected >0.6, length >200bp. Usually the primers for BSP are then placed within CpG islands. However, for this study, the primers were designed for targeted regions. These regions presented differential methylation that were identified with WGBS for the different groups and experiments.

A list of all the primers used for BSP analysis are presented in Tables 2.3 and 2.4.

Table 2.3. Primers sequences for BSP analyses and analyses of methylation level cut-off.

DMR ID	Forward Primer	Reverse Primer	Product size
C001	5'-GTAGTGAGATATTTTTATAGGTTGT-3'	5'-CTATTCAATTAAATTCTTTAAACTC-3'	178
C002	5'-TTTTATAATTGTGTTAGTTATTTGTTAAAA-3'	5'-AAACTCATTAAAAATAAATTTATTAAAATC-3'	278
C003	5'-GTAGAGAAGTTTTTTTGTTTAGTAGAAAGT-3'	5'-AACCCACAATACAACCTAATACTTCTT-3'	243
C004	5'-GGATTTTTTTTATGGAAGGTT-3'	5'-AAAACCAAAAATATTTTCAAAAAC-3'	238
C005	5'-TTTTTTAATTTGGGTGGATGAAAT-3'	5'-AAATAAAATAAACAAAACCCTAAATC-3'	268
C006	5'-GGAGGATTATTTAGGAGATTAAATAAAT-3'	5'-TAAACATAAACATATTCAAAAACCC-3'	261
C007	5'-AAGTTGTTTGATTTTTTATTTTTAT-3'	5'-TTTTTATTTTACTACTTAACTATCTCC-3'	237
C008	5'-TAGGTATTTATATGGATAGGAATGT-3'	5'-CACCTACATAATATTTTAAAAATTAAA-3'	175
C009	5'-GGTTTTAGTTGATTTTTGGTTTTTA-3'	5'-CCATAAAAATCCTCTTTATATACCTATC-3'	233
C010	5'-TTTTTTTTGAGTTTGTTGAATTA-3'	5'-AAAACCATAACATTATACAATACTTTAC-3'	214
C011	5'-TTTTTAGATTGATTTTTGTAGGGTTAAAA-3'	5'-ATCAAATCTTCCCAAAAAATAAAAAAT-3'	272
C012	5'-TAATTGTTAAAATAAATATTTTAGGTGTAA-3'	5'-ACCTACTAAACAACTACTAAATCAACTTAA-3'	267
C013	5'-TAGTTTGGAAAAGGATGTAAAAATAGTTA-3'	5'-AACTAATCAAACAATAAAAACCTTAAAC-3'	233
C014	5'-TAATTTTGGTGGTAGTAATATTGTATAATG-3'	5'-TTCTACTTTCAACTTCCCTATAACC-3'	189

Table 2.3. Continued from previous page

C015	5'-TTATTGTTTTTTGTGTGAGTGTTGT-3'	5'-CACTTAAAAAAATTATTATTCTAACTAAAA-3'	254
C016	5'-TTTTGGTAATTGTTGATTTTGTAAATT-3'	5'-ACTACTATTCCTACAACCCCAAATC-3'	280
C017	5'-TAGGGATATTAGGTAATAGGGTAGGGA-3'	5'-TTTCATTATTAAAACAAAACACCAACA-3'	257
C018	5'-AATGTGGTTAGTTTAAAGGTGATTG-3'	5'-AACCACAATAAAAAACAAAAC-3'	284
C019	5'-TTGTTTGTAATAGTATAGAATTATGGAATT	5'-AAAAAACTTAAACTCCCTCTTACCC-3'	204
C020	5'-AGAAAATAAGTTTAATATGAATGTATGTTA	5'-TACTACTTTATTAAAAAACCCAAAA-3'	197
C021	5'-TTTTTTGAATTTGAGTTTAGTATTAATTA	5'-ATCAACACACTCTAAACCACCATAC-3'	189
C022	5'-TTATTTAAGTATTTAGGGATGTTATTTTTT	5'-AAACTATATAAACTCCAAACTAACC-3'	191
C023	5'-AGTTTGATGGTTAAATGTTATTTGA	5'-CATCTTCCTTACCAATAATCAACTACTC-3'	262

Table 2.4. Primers sequences for BSP analyses and confirmation of WGBS data.

DMR ID	Forward Primer	Reverse Primer	Product size
19	5'-TTATTGTTTTTTGTGTGAGTGTTGT-3'	5'-CACTTAAAAAAATTATTATTCTAACTAAAA-3'	254
223	5'-TTTTTTGAATTTGAGTTTAGTATTAATTA-3'	5'-ATCAACACACTCTAAACCACCATAC-3'	189
337	5'-TTATTTAAGTATTTAGGGATGTTATTTTTT-3'	5'-AAACTATATAAACTCCAAACTAACC-3'	191
341	5'-AATGTGGTTAGTTTAAAGGTGATTG-3'	5'-AACCACAATAAAAAACAAAAC-3'	284
382	5'-AGAAAATAAGTTTAATATGAATGTATGTTA-3'	5'-TACTACTTTATTAAAAAACCCAAAA-3'	197
422	5'-TAGGGATATTAGGTAATAGGGTAGGGA-3'	5'-TTTCATTATTAAAACAAAACACCAACA-3'	257
487	5'-GTGATTTTGTGTTGTAATGAGTTAGGA-3'	5'-AAAAAAAACTAAACTACCTAATAACTTC-3'	276
2176	5'-AGTTTGATGGTTAAATGTTATTTGA-3'	5'-CATCTTCCTTACCAATAATCAACTACTC-3'	262
2398	5'-ATGTTATTTGGTATTAGGTTTTTGG-3'	5'-CATTTATCAATATCTACATAAAACAATTTA-3'	208

2.9.2 Amplification of bisulfite treated DNA

Amplification of sodium bisulfite treated DNA is known to be difficult due to oversimplification of base composition of DNA after treatment, often resulting in primer dimer and non-specific product formation. Therefore, ZymoTaq DNA Polymerase (Zymo Research, USA), a hot-start polymerase, was used for all the work requiring bisulfite sequencing PCR in this study.

The reaction was performed with a mastermix containing 25 µL of 2x reaction buffer, 0.5 µL of dNTP mix, 0.4 µL of Zymo Taq DNA polymerase, 10 pmol of forward and reverse primer and 1 µL of bisulfite treated DNA. Total volume of the PCR reaction was 50 µL.

The samples were amplified using a thermocycler (Mastercycler nexus, Eppendorf, USA) with the following steps: 95°C for 10 minutes, 38 cycles of 95°C for 30 seconds, variable annealing temperatures for 35 seconds and 72°C for 45 seconds. After completion of the cycling phase, a final extension step was performed at 72°C for 7 minutes.

The primers were optimised prior the use for BSP analysis. Validation was performed running the PCR products on agarose gel for size confirmation and by sequencing and comparison to the expected sequences.

2.9.3 DNA gel electrophoresis

DNA gel electrophoresis was used to separate amplified products based on their size. Agarose concentration varied from 1 to 2.5% based on expected product size and resolution needed. Molecular grade agarose (Bioline Ltd., UK) was added to 1XTBE buffer (89 mM Tris base, 89 mM boric acid and 2 mM EDTA, adjusted to pH 8.0) according to the required gel concentration. Midori Green Advance DNA Stain (Nippon Genetics, Germany) was used for gel

staining. Then, 100 bp or 1 kb DNA molecular weight markers (New England Biolabs, USA) and DNA samples were mixed with 6X loading dye (New England Biolabs, USA) and loaded into the wells. Gel electrophoresis was performed at 80 V for 30 to 60 minutes and the gel was visualised in an UV transilluminator.

2.9.4 DNA purification and sequencing

After amplification the PCR products were purified using QIAquick PCR purification kit (Qiagen Ltd., UK). Briefly, 5 μL of sodium acetate (3M, pH5.2) and 250 μL of PB buffer were added to 50 μL of amplified DNA samples. Samples were mixed by pipetting, transferred to QIAquick spin columns and centrifuged (1 minute, 5900 x g). Flow through was discarded and 750 μL buffer PE was added to the columns and centrifuged for 1 minute at 5900 x g. Columns were dried for an additional 1 minute at 5900 x g. Columns were placed in 1.5 mL eppendorfs and water (20 μL) was added to the centre of the column and centrifuged for 1 minute to elute the DNA sample.

The purified DNA samples were sequenced by the Functional Genomics and Proteomics facility, School of Biosciences, University of Birmingham, Birmingham, UK using an ABI3730 DNA analyser (Applied Biosystems, UK). Each sequencing reaction was prepared using 10 ng of the purified DNA samples and 0.4 μL of either reverse or forward primers (10 pmol) adjusted to the final volume of 10 μL using nuclease-free water. To analysis of the BSP data was performed using the peak heights for C and T bases at each CpG site obtained from the electropherogram.

Methylation level was calculated using the formula below:

$$\% \; methylated \; C \; at \; each \; CpG \; site = \left(\frac{C \; peak \; height}{C \; peak \; height + T \; peak \; height} \right) * 100$$

2.9.5 Generation of artificially methylated and un-methylated DNA

In order to assess the sodium bisulfite conversion efficiency, artificially methylated and unmethylated DNA was generated.

DNA samples were amplified using GenomePlex Complete Whole Genome Amplification (WGA) Kit (Sigma-Aldrich, UK). The amplification of DNA fragments passively removes the methyl group from the cytosine nucleotide since it will not be added to the new DNA strand. This reaction started with the fragmentation step where DNA at 1 ng/µL and 10X fragmentation buffer (1 µL) were mixed and incubated for 4 minutes at 95°C. Fragmentation was followed by library preparation. 1x library preparation buffer (2 µL) and library stabilisation solution (1 µL) were added to the chilled sample and incubated at 95°C for 2 minutes. Then, the sample was chilled again and the library preparation enzyme (1 µL) was added. The reaction was incubated on a thermocycler at 16°C for 20 minutes, 24°C for 20 minutes, 37°C for 20 minutes and 75°C for 5 minutes. After generation of the library, the sample was placed on ice and a mastermix containing 10x Amplification Master mix, water and WGA DNA polymerase were added to the sample. The sample was incubated on a thermocycler with the following conditions: initial denaturation at 95°C for 3 minutes, then 14 cycles of 94°C for 15 seconds (denaturation) and 65°C for 5 minutes (annealing/extension).

Methylated DNA was generated with CpG methylase (New England Biolabs, USA) following the protocol provided. Briefly, 1 µg of DNA was mixed with 10X NEBuffer 2 (2 µL), S-

adenosylmethionine (2 µL), 1 µL of SssI methylase (4 U/µL), in a final volume of 20 µL. Sample was incubated at 37°C for 4 hours followed by incubation at 65°C for 20 minutes. Unmethylated and fully methylated DNA samples were purified using QIAquick spin columns (Qiagen Ltd, UK), as described in section 2.8.4, and stored at -20°C until further use.

2.9.6 Sodium bisulfite conversion efficiency

Purified methylated and un-methylated genomic DNA samples were used to assess the sodium bisulfite conversion efficiency. Both samples were treated with sodium bisulphite as described in section 2.7.1. One region was selected to analyse the efficiency of conversion. The fragment sequenced contained 37 cytosines, of these, 8 were within a CpG context, 6 in CHG and 23 in CHH. Fully methylated and un-methylated samples were sequenced and results are in Figure 2.3. Sodium bisulfite conversion efficiency was virtually 100%.
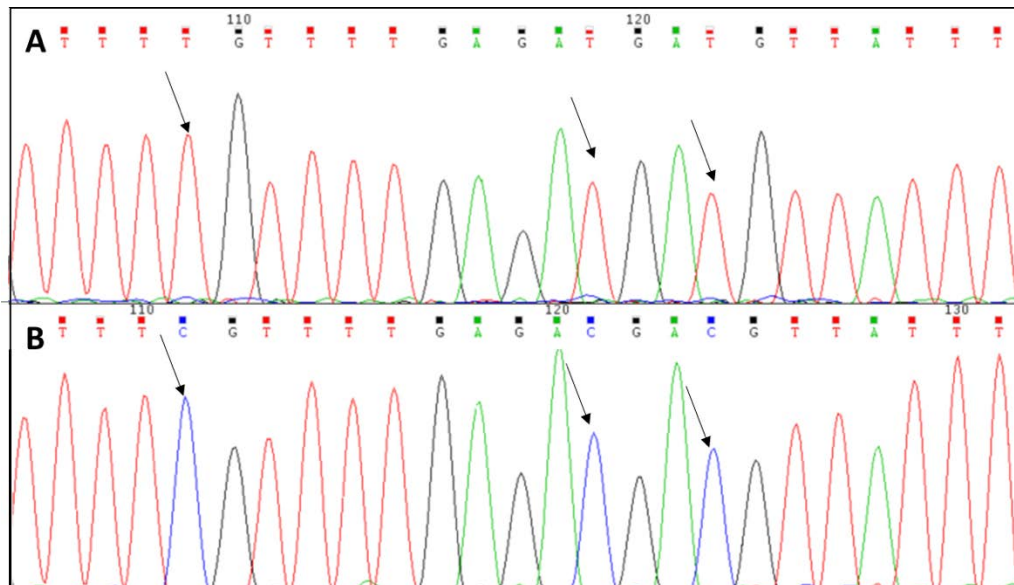
Figure 2.3 Artificially fully methylated and un-methylated DNA fragments sequenced to assess the conversion efficiency from un-methylated C to T. A) Artificially un-methylated DNA due to whole genome amplification. B) Fully methylated DNA using SssI methylase. Arrows indicate CpG nucleotides.

2.10   Gene expression analysis

2.10.1   Primer design and validation

The forward and reversed primers were designed with Primer3 http://simgene.com/Primer3) and are described in Table 2.5. Primer were synthesised by Integrated DNA technologies (Belgium).

Primers were validated and sequenced for confirmation. BIOTAQ DNA polymerase (Bioline, UK) was used for product amplification. Each reaction contained: 10x NH4 Reaction buffer (5 µL), 50mM $MgCl_2$ solution (1.5 µL), 100mM dNTP Mix (1 µL), BIOTAQ (0.5 µL), 1.5 µL of each primer at 10 µM, 3 µL of cDNA (40 ng/µL) and water to a final volume of 50 µL. The samples were run on a thermocycler using the following program: initial denaturation at 95°C for 10 minutes, 40 cycles of 95°C for 30 seconds, 60°C for 35 seconds and 72°C for 45 seconds. Following completion of the cycling phase, a final extension step was performed at 72°C for 7 minutes. Samples were analysed on a 1.5% agarose gel electrophoresis and sequenced for confirmation. Methods are described in sections 2.8.3 and 2.8.4.

2.10.2   cDNA synthesis

The cDNA was synthesised using Tetro cDNA Synthesis Kit (Bioline, UK). Each reaction contained 2 µg of RNA, 1 µL of random hexamer, 1 µL of 10mM dNTP mix, 4 µL of 5x RT buffer, 1 µL of RiboSafe RNase inhibitor and 1 µL of Tetro reverse transcriptase (200 U/µL). RNase free water was added to a final volume of 20 µL. Samples were incubated for 10 minutes at 25°C followed by 30 minutes at 45°C. Then, the reaction was terminated by incubation at 85°C for 5 minutes and chilled on ice. Samples were stored at -80°C until processed.

2.10.3   RT-PCR

RT-PCR analysis was performed on an AriaMx Realtime PCR system (Agilent technologies, UK) using SensiFAST SYBR Lo-ROX kit (Bioline, UK). Three biological replicates, with three technical replicates were run for each group. Each sample contained 80 ng of synthesised cDNA, 10 µL of SensiFAST mix, forward and reverse primer (2 to 5 pmol depending on primer efficiency) and nuclease free water (to a final volume of 20 µL). The amplification was performed with a 2-step cycle: 95°C for 5 seconds (denaturing) and 60°C for 30 seconds (annealing and extension). Melting curves were generated to ensure single product amplification. ROX was used as a reference dye. After correction and baseline setting the threshold cycle (CT) values were exported. The geometrical average of *Actin* and *Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* genes were used as internal reference for normalisation (Vandesompele *et al.*, 2002). Data were analysed using delta-delta CT method of relative quantification (Livak and Schmittgen, 2001).

Table 2.5. Primer sequences for real-time PCR used for gene expression analyses of the one-carbon and demethylation pathways.

Symbol	Gene ID	Forward Primer	Reverse Primer	Product size
DNMT1	Dapma7bEVm005001	5'-CCTGATCCGTTCTGCATTGG-3'	5'-TCCAGTCAGAGCGAAAACCT-3'	157
DNMT2	Dapma7bEVm011900	5'-GCGGAAGATCAATGGGCAAT-3'	5'-AACAGTTCCCCATCCATCGT-3'	225
DNMT3	Dapma7bEVm006722	5'-AGTTACAGCGTTGGGGAAGA-3'	5'-ATTTAAGGGCCCAGTCGGAA-3'	247
MAT	Dapma7bEVm004771	5'-CGACATGCCGGGTAAAGAAG-3'	5'-GACCAACACCATGCCAGTTT-3'	214
SAHH	Dapma7bEVm024816	5'-TGCAAGCACTCTCTTCCTGA-3'	5'-AGAGCGCAGATAGGATCGAC-3'	167
MTRR	Dapma7bEVm003609	5'-TGTATGCCGCTTCATTGGTG-3'	5'-CAACTGCCTTCTTCGCTTGT-3'	151
BHMT	Dapma7bEVm018566	5'-CGGTTAGAATTGCTCGCGAA-3'	5'-ACGCCAAATAATCCACACCG-3'	234
MS	Dapma7bEVm002113	5'-TGAGCGGTGGTGTGTCTAAT-3'	5'-AGAGTTGCAGCAATTTGGGG-3'	174
GNMT	Dapma7bEVm001624	5'-GTACTGCTGCGGAAGGATTG-3'	5'-TTGAATCGATCCCTGTGCCA-3'	183
TET1	Dapma7bEVm018501	5'-GGTTAGAAAGTTGGGCGCAA-3'	5'-GGACGAACGACCTATTTGCC-3'	233
TET2	Dapma7bEVm029206	5'-CGAATAGAGCAACAACGGCA-3'	5'-ACGGAAATGCGTGATGGATG-3'	178
ACTIN	Dapma7bEVm019018	5'-GGTATGTGCAAGGCTGGATT-3'	5'-GGTGTGGTGCCAGATCTTTT-3'	225
GAPDH	Dapma7bEVm015323	5'-GGGGACAGACGTTTCCTGTA-3'	5'-AAGGGGTCATTGACAGCAAC-3'	168

2.11   Target quantification of one-carbon pathway metabolites

The targeted metabolomics study was funded by the NERC Biomolecular Analysis Facility - Birmingham node (NBAF-B) in the School of Biosciences at the University of Birmingham, UK. Metabolites samples were prepared as described in section 2.4.3. Target metabolites quantitation was performed according to Mirbahai *et al.* (2013) with modifications. S-adenosyl-L-methionine-d3 (S-methyl-d3) tetra (p-toluenesulfonate) salt (CDN Isotopes, UK) was used as the internal standard.

Samples were re-suspended in 5 µL of acetonitrile:water mixture 1:1 containing SAM-d3 at the concentration of 0.125 µmol/mL. The re-suspended samples were vortexed and centrifuged for 10 minutes at 13,000 x g at 4°C. Then, the samples were transferred to a 96 well plate for analysis. Six biological replicates were analysed (2 µL injections; acetonitrile was used for injection loop). Negative controls and quality control samples were run at the beginning and intercalated with the sample runs to correct the background noise.

The samples were analysed using Dionex UltiMate 3000 liquid chromatography system with micropump coupled to a triple stage quadrupole (TSQ) tandem mass spectrometer (Thermo Fisher Scientific, UK) with Ion Max-S atmospheric pressure ionisation (API) spray source (Thermo Fisher Scientific, UK). Separation was achieved using a reverse phase column with weak anion exchange properties (Acclaim Mixed-Mode WAX column, 250 x 0.3 mm internal diameter, 5 µm particle size, 120 Aº pore size, Dionex, Idstein, Germany) with column oven temperature of 18°C (minimum and maximum temperature of 16°C and 22°C), and under a gradient running buffer including buffers A, B and C (Buffer A: 10mM ammonium formate, pH

6.2; Buffer B: 10mM ammonium formate, pH 4.2, 75% acetonitrile and 25% water; Buffer C: acetonitrile:water (1:1))(Table 2.6).

Table 2.6. Buffers gradient used for liquid chromatography. Buffer A: 10mM ammonium formate, pH 6.2; Buffer B: 10mM ammonium formate, pH 4.2, 75% acetonitrile and 25% water; Buffer C: acetonitrile:water (1:1).

Steps	Retention time (minutes)	Flow (µL/min)	%B	%C	%A
1	0	8	100	0	0
2	5	8	100	0	0
3	10	8	40	0	60
4	14	8	10	0	90
5	17	6	10	0	90
6	17.01	6	0	100	0
7	18	6	0	100	0
8	18.01	6	100	0	0
9	24	8	100	0	0
10	28	8	100	0	0

The 10 metabolites of interest were: SAH, methionine, adenosine, betaine, sarcosine, SAM, glycine, dimethylglycine, choline and stachydrine. The masses of the precursor and product ions used for detection of the 10 metabolites and the internal standard are described in Table 2.7. Data acquisition was performed as multiple reaction monitoring.

Table 2.7. The masses of the precursor and product ions used for detection of the 10 metabolites of interest and internal standard using LC-MS/MS. The chemical formula, ion mode, S-lens value, collision energy parent (precursor) masses and product masses used for detection.

Compound	Formula	Ion form	Parent mass	S-lens	Collision energy (%)	Product mass
Adenosine	$C_{10}H_{13}N_5O_4$	$[M+H]^+$	267.935	88	45	118.981
					19	135.993
Betaine	$C_5H_{12}NO_2$	$[M+H]^+$	118.000	78	53	42.041
					26	58.045
					19	59.077
Sarcosine	$C_3H_7NO_2$	$[M+H]^+$	89.985	43	51	30.103
					25	42.064
					12	44.068
Stachydrine	$C_7H_{13}NO_2$	$[M+H]^+$	143.973	93	36	42.056
					25	58.052
					22	84.040
Methionine	$C_5H_{11}NO_2S$	$[M+H]^+$	150.045	56	16	56.107
					23	61.052
					10	104.076
					6	133.060
Glycine	$C_2H_5NO_2$	$[M+H]^+$	76.000	43	11	30.100
					9	48.100
Dimethylglycine	$C_4H_9NO_2$	$[M+H]^+$	104.07	61	39	42.100
					10	58.160

Table 2.7. Continued from previous page

Choline	C5H15NO2	[M+H]+	104.028	70	21	45.060
					34	58.066
					17	60.080
S-(5'-Adenosyl)-*L*-homocysteine	$C_{14}H_{20}N_6O_5S$	[M+H]$^+$	384.937	106	37	87.962
					19	133.950
					20	135.987
					12	249.982
S-(5'-Adenosyl)-*L*-methionine	$C_{15}H_{23}N_6O_5S$	[M+H]$^+$	398.944	110	30	96.953
					28	136.022
					15	250.014
					13	298.015
S-Adenosyl-*L*-methionine-*d3*	$C_{15}H_{20}D_3N_6O_5S$	[M+H]$^+$	402.063	114	31	96.998
					15	250.037
					16	267.092

2.12   Statistical analyses

The statistical analysis of the data was performed using SPSS 21. Normal distribution of the data was evaluated via Shapiro-Wilk's test and homogeneity of variance was analysed with Levenes' test. For comparison of two or more groups with normal distribution and homogenised variance 2-tailed independent student's t-test and one-way ANOVA with Tukey's post-hoc test were used, respectively.

When the requirements for normal distribution and homogeneity of variance were not met, data were analysed by applying non-parametric statistics, using a Kruskal-Wallis test (more than two independent groups) or Mann-Whitney test (two independent groups).

Statistical approaches for differential methylation analyses are described in Chapter 4.

Chapter 3

Distribution and levels of DNA methylation across the genome of *Daphnia magna*

3.1 Introduction

DNA methylation is involved in many biological processes. In general, DNA methylation is considered an important regulator of gene expression, acting as a system of cellular memory especially for long-term silencing of genes (Jaenisch and Bird, 2003). In vertebrates it is also important for X-chromosome inactivation, imprinting and tissue-specific gene expression (Crider *et al.*, 2012).

Interestingly, DNA methylation patterns differ between vertebrate, plants and invertebrates. DNA methylation is generally widespread and evenly distributed in vertebrates, occurring mainly where a cytosine is directly followed by a guanosine (CpG) except for regions known as CpG islands (CGI). CGIs are GC-rich regions, often unmethylated, that show a high density of CpG dinucleotides relative to the rest of the genome and are positioned at the 5' ends of many vertebrate genes. On the other hand, invertebrates, plants and fungi present a sporadic pattern of DNA methylation, where regions of heavily methylated DNA are interspersed with regions that are unmethylated (Suzuki and Bird, 2008).

Although DNA methylation is present in invertebrates, its function has not been comprehensively studied. As distribution of DNA methylation across the genome is different between vertebrate and invertebrates (reviewed in section 1.3), it is essential to determine if DNA methylation machinery and its function is conserved between vertebrates and invertebrates. In addition, it is important to determine and understand the significance and functional implications that arise from these differences between the two groups. In addition, epigenetic reprogramming events need to be investigate in invertebrates.

The traditional invertebrate model organisms for genetic studies, *Drosophila melanogaster* and *Caenorhabditis elegans* do not contain DNA methylation or it is restricted to a few sites (Capuano *et al.*, 2014; Park *et al.*, 2011; Simpson *et al.*, 1986). Therefore, epigenetic studies in invertebrates have been limited to alternative organisms, such as the honey bee *Apis mellifera* (Lyko *et al.*, 2010; Rasmussen and Amdam, 2015), the wasp *Nasonia vitripennnis* (Beeler *et al.*, 2014; Wang *et al.*, 2013; Zwier *et al.*, 2012), the ant *Camponotus floridanus* (Glastad *et al.*, 2015), and the oyster *Crassostrea gigas* (Gavery and Roberts, 2010; Rivière, 2014; Wang *et al.*, 2014). These environmentally relevant species are often difficult to maintain in the laboratory and it is challenging to induce phenotypic plasticity in these species in response to environmental conditions under laboratory conditions. In contrast, *Daphnia* species have great potential as an environmentally-relevant invertebrate model organism for epigenetic research (Harris *et al.*, 2012; Vandegehuchte and Janssen, 2013). They are widely distributed across the globe with well-known ecology (Lampert and Kinne, 2011). Also, *Daphnia* have been used as a model organism in ecotoxicology and have been maintained under laboratory conditions for many years. In addition, in favourable conditions their reproduction happens through parthenogenesis, producing clonal offspring (see section 1.5.1).

Despite the great potential of *Daphnia* for epigenetic studies, its DNA methylation toolkit (i.e. genes involved in DNA methylation and demethylation and one-carbon pathway) has not been profiled comprehensively. This is partly due to lack of a fully annotated genome. Although limited, some information regarding *Daphnia's* methylome is already available. Global levels of DNA methylation, measured by LC-MS, are described for two different inbred strains. For the Iinb1 strain the global cytosine methylation was 0.49 ± 0.19% on average. The Xinb3 strain presented a global methylation level of 0.52 ± 0.16% (Asselman *et al.*, 2015). Global DNA

methylation level was also measured for *D. magna* bham2 and is presented in section 3.3.3. Furthermore, several chemicals are known to affect the global DNA methylation levels in *Daphnia*, including 5-azacytidine, genistein and vinclozolin (Vandegehuchte *et al.*, 2010a, 2010b, 2009a, 2009b). Gene specific methylation was also analysed in *D. magna* and showed relationships with gene family size and diversification of genes (Asselman *et al.*, 2016).

The methyl donors, essential for DNA methylation, are obtained from a series of reactions, part of the one-carbon pathway that leads to the production of S-adenosylmethionine (SAM). The methyl group is transferred from SAM to the cytosine resulting in 5-methylcytosine and S-adenosylhomocysteine (SAH). DNA methytransferases catalyse the methylation of the DNA molecule. The additional enzymes of the one-carbon pathway are essential for the maintenance of normal levels of methylation (Herceg and Vaissière, 2011; Ulrey *et al.*, 2005) (see section 1.2.2).

DNA demethylation can either occur through active or passive pathways. TET enzymes are responsible for the active removal of methylation by a multistep reaction, while DNA methylation can be passively lost during replication of DNA, due to malfunction of enzymes from the one-carbon pathway or absence of methyl donors (Piccolo and Fisher, 2014; Song *et al.*, 2013; Tahiliani *et al.*, 2009) (see section 1.2.5).

The use of *Daphnia* for epigenetic studies can potentially have a significant impact on the current approaches for risk assessment and environmental monitoring. Currently, epigenetic mechanisms are not considered during risk assessment of substances. However, several studies have demonstrated the importance of epigenetic mechanisms, such as DNA methylation, in mediating chemical effects upon the phenotype and health of organisms.

However, the lack of knowledge about *Daphnia's* methylome is still a barrier for epigenetic studies in this species.

To overcome this problem, a comprehensive overview of the DNA methylation profile in *Daphnia magna* was necessary. This would facilitate investigation of the induction of gene specific DNA methylation changes in response to stressors, and analyses of the persistence and accumulation of the induced changes. For these studies it was most important to use environmentally relevant stressors with reported effects on DNA methylation at concentrations detected in the environment.

To meet these aims, the first objective was to obtain a reference genome for *D. magna* strain Bham2. At the time this part of the project started, no publicly accessible published genome sequence was available for *D. magna*. However, recently the genome and transcriptome of *D. magna* strain Xinb3 have been released (Orsini *et al.*, 2016). Still, the currently available draft genome sequence is incomplete, based on its genome size of 129 Mb, and not fully annotated (Orsini *et al.*, 2016; Routtu *et al.*, 2014). Therefore, it was decided for this study to generate a new draft genome sequence for the *D. magna* Bham2 strain and to use the available gene sets for annotation of the assembled genome.

In the second stage of the project the aim was to identify and describe the DNA methylation toolkit and the methylome of *D. magna* Bham2 to establish the "normal" DNA methylation profile. Furthermore, homology searches were employed to identify *D. magna* enzymes and metabolic pathways potentially involved in DNA methylation. The expression levels of genes and concentrations of metabolites in the one-carbon pathway were analysed for different ages of organisms and after their exposure to several stressors to characterise dynamic changes in DNA methylation machinery in response to age and stress. Changes in the

methylome distribution and levels were also analysed using whole genome bisulfite sequencing (WGBS) and gene specific bisulfite PCR.

Overall, the aims of chapter 3 were to describe: i) the overall pattern of DNA methylation across the genome of *D. magna*, ii) the DNA methylation machinery in *D. magna*, iii) the dynamic changes in DNA methylation machinery in response to age. Accomplishing these three aims was necessary to achieve a basic knowledge of *Daphnia's* methylome, enabling specific investigations regarding the role of DNA methylation in *Daphnia* undertaken in chapters 4 and 5.

3.2   Overview of experimental design

DNA samples extracted from 5 and 14 days old whole *Daphnia magna* Bham2 strain were used for the analyses presented in this chapter (Figure 3.1). Methods of sample preparation are described in section 2.4. DNA extraction was performed as described in section 2.4.1.1. Genome sequencing and WGBS procedures are described in section 2.7 and were performed with three biological replicates per age. Global methylation levels were analysed using LC-MS (n=6) and the methods are presented in section 2.6.

RNA samples were obtained as shown in section 2.4.2 for *Daphnia* at different ages (1, 5, 12, 21 and 28 days old). Three biological replicates and three technical replicates each were used for gene expression analysis using RT-PCR. All procedures, including the primers designed for each gene are described in section 2.9.
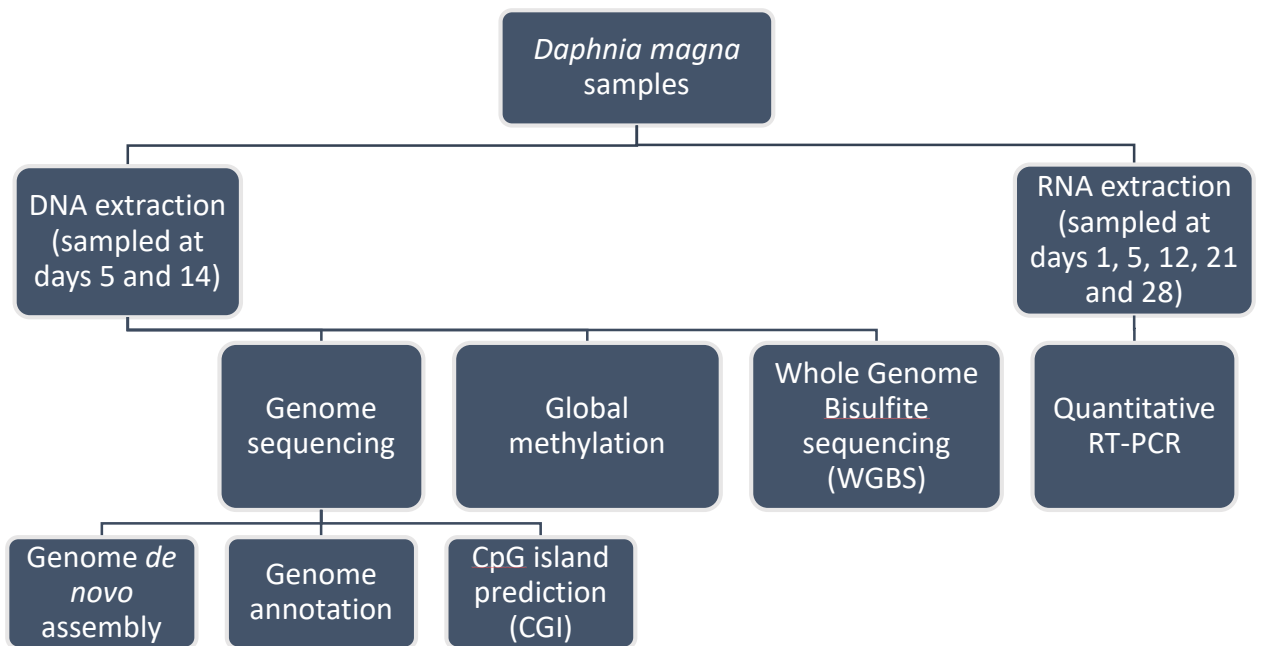
Figure 3.1 Workflow of analyses performed in chapter 3. Distribution and level of methylation across the *D. magna* genome and age-related changes.

3.3   Results

3.3.1 Characterisation of DNA methylation machinery

The first step in the characterisation of the *Daphnia magna* methylome was performed by identifying the genes potentially involved in DNA methylation and demethylation pathways and the one carbon pathway. To achieve this, *Homo sapiens* protein sequences for the genes listed in Table 3.1 were obtained from NCBI and homologous searches against the *Daphnia magna* database (v2.4) was performed using BLAST available at (http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna/BLAST/).

Similarity of *Daphnia's* protein sequences to human, honey bee and zebrafish protein sequences were assessed using EMBOss Matcher software, a pairwise sequence alignment tool (Rice *et al.*, 2000). Conserved domains for DNMTs were identified using NCBI's conserved domains database with default settings (Marchler-Bauer *et al.*, 2015) and Pfam (Finn *et al.*, 2015) and analysed in Jalview (Waterhouse *et al.*, 2009).

Table 3.1. *Daphnia magna* sequences encoding enzymes involved in DNA methylation mechanisms and protein similarity to *Homo sapiens*, *Danio rerio* and *Apis mellifera*. All similarity comparisons were made at protein level. *D. magna* protein ID for geneset (finloc9b).

Enzyme symbol	Sequences ID	*D. magna* protein ID	*H. sapiens* protein accession	Similarity to *H. sapiens*	*D. rerio* protein accession	Similarity to *D. rerio*	*A. mellifera* protein accession	Similarity to *A. mellifera*
DNMT1	dmDNMT1A	Dapma7bEVm005001	P26358	66%	AI63894	65.3%	XP_006562865	67.5%
	dmDNMT1B	Dapma7bEVm024669		90%		76.7%		59.6%
DNMT2	dmDNMT2	Dapma7bEVm011900	Q6ICS7	47.1%	AAI14323	62.2%	XP_006563008	49.5%
DNMT3	dmDNMT3	Dapma7bEVm006722	Q9Y6K1	43.9%	AAI62467	44.9%	XP_006568730	41.2%
MAT	dmMAT	Dapma7bEVm004771	Q1JL80	72.1%	NP_956165	86.3%	XP_006564332	98.1%
SAHH	dmSAHH	Dapma7bEVm024816	P23526.4	75.4%	AAI65366	72.5%	XP_391917	74.3%
MTRR	dmMTRR	Dapma7bEVm003609	Q9UBK8.3	57.4%	XP_689157	58.3%	-	-
BHMT	dmBHMT	Dapma7bEVm018566	Q93088.2	43.4%	AAI09473	47.1%	XP_003250116	57.1%
MS	dmMS	Dapma7bEVm002113	Q99707.2	81.9%	NP_932338	81.8%	-	-
GNMT	dmGNMT	Dapma7bEVm001624	Q14749.3	68.2%	AAH62527	68.4%	-	-
MTHFR	dmMTHFR	Dapma7bEVm002622	P42898.3	78.7%	NP_001268769	77.5%	XP_006566979	65.9%
TET	dmTET1	Dapma7bEVm018501	Q8NFU7	68.3%	XP_005156766	58.9%	A0A088ALU5	74%
	dmTET2	Dapma7bEVm029206		50.9%		51.3%		81.6%

3.3.1.1 DNMTs homologs and conserved domains

Homology searches against the *D. magna* genome identified one complete copy of each *DNMT1*, *DNMT2*, and *DNMT3* genes and a partial sequence similar to *DNMT1*.

Comparing *D. magna*'s DNMT1 protein sequence to human, zebrafish and honey bee *dmDNMT1A* revealed 66%, 65.3%, 67.5% similarity, respectively. Protein alignment revealed highly conserved sequences within the domain regions (Figure 3.2). For *dmDNMT1B* the similarity was 90%, 76.7%, and 56.6% compared to human, zebrafish and honey bee, respectively. However, this is described as a partial protein and it is not mapped to the *Daphnia* genome, indicating that it could be due to contamination or misassembled sequencing reads. Therefore, *dmDNMT1B* was excluded from further analyses.

DNMT2 presents the structural characteristics of DNA methyltransferases, the DNA methylase domain, however it has been shown to methylate a small tRNA instead, and this function is likely conserved from plants to mammals (Goll *et al.*, 2006; Schaefer and Lyko, 2010). For *D. magna DNMT2* the similarity is low when comparing the full protein sequence to human (47.1%), honey bee (49.5%) and zebrafish (62.2%) DNMT2, but DNA methylase domain sequence showed high conservation (Figure 3.3).

A homolog was found for *DNMT3* and, as for *DNMT2*, the similarity with the human protein is very low (43.9%) (41.2% similarity to honey bee and 44.9% to zebrafish) when comparing the full sequence. However, conserved domain regions presented higher similarity, as evidenced in Figure 3.4. Nevertheless, comparing DNMT1 and DNMT3, showed more variability for DNMT3 protein sequence.

Overall, the number of *DNMT* genes in *Daphnia* appears to be restricted to a single copy of each gene. As demonstrated before for invertebrates, there is no correlation between the evolutionary relationship of two species and the number of genes encoding DNMTs, therefore there is no standard number of enzymes that is characteristic to the invertebrate groups (Lyko and Maleszka, 2011; Glastad *et al.*, 2011).

As presented before, most of the conserved domains found in humans were identified in *D. magna* (Figure 3.5). The first domain, DMAP biding, was not found in dmDNMT1. DMAP biding mediates the interactions of DNMT1 and the transcriptional repressor DMAP1 (Rountree *et al.*, 2000). The same organization was observed for the two honeybee enzymes (DNMT1a and DNMT1b), the three enzymes of Nasonia vitripennis (DNMT1a, DNMT1b and DNMT1c), and DNMT1 for silkworm (Mitsudome *et al.*, 2015; Werren *et al.*, 2010). The lack of the first domain for honeybee and silkworm could explain the low methylation levels of transposable elements and repeated sequences (Lyko *et al.*, 2010; Xiang *et al.*, 2010). The same probably happens for *D. magna*, but analyses need to be performed to confirm this.

DNMT2 and DNMT3 have similar conserved domains organization as the other organisms analysed. DNMT2 presents a DNA methylase domain despite its function methylating a small RNA. DNMT3 has two conserved domains, PWWP and DNA methylase characteristic for *de novo* methyltransferases. In mammals, DNMT3 is responsible for the establishment of new pattern of methylation and the same functions is hypothesised for *D. magna* (Klose and Bird, 2006; Okano *et al.*, 1999).

Regarding the enzymes involved in the one-carbon pathway and DNA demethylation, the similarity is low for most of them (Table 3.1). However, it was possible to find homologs for all proteins of interest. This indicates that *D. magna* has the complete toolkit for DNA

methylation and that the pathways are likely to be ancestrally conserved for DNA methylation and active demethylation. However, further analyses still need to be conducted to assure the functions of the identified proteins.
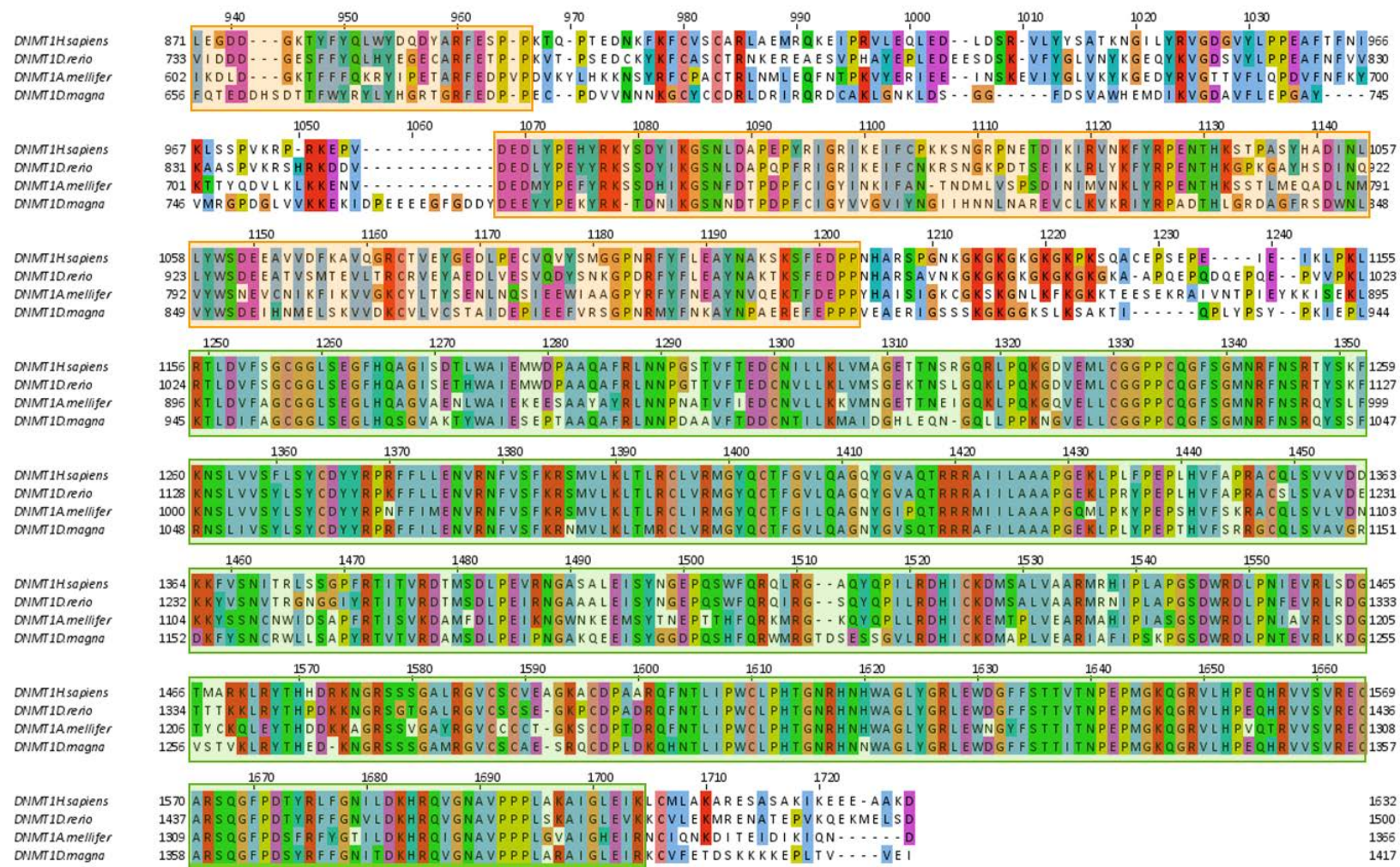
Figure 3.2 Protein alignment for DNMT1 sequences for human, zebrafish, honey bee and *Daphnia*. Only DNMT1a sequence is presented for honey bee. Conserved domains are highlighted within the sequences; more information on domain organisation is presented in Figure 3.5.
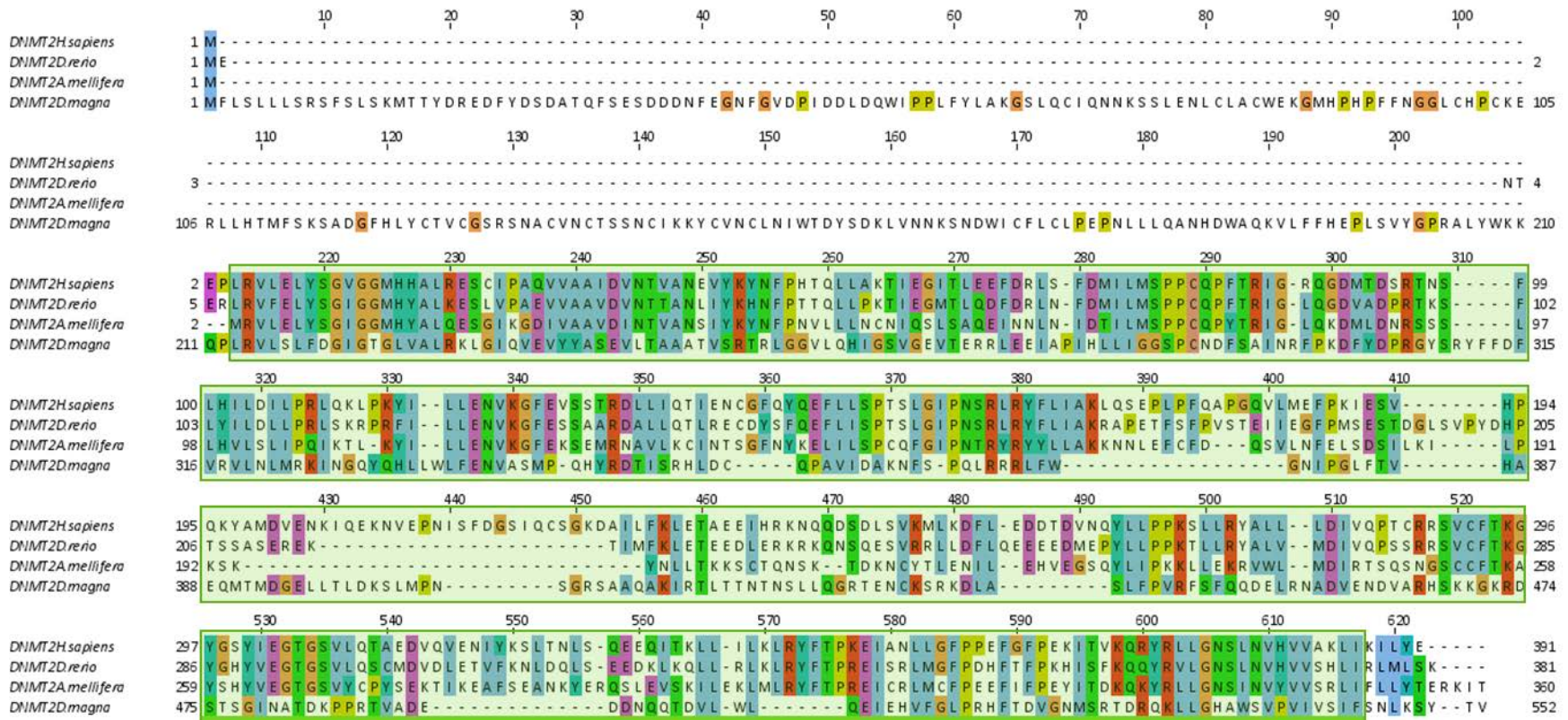
Figure 3.3 Protein alignment for DNMT2 sequences for human, zebrafish, honey bee and *Daphnia*. Conserved domain is highlighted within the sequences; further information on domain organisation is presented in Figure 3.5.

```
                        920        930        940        950        960        970        980        990       1000       1010
DNMT3H.sapiens    352 SAFHQATYN-KQPMYRKAIYEVLQVASSRAGKLFPVCHDSDESDTAKAVEVQNKPMIEWALGGFQPSGPKGLEPPEE--------------EKNPYKEVY------ 436
DNMT3D.rerio      849 KCFCSNSFA-TVMAYKDAIFSSLQVASRRSRMFF-----SPESESKDELL---RVMLNWAFGGFEPLGADGLQPQAEYSVKVKKGKRKNPTGKLF----------- 934
DNMT3A.mellifer   663 EKMRDYTSNTKKHIYLVGVLQAS-------------------------------------------------------------------KDYCSRLGFD 695
DNMT3D.magna       95 RLLNQAHNN----LYEKGVLEAL-------RMLA-----HSKNQPK-------------------------------------LVKPTAGKLKSWAIKIFSSVKGNT 148

                       1030       1040       1050       1060       1070       1080       1090       1100       1110       1120
DNMT3H.sapiens    437 -------------------------TDMWVEPEAAAYAPPPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVR---QKCRNIEDICISCGSLNVT 503
DNMT3D.rerio      935 NLTVPLNKIPESLDLNNGLVDLGTTDADKKRLYSKWNGRSMQTVKIRRKYKQRNKNIIPTVQIESRQNSQKRHQMVHEFL---KNKRKIEDFCLSCGSMSVD 1033
DNMT3A.mellifer   696 TSNWTLDDAFEY--FSK-----PNHYDYASSANTWRREDSVKIYDKYSARIAEKLNELKDNPNVDDQRANDI-NNSDDLRSAIKGEISFDSLCLKCLRVSND 789
DNMT3D.magna      149 NRDVVLDDFPEW--LVRCISRIKDTNDKLKRISNEKKESRLSAFQGGQKKKEKKKNNAEDDDVTYGPRDCDI-AVVDRIK---AGTLDMDYFCLGCHVEIMG 244

                       1130       1140       1150       1160       1170       1180       1190       1200       1210       1220
DNMT3H.sapiens    504 --LEHPLFVGGMC-QNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFCVECVDLLVGPGAAQAAIKEDPWNCYMCGHKG--TYGLLRRRED 600
DNMT3D.rerio     1034 --IIHPLFEGKLC-TNCKFNFTETLYRYDEDGYQSYCTVCCSGMEVILCGHDSCCRSFCVDCLDILVCQGTFDQLKNVDPWTCYLCAPET--SSGALKPRHD 1130
DNMT3A.mellifer   790 EMDIHPFFEGSLC-KDCSERYKPCMFVFGNDSKCFYCTVCAASGMVIICDKEDCPRVYCTACMKHLLCPTTYEQVLQEDPWECFLCKSRSFTTDTIVRPRAN 890
DNMT3D.magna      245 PCTEHPLFLGSLCNQEECKNELLGTSYSVGEDDIHVGCAICGSTGDCLVCENGSCSRVYCSYCVELLVGEGSMSAIMQKTPWFCFLCAHFATESHGLLQPRSD 346

                       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320
DNMT3H.sapiens    601 WPSRLQMFFANNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWG--- 699
DNMT3D.rerio     1131 WSIRVQEFFANDTGMEFEPHRVYPSIPAIQRRPIRVLSLFDGIATGYLVLRDLGFKVEKYVASEIDEESITISMVNHDGKITQVDDVKNITKKHIEQWG--- 1229
DNMT3A.mellifer   891 WKDKIINMFRTSCDSNVEHLVAKHN--SEK-RKIRVLSLFDGLGTGLLVLLKLGF------------------------------------------------ 942
DNMT3D.magna      347 WALKLRNFLHQEFPLQLNRHLSKNGQPS-----LRVVVLHDNLGAVNYALGSLRLHLDQYMASDRWEEVYELNGVDFV------VLKNANQMKDEEWAKLC 436

                       1330       1340       1350       1360       1370       1380       1390       1400       1410       1420
DNMT3H.sapiens    700 PFDLVIGGSPCN--DLSIVNPA-----RKGLYEG-TGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARY 793
DNMT3D.rerio     1230 PFDLLIGGSPCN--DLSIVNPA-----RKGLYEG-TGRLFFEYYRLLNVLKPKEDDPQPFFWLFENVTFMQTHVKADICRFLECNPVLVDAVKVSPAHRARY 1323
DNMT3A.mellifer   943 ----IFFYSPCV--------------FSSLLLS-YLFIFILYLYILYILLYLYTYIYPIIYIFIYISFIF-------FLIYLPL--------------- 1002
DNMT3D.magna      437 PIHLLVSCLPAKKFDPQRHQPAGREVGRRQFAEGEYGEAFFDFFHIKNCIE-RNNQETPLLWAVENVSCYHIT----ISRFLQMEPIIFDIRDGQGCVKHRL 533

                       1430       1440       1450       1460       1470       1480       1490       1500       1510       1520
DNMT3H.sapiens    794 FWGNLPGMNR------------PLAST-----------------------------VNDKLELQE-------CLEHGRIAKFSKVRTITTRSNSIKQG- 843
DNMT3D.rerio     1324 FWGNIPGMNR------------PIIAS-----------------------------QNDKLCLQE-------CLEPGRTAKYEKVRTITTRQNSLKQGT 1374
DNMT3A.mellifer  1003 ----------------------------------------------------REDR-CFQ------------------------------- 1009
DNMT3D.magna      534 VWTNIPEQNFKPQLEQVINNDCPSISSLPLSREPFFIKYPWILEDIVAIAQPRPANWQEKHDRMLFQRDTRSKTKIIKRPSYLEHYSVVQLTQKSAKKRRKM 635

                       1540       1550       1560       1570       1580       1590       1600       1610       1620       1630
DNMT3H.sapiens    844 KDQHFPVFM-----NEKED-----ILWCTEM------ERVFGFPVHYTD-VSNMSRLARQRLLGRSWSVPVIRHLFAPL------------------- 905
DNMT3D.rerio     1375 NDAHFPVTM------NGKDD----HIWITEL------EKIFGFPKHYTD-VKSMGRPQRQRVLGKSWSVPVIRHLLAPL------------------- 1436
DNMT3A.mellifer  1010 ---------------------------------------------V---------------------------------------------- 1010
DNMT3D.magna      636 DDEFHPNWRLRENGNKAEDTDDDQTVYVAALTNHATYARSLGFPEPFPEFVSHLDEAAVQERLNRSLPVSLWIHLFQPLLKLARIAEPDCKSTILLLLQRQC 737

                       1640       1650       1660       1670       1680       1690       1700       1710       1720       1730
DNMT3H.sapiens    906 -------------------------------------------KEYFAC------------------------------ 911
DNMT3D.rerio     1437 -------------------------------------KDYFACDEFPVK------------------------------ 1448
DNMT3A.mellifer  1011 --------------------------------------YSYF----------------------------------- 1014
DNMT3D.magna      738 GHGNTRSHRHCHQNSRNSHVGCVAVFSNTISDSDKIGCKKEPVSYSSMTGEMQTEFHSIRPIQTFRQIPTNVLRRLHVEVKAILHFRWTVDASNGSLPNPTR 839

                       1740       1750       1760
DNMT3H.sapiens        -----------------------------------V 912
DNMT3D.rerio          -----------------------------------
DNMT3A.mellifer       -----------------------------------
DNMT3D.magna      840 RDDGSCDFPSARQVRRWRDMRESRIQSSACLKDD 873
```

96

Figure 3.4 Protein alignment for DNMT3 sequences for human, zebrafish, honey bee and *Daphnia*. Only DNMT3a sequence is presented for human and zebrafish. Conserved domains are highlighted within the sequences; further information on domain organisation is presented in Figure 3.5.
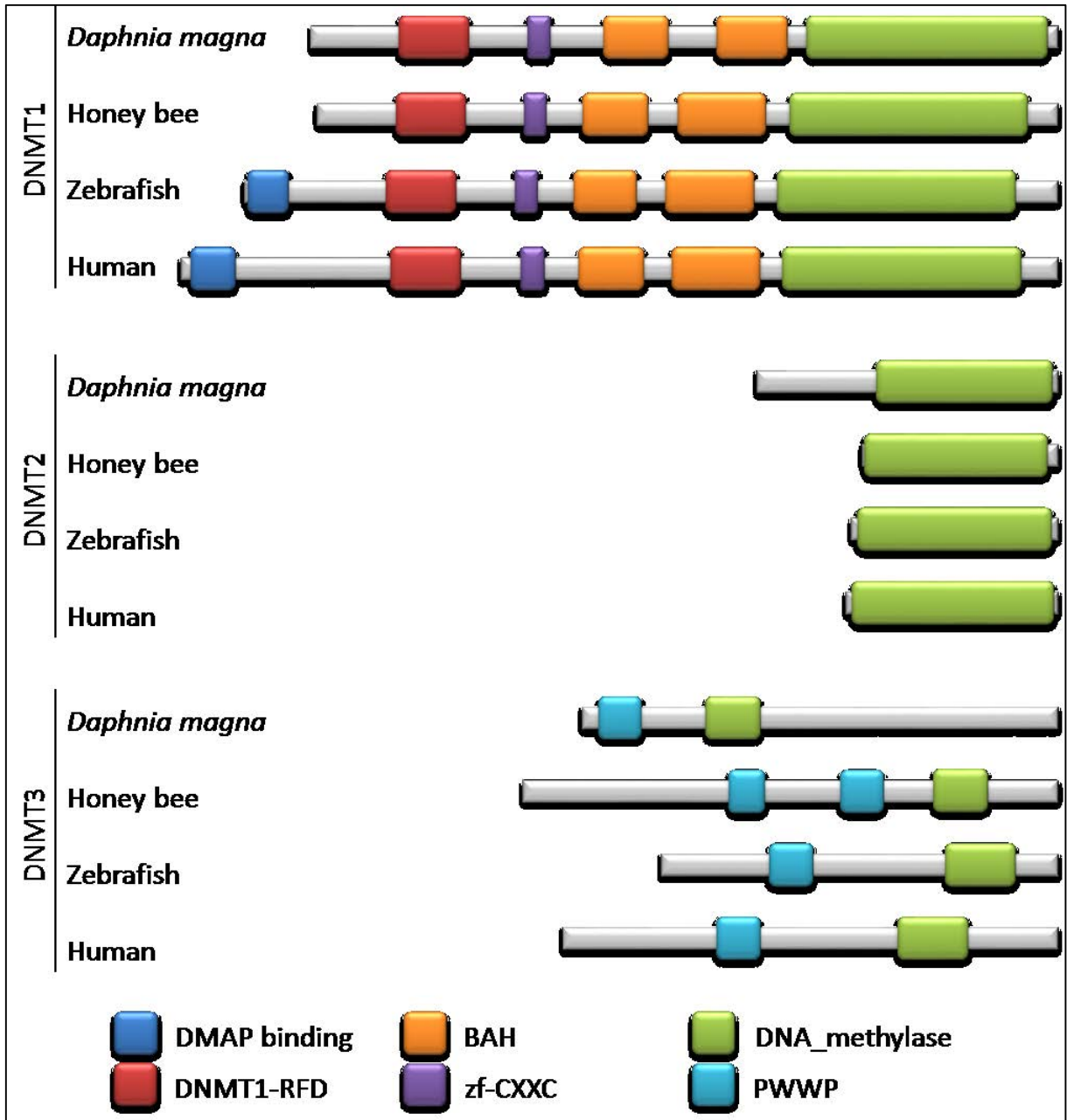
Figure 3.5 Conserved domains structure of the DNMT1, DNMT2 and DNMT3 for *Daphnia magna*, honey bee, zebrafish and human. Only DNMT1a sequence is presented for honey bee and DNMT3a sequence for human. Honey bee DNMT1b and human DNMT3b presented same domain organisation as the enzymes already shown for each species. Different domains are represented by different colours as demonstrated in the figure legend.

3.3.2 Analysis of gene expression levels of enzymes involved in the DNA methylation processes

Following identification of the enzymes potentially involved in DNA methylation and demethylation processes and the one-carbon pathway in *Daphnia* (Table 3.1), their transcription levels were investigated at different time points using RT-PCR. The methods for primer design and analyses of the data are described in sections 2.9.1 and 2.9.3, respectively. RNA was obtained from *D. magna* at different ages (days 1, 5, 12, 21 and 28) in order to characterize the changes that occur throughout the lifespan of the organism. Results are presented as Log2 fold-change comparing the expression level at different ages to the expression at day 1.

DNMTs expression was analysed for different ages (Figure 3.6). *DNMT1*, encoding the enzyme responsible for the maintenance of DNA methylation, presented increased expression ($p < 0.05$) at days 12, 21 and 28. *DNMT2*, encoding an RNA methyltransferase, was downregulated only after day 5. The *de novo* methytransferase, *DNMT3*, was downregulated at days 5 and 12. Expression returned to day 1 levels at day 21 and increased at day 28 (Figure 3.6).

Furthermore, the transcripts encoding enzymes involved in the one-carbon pathway were also analysed. *GNMT* followed the same expression pattern as *DNMT3* (Figure 3.7). *SAHH* expression was increased at day 12, while *MTRR* expression was decreased at day 5. *MS* presented increased expression at all ages compared to day 1, although not statistically significant at day 21. *BHMT* and *MAT* were upregulated at days 5 and 12 (Figure 3.8).

Ten-eleven translocases (*TETs*) expression was also quantified. The two homologs identified were downregulated at day 5 compared to day 1 and *TET*_1 was also downregulated at day 21 (Figure 3.9).
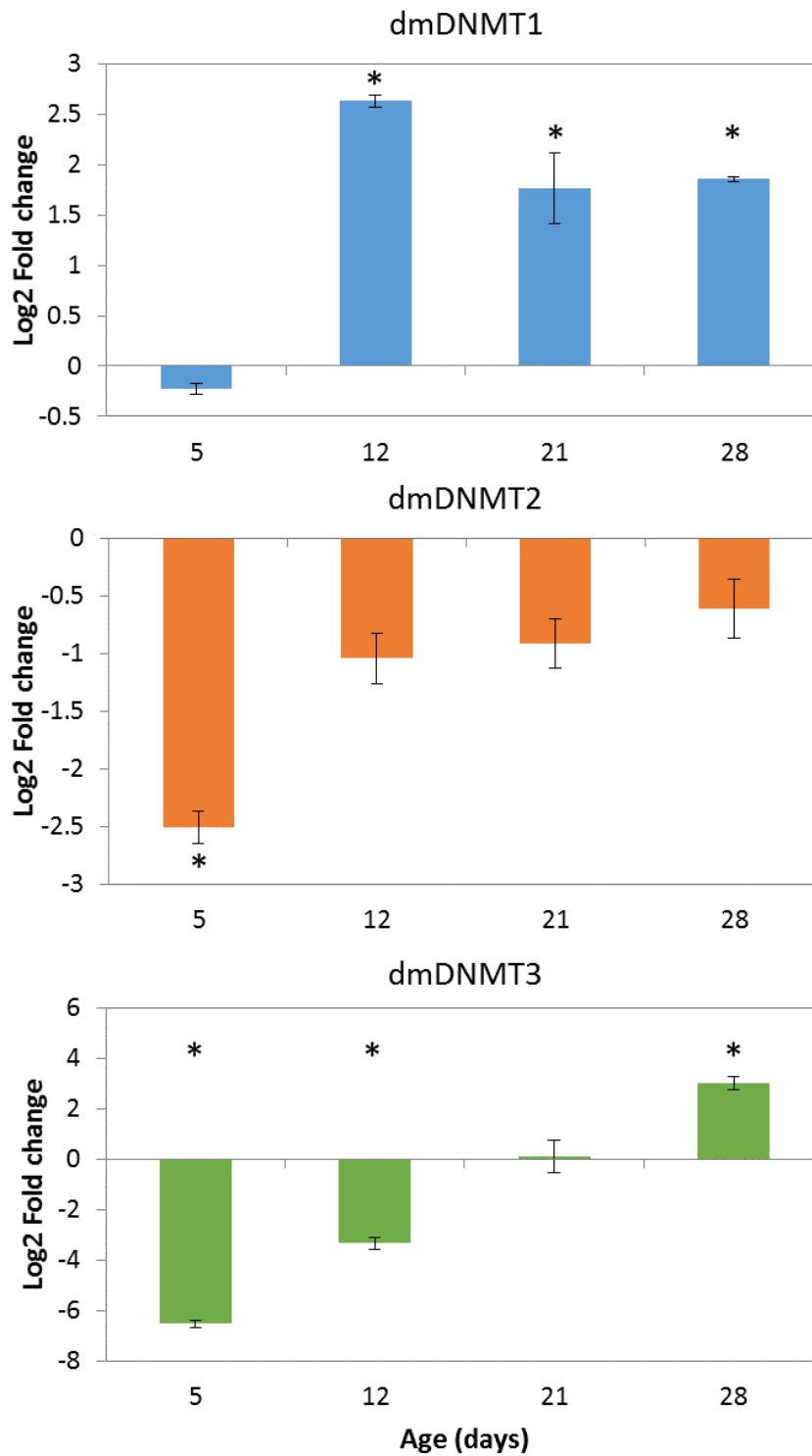
Figure 3.6 Gene expression analysis of DNA methyltransferases (DNMTs). Relative log2 fold change to day 1 expression. * Significantly different from day 1 (t-test; p<0.05). Error bars indicate standard error of the mean. Three biological replicates were analysed with three technical replicates.
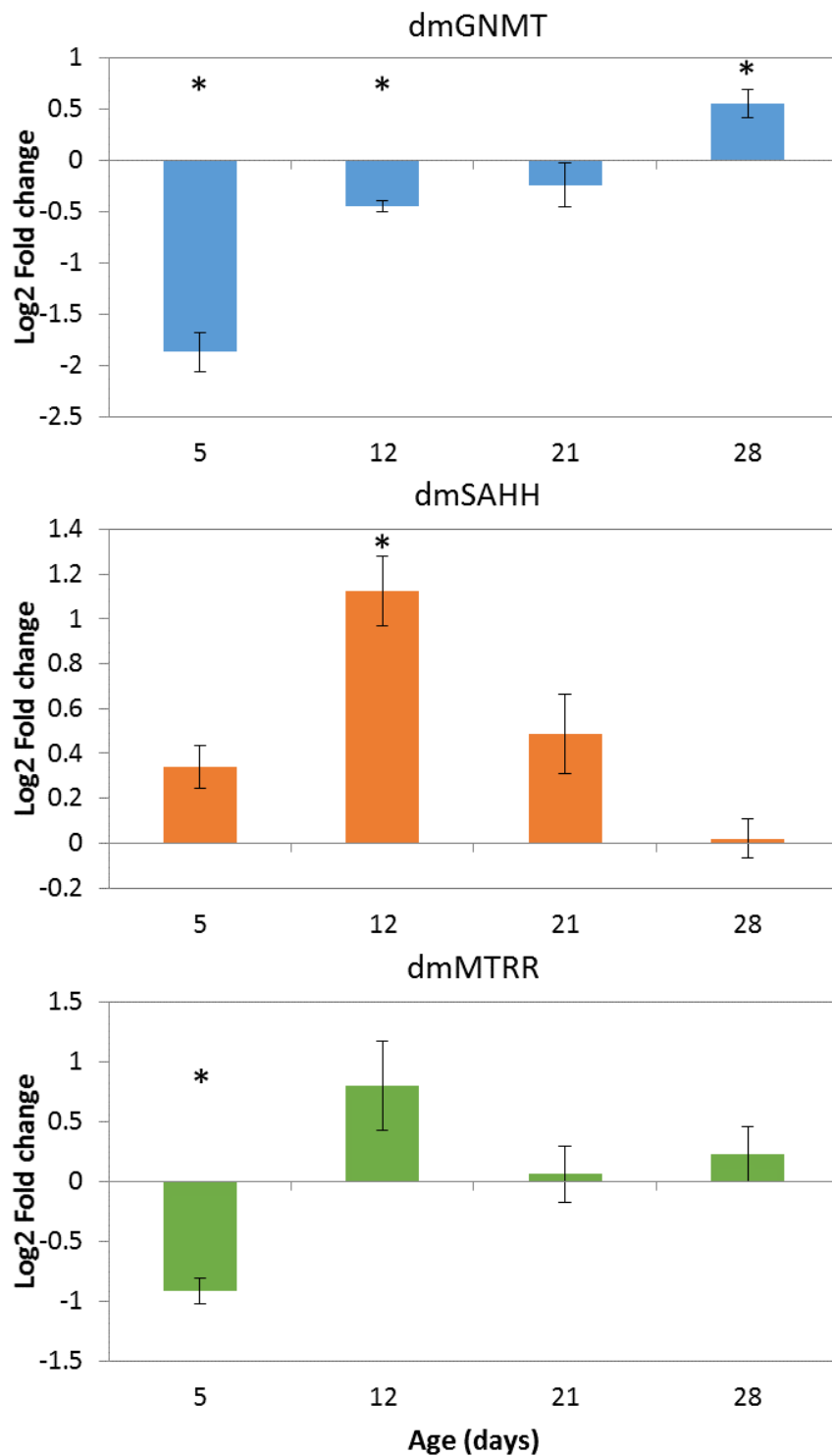
Figure 3.7 Gene expression analysis of *GNMT*, *SAHH* and *MTRR*. Relative log2 fold change to day 1 expression. * Significantly different from day 1 (t-test; $p < 0.05$). Error bars indicate standard error of the mean. Three biological replicates were analysed with three technical replicates.
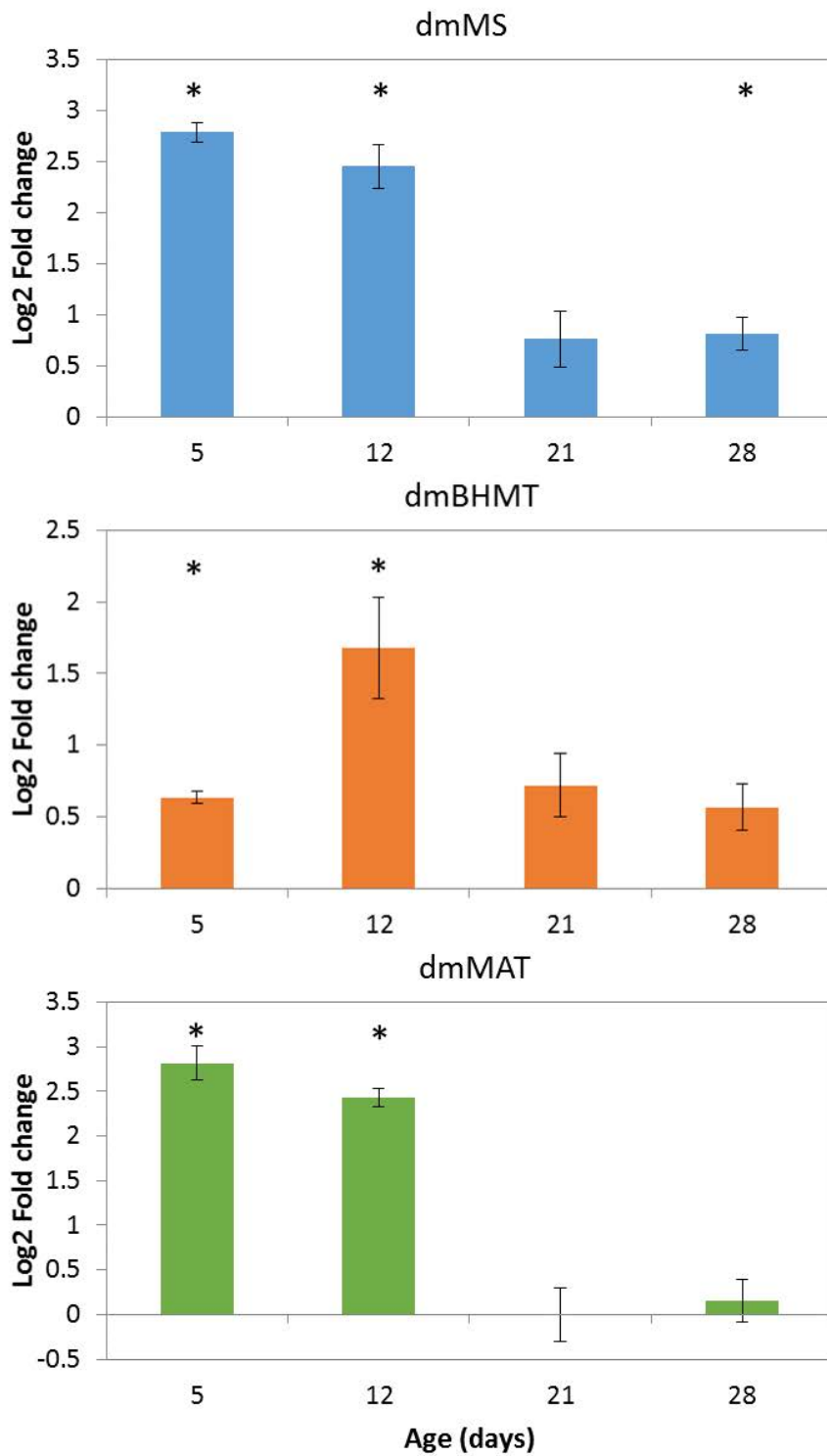
Figure 3.8 Gene expression analysis of *MS*, *BHMT* and *MAT*. Relative log2 fold change to day 1 expression. * Significantly different from day 1 (t-test; $p < 0.05$). Error bars indicate standard error of the mean. Three biological replicates were analysed with three technical replicates.
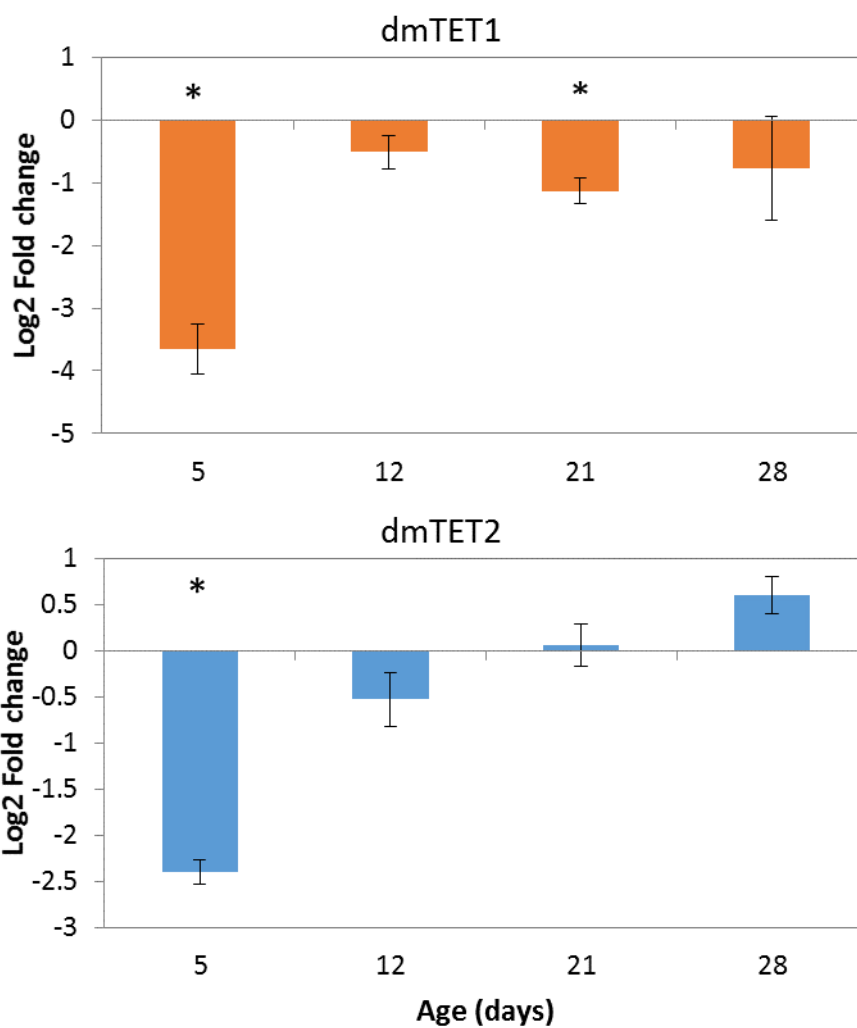
Figure 3.9 Gene expression of *TET* homologs. Relative log2 fold change to day 1 expression. * Significantly different from day 1 (t-test; p<0.05). Error bars indicate standard error of the mean. Three biological replicates were analysed with three technical replicates.

3.3.3 Global DNA methylation

Global DNA methylation was measured with liquid chromatography coupled to mass spectrometry (methods are presented in section 2.6). Results are presented as percentage of methylated cytosine. For 14 days old daphniids, the overall DNA methylation level was measured as 0.14 ± 0.007% (mean ± SEM). Overall DNA methylation for Bham2 is lower than the other *D. magna* strains (0.49 and 0.52%), but it is in accordance with the values reported for other invertebrates.

3.3.4 The draft genome of *Daphnia magna* Bham2 strain

Daphnia species play an important role in freshwater ecology. Also, several *Daphnia* species are well established as model organisms for ecotoxicology. Thus, *Daphnia* is a model organism used for linking laboratory-based studies directly to field studies. Unravelling the genome sequence of *D. magna* will allow determination of the relationships between genotype and phenotype of organisms and aid understanding the effects the environment on populations and communities. It will be a powerful tool to answer several questions from an ecological, ecotoxicological and evolutionary perspective. An effort to sequence the *D. magna* genome (Xinb3 strain) is already being made by the *Daphnia* Genomics Consortium, however it was only published in April of 2016, and still as a draft genome. It is available via NCBI BioProject PRJNA298946.

The genome sequencing was performed at the University of Birmingham using a Hiseq 2500 platform on Rapid run mode, producing 24.1 Gb of raw data. The data analyses, including

quality control, genome assembly and DNA methylation calls, were performed in collaboration with Genotypic Technologies.

3.3.4.1   Raw data quality control and processing

The Illumina HiSeq paired end raw reads were quality checked and processed using TrimGalore to remove adapters and poor quality reads. The SeqQC report for the Illumina raw data and processed reads statistics are presented in Table 3.2. Processed reads generated 19.8 Gb of data corresponding to approximately 86X coverage (considering a genome of 200 Mb).

Table 3.2. Quality control statistics for raw and processed sequencing reads.

Sample	Raw reads		Processed reads	
	1	2	1	2
Mean Read Length	151	151	138	138
Total Number of Reads	(71.13 millions)	(71.13 millions)	(62.79 millions)	(62.79 millions)
Total Number of HQ Reads 1*	(68.22 millions)	(59.68 millions)	(63.31 millions)	(63.31 millions)
Percentage of HQ Reads	95.91%	83.91%	100.00%	100.00%
Total Number of Bases	10740134267 bases	10740134267 bases	8713888298 bases	8713888298 bases
Total Number of Bases in Mb	10740.13427 Mb	10740.13427 Mb	8713.88830 Mb	8713.88830 Mb
Total Number of HQ Bases 2*	10275779095 bases	9138945739 bases	8591320524 bases	8330093451 bases
Total Number of HQ Bases in Mb	10275.77909 Mb	9138.94574 Mb	8591.32052 Mb	8330.09345 Mb

Table 3.2. Continued from previous page

Percentage of HQ Bases	**95.68%**	**85.09%**	**98.59%**	**95.60%**
Total Number of Non-ATGC Characters	89409 bases	3441781 bases	42071 bases	492964 bases
Total Number of Non-ATGC Characters in Mb	0.089 Mb	3.4418 Mb	0.042 Mb	0.493 Mb
Percentage of Non-ATGC Characters	0.00%	0.03%	0.00%	0.01%
Number of Reads with Non-ATGC Characters	59925	339586	41830	258023
Percentage of Reads with Non-ATGC Characters	0.08%	0.48%	0.07%	0.41%

* >70% of bases in a read with >20 phred score and reads which are of low quality can be trimmed and used 2* bases with >20 phred score.

3.3.4.2   Draft *De novo* assembled genome of *Daphnia magna* Bham2

De novo assembly of Illumina HiSeq data was performed using ABySS 3.8 assembler, followed by scaffolding using paired-end data with SSPACE scaffolder. Then, the intra-scaffold gaps were closed using GapCloser tool to the maximum level possible. The *de novo* assembly QC statistics at each step are presented in Table 3.3.

The genome was assembled to 1,828,469 contigs, later grouped to 124,614 scaffolds with a total length of 122 Mb. The maximum scaffold length is 288,378 bp and the minimum length 200 bp. The average scaffold length is 985.4 ± 2,281.8 bp. The number of generated scaffolds with less than 500 bp is 70,043, while only 31,116 scaffolds were more than 1kb. The N50 value, the length for which 50% of all bases in the assembly are in a contig of specified length, for the GapClosed scaffolds is 2,014. In other words, this means that 50% of the assembly contains contigs with length equal or greater than 2,014 bp (Additional File 3.1).

The high number of small contigs and scaffolds is not ideal for assembling a genome, making it a challenging task, especially for genomes with a high number of repetitive elements, such as in *Daphnia*. Nevertheless, the assembled genome provides a good starting point for mapping DNA methylation data and conducting DNA methylation experiments in this invertebrate model organisms.

Table 3.3. Statistics for the different steps of genome construction.

Description	ABySS contigs	Scaffolds	GapClosed scaffolds
Contigs Generated	1828469	124614	124614
Maximum Contig Length	78705	287464	288378
Minimum Contig Length	64	200	200
Average Contig Length	156.1 ± 474.9	985.4 ± 2278.9	985.4 ± 2281.8
Median Contig Length	117	421	421
Total Contigs Length	285476057	122796826	122791038
Total Number of Non-ATGC Characters	638727	1060445	210008
Percentage of Non-ATGC Characters	0.224	0.864	0.171
Contigs >= 100 bp	867472	124614	124614
Contigs >= 200 bp	128114	124614	124614
Contigs >= 500 bp	54166	54571	54571
Contigs >= 1 Kbp	30689	31117	31116
Contigs >= 10 Kbp	587	616	614
Contigs >= 1 Mbp	0	0	0
N50 value	140	2013	2014

3.3.4.3   CpG island prediction

The draft *de novo* assembled genome was used for prediction of the GpG islands (CGI) using EMBOSS newcpgreport software. A total of 30,600 CGIs were identified using this method. The size ranged from 200 bp to 3,606 bp with an average size of 385 bp. The identified CpG islands were used for analyses of DNA methylation profiles described in this chapter (Additional File 3.2).

3.3.4.4   Annotation of transcription start sites (TSS)

The first attempt to annotate the *de novo* assemble genome was by blasting the *D. magna* genome v2.4 to the Bham2 *de novo* genome. The first gene set available for *D. magna* v2.4 was based on gene prediction on the draft genome assembly, therefore it was not a high quality gene set. However, at the time, it was the only gene set available for the *D. magna* genome.

Also, due to the length of the scaffolds and contigs, several regions of the genome were only partially annotated to the new Bham2 genome. To try to overcome this problem the annotation was centred around the transcription start sites (TSS), however due to the nature of the methods employed, the hits did not represent single regions or genes. This method resulted in the identification of 25,820 TSSs in Bham2 genome (Additional File 3.3).

Methylation is known to occur around TSS located at promoter and first exon regions. Thus, the identified TSS represent an option that can be used to focus the DNA methylation profiling of *Daphnia's* genome.

3.3.4.5 Annotation of transcripts based on *Daphnia* evidential genes

With the release of the new *D. magna* gene set (finloc9b) in April of 2016, this one based on mRNA assembly and therefore much more complete, the draft genome for Bham2 was annotated again.

The major difference between the gene sets is the method of gene prediction. The first one, used for TSSs annotation, is based on gene prediction using the *D. magna* draft genome (v2.4), while the second gene set was achieved mostly through mRNA assembly, with a minor portion being from genome prediction. This makes the second gene set (finloc9b) much more reliable.

It is important to note that the draft genome was not improved with the release of this new gene set, therefore the choice of sequencing and *de novo* assembling of the Bham2 genome was still appropriate.

From the 29,121 genes predicted for *D. magna* genome based on the transcripts, 5,831 were annotated to the Bham2 draft genome, representing 20% of the total genes. The parameters for blast search in this case were more severe than TSSs and could be influenced by the fragmentation of the *de novo* assembled genome, explaining the low number of genes that were annotated (Additional File 3.4).

It is worth mentioning that only 65% of the predicted genes from the finloc9b dataset were mapped to the reference genome v2.4 (Orsini *et al.*, 2016).

3.3.5 DNA methylation profiling in *Daphnia magna*

3.3.5.1   Analysis of Methylation

For methylation mapping, firstly the reference genome was transformed to a bisulfite-converted version. Then, sequences were aligned to these reference using the short read aligner Bowtie2 (Langmead *et al.*, 2009), searching for unique alignments. After that, Bismark determines the methylation status of each DNA strand for different contexts (CpG, CHG and CHH, where H can be either A, T or C) (Krueger and Andrews, 2011).

Mapping efficiency of the bisulfite treated samples was around 35%. Low mapping efficiency could be attributed to the genome used as reference coupled with the already difficult read alignment of bisulfite treated sequences due to decreased complexity.

3.3.5.2   DNA methylation profile

After DNA methylation mapping, the output files were visualised using SeqMonk software. Each cytosine site was represented by different reads that are either methylated or not. The proportion of methylated reads for the same site is used to measure the methylation levels. The overall levels and site specific methylation can be identified for different contexts as CpG, CHG and CHH.

The output files were loaded to SeqMonk for visualisation of the methylation calls for the different contexts. Some scaffolds were identified with unusually high levels of methylation in CHG and CHH context and high numbers of reads. Based on these results and what was previously reported for arthropods, such as *A. mellifera*, *N. vitripennis* and *B. mori*, the occurrence of DNA methylation outside CpG context is very rare. Therefore, the scaffolds

presenting a non-expected pattern of methylation and high number of reads were excluded from the analysis (Additional File 3.5).

The initial genome size was 122Mb. After the removal of the excluded scaffolds the genome size was reduced to 119Mb, representing 92% of the length of the *D. magna* draft genome (v2.4).

The removal of the excluded scaffolds improved the analysis of global patterns of DNA methylation for *D. magna*. This allowed the proper identification of overall methylation levels, without the interference of regions presenting anomalous high methylation. In addition, two different steps (non-CpG methylation and biological replicates) to avoid false positives were applied to these data and the results are described below.

Non-CpG methylation is often described as a rare event for related invertebrates (Lyko *et al.*, 2010). Therefore, the amount of reads represented as methylated can be used as an estimation of bisulfite C-to-T conversion efficiency and to estimate the false positive rate for DNA methylation quantitation (Xiang *et al.*, 2010). Methylation values at non-CpG contexts were given by the ratio of the total number of reads and the methylated reads at CHG and CHH sites. The false positive rate for the 3 replicates was measured as 1.14%. It indicates that those reads are likely derived from non-converted cytosines, sequencing errors or can be from contaminating DNA. Xiang *et al.* (2010), when describing the silkworm methylome also encountered the same problem, obtaining higher methylation outside CpG contexts than expected. The use of biological replicates was able to solve this problem, as often mCpGs are conserved among replicates while mCs at non-CpG context are discordant. Therefore, independent biological replicates were used for the methylome analyses.

A methylated site was defined by Wang *et al.* (2013) as a site containing 10% methylated Cs and coverage ≥10. For this study, only sites containing 3 or more reads in 3 biological replicates were considered for the measurements. The cut-off value for methylated sites was set as ≥50%. This way, it requires at least two unconverted reads at minimum coverage for that site to be called as a methylated site, as one methylated read only accounts for 33.33% of methylation level, and it needs to be a consistent value for the replicates.

In summary, filtering for methylation profiling was done within and between samples. Only sites containing 3 or more reads in each of the three replicates were used for methylation quantitation. From those, only sites with methylation level higher than 50% were considered as methylated. These results are displayed in Table 3.4.

Table 3.4. Methylated cytosines in CpG, CHG, and CHH genomic context (H = A, T, or C)

Cytosines	Sites in Genome	Methylated Sites	% of All mCs
CpG	2425520	18228	98.58
CHG	1969049	151	0.817
CHH	5213998	112	0.606
Total	9608567	18491	

After the filtering step, only 263 Cs were methylated in non-CpG context, corresponding to 0.0027% of total cytosine sites and 0.0036% of non-CpG sites. Therefore, the use of biological replicates and strict cut-off values was effective in removing the false-positive methylated cytosines in the final methylation profile.

Overall, only a relative small proportion of the *D. magna* genome was found to be methylated (Additional File 3.6). For the cytosines sites covered during the analysis, 0.19% of these sites

were methylated, while for all CpGs, 0.63% were methylated. Primarily, methylation occurs at

CpG sites accounting for approximately 98.6% of the methylated cytosines (Figure 3.10).
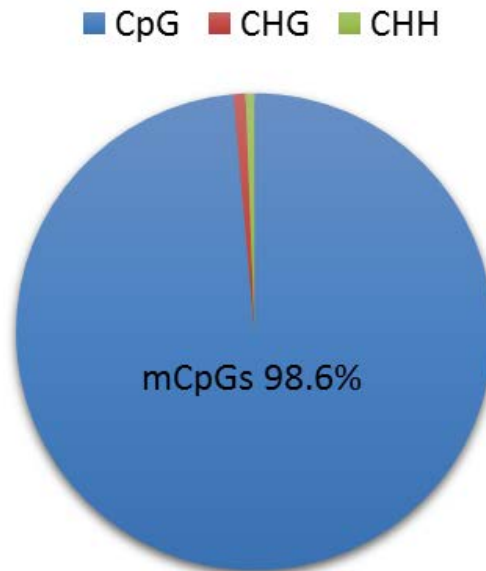
Figure 3.10 Total DNA methylated sites in different context along the *D. magna* Bham2
genome.

The distribution of methylation along the genome is not random. The analyses indicate that

DNA methylation is located within gene bodies, especially at the beginning of the genes

(Figure 3.11). Lower levels of methylation were observed for the predicted CGIs, compared to

the flanking areas (Figure 3.12).
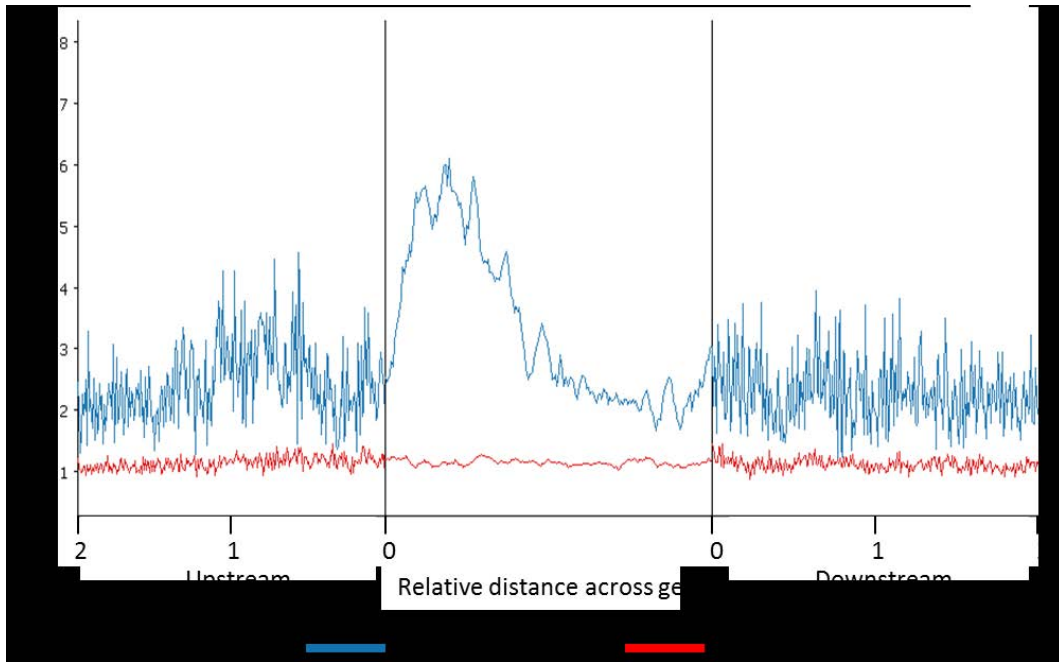
Figure 3.11 DNA methylation distribution along genes and flanking areas for CpGs and at non-CpG context. Flanking areas are set 2kb upstream and downstream the gene. Due to the different length of genes, they are presented with relative distance. Methylation in CpG context is presented in blue, non-CpG methylation is presented in red.
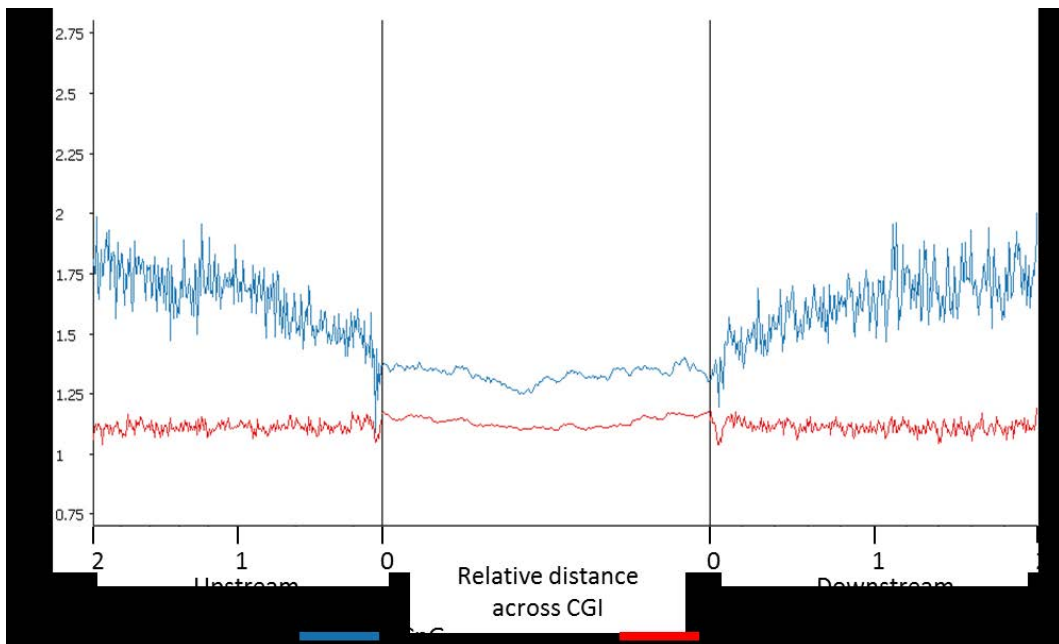
Figure 3.12 DNA methylation distribution along CGIs and flanking areas for CpGs and at non-CpG context. Flanking areas are set 2kb upstream and downstream of the CGI. Due to the different length of CGIs, they are presented with relative distance. Methylation in CpG context is presented in blue, non-CpG methylation is presented in red.

116

Comparisons between methylation profiles in 14 days old *Daphnia* and 5 days old were performed following the same procedure described above (Additional File 3.7). A scatter plot containing all covered cytosine sites in 14 and 5 days old samples revealed a widespread distribution of methylated probes (Figure 3.13). In total 26,283 probes were methylated for at least one of the samples. Both samples shared 10,432 methylated probes, while 8,055 were exclusive of 5 days old samples and 7,796 were considered methylated only in 14 days old *Daphnia*. Figure 3.13 shows that most of the probes were present in both groups, however many did not pass the cut-off value of 50% methylation. This indicates that the methylation level of those probes can be affected by age and could be related to developmental mechanisms in one or several different tissues, through tissue-specific methylation.
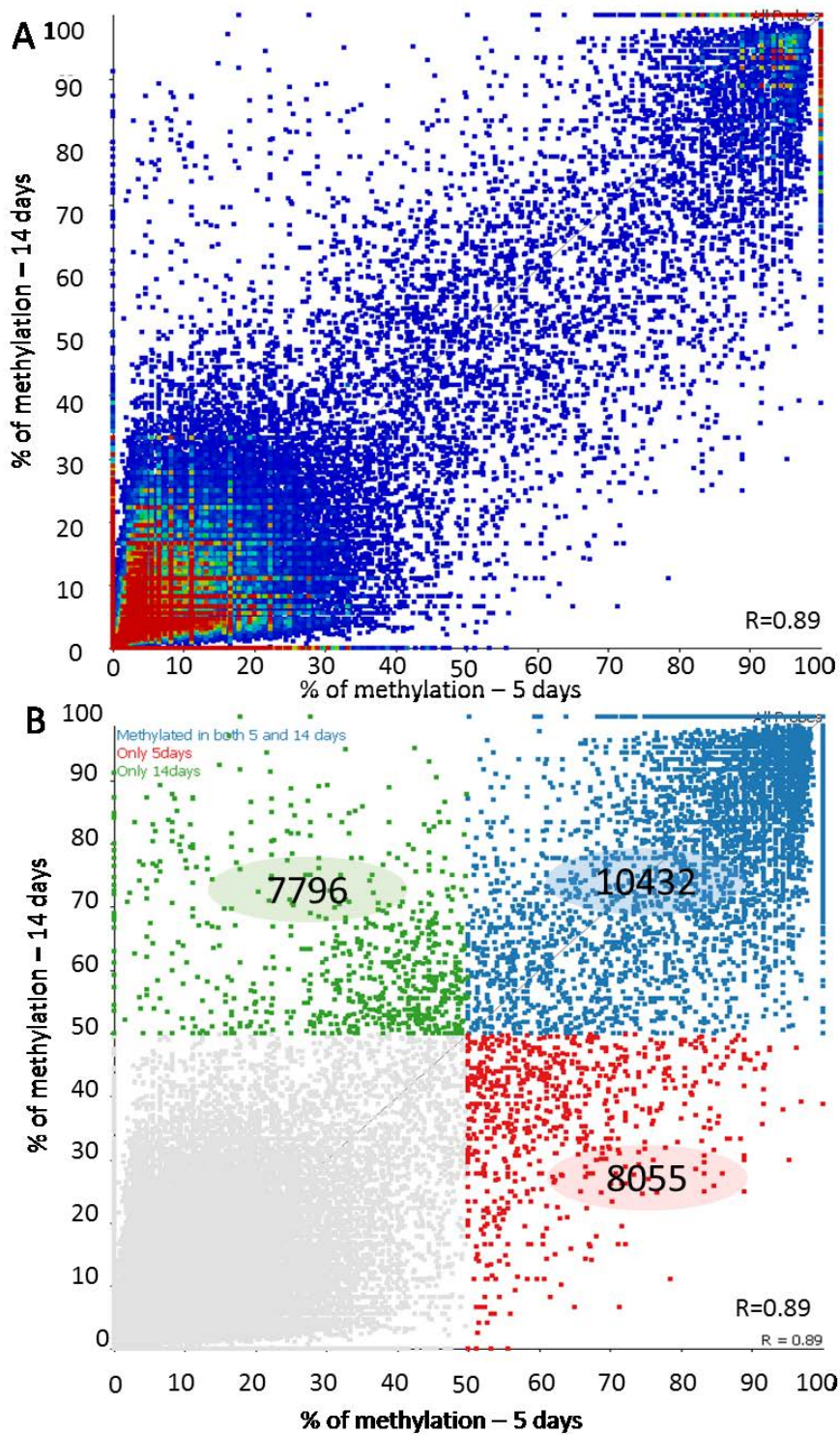
Figure 3.13 Scatter plot of cytosine probes (single Cs) at CpG context for 5 and 14 days old daphniids. A) All probes for 5 days old *Daphnia* and 14 days old *Daphnia*. Each probe is represented by a dot. Density of probes is represented by red (high density) and blue (low density). B) Light blue probes were classified as methylated in both samples. Red probes were classified as methylated only in 5 days old daphniids and green probes were methylated only for 14 days old daphniids. Probes in grey were not methylated for both groups.

Yet, when looking at probes within gene bodies, it was possible to observe that the methylation is more consistent between 5 days and 14 day groups (Figure 3.14). The figure clearly indicates that some probes fall in the regions highlighted in green and red in Figure 3.13. However, no differences in the overall distribution of methylation across the gene bodies was identified (Figure 3.15).
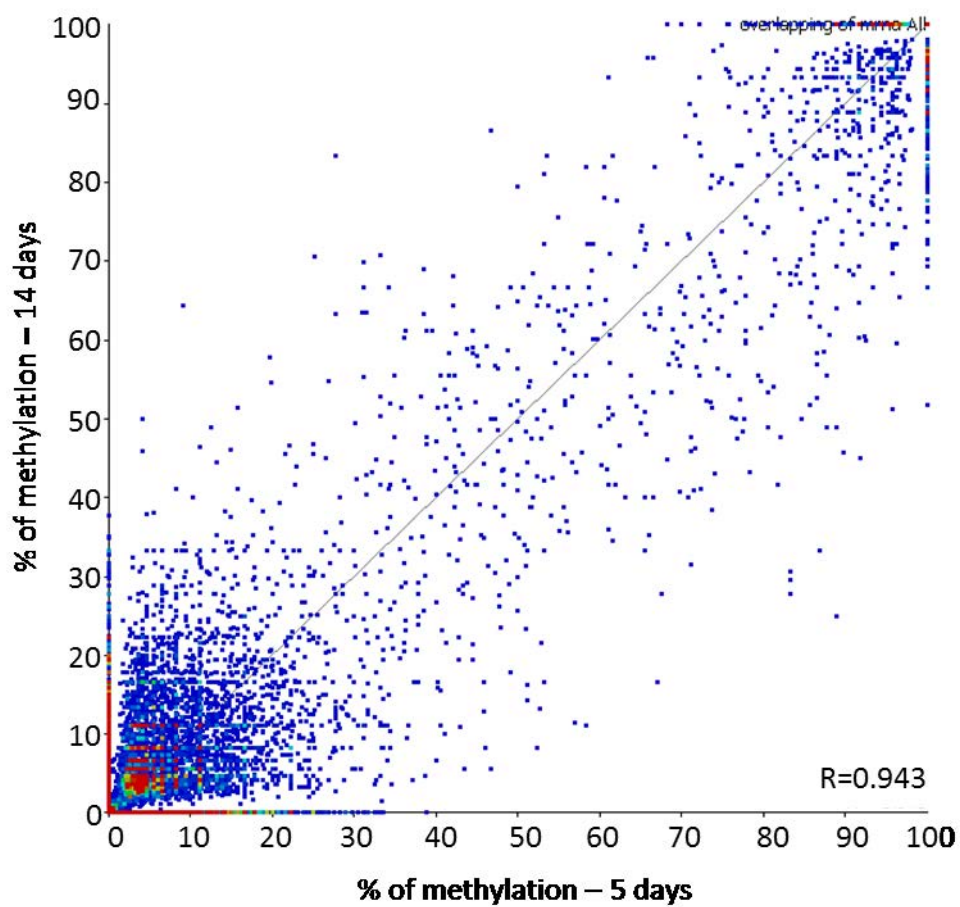
Figure 3.14 Scatter plot of cytosine probes (single Cs) at CpG context for 5 and 14 days old daphniids that are overlapped by an annotated gene. Each probe is represented by a dot. Density of probes is represented by red (high density) and blue (low density). R=0.943

119
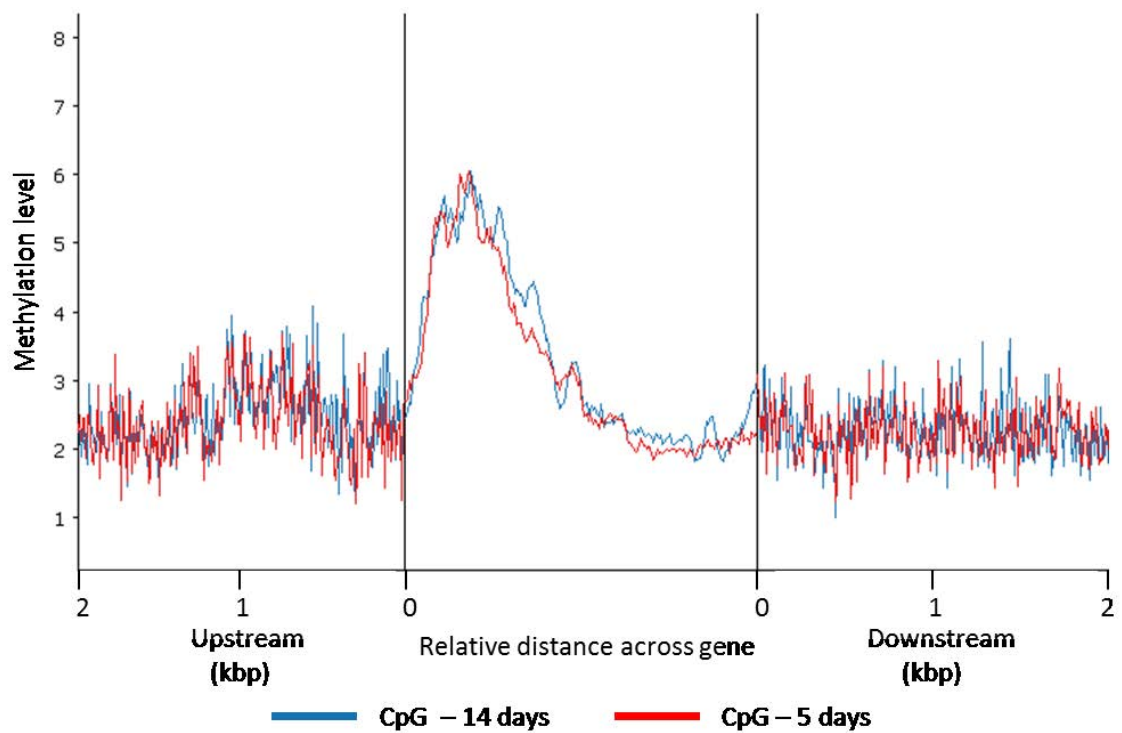
Figure 3.15 DNA methylation distribution along genes and flaking areas at CpGs sites in 5 days old and 14 days old *Daphnia*. Flanking areas were set as 2kb upstream and downstream the gene. Due to the different length of genes, they are presented with relative distance. Methylation in CpG context is presented in blue, non-CpG methylation is presented in red.

3.4 Discussion

DNA methylation has been investigated across a wide range of species. Originally the efforts were focused on the use of vertebrate species such as human and mouse, specifically due to medical application or their use as model organisms. Lately many studies have concentrated on uncovering the DNA methylation profiles of invertebrates. Figure 3.16 presents a summary of information available for a number of species, including plants, invertebrates and vertebrates, that will be discussed within this section.

It is now known that the level and distribution pattern of DNA methylation can vary dramatically between species (Jiang *et al.*, 2014). In vertebrates, especially mammals, the DNA methylation pattern has a "global" distribution, where candidate sites are methylated across the entire genome, excluding promoter regions that remain largely unmethylated and are associated with gene transcription. On the contrary, invertebrates present a distinct distribution of methylated cytosines, having in general a "sporadic" pattern of DNA methylation. However, some invertebrate species have a very low level of DNA methylation or their DNA methylation is restricted to a specific life stage (Breiling and Lyko, 2015; Feng *et al.*, 2010a; Jiang *et al.*, 2014).

The nematode *C. elegans* does not present any enzyme homologous to DNA methyltransferases, and only 0.0033% of its cytosines were found to be methylated (Hu *et al.*, 2015; Simpson *et al.*, 1986), although epigenetic control is still important through other mechanisms. The model organism *D. melanogaster* also exhibits a very low level of DNA methylation (0.03%) but from the known DNMTS, it only encodes the gene for DNMT2, a RNA methylase (Capuano *et al.*, 2014; Glastad *et al.*, 2011). Other invertebrates, like *A. mellifera*,

B. mori and *N. vitripennis*, have higher levels of DNA methylation than *C. elegans* and *D. melanogaster* (0.11-0.18%), but still very low when compared to vertebrates (~4-8%) (Beeler *et al.*, 2014; Lyko *et al.*, 2010; Rasmussen and Amdam, 2015; Xiang *et al.*, 2010). The global DNA methylation level in *Daphnia* was measured in different strains and found to vary from 0.14% to 0.52% (Asselman *et al.*, 2015). This is in accordance with the overall values for other arthropods.

Despite the conservation of overall methylation levels among invertebrates, the presence of DNMTs homologs is not distributed along the evolutionary tree (Lyko and Maleszka, 2011). The DNMTs encoded on the genomes of different invertebrates do not occur according to the phylogenetic divergence of these organisms. Also the number of DNMTs within each class is not conserved even with a taxonomy group (i.e. hymenoptera). *A. mellifera* and *N. vitripennis* belong to the same order in the Insecta class and differ in the number of DNMT1 (Glastad *et al.*, 2011; Lyko and Maleszka, 2011). Therefore, while it is possible to say that mammals encode the same number of DNMTs, there is not a number or pattern of DNMTs that can be associated with invertebrates, or even to a more limited group, such as arthropods.

Another important difference identified between vertebrates and invertebrates was the lack of DMAP domain in DNMT1. This was identified for *A. mellifera*, *B. mori* and *D. magna* and it is suggested that, as the result of lacking of this domain, transposable elements on these organisms are largely unmethylated (Lyko *et al.*, 2010; Mitsudome *et al.*, 2015; Rountree *et al.*, 2000; Werren *et al.*, 2010; Xiang *et al.*, 2010).

DNA methylation machinery includes the enzymes involved in DNA methylation and demethylation pathways. The mechanisms of DNA methylation are mostly conserved across different species. The enzymes of the one-carbon pathway are the responsible for providing

the necessary metabolites and to maintain the reaction of DNA methylation. For some species, the complete set of enzymes failed to be identified. However, for several of them, this problem is more likely linked to a lack of high quality genome construction and annotation.

On the other hand, the pathway of DNA demethylation does not appear to be conserved along different groups. The presence of 5hmC was analysed in many species, and for example, it was not possible to identify 5hmC in *C. elegans*, even with an extremely low detection limit (Erdmann *et al.*, 2014; Hu *et al.*, 2015). Several species still lack information about the presence of 5hmC as illustrated in figure 3.16.

The lack of a complete and annotated genome is one of the complications for the analyses of the DNA methylation profile in any species. For *D. magna* the quality of the available genome is very poor, although the recently published gene set adds a lot to it (Orsini *et al.*, 2016). We aimed to produce a *de novo* assembled genome for *D. magna* bham2 strain. However, due to many aspects described in this chapter, it was only possible to assemble a partial genome, with incomplete annotation. Although far from ideal, this draft genome allowed the analyses of the DNA methylation profile in *D. magna*.
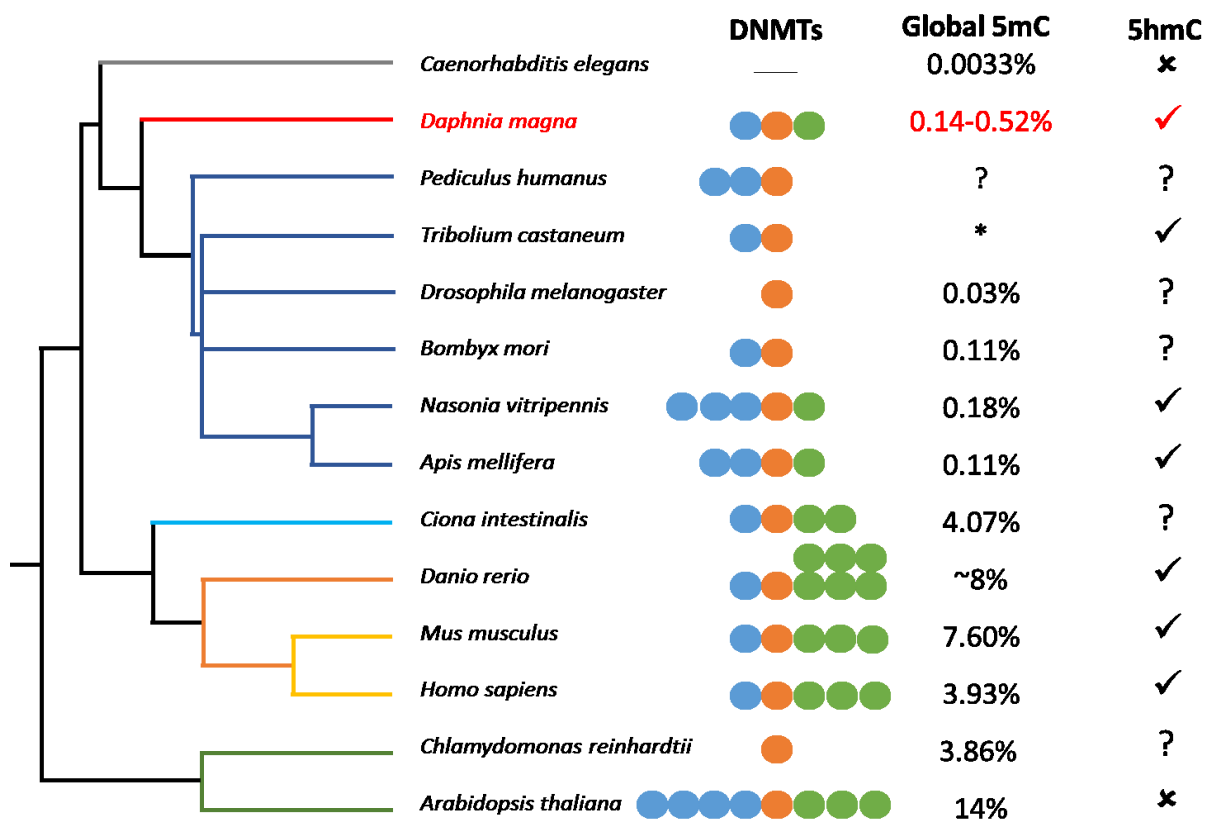
Figure 3.16 Phylogenetic distribution of DNA methylation in vertebrates and invertebrates. Phylogenetic tree was generated with NCBI Taxonomy (http://www.ncbi.nlm.nih.gov/guide/taxonomy/). Branches are named for species for which DNA methylation information has been obtained. Branch colour indicate species taxa (grey: Nematoda; red: Crustacea; blue: Insecta; light blue: Tunicata; orange: Vertebrata; yellow: Mammalia; green: Plantae). The number of DNA methyltransferases (DNMTs) are represented as dots, with blue representing DNMT1, orange representing DNMT2 and green for DNMT3. The percentage of cytosine methylation is shown for each species. "*" to indicate the presence of DNA methylation in embryonic stages of T. castaneum. Presence of 5hmC is indicated by a check mark. An "✖" marks the species where 5hmC was not identified. Question marks indicate no data for 5mC and 5hmC.

References are listed per species as follows: *C. elegans* (Hu *et al.*, 2015; Simpson *et al.*, 1986); *D. magna* (Asselman *et al.*, 2015); *P. humanus* (Glastad *et al.*, 2011); *T. castaneum* (Cunningham *et al.*, 2015; Feliciello *et al.*, 2013; Zemach *et al.*, 2010); *D. melanogaster* (Capuano *et al.*, 2014; Lyko *et al.*, 2000; Rasmussen and Amdam, 2015); *B. mori* (Xiang *et al.*, 2010); *N. vitripennis* (Beeler *et al.*, 2014; Pegoraro *et al.*, 2016); *A. mellifera* (Lyko *et al.*, 2010; Wojciechowski *et al.*, 2014); *C. intestinalis* (Ponger and Li, 2005; Zemach *et al.*, 2010); *D. rerio* (Kamstra *et al.*, 2015b; Smith *et al.*, 2011); *M. musculus* (Capuano *et al.*, 2014); *H. sapiens* (Li *et al.*, 2010); *C. reinhardtii* (Feng *et al.*, 2010a; Ponger and Li, 2005); *A. thaliana* (Capuano *et al.*, 2014; Erdmann *et al.*, 2014).

Although the common approaches for DNA methylation profiling could not be applied to the *de novo* assembled genome, it allowed an overview of DNA methylation distribution and patterns in *D. magna*. Therefore, future studies need to address the link between DNA methylation and regulatory regions on the genome. Future analyses will be reliant on the construction of the full sequence and annotation of the *D. magna* genome.

The overview of *D. magna* methylation profile confirmed what was already predicted. DNA methylation is targeted to gene bodies, especially at the beginning (5' end) of the genes. Additionally, the overall pattern of DNA methylation suggests that CGIs are generally unmethylated in *Daphnia*. Changes in the global levels of DNA methylation have already been reported in *Daphnia* exposed to chemicals and environmental stressors (Asselman *et al.*, 2015; Menzel *et al.*, 2011; Vandegehuchte *et al.*, 2010b, 2009a, 2009b). This information coupled to the findings in this study classifies *D. magna* as a useful model organism for epigenetic studies, due to its responsive epigenome, extensive knowledge on its ecology and the easy maintenance in laboratory conditions allowing the manipulation of different conditions and assessment of responses. To this point, gene specific DNA methylation changes were not investigated in *D. magna*.

As presented above, *D. magna* encodes the full toolkit for DNA methylation and demethylation. The DNMTs transcription varies substantially with ageing. DNMT1 gene expression increases when the animals achieve maturity (approximately at day 12) and the expression level is maintained until at least day 28. Expression of DNMT3 decreases after day 5 in comparison with day 1. At day 28 DNMT3 expression is increased once more. This pattern of gene expression was already demonstrated to occur in humans (Xiao *et al.*, 2008), hence these differences related to ageing could be significant for the global hypomethylation and

targeted hypermethylation that are shown to occur during across a lifetime (Zampieri *et al.*, 2015).

Regarding DNA methylation profiling for *D. magna* at different timepoints, the analysis of overall distribution of DNA methylation suggests some differences between 5 days old and 14 days old *Daphnia*. Those changes appear not to be occurring in the regions of the annotated genes. It is possible that the timepoints chosen for the comparison in DNA methylation profiles are not optimal, despite being covering two distinct stages of development in *Daphnia* (juvenile and mature *Daphnia*). Nevertheless, the use of organisms with different ages (i.e. neonates, mature and "old" animals) could help to differentiate the effects on DNA methylation more easily. The investigation of the alterations in DNA methylation during development could also add to the knowledge about DNA methylation in *D. magna* and its possible dynamic changes.

3.5   Conclusions

In conclusion, the results presented in this chapter demonstrate that *D. magna* has the complete toolkit for DNA methylation. Homologs of genes encoding the enzymes of the one-carbon cycle were identified in the *D. magna* genome as well as the TETs enzymes involved in the active demethylation pathway. The profile of DNA methylation follows the sporadic distribution described for invertebrates, presenting an increase in DNA methylation for gene bodies, mainly at the beginning of the genes.

The expression of transcripts for the enzymes from the one-carbon pathway, especially DNMTs, changes with ageing. The overall distribution of DNA methylation also suggests age-

specific changes; however further analyses need to be done to identify the regions of the genome where DNA methylation is altered.

Therefore, based on the presented results, *D. magna* could represent a good model organism for epigenetic studies. However, to achieve full use of the information provided by this organism an improved genome has to be constructed. It will also be important to define the strain to be used and age of the exposed animals.

It is proposed that, in contrast to studies on the genome that use inbred strains (Xinb3, Iinb1), epigenomic studies should be performed in non-inbred strains (i.e. Bham2). Besides the possible deleterious effects of inbreeding (Charlesworth and Willis, 2009), the differences in global methylation found for inbred strains (up to 0.52%) and the strain used in this study (0.14%) could indicate an alteration of the DNA methylation profile when comparing strains. Regarding the age of the organisms, we have demonstrated that the expression of transcripts encoding the enzymes on the one-carbon pathway already vary with age. Likely, the DNA methylation profile can also vary. Therefore, it is important to establish a standard approach for exposures and analysis of the effects on DNA methylation profile.

Chapter 4

Age-related and stress-induced Differentially Methylated Regions (DMRs) in *Daphnia magna*

4.1   Introduction

DNA methylation studies in *Daphnia* have shown the great potential of these organisms as an environmentally relevant invertebrate model species for epigenetic studies. The results in chapter 3 describe the "normal" status of DNA methylation in *Daphnia magna*. Also, age-related changes in the overall DNA methylation status and in expression of genes involved in the DNA methylation process were reported.

It is recognised that the "normal" status of DNA methylation may fluctuate to some extent but, having this as a starting point, the analysis of stress-induced changes was the subsequent analysis to be performed. The aim of the work reported in this chapter was to identify, at a single nucleotide resolution, the alterations in DNA methylation between control and stressor-exposed groups.

Sodium bisulfite treatment is the standard method for DNA methylation analysis at a single nucleotide resolution, although it presents some limitations as shown in section 2.7. When DNA is treated with sodium bisulfite, the unmethylated cytosines are converted to uracils while the methylated cytosines are not converted. After several PCR amplifications, the methylated cytosines remain as cytosines and the unmethylated ones appear as thymines (Frommer *et al.*, 1992).

Bisulfite treatment can be coupled with whole genome high throughput sequencing allowing the analysis of the methylation status across the entire genome at single nucleotide-resolution (Cokus *et al.*, 2008). This is an extremely powerful approach for achieving high resolution DNA methylation profiling and detailed analysis of distribution of DNA methylation across various sections (e.g. regulatory regions, intergenic and intragenic regions) of the genome. The

comparison of DNA methylation profiles of two or more different sample types will result in identification of Differentially Methylation Regions (DMRs).

Differential methylation is often associated with diseases or disorders, moreover it has been identified to contribute to tissue-specific gene expression (Slieker *et al.*, 2013; Song *et al.*, 2009; Wan *et al.*, 2015). As detailed in section 1.2.4, the relationship with gene expression is not simple as previously believed. The context where DNA methylation occurs appears to affect gene expression either negative or positively (Wan *et al.*, 2015). Therefore, DMRs are part of the normal DNA methylation pattern and contribute to biological diversity, even within cell types. Thus, problems can arise when the established normal DNA methylation pattern is affected in one or more cell type of the organism.

Abnormal DNA methylation is associated with the development and/or progression of several diseases, such as cancer and neurological disorders (Bird, 2002; Jones and Baylin, 2002; Jones, 2012). Most importantly, it has been identified that several of the differentially methylated regions can be used as biomarkers of exposure to certain stressors as well as for early detection and monitoring of the progression of diseases (Mikeska and Craig, 2014). Thus, DNA methylation profiling and identification of DNA methylation biomarkers and DMRs that are indicative of a stressor category can prove to be a useful resource.

Different approaches can be used to identify DMRs. Several statistical methods and software packages have been developed to compare and detect regions with altered DNA methylation (Rackham *et al.*, 2015). Consequently, it is extremely important to select the most appropriate statistical approach based on the experimental design and the model organism in order to identify and analyse DMRs.

According to Rackham *et al.* (2015), the choice of software package for DMRs analysis is a difficult task since it is difficult to judge the differences between the methods as there are no benchmarks for direct comparison. For the work presented in this thesis, the SeqMonk software was used for identification of DMRs (www.bioinformatics.babraham. ac.uk/projects/seqmonk/). SeqMonk, developed by the Babraham institute, is a research tool that allows the visualisation of the data and can be used to analyse the mapped DNA methylation high throughput sequencing data. It allows the use of custom genomes and annotation tracks, allowing analyses of non-annotated, or incomplete, genomes for DNA methylation studies. Another advantage is that data can be directly imported from Bismark outputs (Chatterjee *et al.*, 2012).

In SeqMonk software the first step in quantifying the level of DNA methylation is to define probes. Probes can be defined as either regions with different length (*e.g.* genes, CGIs, running windows) or single cytosine sites, set around predefined annotation tracks (e.g. mRNA, CDS, CpG islands), or unbiased (i.e. entire genome). When annotation tracks are used, the analysis is considered to be "biased", since the comparison will be done in specific regions that can have a different CpG composition and distribution. When the analysis is "unbiased", every region with enough coverage will be analysed.

Despite the majority of studies focusing on methylation pattern around promoters or CGI, problems can arise with the use of biased methods for DNA methylation analysis. Firstly, CpGs are not randomly distributed across the genome, as evidenced by regions containing higher concentration of CpGs, termed CpG islands. DNA methylation does not appear to be randomly distributed either. Also, the previous knowledge about DNA methylation distribution and profiling, mostly built based on mammalian studies, suggested a simple relationship of

presence of methylation and transcription repression (Wan *et al.*, 2015). In fact, DNA methylation at several sites has been positively correlated with transcription activation and these sites are sometimes located far from the responding gene (Irizarry *et al.*, 2009). Therefore, the use of unbiased genome-wide analyses, without the focus on promoter regions or CGI, can help to elucidate the effects of DNA methylation on gene expression.

Based on what was discussed previously, the aims of chapter 4 are: i) To test different methods of DMRs identification (biased *vs.* unbiased methods) using the visualisation software SeqMonk. To achieve this aim, *Daphnia* exposed to 5-azacytidine and respective controls were used for analysis; 5-azacytidine is a chemical with known demethylation effects and therefore it can be used as a positive control; ii) To identify DMRs induced in *Daphnia* as a result of exposure to 5-azacytidine (3.7 mg L^{-1}, 5 days exposure), arsenic (100 µg L-1, 14 days exposure) and hypoxia (<2 mg L^{-1} of dissolved oxygen, 14 days exposure) using the selected method of analysis in aim (i).

4.2   Overview of experimental design

The data used in this chapter were obtained after high throughput bisulfite sequencing of the DNA samples extracted from *Daphnia* exposed to three different stressors, arsenic, hypoxia, 5-azacytidine and respective controls (n=3). The stressors and concentrations were defined based on literature research and previous studies with *Daphnia* (details in sections 1.4 and 2.3.2). The DNA was extracted using CTAB method as described in section 2.4.1.1. Samples were bisulfite treated using EZ DNA methylation gold kit (see section 2.7.1). The method for library construction, quality control, quantitation and sequencing on Illumina HiSeq 2500 are described in sections 2.7.2, 2.7.3 and 2.7.4.

The sequencing data were subjected to quality checks and were aligned to the *D. magna de novo* assembled genome generated in Chapter 3. Alignment and methylation call were performed with the software Bismark (Krueger and Andrews, 2011). After, methylation data were visualised and analysed on SeqMonk.

The workflow for analysis of DNA methylation profiles after WGBS is described in Figure 4.1.
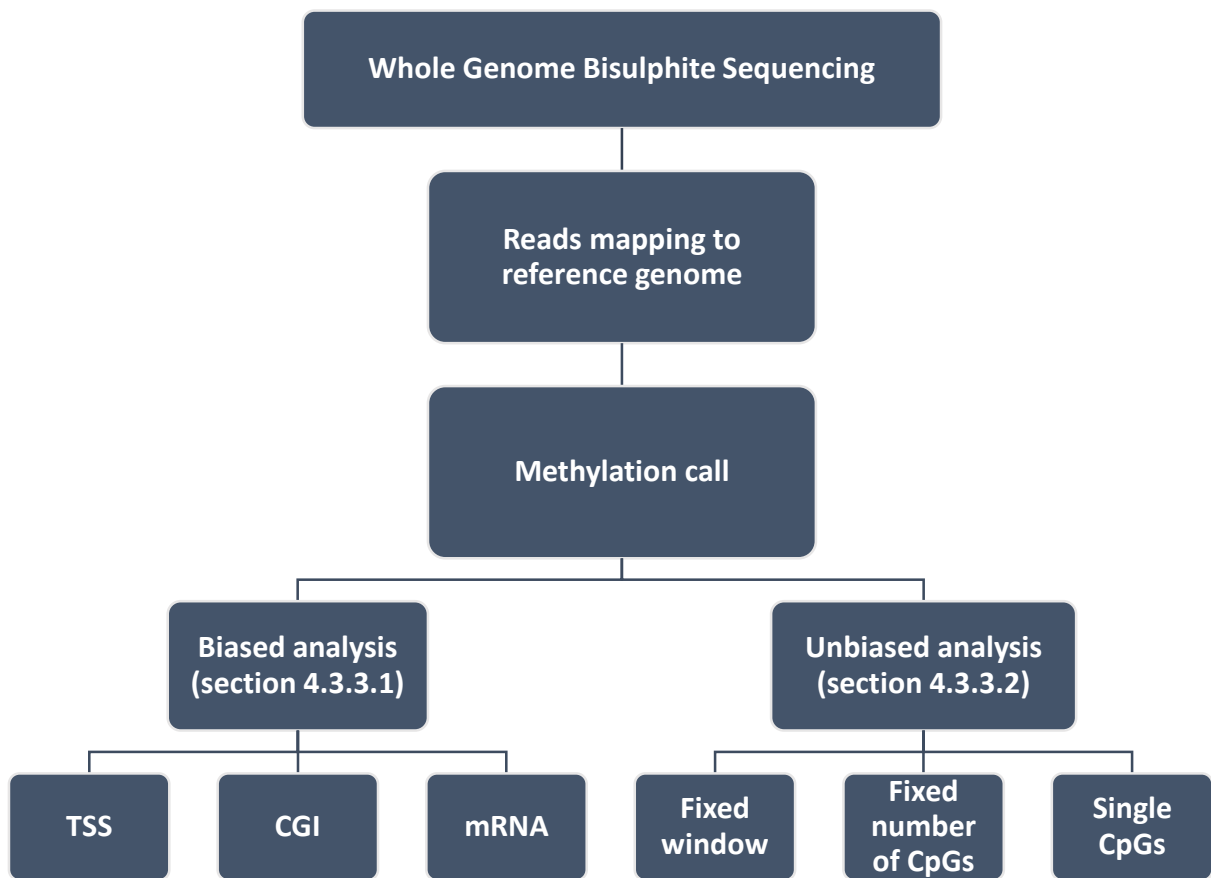
Figure 4.1 Workflow of the analysis performed in chapter 4. Comparison of unbiased and biased method of DNA methylation quantification and identification of Differentially Methylated Regions (DMRs).

4.3   Results

After the mapping (section 3.3.5) of the sequencing reads the analysis can either be done visually or using a software package for DMRs identification. To identify stress-induced DNA methylation changes the methylation status was compared between the non-exposed (control) and exposed (treated) groups at determined loci or regions, and when statistically different, they were characterised as a DMRs.

Ideally, every cytosine should be compared between control and treated groups. However, the high number of comparisons make this very difficult or impossible, both in terms of computational resources and statistical power, since an enormous amount of tests will be performed. This requires a multiple testing correction, and due to the high number of statistical tests, it is subject to a large number of false negatives. Therefore, a reliable method of analysis needs to be used in order to extract maximum information, without compromising the quality of the data.

4.3.1 Read mapping and DNA methylation call

Read mapping and methylation calls were performed using the software Bismark. After quality control and trimming of the poor quality reads, the sequence reads were aligned to the *de novo* assembled genome for *Daphnia magna* Bham2. DNA methylation calls were performed generating a list containing all covered cytosine sites with the methylation status for each site represented either as a '+' for methylated or '-' for unmethylated.

4.3.2 Pre-treatment of DNA methylation reads

Sample files were loaded to SeqMonk software for visualisation. Similar to what was observed in chapter 3, section 3.3.5.2, some regions presented a high number of reads conferring an abnormal coverage when compared to adjacent regions. For the analysis presented in chapter 3, it was possible to identify which scaffolds contained the unusually high number of reads, due to the methylated sites presented in non-CpG context (a rare event in invertebrates). Using this information, the scaffolds were removed from the final analysis of the normal profile of DNA methylation.

However, since only the reads in CpG context were used for the DMRs identification a different approach needs to be used. To remove the regions with high coverage a filter was applied for outliers. Firstly, the probes were defined with the window size of 3kb. Then, they were quantified with read count and corrected with the total read count (Figure 4.2). Probes were then filtered using box whisker test for outliers using a stringency of 10 above the median for at least one of the data stores. The identified outliers were converted to an annotation track and removed from the data.
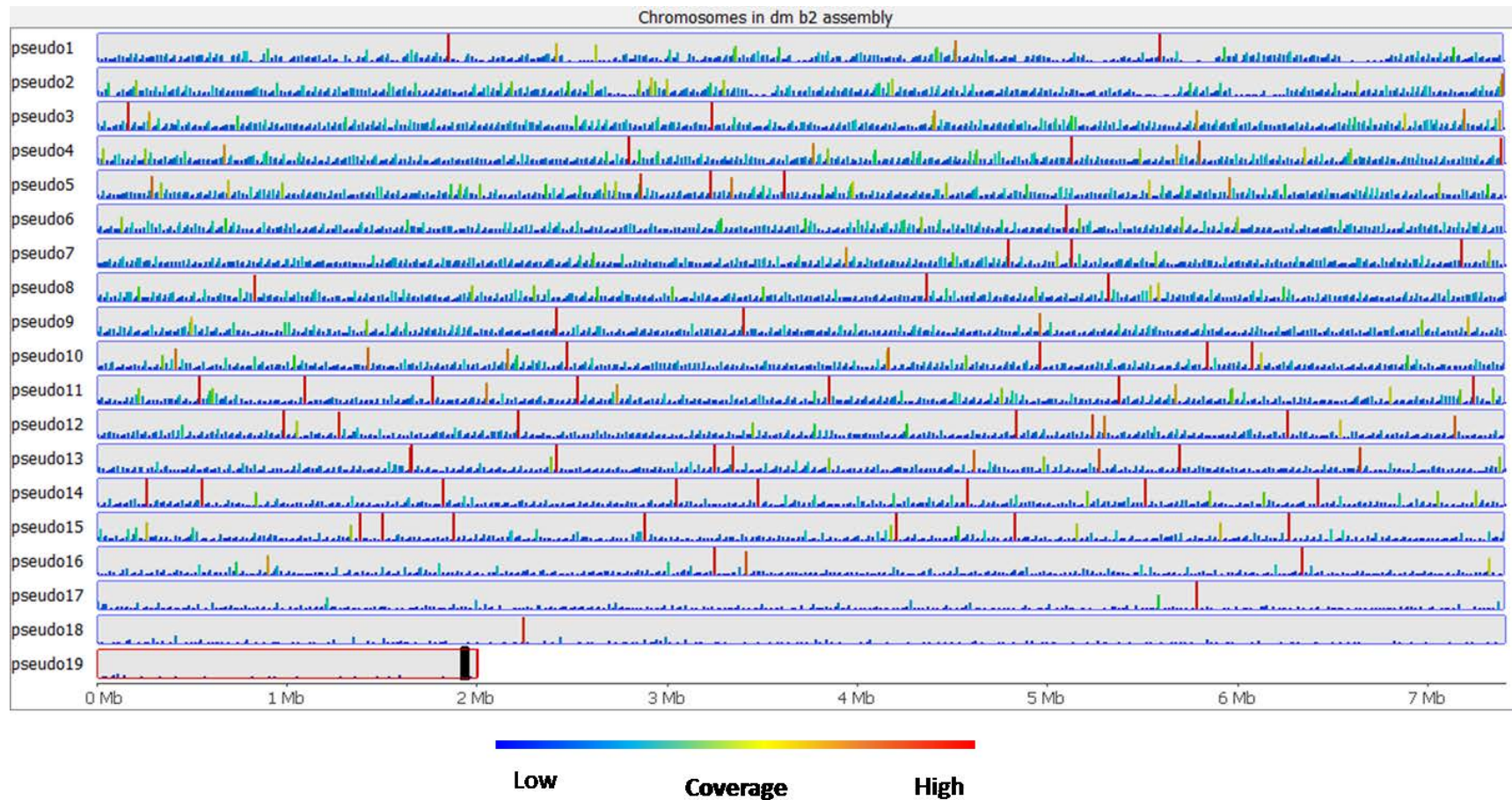
Figure 4.2 Read count quantitation over probes with 3kb in length for one replicate of control group. Only reads in CpG context were used. Coverage is shown with colour scheme ranging from blue to red. Red bars indicate probes with abnormally high coverage compared to adjacent regions. Abnormal regions were removed with box whisker test for outliers using a stringency of 10 above the median for at least one of the three replicates.

136

4.3.3 Methods for identification of Differentially Methylated Regions (DMRs)

A DMR is defined as a region with significant difference in methylation levels between two or more samples (Rijlaarsdam *et al.*, 2014). In order to identify which method of DMRs identification was the best for this experimental design, six different approaches were taken. They are divided into 'biased analyses', where the regions are set based on an annotated track, and 'unbiased analyses', where the regions are compared across the genome without the use of any annotation track to set the regions (see Figure 4.1).

Independent of the method used for identification of DMRs, the methylation levels for the sets of probes were quantified using the 'bisulfite methylation over features' pipeline available on SeqMonk. This pipeline measures the methylation levels of individual cytosines. If probes were set containing more than one cytosine, the DNA methylation value is presented as the average value.

4.3.3.1   Biased analyses

4.3.3.1.1 Probes over transcription start sites

Based on the annotation tracks available for the *Daphnia magna* Bham2 genome, an option was to set the methylation quantitation around the transcription start sites (TSS), since methylation alteration is often described around the promoter regions and first exon of the genes and controls transcription (Brenet *et al.*, 2011; Suzuki and Bird, 2008).

The positions for TSS described in section 3.3.4.4 were loaded to SeqMonk. The probes for methylation quantitation were set around TSS with 1kb upstream and 1kb downstream to account for promoter and first exon region (Figure 4.3).
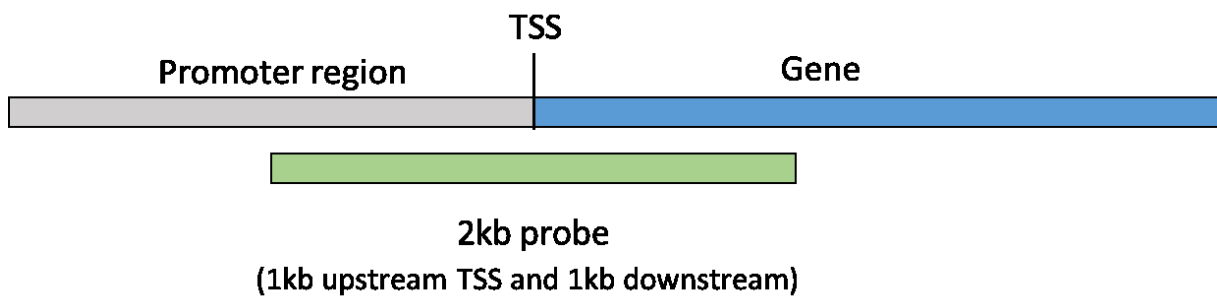
Figure 4.3 Schematic representation of TSSs location and the probes set around it. Gene is represented in blue and promoter region in grey. 2kb probe is represented in green with the TSS located in the middle of the probe.

25,820 probes were set around the TSS, the methylation was quantified and the 2 kb regions were compared between control and treatment using replicate statistical test (Figure 4.4). The replicate statistical test looks for a consistent effect across the biological replicates. It uses a t-test to assess whether a set of replicates (exposed group) shows a significant difference to the other (control group). With the p-value of < 0.1, 2,094 probes passed, while with a p-value of 0.05 and 0.01, the number of probes different probes were 1,051 and 245, respectively. However, no probes passed the multiple testing correction (p<0.1). Therefore, using this approach no region was identified as a DMR.
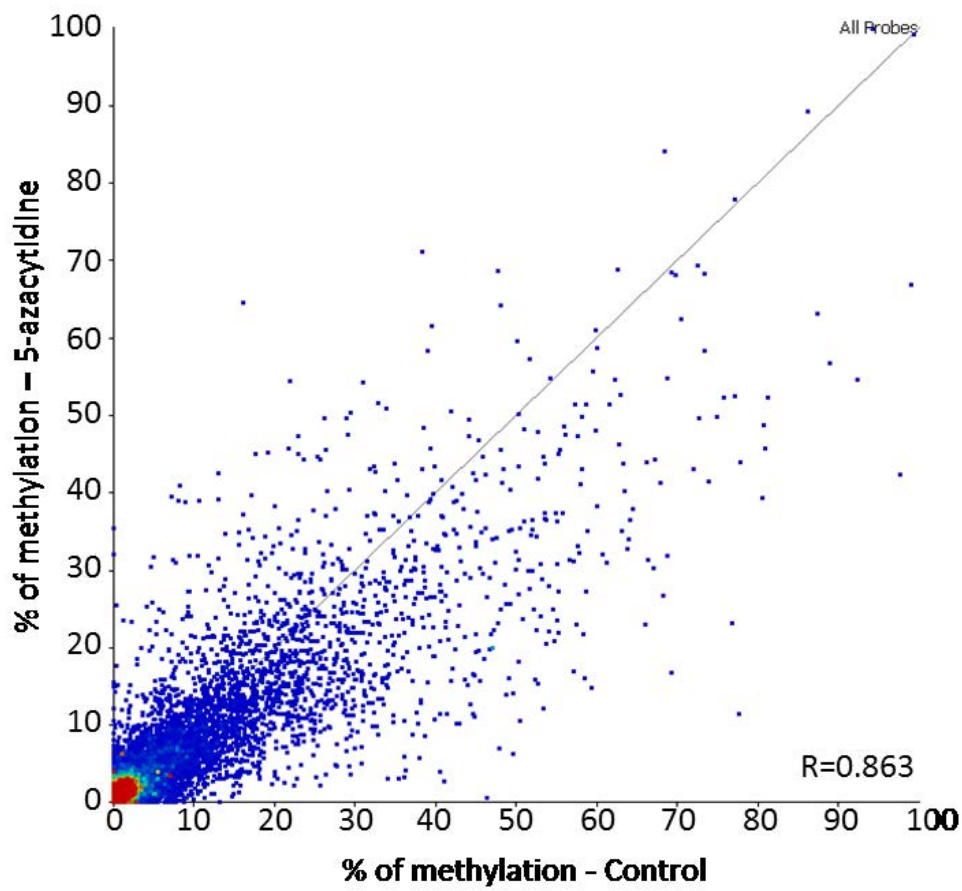
Figure 4.4 Scatter plot of probes set around TSS. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis and 5-azacytidine in the y axis.

4.3.3.1.2 Probes over CpG islands

For this analysis the probes were set on CpG islands (CGI). CGI detection is detailed in chapter 3, section 3.3.4.3. The total number of probes is equal to the number of identified CGIs, therefore the number of probes is 30,600. Lengths of the probes are variable due to the differences in CGI length. They vary from 200 to 3,606 bp but with the majority (82%) being between 200 and 500 bp (Figure 4.5).

Probes were tested with replicate statistical tests (t-test comparing control and treatment) and no probes passed the multiple test correction. In general, methylation level was very low

for the probes analysed (Figure 4.6). This is due to the averaging of the values per probe, which

was affected by the large size of the probes. Thus, the quantitation of methylation over CGI

was not a good option for this study.

Figure 4.5 Length of probes set over CGIs. Number of probes in the y axis, length in bp in the x axis.

Figure 4.6 Scatter plot of probes set over CGI. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis and 5-azacytidine in the y axis. R=0.876.

4.3.3.1.3 Probes over genes

The *de novo* assembled genome of *Daphnia magna* Bham2 was annotated against the publicly available 29,121 primary transcripts (finloc9b) from the *Daphnia* genome v2.4 database (http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna/), as shown in section 3.3.4.5. In total 5,831 genes were identified. Probes were set to cover the entire gene (Figure 4.7) and used for quantification of DNA methylation level.

141
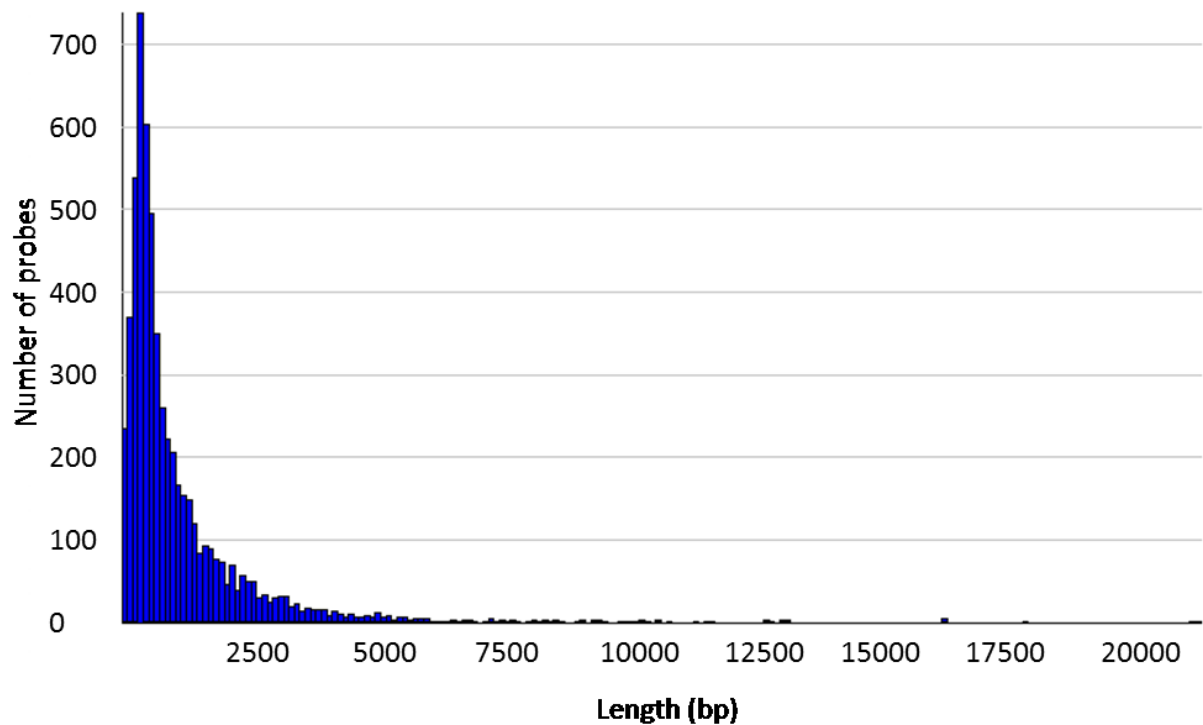
Figure 4.7 Length of probes set over annotated genes. Number of probes in the y axis, length in bp in the x axis.

In general, the methylation levels of the majority of the probes in the 5-azacytidine treatment group compared to the control group were decreased (Figure 4.8). However, this trend was not statistically significant after applying the Benjamini and Hochberg multiple test correction. This is mainly caused by the fact that regions defined as probes in this setup cover a large section of the genome with multiple CpG sites. Although the methylation levels of some of the CpG sites may have been affected and altered as a result of the treatment, these changes can be averaged out by large sections in the probe with no change between the control and treatment. Therefore, it is thought that defining very large regions as probes can often mask truly important changes at certain CpG sites.
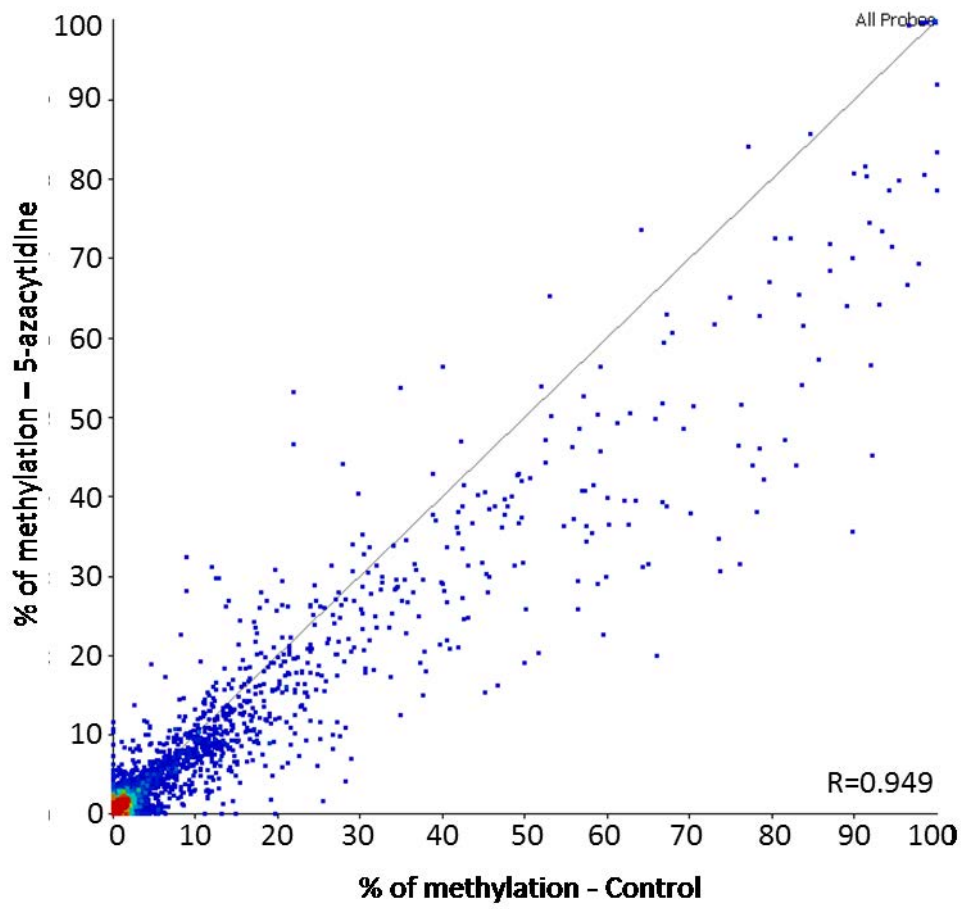
Figure 4.8 Scatter plot of probes set over annotated genes. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis and 5-azacytidine in the y axis. R=0.949

4.3.3.2 Unbiased analyses

For the unbiased analyses the regions were set without the use of annotation tracks. Three methods were tested: i) fixed windows; ii) fixed number of CpGs; iii) Single CpGs applying running window statistical test.

4.3.3.2.1 Fixed window

For the fixed window analysis, the probes were set as fixed windows with 100 bp in length at regions with detectable CpG sites. A total of 1,352,514 probes were set for the analysed samples (Figure 4.9). The set probes presented a very uneven coverage along the genome, caused by the unequal distribution of CpG sites across the genome. Therefore, the set probes containing a high number of CpG sites will have a larger averaging effect than regions with few cytosine sites causing problems in downstream analyses. Nevertheless, the replicate set statistical test was applied to search for statistically different regions. The p-values of 0.1, 0.05 and 0.01 were considered.

Applying p-value cut-off points of less than 0.1, 0.05 and 0.01, the number of probes that were identified as statistically significant were 19,443, 10,631 and 3,107, respectively. However, only 9 probes passed the multiple testing correction at a p-value of 0.1. Therefore, very few regions were identified as DMRs using this design of analysis.

Figure 4.9 Scatter plot of probes set with fixed length of 100 bp. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis and 5-azacytidine in the y axis. R=0.907

4.3.3.2.2 Fixed number of CpGs

One option to overcome the problem of uneven coverage detected when using fixed length of probes, is to define probes as regions with a fixed number of 100 CpGs. In this approach the probes will differ in length but will contain same number of CpGs, although the number of reads per probe can differ (Figure 4.10).

As demonstrated in figure 4.11, a trend of decreasing methylation prevalence was observed in the treatment group compared to control group with 8,244, 4,183 and 946 probes identified as hypomethylated in the treatment group with p-values less than 0.1, 0.05 and 0.01, respectively. However, the identified probes did not pass the Benjamini and Hochberg multiple testing correction.
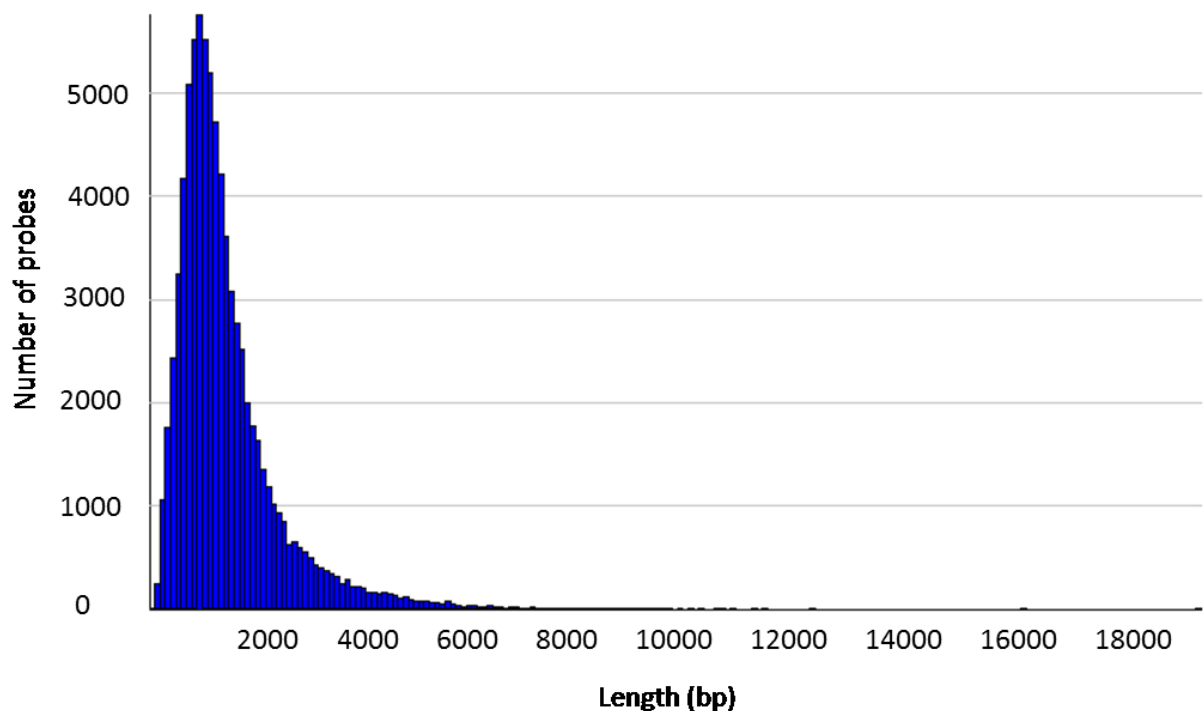
Figure 4.10 Length of probes set over probes with a fixed number of 100 CpGs. Number of probes in the y axis, length in bp in the x axis.
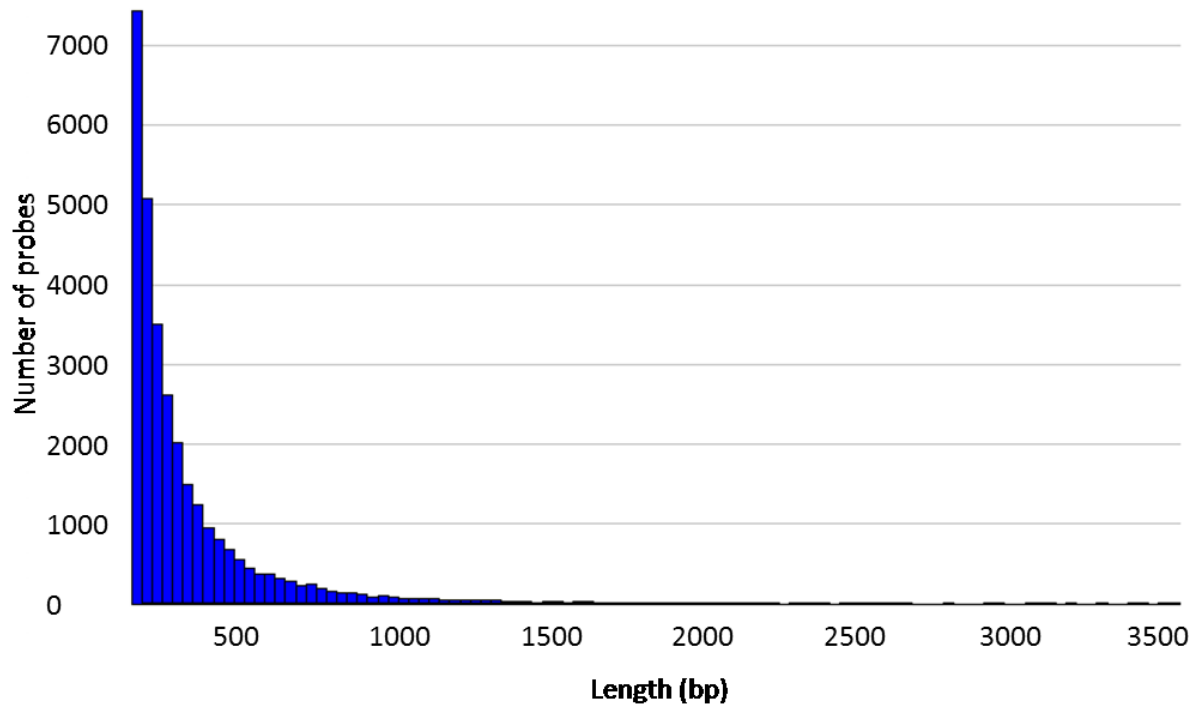
146
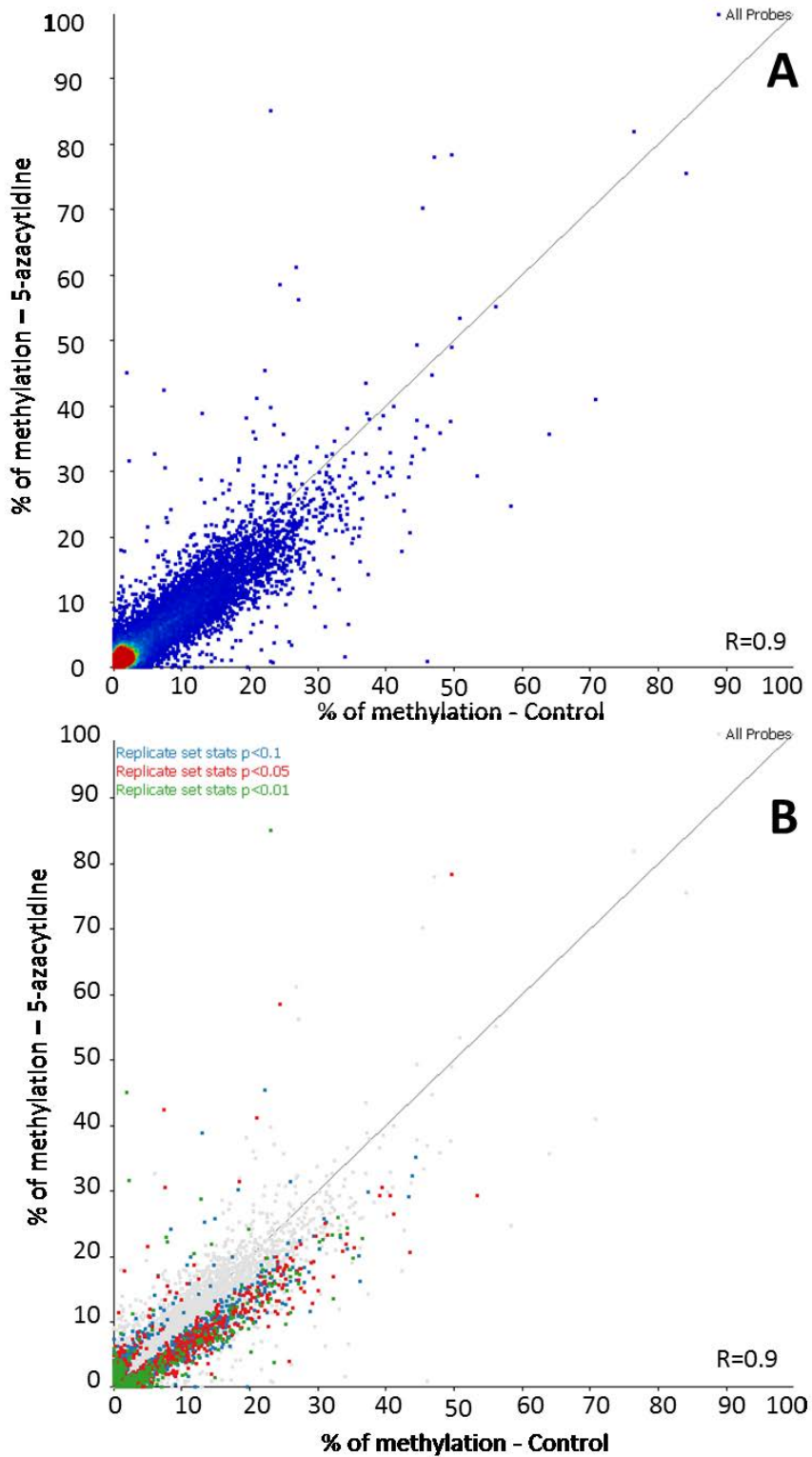
Figure 4.11 Scatter plot of probes set with fixed number of CpGs. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis an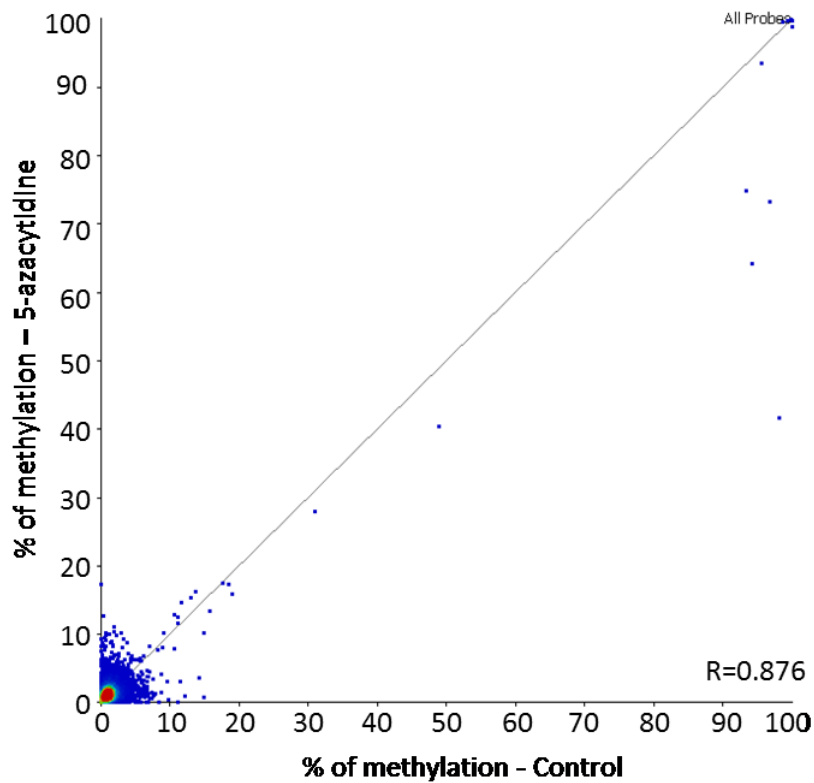d 5-azacytidine in the y axis. A) All the probes are plotted for control and 5-azacytidine. B) Probes that passed the replicate statistical test are highlighted (blue p<0.1; red p<0.05; green p<0.01) (no multiple test correction was applied). R=0.9

4.3.3.2.3 Single CpGs with windowed statistical test

As discussed previously, ideally the identification of differentially methylated regions should be achieved by comparing the methylation levels of every single cytosine base between the different experimental conditions. However, it is impossible to collect enough data, for several replicates or in sequencing depth, to analyse each cytosine and have p-values that can survive the multiple testing correction since the p-value is adjusted based on the number of tests performed (Noble, 2009).

Thus, one way to overcome this problem is to set the probes for each cytosine position but instead of comparing the cytosines individually, the comparison can be set within windows (Baumann and Doerge, 2014). Data interpretation using windows compared to individual locations increases statistical power, simplifies computational resources, reduces sampling noise, and reduces the total number of tests performed (Beissinger *et al.*, 2015).

The window size is an arbitrary choice and will have an effect on the results and outcome interpretation. Larger windows will have more data and more statistical power, generating low p-values. On the other hand, the averaging effect will be higher. Small windows will generate good resolution for better interpretation of the biological significance of the results, however will include lesser observations and will have higher p-values. In the end, it is a trade-off between statistical power and more specific biological effects identification (Beissinger *et al.*, 2015). Therefore, all cytosine sites with 3 or more reads were set as probes and the comparison was performed with windows of 100 bp (Zhong *et al.*, 2013). The probes where filtered to include only the probes present in both groups, with a total of 1,217,496 probes. The 'windowed replicate test' with Benjamini and Hochberg multiple test correction (p-value <0.1) resulted in identification of 6450 differentially methylated probes between treatment

and control groups. Furthermore, 5434 and 3676 probes were differentially methylated in the treatment group with p-value < 0.05 and 0.01, respectively (Figure 4.12).

This method was chosen to be used for the further analyses in this study. It was able to identify differentially methylated cytosines when applying a conservative strategy and p-value, even after FDR. Probes still need to be grouped into DMRs, nonetheless it is done by simply combining the adjacent probes presenting concordant methylation levels. DMRs methylation level is expressed as the average value of each cytosine within the region.
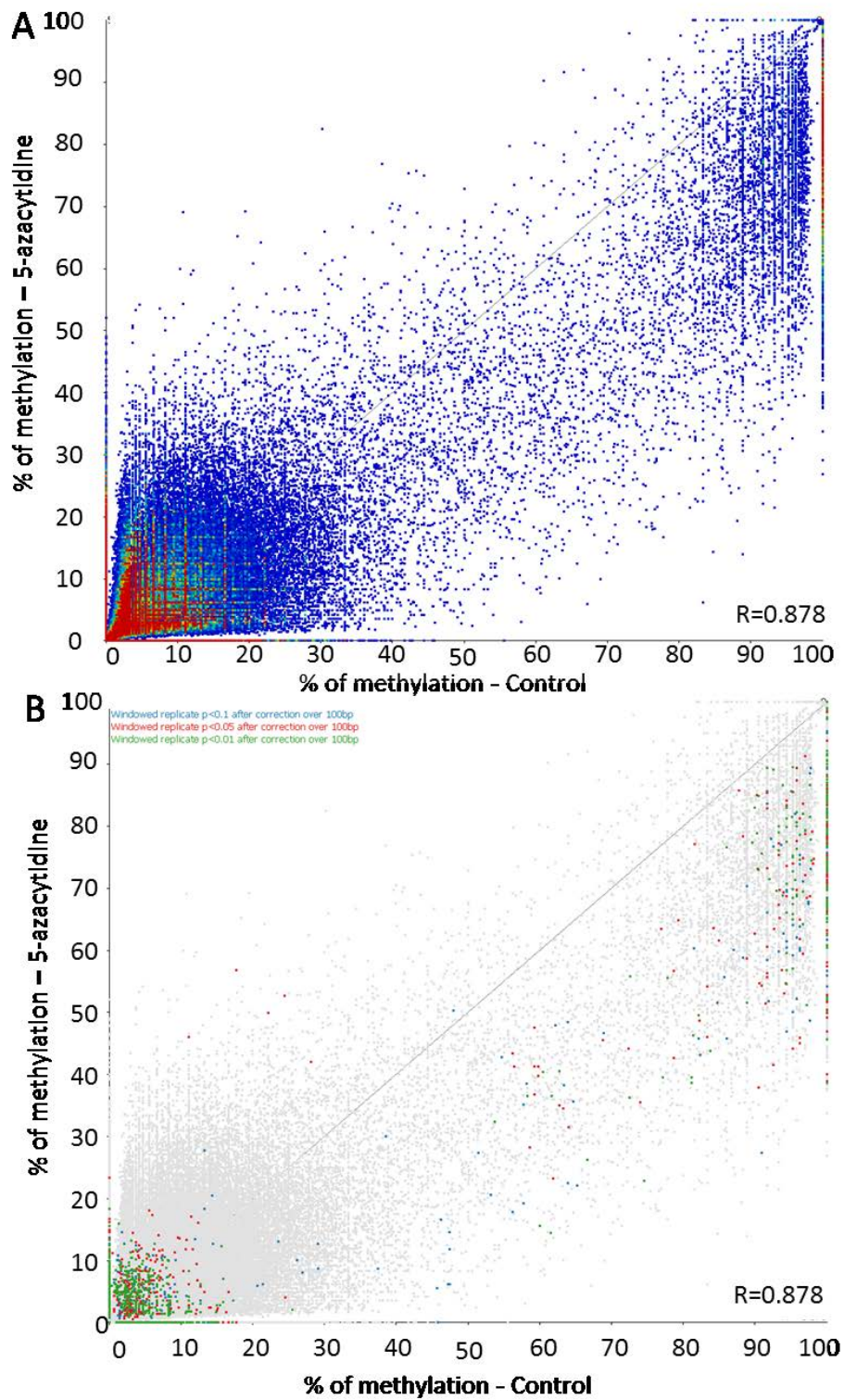
Figure 4.12 Scatter plot of probes in single CpGs. Probes set for 3 replicates for each control and 5-azacytidine exposure. Control probes in the x axis and 5-azacytidine in the y axis. A) All the 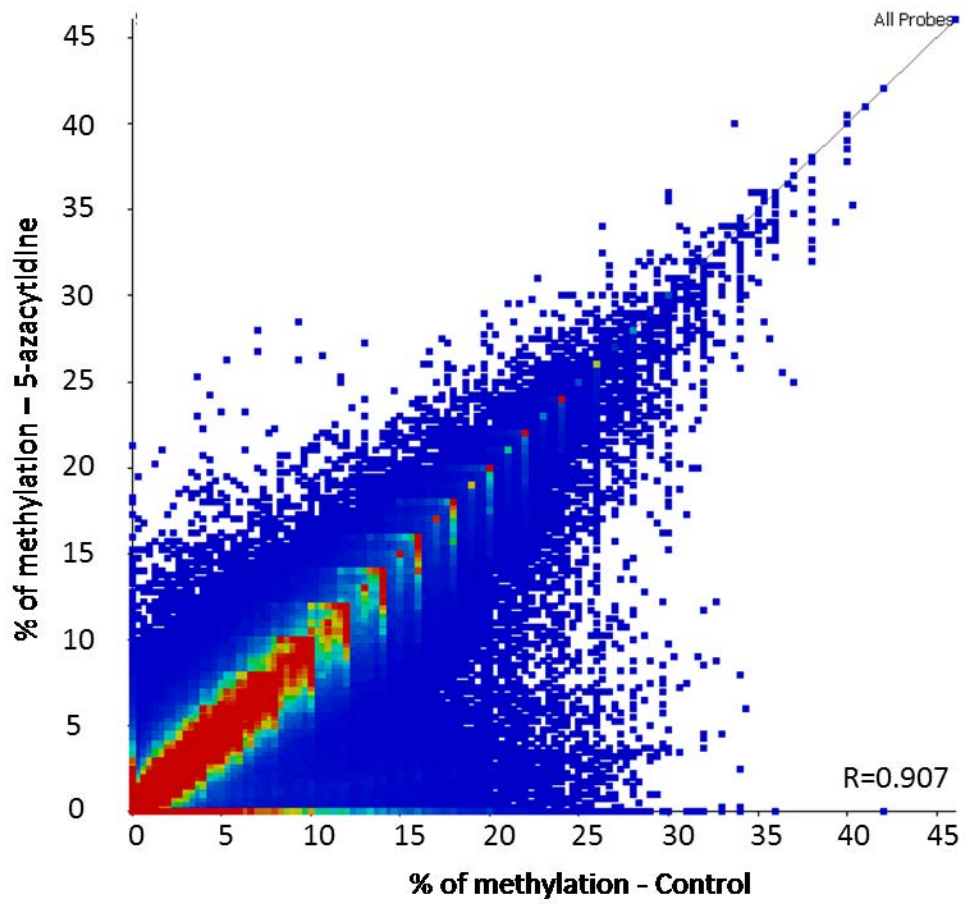probes are plotted for control and 5-azacytidine. B) Probes that passed the windowed replicate statistical test with Benjamini and Hochberg multiple test correction are highlighted (blue p<0.1; red p<0.05; green p<0.01), R=0.878.

4.3.3.3   Determining the appropriate DNA methylation level cut-off for bisulfite treated invertebrate samples

As demonstrated in chapter 3, section 3.3.5.2, sequencing artefacts are present in these data due to bisulfite treatment, the sequencing procedure and alignment. Consequently, it is necessary to determine the levels of false positive methylated sites, and assess the effects on DMRs.

Despite a high bisulfite conversion efficiency (section 2.8.6, Figure 2.4), the inner characteristics of the WGBS analyses can lead to mismapping of sequencing reads and miscalculations of the methylation levels since it is calculated by the ratio between methylated and unmethylated reads.

In chapter 3, false positive methylated sites were eliminated by selecting an arbitrary methylation level greater than 50% coupled to the use of biological replicates. The value of 50% was selected based on the coverage (minimum of 3 reads). This way, one methylated read present on a non-methylated site (possibly a sequencing artefact) will not call that cytosine as methylated (false positive).

For DMRs identification, biological replicates were already used for statistical test. Probes listed as differentially methylated passed the statistical test showing that there is little divergence in DNA methylation level among biological replicates for those sites. However, no cut-off on the level of methylation was applied.

In order to determine the DNA methylation level to be used as a cut-off point for these samples, different regions were selected across the genome for direct bisulfite PCR. The regions contained variable number of probes with different levels of DNA methylation varying

from 100% to 0%. There is a low occurrence of regions with percentage of DNA methylation from 20% to 60%. Therefore, when available, specific primers were designed and the PCR products were sequenced to quantify the methylation levels for each CpG position. Methylation values for the cytosines in the region were averaged and the results are presented as percentage of methylation for the region.

Table 4.1 presents the value for each region obtained with WGBS and BSP. It was possible to observe that regions with high methylation level were consistent between WGBS and BSP results. However, regions with low methylation levels measured with WGBS were not confirmed with BSP. Therefore, these regions were considered as false positives.

Breiling and Lyko (2015), in the study comparing the methylation distribution of vertebrates and invertebrates, considered the cut-off point of 20% of DNA methylation as bisulfite conversion artefacts. Based on the measurements made, the cut-off value of DNA methylation level was set as 40%. This value is higher than reported previously; however, it was set according with the obtained results and in a conservative way to avoid misinterpretation of the effects of the stressors on the methylation profiles.

Table 4.1. Cut off value for methylation quantitation with WGBS and BSP.

Region ID	% 5mC WGBS	% 5mC BSP
C001	0.31	0
C002	4.84	0
C003	7.2	0
C004	8.54	0
C005	10.41	0
C006	10.41	0
C007	10.97	0
C008	11.11	0
C009	11.11	0
C010	12.31	0
C011	23.97	46.67
C012	41.52	68.20
C013	48.33	69.35
C014	49.00	69.16
C015	60.79	66.55
C016	94.89	100
C017	95.89	100
C018	96.21	100
C019	97.52	100
C020	97.71	100
C021	98.55	100
C022	99.52	100
C023	100	100

4.3.4 Age-related DNA methylation changes

After deciding on the most appropriate method to be applied to further DMR identification, two datasets were analysed, to investigate the age-related and stress-induced alterations on the DNA methylation profile.

For the age-related changes analyses, the dataset presented in chapter 3, section 3.3.5 was used. Previously it was used for an overview of DNA methylation profile in 5 days old and 14 days old *Daphnia*.

Probes were defined as single cytosines for sites containing at least 3 reads. Methylation was quantified using bisulfite quantitation pipeline. Then, probes were filtered to include only probes that were measured in both groups (5 days old and 14 days old). The windowed replicate test was applied, followed by Benjamini and Hochberg multiple test correction. 6955 probes were identified as statistically significantly different between groups (p-value < 0.01) (Additional File 4.1). After setting the cut-off value of DNA methylation level of at least 40% for one group, only 38 probes passed. Using the genes annotated to the Bham2 genome (section 3.3.4.5) no annotation was identified. The 38 probes were grouped into 9 DMRs.

The sequences were then extracted and blasted to the *D. magna* reference genome (available at: http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna/BLAST/). From the 9 DMRs, 6 were annotated to the *D. magna* genome. Results are presented in Table 4.2.

Table 4.2. DMRs detected for the different ages (5 days and 14 days) after use of 40% of DNA methylation as a cut-off point. DMRs were annotated to the *Daphnia magna* genome (version 2.4, geneset: finloc9b).

DMR	Name	Gene ID	Location for SeqMonk	Length	CpG count	%5mC 5 days	%5mC 14 days
1	60S ribosomal protein L7a (100%D)	Dapma7bEVm009851t1	Chrpseudo5:22729-22784	55	7	47.04	15.99
2	RING finger protein (100%T)	Dapma7bEVm012692t1	Chrpseudo6:535527-535567	40	3	97.14	100
3	Ubiquitin-conjugating enzyme E2 (95%D)	Dapma7bEVm011441t1	Chrpseudo7:7376618-7376649	31	3	95.99	100
4	Uncharacterized protein (98%D)	Dapma7bEVm028259t1	Chrpseudo8:5513286-5513310	24	3	42.93	60.73
5	no hits	no hits	Chrpseudo11:1094393-1094414	21	3	14.41	42.96
6	no hits	no hits	Chrpseudo11:1094580-1094591	11	4	13.61	42.35
7	no hits	no hits	Chrpseudo11:1094750-1094781	31	3	16.65	44.81
8	Uncharacterized	Dapma7bEVm026673t1	Chrpseudo12:2990773-2990859	86	6	53.91	72.02
9	40S ribosomal protein S4 (100%P)	Dapma7bEVm001645t1	Chrpseudo14:6498000-6498017	17	3	23.44	68.02

4.3.5 Stress-induced DNA methylation changes

Genomic DNA extracted from control and three treatment groups (arsenic, hypoxia and 5-azacytidine) were subjected to whole genome bisulfite sequencing (described in section 2.7) to obtain stressor-induced DNA methylation profiles. After conducting the initial quality control steps (described in 4.3.2), the sequences were mapped to the *de novo* assembled genome using the software Bismark.

Again, based on the results discussed in section 4.3.3, the best option for DMRs identification was to quantify the methylation for single cytosines and to compare the different samples using the running window statistical test. Therefore, this method was applied to the different sample groups (arsenic, hypoxia and 5-azacytidine) in order to identify the DMRs for each treatment group compared to the control group.

After applying Benjamini and Hochberg multiple test correction, a total of 9,834 unique probes were identified as statistically significantly different between treatment and control groups (p-value < 0.01) (Additional File 4.2). The majority of the identified DMRs were unique to each treatment category, however some were shared between 5-azacytidine and arsenic (24), 5-azacytidine and hypoxia (11) and arsenic and hypoxia (464) (Figure 4.13). These results indicate that the majority of effects are stressor specific and potentially linked to their different modes of action (MoA).
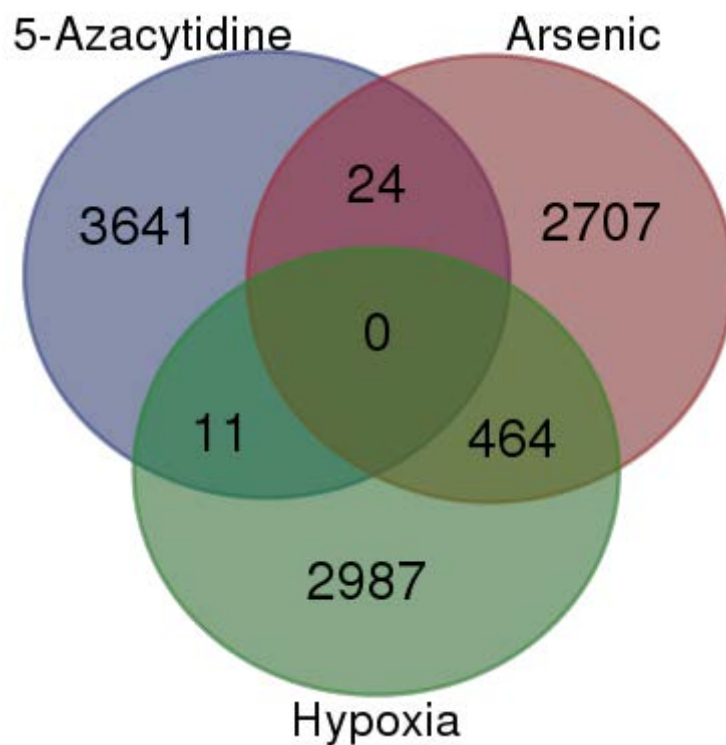
Figure 4.13 Venn diagram of stress-induced differentially methylate probes. Probes set for single CpGs, using 3 biological replicates for each treatment and respective control. Probes were identified using windowed replicate test with multiple test correction.

Figure 4.14 A and B demonstrates that the DMRs in the 5-azacytidine treatment group can be separated into two distinct groups of probes, some presenting less than 30% and other with methylation levels over 60%. Most of the probes have less than 30% methylation level and only a smaller proportion of the DMRs have methylation levels higher than 60%, at least for one of the groups, showing a bimodal distribution. The same did not apply to arsenic (Figure 4.14, C and D) and hypoxia (Figure 4.14 E and F), where the vast majority of the significantly different probes presented a low level of methylation (< 30%).
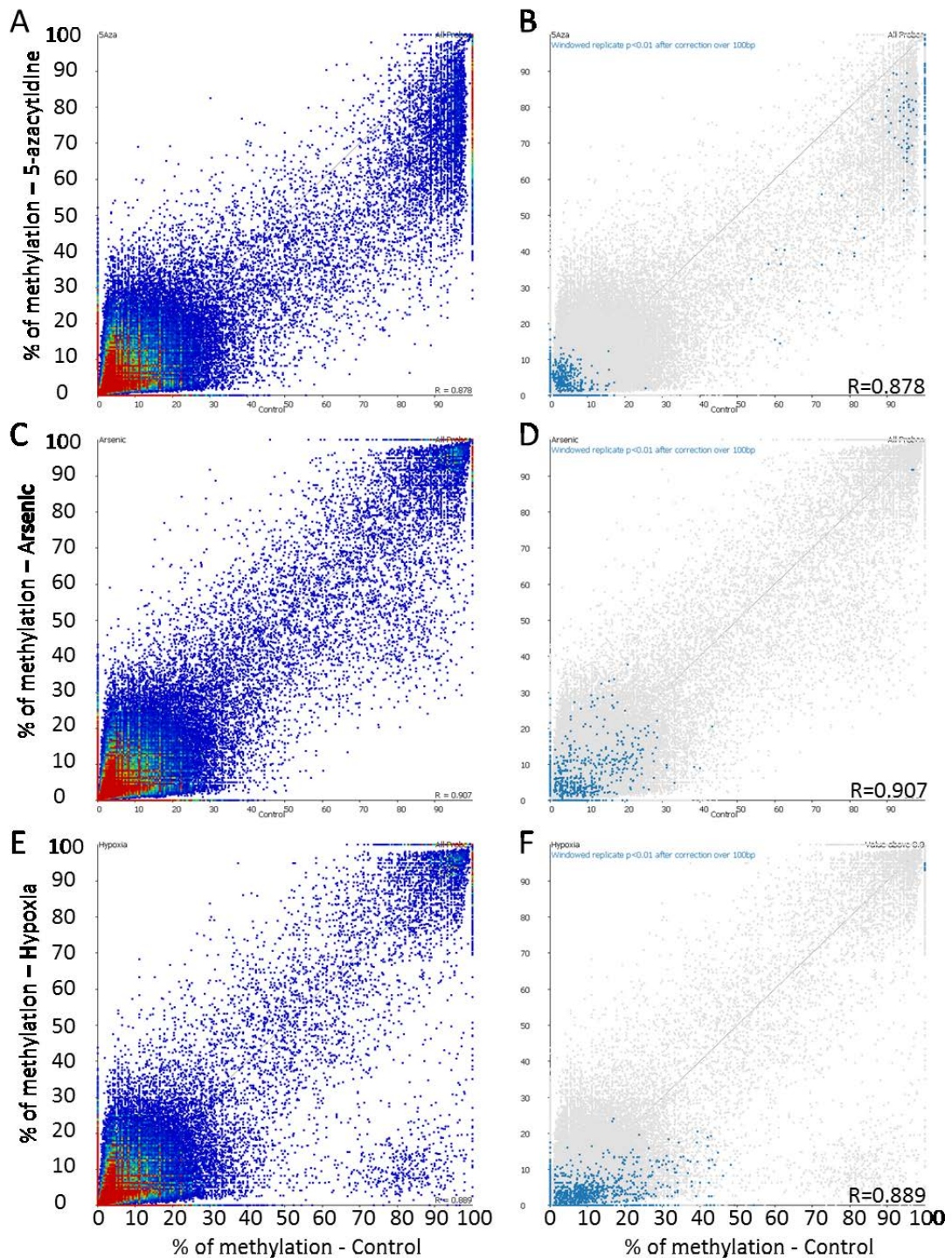
Figure 4.14 Scatter plots comparing exposures (5-azacytidine, arsenic and hypoxia) and controls. A) All probes measure for 5-azacytidine treatment and control; B) Probes that passed the windowed replicate test with multiple test correction p<0.01 for 5-azacytidine exposure; C) All probes measure for arsenic treatment and control; D) Probes that passed the windowed replicate test with multiple test correction p<0.01 for arsenic exposure; E) All probes measure for hypoxia treatment and control; F) Probes that passed the windowed replicate test with multiple test correction p<0.01 for hypoxia exposure;

4.3.5.1　　Differentially methylated CpGs grouping and annotation

After the identification of statistically significant different probes and using the established percentage of DNA methylation as a cut-off value, the probes, originally single CpGs, were grouped into DMRs. The criteria used were based on proximity and methylation value. Grouping of probes was performed manually and they were individually numbered.

Before applying the cut-off value, the 9834 differentially methylated CpGs were grouped to a total of 1622 DMRs for all the treatment groups. 607 for 5-azacytidine treatment, 505 for arsenic and 510 for hypoxia. However, when the 40% cut off value was applied, only 27 DMRs were detected (Table 4.3 and 4.4). Being 22 for the 5-azacytidine treatment, 1 for arsenic and 4 for hypoxia.

The 27 DMRs were blasted against the *D. magna* reference genome (available at: http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna/BLAST/) and regions were annotated to the closest transcript.

The majority of DMRs presented very low methylation percentage and could represent bisulfite conversion artefacts, represented by the number of DMRs excluded with the cut-off. Therefore, a confirmation of these results through gene-specific bisulfite sequencing is an important step in order to verify these results.

Table 4.3. DMRs detected for the three treatments after use of 40% of DNA methylation as a cut-off point.

Treatment	DMR	Location for SeqMonk	Length	CpG count	Control %5mC	Treatment %5mC
5-azacytidine	7	pseudo1:1100940-1101033	93	9	95.38	73.72
	8	pseudo1:1126959-1127051	92	6	79.29	44.93
	19	pseudo1:3827448-3827489	41	4	60.79	38.48
	26	pseudo1:4394907-4395003	96	8	97.50	77.32
	29	pseudo1:4566232-4566288	56	3	100	91.16
	62	pseudo2:3039863-3039938	75	5	100	67.55
	154	pseudo3:6175914-6176005	91	12	97.25	71.70
	193	pseudo4:2884026-2884121	95	9	96.55	78.46
	223	pseudo4:5935071-5935128	57	5	97.71	73.97
	305	pseudo6:1949624-1949722	98	10	94.90	79.83
	337	pseudo7:2426207-2426304	97	5	99.52	85.96
	341	pseudo7:2858392-2858477	85	6	96.22	74.51
	382	pseudo8:3547888-3547960	72	12	97.53	77.08
	422	pseudo9:1977525-1977612	87	4	95.89	71.89
	470	pseudo10:5051355-5051432	77	7	97.54	75.36
	487	pseudo11:2128683-2128782	99	8	93.21	51.73
	514	pseudo12:636836-637003	167	13	100	98.85
	566	pseudo14:4620314-4620374	60	6	64.84	24.63
	584	pseudo15:4043969-4044079	110	12	100	88.58
	591	pseudo16:1132424-1132484	60	5	98.89	59.05
	595	pseudo16:2774168-2774223	55	7	97.96	65.81
	607	pseudo18:1598357-1598427	70	6	98.10	80.99

Table 4.3. Continued from previous page

Arsenic	1085	pseudo15:3321858-3321879	21	4	96.82	91.67
Hypoxia	2010	pseudo1:428692-428765	73	5	42.14	13.01
	2128	pseudo3:1888486-1888575	89	6	40.68	13.48
	2176	pseudo4:1204565-1204603	38	3	100	94.81
	2398	pseudo10:3449455-3449537	82	3	100	93.33

Table 4.4. DMRs that were annotated to the *Daphnia magna* genome (version 2.4)

Treatment	DMR	Name	Gene ID
5-azacytidine	7	Serine/threonine-protein kinase NLK (80%D)	Dapma7bEVm003687
	8	Glucose dehydrogenase precursor (99%D)	Dapma7bEVm007219
	19	sp zinc finger transcription factor (64%H)	Dapma7bEVm011563
	26	WD repeat, SAM and U-box domain-containing protein	Dapma7bEVm002638
	29	Proteasome subunit alpha type-2 (100%D)	Dapma7bEVm004440
	62	Argonaute-2 (90%D)	Dapma7bEVm009641
	154	Protoporphyrinogen oxidase (100%R)	Dapma7bEVm010840
	193	Inturned (67%H)	Dapma7bEVm000179
	223	protein serine/threonine phosphatase (100%D)	Dapma7bEVm004815
	305	WD repeat and FYVE domain-containing protein (100%M)	Dapma7bEVm004992
	337	Host cell factor (57%H)	Dapma7bEVm010207
	341	Guanine nucleotide-binding protein G(s) subunit al..	Dapma7bEVm007254
	382	Cyclin-dependent kinase (100%R)	Dapma7bEVm015316
	422	Galactose-1-phosphate uridylyltransferase (100%P)	Dapma7bEVm006311
	470	-	no hit
	487	LIM and calponin domains-containing protein (100%D)	Dapma7bEVm000208
	514	-	no hit
	566	Uncharacterized	-
	584	Uncharacterized protein (77%P)	Dapma7bEVm017325
	591	Dynactin subunit (96%D)	Dapma7bEVm015177
	595	Uncharacterized protein (68%P)	Dapma7bEVm014899
	607	Uncharacterized protein (98%P)	Dapma7bEVm021583

Table 4.4. Continued from previous page

Arsenic	1085	Uncharacterized protein (100%P)	Dapma7bEVm021083
Hypoxia	2010	-	no hit
	2128	-	no hit
	2176	Tubulin alpha-1 chain (100%D)	Dapma7bEVm009691
	2398	Calcium-transporting ATPase type 2C member (94%H)	Dapma7bEVm004556

4.3.6 Gene specific bisulfite sequencing

In order to confirm the methylation values obtained with WGBS, regions were selected to be tested using gene specific bisulfite sequencing. In total 6 regions were tested for 5-azacytidine (19, 223, 337, 341, 382, 1085). Region 1085 and 337 were also tested for arsenic treatment and 337 was tested for hypoxia (Table 4.3). Methylation percentage was quantified using the relative peak height for each cytosine in the CpG context on the sequencing chromatograms. All regions tested for the 5-azacytidine treatment presented hypomethylated sites when comparing to control samples. Region 337 was annotated as a host cell factor gene and it was tested for the three treatments. For the control samples, the region was fully methylated. A decrease of 11.9% in methylation was observed for 5-azacytidine, while for arsenic and hypoxia no changes were observed (Figure 4.15).

For 5-azacytidine samples, DMR 19 identified as a zinc finger transcription factor presented a 6% decrease in methylation while region 223, annotated as protein serine/threonine phosphatase showed a decrease of 10.3% (Figure 4.16 and 4.17). The largest decrease in methylation was observed for DMR 382 (19%), annotated as cyclin-dependent kinase (Figure 4.18).

Region 341 presented a different distribution of methylation than the others tested. In total, 9 CpGs were covered during gene specific sequencing, however, only 4 were methylated, and affected by 5-azacytidine exposure (decrease of 11%) while 5 were completely demethylated for both control and treatment samples (4.19).

Region 1085 was sequenced for both 5-azacytidine and arsenic groups. Despite being annotated to the *D. magna*'s genome it was likely composed of bacterial DNA as only one

methylated site was found in a CHG context. With this possibility raised, homology search was

performed and the sequence for the transcript Dapma7bEVm021083 presented similarities to

several Proteobacteria sequences, including *Variovorax* sp., known to occur as symbionts in

Daphnia (Qi *et al.*, 2009). Therefore, this DMR was eliminated from further analyses.
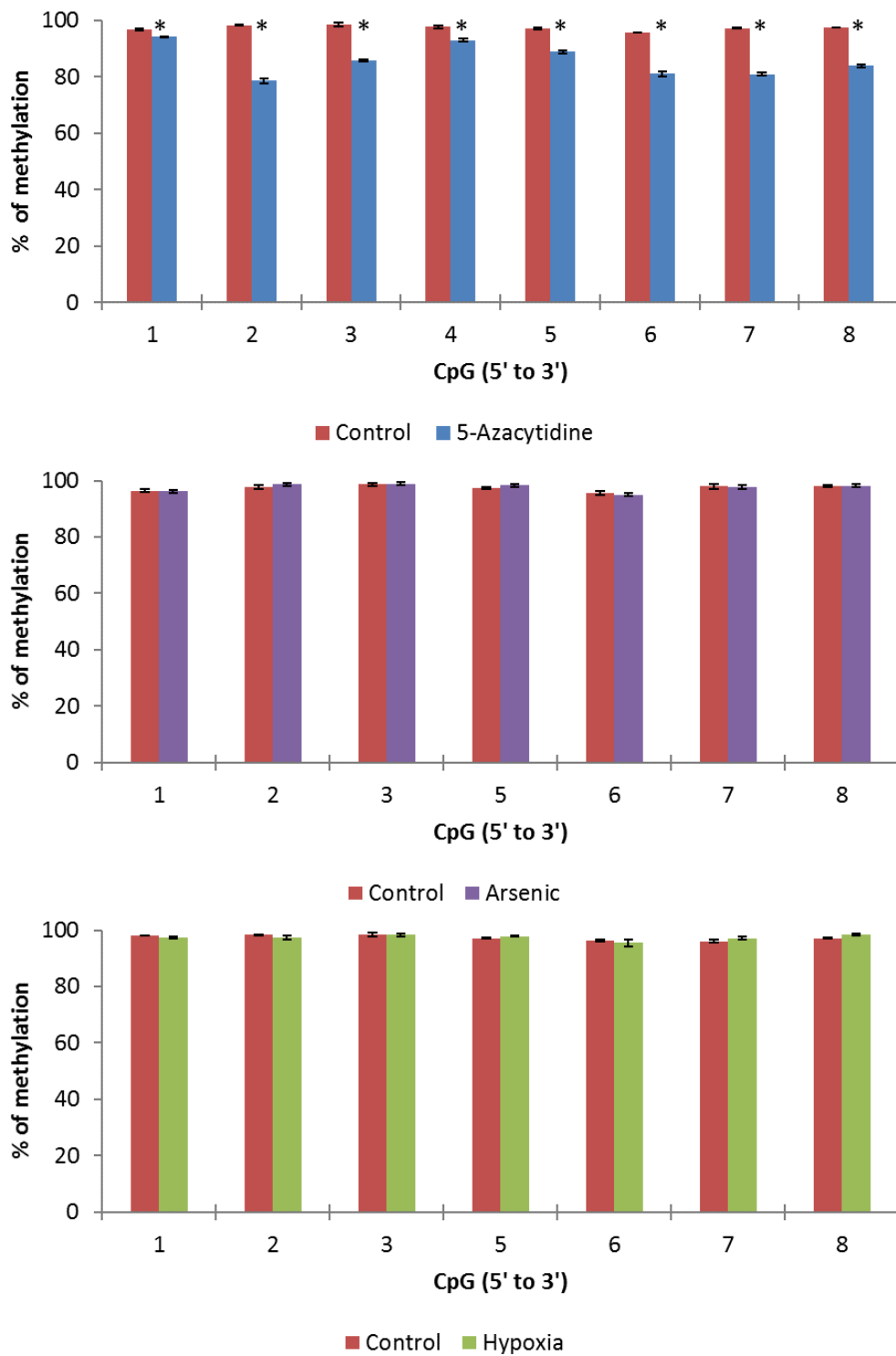
Figure 4.15 Direct bisulfite sequencing PCR data for DMR 377. To confirm the DMR identification using WGBS, three biological replicates were analysed for each treatment group and control using direct BSP. Methylation percentage is shown in the y axis and CpG sites in the x axis. Treatment and control values were compared using t-test and * represent the sites that were statistically significant different from control (p < 0.05).
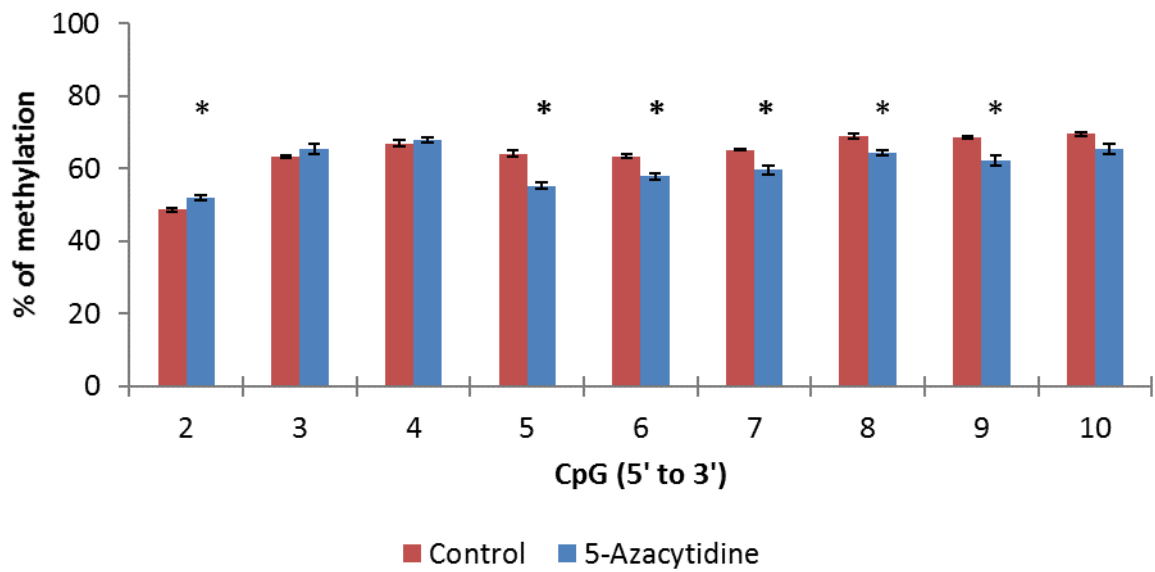
Figure 4.16 Direct bisulfite sequencing PCR data for DMR 19. To confirm the DMR identification using WGBS, three biological replicates were analysed for 5-azacytidine group and control using direct BSP. Methylation percentage is shown in the y axis and CpG sites in the x axis. Treatment and control values were compared using t-test and * represent the sites that were statistically significant different from control (p < 0.05).
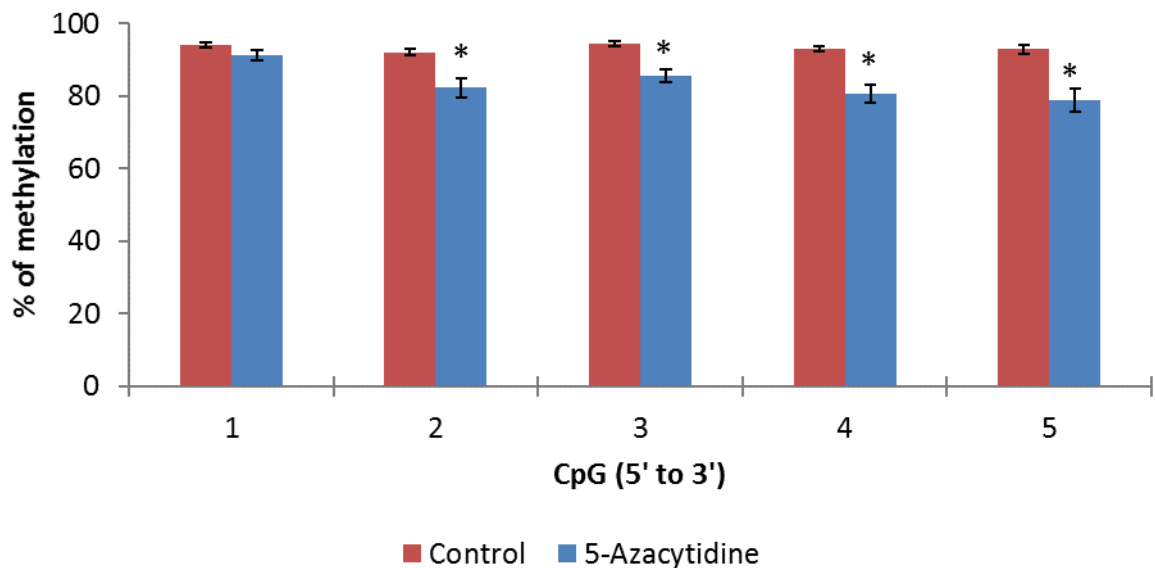
Figure 4.17 Direct bisulfite sequencing PCR data for DMR 223. To confirm the DMR identification using WGBS, three biological replicates were analysed for 5-azacytidine group and control using direct BSP. Methylation percentage is shown in the y axis and CpG sites in the x axis. Treatment and control values were compared using t-test and * represent the sites that were statistically significant different from control (p < 0.05).
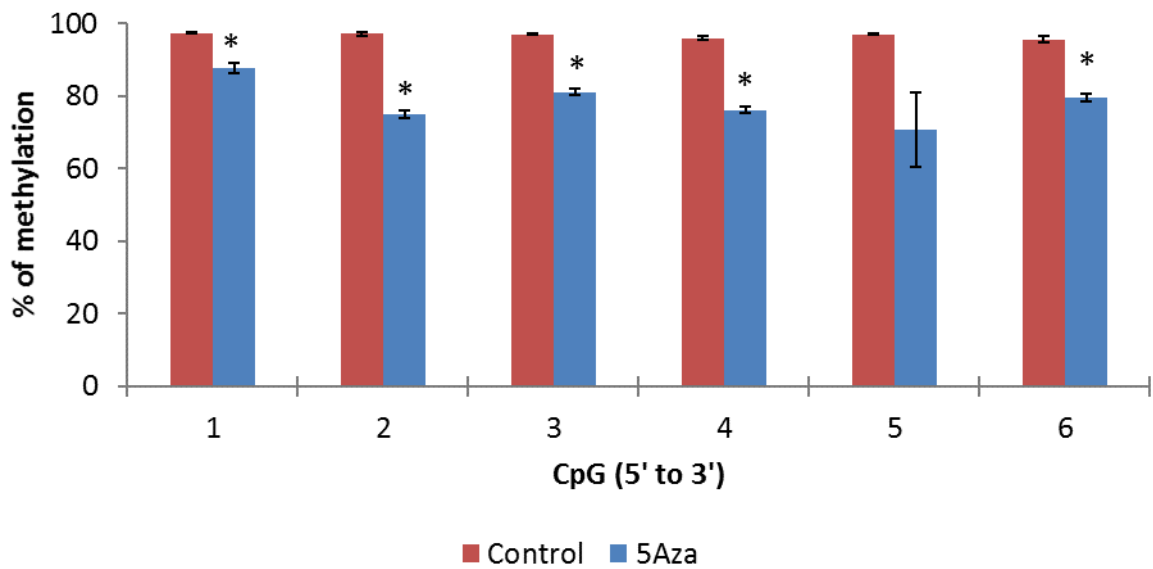
Figure 4.18 Direct bisulfite sequencing PCR data for DMR 382. To confirm the DMR identification using WGBS, three biological replicates were analysed for 5-azacytidine group and control using direct BSP. Methylation percentage is shown in the y axis and CpG sites in the x axis. Treatment and control values were compared using t-test and * represent the sites that were statistically significant different from control (p < 0.05).
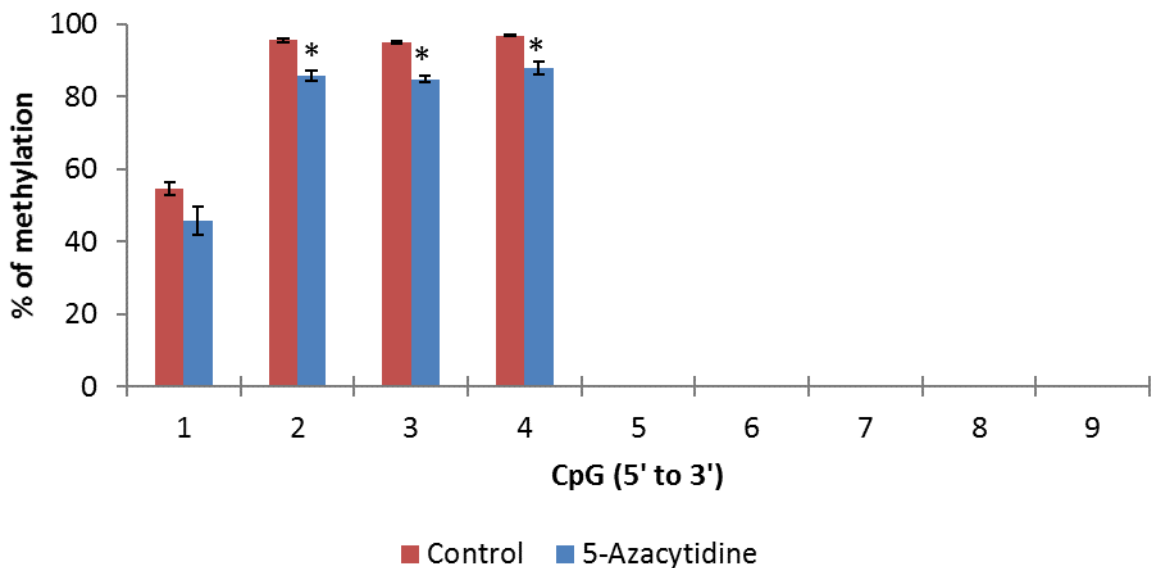
Figure 4.19 Direct bisulfite sequencing PCR data for DMR 341. To confirm the DMR identification using WGBS, three biological replicates were analysed for 5-azacytidine group and control using direct BSP. Methylation percentage is shown in the y axis and CpG sites in the x axis. Treatment and control values were compared using t-test and * represent the sites that were statistically significant different from control (p < 0.05).

4.4   Discussion

4.4.1 Challenges for differential methylation analyses

The genomic resources available for *D. magna* are still scarce. Although efforts have been made to construct and annotate its genome, it is still lacking important information for a complete DNA methylation profiling and DMR analyses. However, the available resources allowed overall DNA methylation profiling (as shown in chapter 3) and the identification of DMRs.

In this chapter, several approaches for DMR analyses were presented. The main decision was either to use biased analysis (using known regions of the genome) or unbiased analysis (no predefined regions for the level of DNA methylation comparison). Both strategies have advantages and disadvantages. As an example, the use of predefined regions for methylation comparison in human presents the great advantage of limiting the number of regions to be analysed, and the possibilities to target the analyses to specific features (i.e. genes, CGIs, promoters, exons/intron, enhancers) (Baumann and Doerge, 2014). It is only possible due to the completeness of the human genome project since, additionally to the annotation of genes, the location of several regulatory regions are already mapped to the genome. The possibility to limit the number of regions analysed could consequently decrease the number of statistical tests performed, increasing statistical power for the detection of DMRs. In this case, the disadvantage of a biased analysis targeted to already known regions for detection of DMRs, is to limit the discovery of novel regions presenting differential methylation (DM) (Robinson *et al.*, 2014).

For *D. magna*, due the size of the genome (predicted ~200Mb) and the lack of a high quality genome sequence and annotation, the best approach for DMR identification is to analyse it covering the entire genome, without targeting the analyses to specific regions. Once the DMRs are identified, they can be annotated using the information available for *Daphnia* or using the general blast search.

Therefore, the best method for DMR analyses for the conditions used in this study was to use the level of methylation of single cytosines and compare the groups using windowed replicate test. It allows the search for DMRs in the entire genome, without bias to annotated regions, and the identification of consistent effects on DNA methylation along short sequences (100bp) (Zhong *et al.*, 2013).

The choice of short windows for the statistical tests was arbitrary. It allows greater resolution of the effects and a targeted identification of the affected region. Also a smaller averaging effect on the methylation levels is observed. Consequently, it will give less statistical power for identification of effects on DNA methylation, however it will provide greater biological relevance (Beissinger *et al.*, 2015; Robinson *et al.*, 2014; Wang *et al.*, 2015).

As discussed in chapter 3 and in section 3.3.5.2, the presence of artefacts in WGBS data is expected. Firstly, the method decreases the complexity of DNA sequences increasing the difficulty for sequencing alignment and assembly. Moreover, the conversion efficiency of non-methylate cytosines, although close to 100%, can introduce errors on methylation quantitation of mapped reads (Breiling and Lyko, 2015; Warnecke *et al.*, 2002). To overcome this problem, quality control steps need to be applied to WGBS data. Different approaches can be taken; in this study the DMR analyses was performed using three biological replicates for each group. The DNA methylation quantitation was performed on sites containing at least

3 reads and present in all three replicates for both control and treatment. Also, regions containing unusual number of reads were excluded from the analysis in order to avoid the effects of overrepresented regions. The actions taken to avoid artefacts related to bisulfite sequencing and analyses were efficient and allowed the use of these datasets for DMR analyses.

After DMR identification another step of filtering was applied to the results. Using direct bisulfite sequencing it was only possible to confirm the level of DNA methylation on the regions with DNA methylation higher than 40%. This was the level of DNA methylation chosen to be applied as a cut-off value for further DMRs analyses to remove false positives.

Therefore, the methods used for DNA methylation quantitation and DMR identification used in this study were very conservative in order to avoid false positive results and increase confidence in the obtained results.

4.4.2 Differential methylation related to ageing and stressors exposures

It was possible to identify a number of effects of ageing and stressors on the DNA methylation profile. Age-related DMRs were analysed comparing the DNA methylation profile of 5 days old and 14 days old animals.

Only 38 probes were statistically significantly differentially methylated between 5 and 14 days old *Daphnia*. They were grouped in 9 DMRs, however only 6 were annotated to the *D. magna* genome. From the annotated DMRs two were annotated to regions coding for the ribosomal proteins that are part of ribosomal structure, linked to protein biosynthesis. The *60S ribosomal protein L7a* gene was hypomethylated in day 14 while *40S ribosomal protein S4* gene presented increased DNA methylation level in day 14 (in comparison to day 5). The other

genes presenting differential methylation were *RING finger protein* and *Ubiquitin-conjugating enzyme E2*. Both are known to play a role in the ubiquitination pathway, targeting substrate proteins for degradation (Lorick *et al.*, 1999).

Often, ageing is known to affect DNA methylation levels leading to overall hypomethylation and site-specific hypermethylation. In general, the identified age-related DMRs presented an increased DNA methylation level for 14 days old animals. It is important to note that the animals at this age had just reached sexual maturity, therefore, they are still "young". It explains the results found for the analysed samples and indicates that studies aiming to link DM and ageing should focus on older animals, since a life span of 60 days can be estimated for *D. magna* cultured at 20°C (Smith, 1963).

The analyses of stress-specific DMRs followed three steps: i) identification of DM cytosine sites; ii) grouping of DMRs based on methylation level and proximity; iii) filtering of DMRs based on cut-off value. In the end, a list of DMRs was built.

The exposures to arsenic, hypoxia and 5-azacytidine caused the alteration in the DNA methylation level of several cytosine sites. It has already been shown that the distribution of DNA methylation in invertebrates, including *D. magna*, presents few genes that are classified as highly methylated while others present low levels of methylation (Asselman *et al.*, 2016; Pegoraro *et al.*, 2016). The identified DM sites followed the same distribution, especially for 5-azacytidine.

The DM sites were grouped into DMRs based on DNA methylation level and proximity and filtered using the cut-off value of 40%. Again, the majority of DMRs that passed the filtering were caused by the exposure to 5-azacytidine, meaning that very little DNA methylation change was observed for the animals exposed to arsenic and hypoxia.

Alterations in the DNA methylation level have been reported to occur due to exposure to arsenic. The effects could be due to the competition of the mechanism of arsenic detoxification and the DNA methylation for the same methyl donor, therefore it is suggested that arsenic exposures can affect the overall level of DNA methylation (Reichard and Puga, 2010; Reichard *et al.*, 2007; Zhao *et al.*, 1997). Several studies suggest that arsenic exposure results in effects on DNA methylation, however only one DMR was identified for *Daphnia* using WGBS and the filtering approaches. Later direct bisulfite sequencing revealed that this DMR, despite being annotated to the reference *D. magna* genome was, in fact, a DNA fragment of a common symbiont of *Daphnia*. The lack of effects caused by arsenic suggests that the mechanisms of toxicity of arsenic require an exposure to a higher concentration, since it is based on the competition for the same methyl donors.

Hypoxia is an important stressor for aquatic organisms. The depletion of oxygen can impair growth, disturb the reproduction and even cause death of aquatic populations (Long *et al.*, 2015). In water bodies it can be caused by the increase in anthropogenic input of organic matter and nutrients that later will affect the oxygen concentrations by increasing algal growth (Wu, 2002).

Some studies have now linked the organisms' responses to hypoxic conditions to regulation by epigenetic mechanisms (Brown and Rupert, 2014; Hattori *et al.*, 2015; Lachance *et al.*, 2014; Tsai and Wu, 2014; Tudisco *et al.*, 2014). In this study, only two DMRs were found in association to hypoxia exposure using WGBS and could be annotated to the *D. magna* genome. The confirmation of these results using direct bisulfite sequencing will be presented in chapter 5.

From the 22 DMRs identified from organisms exposed to 5-azacytidine, 6 were selected to be analysed by direct bisulfite sequencing. The decrease in DNA methylation levels was confirmed for all the 6 DMRs. DMR 337 was also tested in groups exposed to arsenic and hypoxia, but no effects were observed.

5-azacytidine was selected because it is a chemical with known effects on the epigenome. Global methylation reduction has already been cited to occur in *Daphnia* exposed to 5-azacytidine (Vandegehuchte *et al.*, 2010b). Its effects target DNA methylation, since this chemical is an analogue of cytosine nucleoside and can be incorporated into the DNA during replication. During methylation of the DNA the DNMTs are then sequestrated by 5-azacytidine and remain attached to the DNA being unavailable for further methylation (Stresemann and Lyko, 2008). The exposure of *D. magna* for 5 days to 5-azacytidine presented decreased methylation levels at specific cytosine sites.

The use of WGBS coupled to the proposed DMR detection method was successful to identify these effects on DNA methylation for *D. magna*. Also, these results were confirmed with region specific bisulfite sequencing. It indicates that the methodology chosen for comparisons of the DNA methylation profiles and DMRs identification are reliable and can be applied in future studies.

4.5   Conclusions

In conclusion, the first part of this chapter describes different methods for DMR identification and the main advantages and disadvantages of each method. Biased methods can directly generate information regarding specific regions of the genome and can compare the effects of the exposures to predefined known regions. On the other hand, unbiased methods do not

target any specific region of the genome and can be applied for the identification of novel effects on the DNA methylation profile.

For organisms with complete genomes and with a set of mapped regulatory regions, like humans or mice, the use of biased analyses is the obvious choice. However, for partial genomes, lacking important annotations on regulatory regions (as *D. magna*) biased analyses is not the best option. In this case, unbiased analyses, that do not rely on annotation information, can provide important information on DNA methylation profiles even without a complete reference genome available.

From the proposed unbiased analyses, the use of single cytosines for DNA methylation quantitation and the comparison between groups using window replicated tests proved to be the best option for this study. It presented less averaging effects with the use of short windows (100 bp) and more biological relevance of the identified difference between the groups.

The effectiveness of the proposed method was tested on the dataset from the three groups exposed to arsenic, hypoxia and 5-azacytidine. Very few DMR were identified from the groups exposed to arsenic and hypoxia, however for 5-azacytidine, 22 DMRs were found. In addition, the 6 DMRs were confirmed with direct bisulfite sequencing. Therefore, using this effective method for DMR identification it was possible to determine stress-specific effects on the DNA methylation profile of *D. magna*. The effects of the stressors on the methylome and DNA methylation machinery will be addressed in chapter 5.

Chapter 5

Sensing the environment: multidimensional investigation of effects on DNA methylation

5.1   Introduction

External stressors, such as chemical pollutants, dietary components, predators and temperature changes can have long-lasting effects on the organism's development, metabolism and health. In part, organisms respond to external cues by altering their DNA methylation patterns (reviewed in Feil and Fraga, 2012). However, so far the majority of genome-wide DNA methylation association studies have been conducted on mammals. The extrapolation of findings from these studies to invertebrates, particularly species that are environmentally relevant, is not without challenges. This is partially due to the differences in DNA methylation machinery and its distribution across the genome between invertebrates and vertebrate species (Vandegehuchte and Janssen, 2013). Thus epigenetic studies in environmentally relevant species, such as *Daphnia*, an environmentally relevant emerging model organism, can help to achieve a better understanding of the role of epigenetic factors in regulating the responses of invertebrates to external cues and their highly dynamic environment.

As discussed in previous chapters, *Daphnia magna* is a useful environmentally relevant model organism to investigate the role of DNA methylation in multiple aspects, such as response to stressors, adaptation, phenotypic plasticity and maternal transfer of information (Harris *et al.*, 2012). For example, in *Daphnia*, global DNA methylation changes were observed in response to several chemicals (Vandegehuchte *et al.*, 2010a, 2010b, 2009a, 2009b) and environmental stressors (Asselman *et al.*, 2015; Menzel *et al.*, 2011). Furthermore, studies are emerging which are paving the way to unravelling the distribution of DNA methylation across various genomic regions in *Daphnia* species (Asselman *et al.*, 2016; Strepetkaitė *et al.*, 2015).

Additionally, we have shown that changes in DNA methylation can be targeted to specific regions and occur in a stress-specific manner (see chapter 4). This is consistent with the concept of epigenetic memory, where the stressors could cause specific changes in DNA methylation profile, or a "footprint", that could later be linked to previous exposures (Bird, 2002; Mirbahai and Chipman, 2014).

Evidence for stressor footprints exists, however numerous aspects still need to be elucidated. Firstly, the persistence of DNA methylation changes has to be assessed, since hypothetically, some changes could be conditional to the presence of the stressor and would not be seen after it is removed or the condition is altered. If persistent, the changes could either have an effect on the organisms' health outcome or be useful as an epigenetic mark of exposure.

Secondly, it is essential to analyse the effects of environmentally relevant conditions and concentrations. Several studies have focused on finding effects of chemicals on the transcriptome and metabolome, however, they often do not take into account the relevance of exposure duration or concentration.

Here, the choice was made to expose the animals in a scenario more closely resembling environmental exposure. The animals were chronically exposed, including the development period in the brood pouch, to environmentally relevant levels of stressors (2 mg L-1 of dissolved oxygen for hypoxia and 100 µg L-1 of arsenic).

In addition, to test for the persistence of the alterations in the DNA methylation pattern, after chronic exposure, the animals were kept for 7 days in clean conditions, without the presence of the stressor (see section 2.3.1 for details on exposure design, concentration and duration). After identifying target changes in DNA methylation profile caused by the stressors, gene expression and metabolite concentrations were analysed for part of the one-carbon pathway.

As reviewed in section 1.2.2, the one-carbon pathway comprises a series of reaction that lead to the production of SAM, the metabolite that provides the methyl group for DNA methylation. A series of reactions also occur to convert the product of DNA methylation, SAH back to the metabolic pathway (Herceg and Vaissière, 2011; Ulrey *et al.*, 2005). This pathway has major importance for the maintenance of normal levels of DNA methylation and can either be the target of stressors or be affected at a later stage. Regardless of which, alterations may be useful to elucidate the modes of action of stressors affecting the methylome.

Therefore, the aims of this chapter are:

1) To investigate the sensitivity of the *Daphnia*'s epigenome to three stressors: 5-azacytidine, a known demethylating agent, arsenic and hypoxia. It is important to highlight that so far all studies have focused on the effect of stressors on global DNA methylation. Although valuable, these studies provide no information on the effect of stressor at a gene expression level. Furthermore, majority of the studies have been focused on a unique factor. Therefore, to achieve a better understanding of how the stressors affect DNA methylation the effects of stressors have been investigated on (i) methylation of regulatory regions and gene bodies using Whole Genome Bisulfite Sequencing (WGBS) results and direct bisulfite PCR, (ii) the one-carbon pathway, and (iii) expression levels of selected genes.

2) To test our pipeline described in Chapter 4 for identification of Differentially Methylated Regions (DMRs) in response to stressors.

3) To design more environmentally relevant experimental conditions. The current approaches for toxicity testing often do not account for the differences between acute and chronic exposures. Additionally, especially for *Daphnia*, the tests, both acute and chronic, start with neonates released from the brood pouch. Therefore, they do not face the exposures during

the developmental stage, whereas in the environment they are exposed to the stressors from the beginning of the embryonic stage.

4) To test the concept of epigenetic memory and recovery. The concept of epigenetic memory relies on the persistence of the alterations of the DNA of the organism caused by the stressor. This is proposed based on studies that have demonstrated the divergence of DNA methylation profiles from young and older twin siblings, and it has been proposed that can be linked to increased susceptibility to diseases later in life (Fraga *et al.*, 2005; Skinner *et al.*, 2010). Therefore, we aimed to assess the presence of stress-specific alterations on the DNA and the maintenance of those alterations once the stressor is removed.

5.2   Overview of experimental design

The exposure design is detailed in section 2.3. Briefly, animals were exposed to arsenic (100 µg L-1), hypoxia (2 mg L-1 of dissolved oxygen) or 5-azacytidine (3.7 mg L-1). For arsenic and hypoxia, samples were collected after 1 day and 21 days of continuous exposure and after a 7-day long recovery period where the *Daphnia* were transferred and maintained in exposure-free media. For 5-azacytidine, samples were collected after 1 day and 5 days of continuous exposure and after a 7-day long recovery period. Animals were immediately dissected to remove the embryos in the brood pouch, if needed, flash frozen in liquid nitrogen and stored at -80 °C.

During sampling, photographs were taken using a stereomicroscope to measure body length as an indicator of growth rate. Additionally, *Daphnia* from the hypoxia exposure were also sampled for haemoglobin quantitation (n=6) (see section 2.5.2).

DNA, RNA and metabolites were extracted from the same homogenate as described in sections 2.4.1.2 and 2.4.2. DNA samples were used for gene level DNA methylation analysis (n=3) (section 2.8). The latter was achieved using either WGBS approach (section 2.7) or targeted bisulfite sequencing (section 2.8). RNA samples were used to measure the expression levels of selected genes using real-time qPCR as described in section 2.9 (n=3, three technical replicates). The extracted metabolites were used for a quantitative, target metabolomics study of one-carbon cycle as described in detail in sections 2.4.3 and 2.10. Six biological replicates were used for the metabolomics study.

Parametric and non-parametric statistical tests were used to analyse the results obtained by the different methods. For RT-PCR and BSP the results were compared between control and treatment using t-tests. Furthermore, the differences in phenotypic measurements and metabolites quantitation were assessed using non-parametric statistical analyses (Kruskal-Wallis and Mann-Whitney).
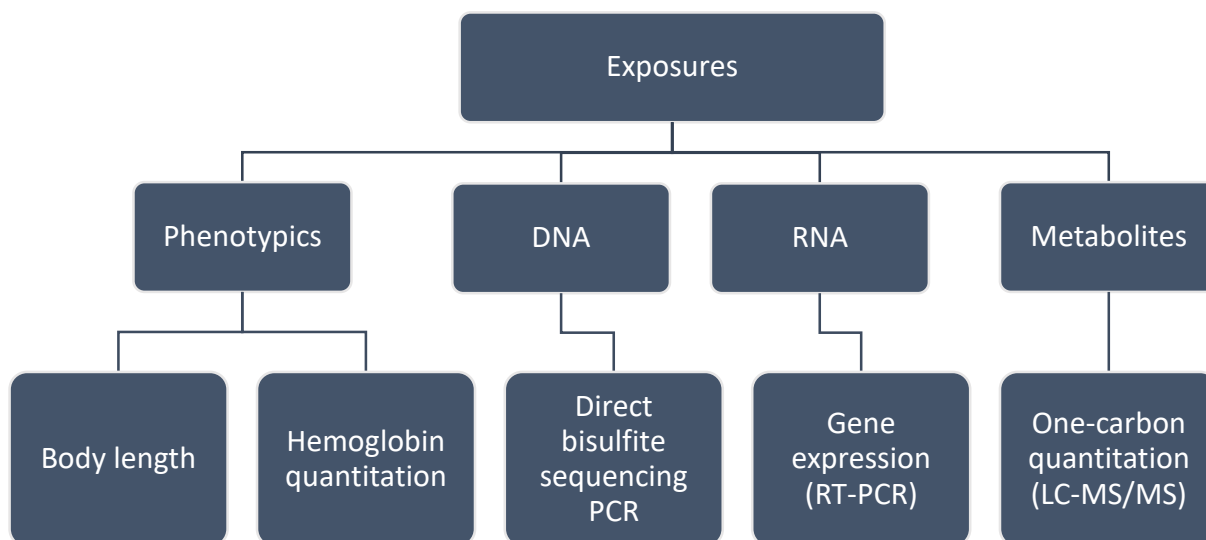
Figure 5.1 Workflow of the analysis performed in chapter 5. Stress induced changes in the methylome, gene expression and metabolites quantitation in *D. magna* after acute and chronic exposure to environmentally relevant concentrations.

5.3   Results

After the identification of altered DNA methylation profiles, described in chapter 4, the aim was to analyse the accumulation and persistence of the effects on the methylome (epigenetic memory throughout lifetime) using two different time points of exposure (day 1 and day 5 or 21), and an additional time point after seven days of recovery. One aim was to determine differences in response to acute (1 day) and chronic (5 or 21 days) exposures.

Results are presented for target DNA methylation alterations, followed by measurements of the effects on phenotypic endpoints; body length and haemoglobin concentration, and effects on the one-carbon pathway indicated by gene expression and metabolite concentration alterations.

5.3.1 Targeted bisulfite sequencing PCR

In Chapter 4 (section 4.3.5) several regions of DNA were identified to be differentially methylated in the treatment groups compared to controls. To test if chronic exposures throughout development at environmentally relevant concentrations can induce differential methylation, a subset of the identified DMRs were selected for further analysis for the three treatment groups at three different time points of 1 day and 5 or 21 days of exposure, and after 7 days of recovery. The latter will provide information to determine if stressor-induced DM can be maintained in the absence of the stressor, potentially throughout the life of an individual.

In total seven regions were selected and analysed for the 5-azacytidine (5 regions) or hypoxia (2 regions) groups (Figures 5.8, 5.9). The sequences and primers used for amplification and sequencing are described in section 2.8.1. The exposure to arsenic did not induce any statistically significant DMR, as shown in section 4.3.5, therefore, samples from this group were not investigated with targeted bisulfite sequencing.

Overall, for most CpG sites the methylation level was less variable between the same CpG sites in different biological replicates in the control group compared to treatment groups (Figure 5.2 as an example). 5-azacytidine treatment caused a decrease in the methylation level of CpG sites located in regions 337, 341, 382, 422 and 487. Statistical differences were not identified for CpGs in regions 337 and 382, however, the average of methylation level, measured for the region at day 1 and day 5 were reduced in 10% and 20%, respectively. Average methylation level for region 341 was decreased at day 1 (7.1%) and day 5 (17.6%). For regions 422 and 487, the average methylation level at day 5 was reduced by 11.6% and 9.6%, respectively.

Nevertheless, the methylation level was restored to the control level after the recovery period for regions 337, 341 and 382. However, for regions 422 and 487 the average methylation level was not fully restored, showing a decrease of 14% and 8.7%. This potentially indicates that certain stressor-induced methylation changes are maintained even in the absence of the stressor (Figure 5.4 and 5.5).

Region 2398 was analysed for both 5-azacytidine and hypoxia treatment groups. Neither exposure affected DNA methylation of this region. For hypoxia, region 2176 was also analysed and did not show any changes after the treatment.
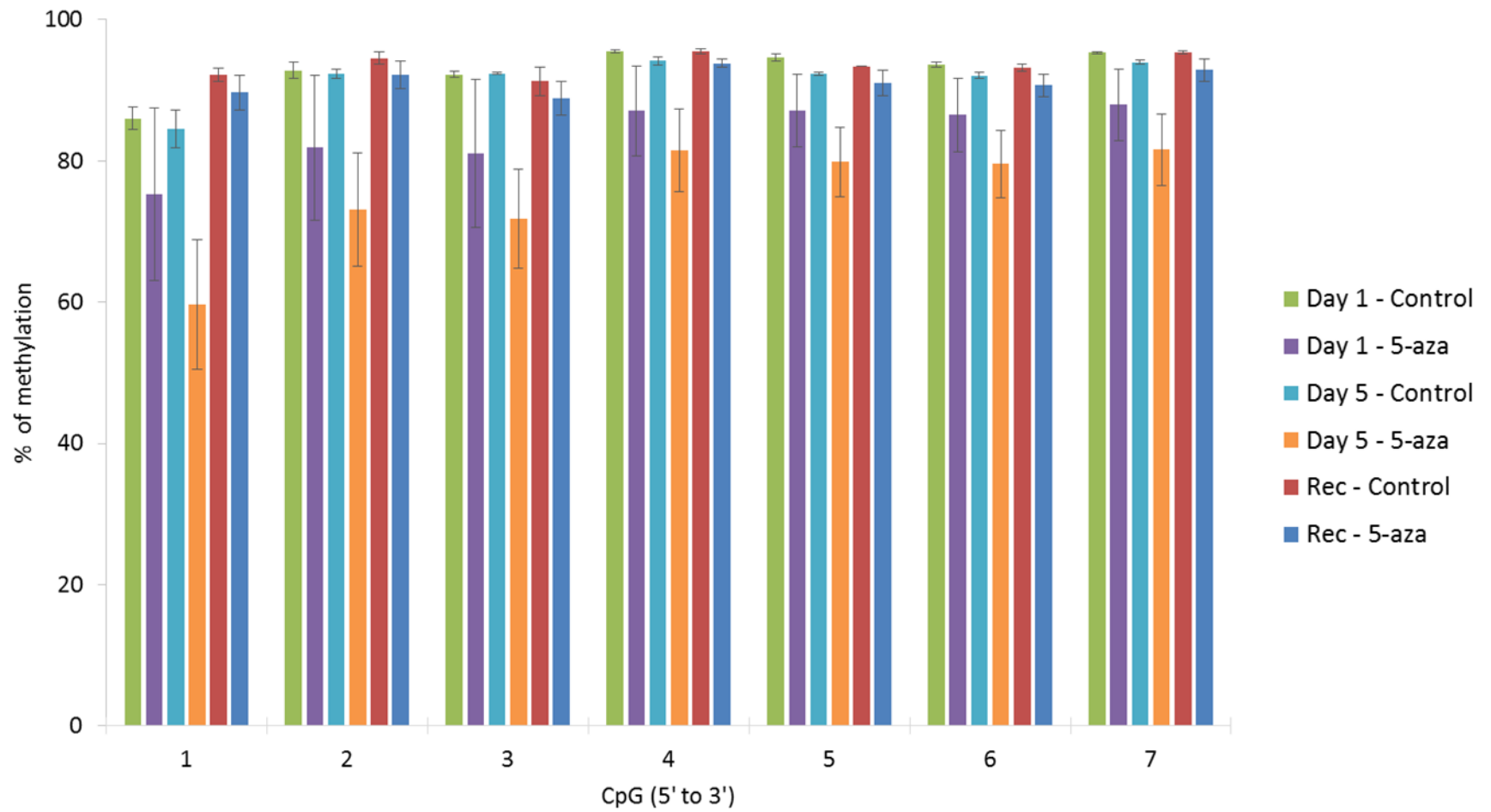
Figure 5.2 Direct bisulfite sequencing PCR data for region 337. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean.
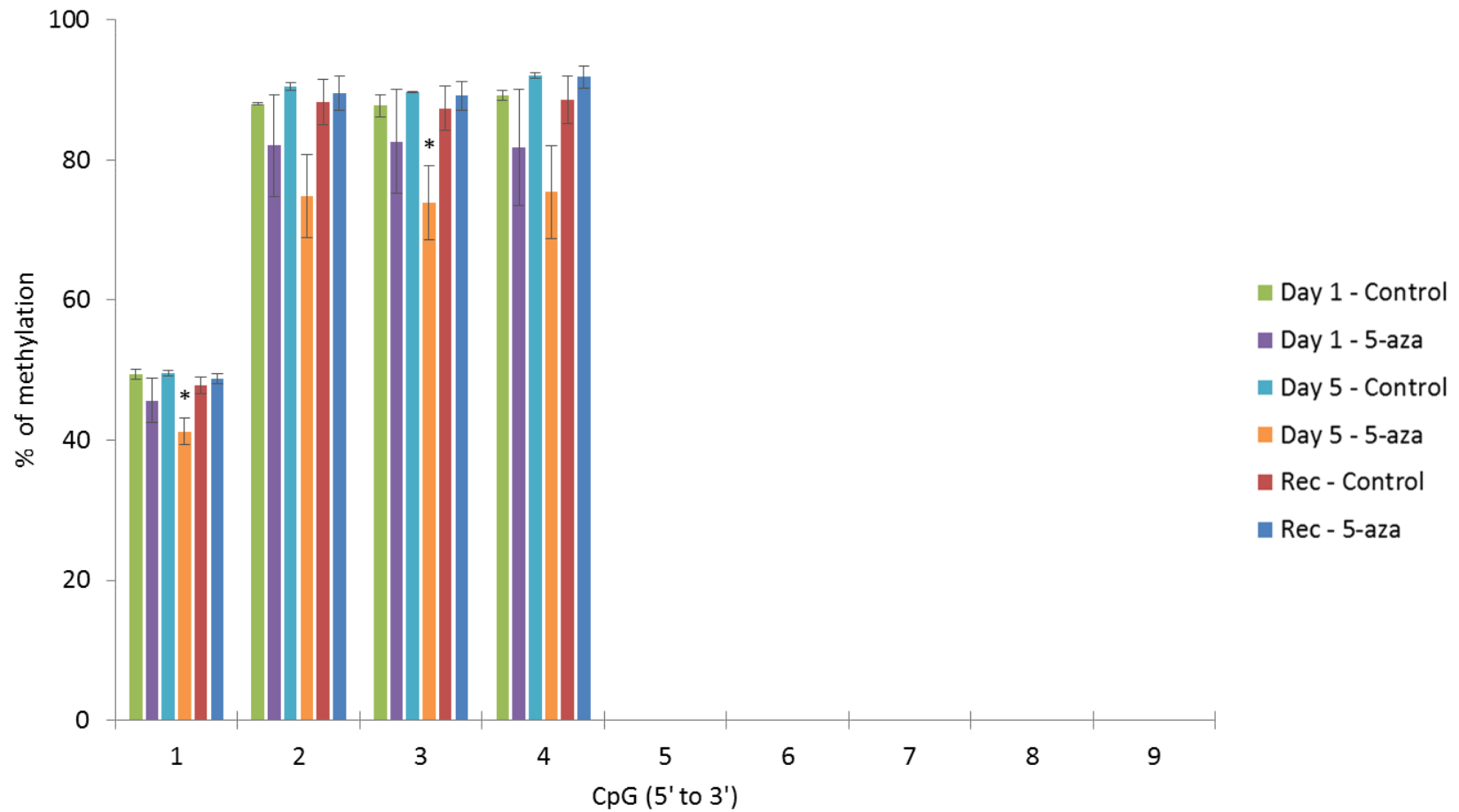
Figure 5.3 Direct bisulfite sequencing PCR data for region 341. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean. * represent the sites that were statistically significant different from respective control (t-test; $p < 0.05$).
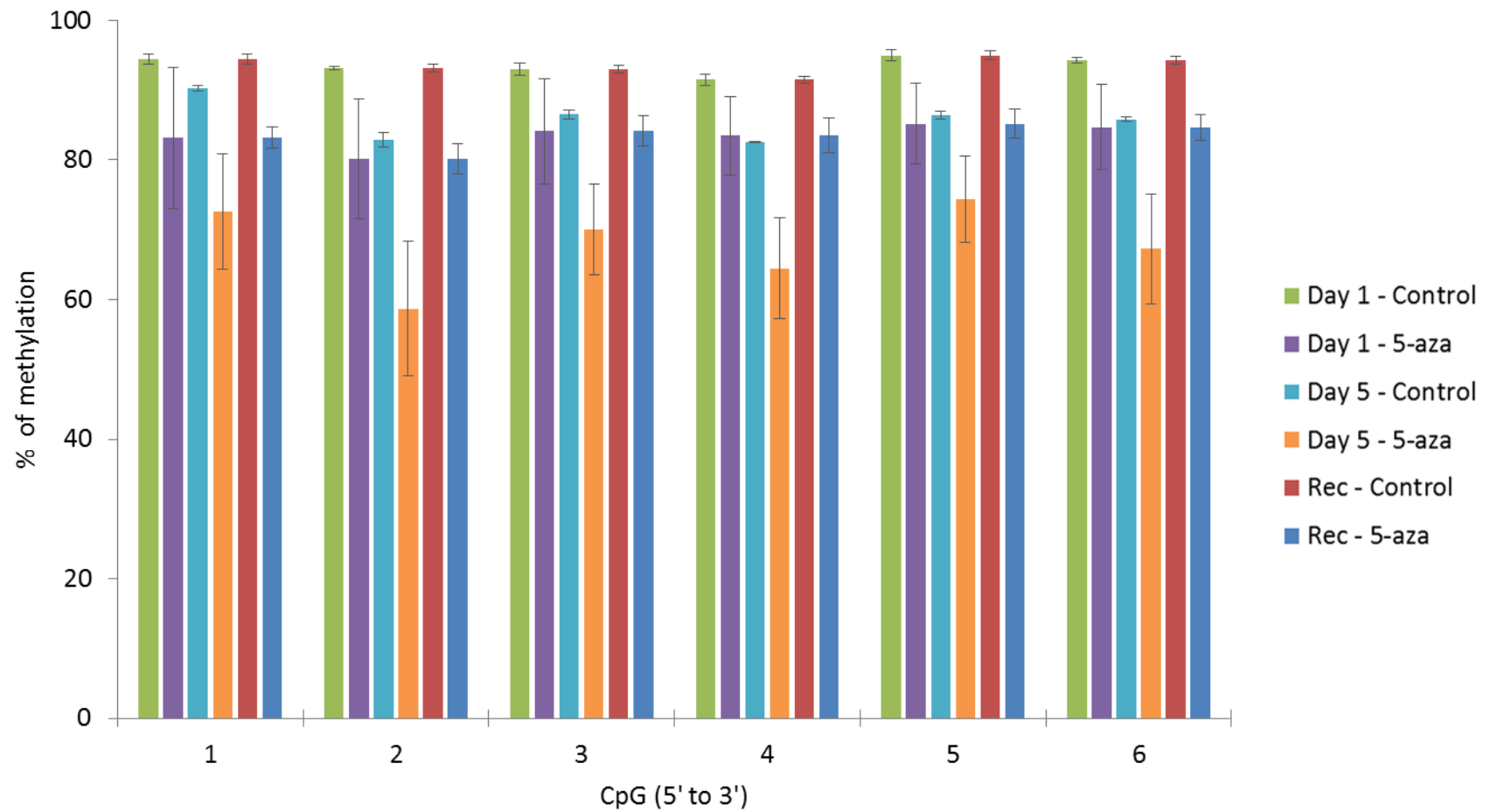
Figure 5.4 Direct bisulfite sequencing PCR data for region 382. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean.
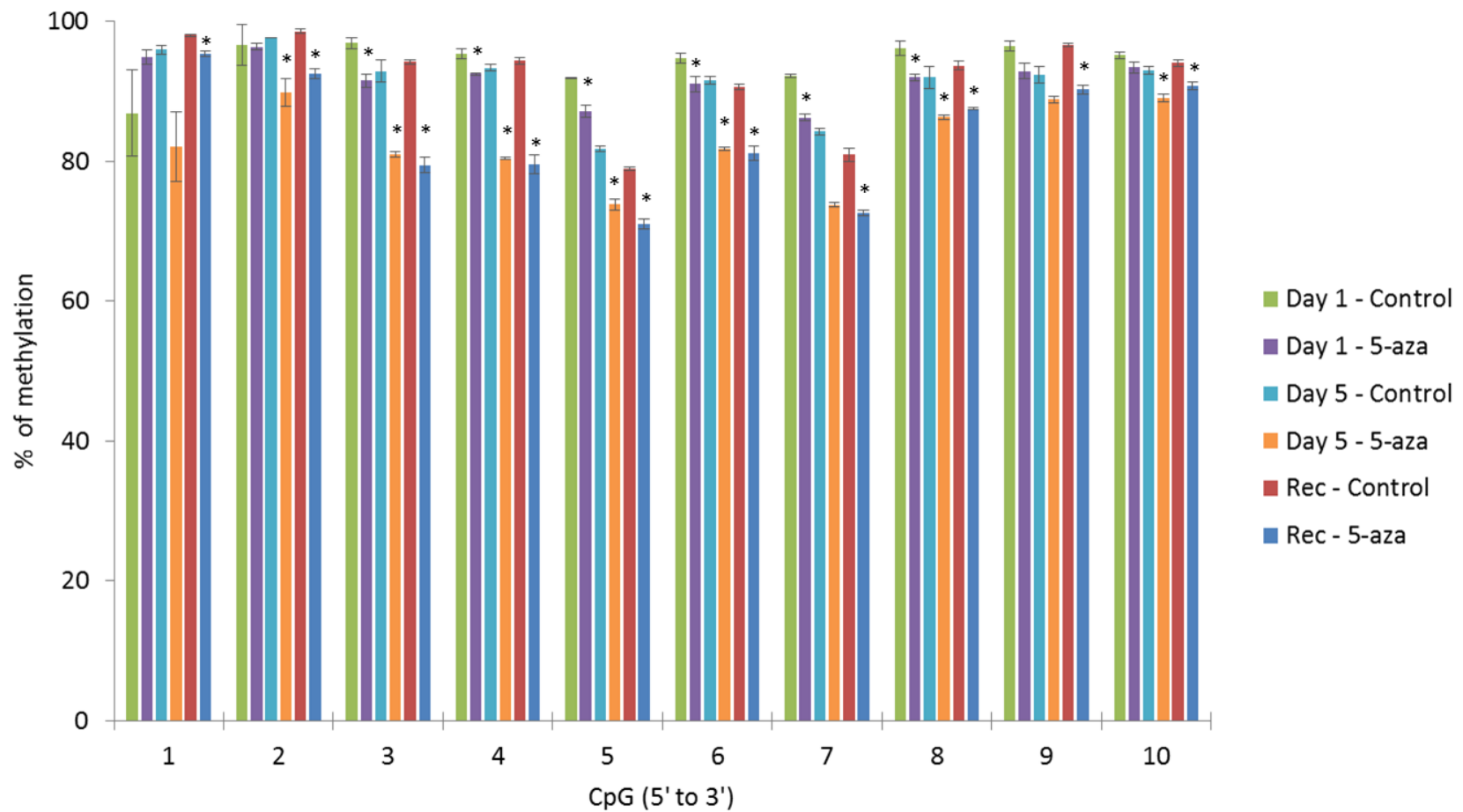
Figure 5.5 Direct bisulfite sequencing PCR data for region 487. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean. * represent the sites that were statistically significant different from respective control (t-test; p < 0.05).
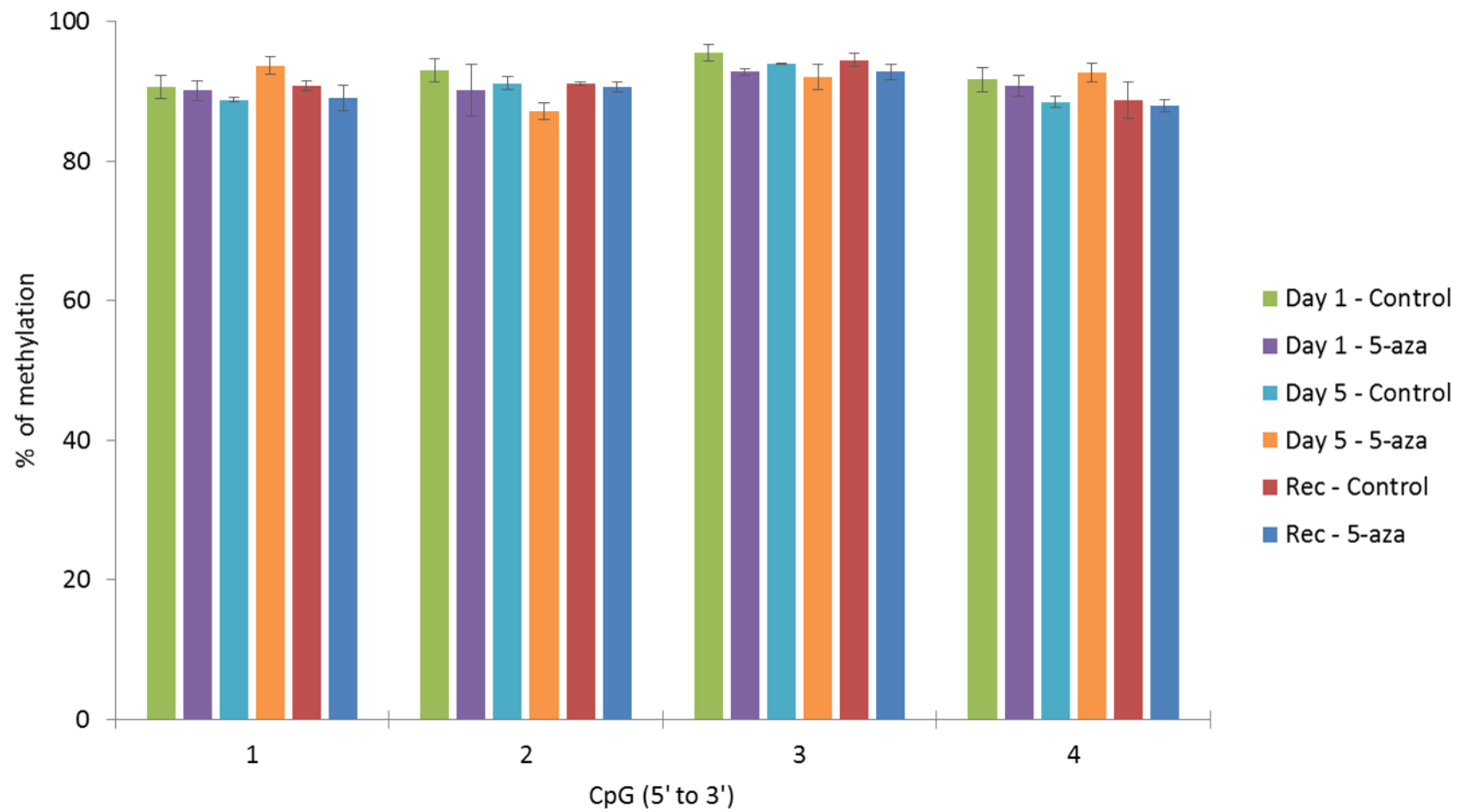
Figure 5.6 Direct bisulfite sequencing PCR data for region 2398. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean.
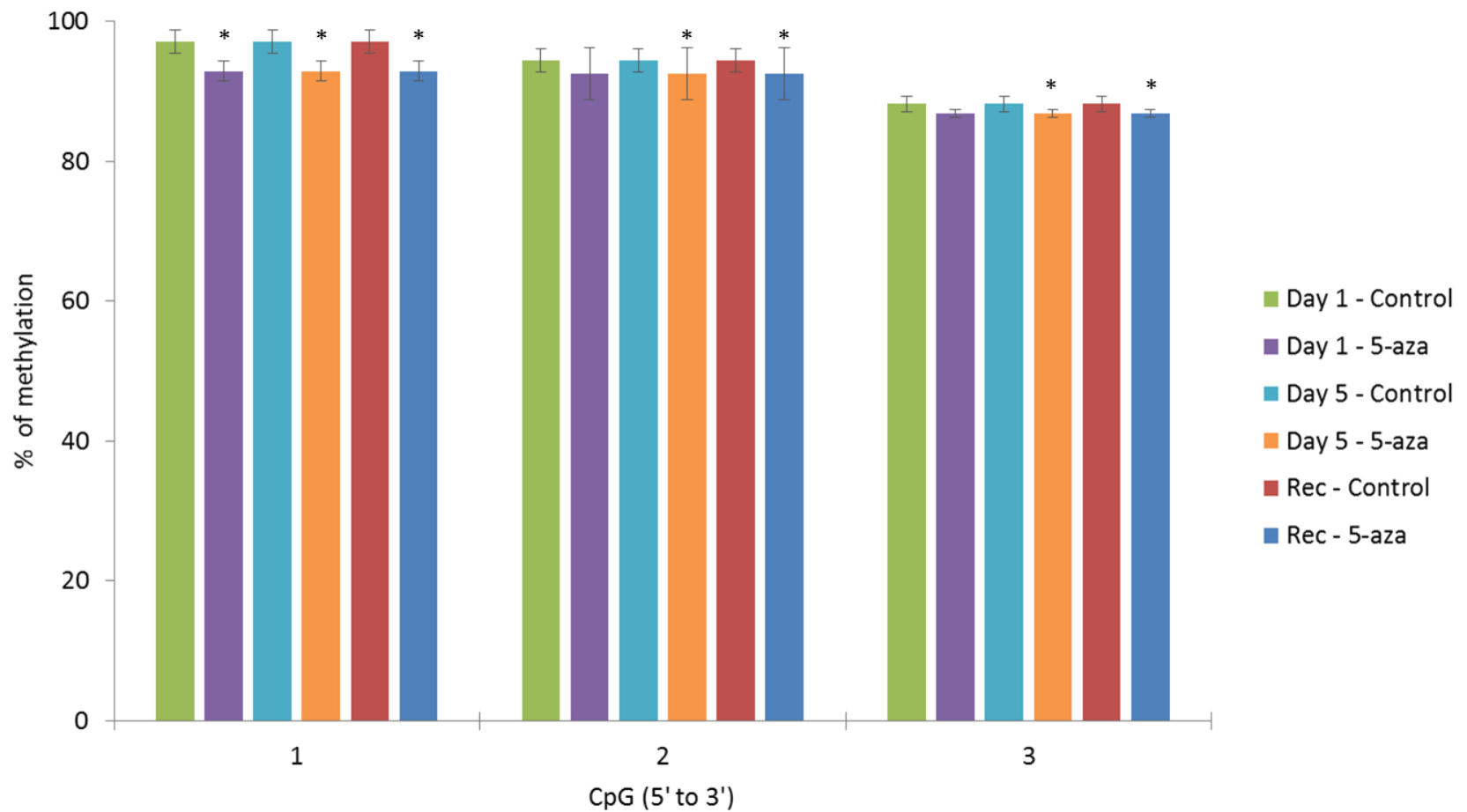
Figure 5.7 Direct bisulfite sequencing PCR data for region 422. Three replicates were analysed for each control and 5-azacytidine groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean. * represent the sites that were statistically significant different from respective control (t-test; $p < 0.05$).
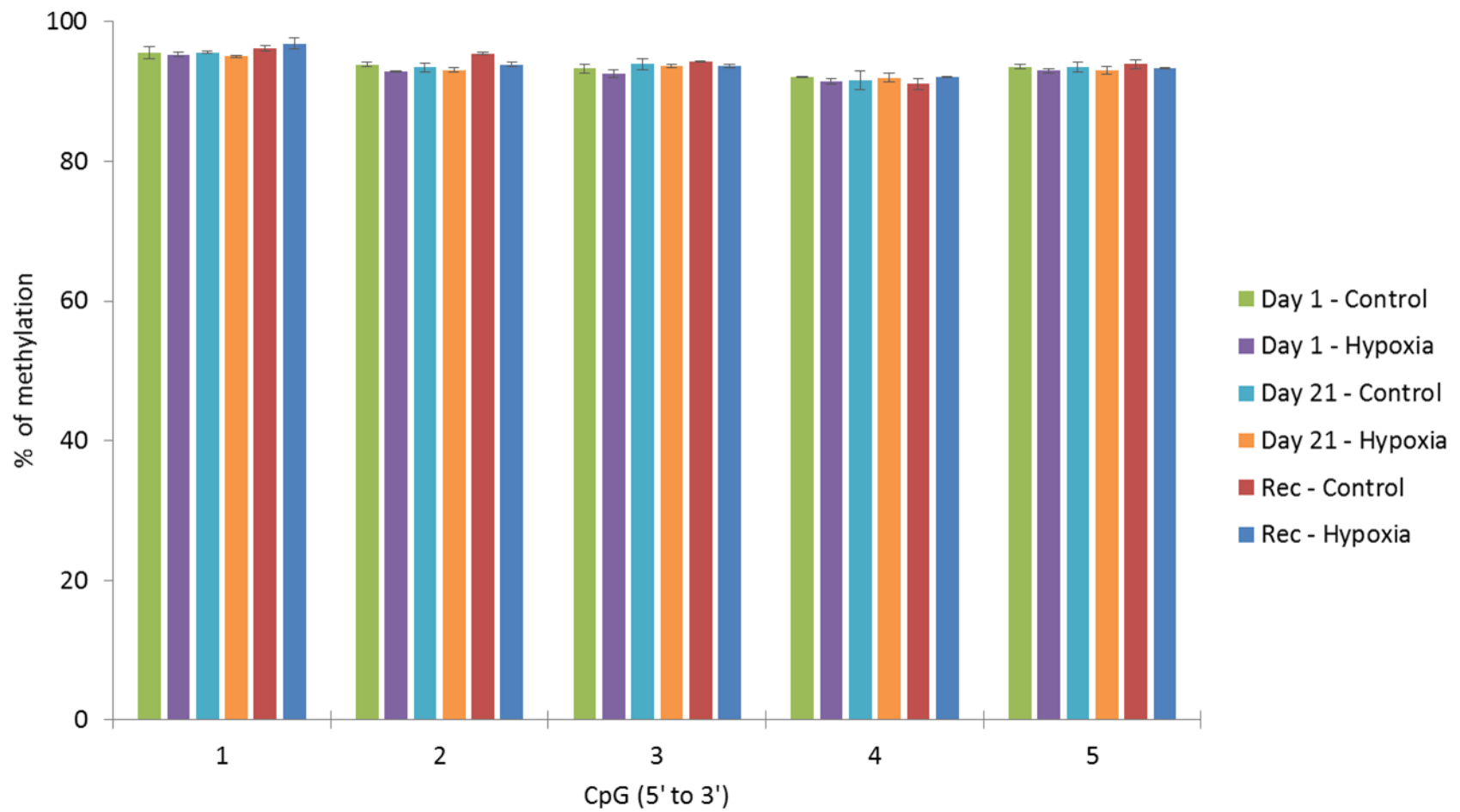
Figure 5.8 Direct bisulfite sequencing PCR data for region 2176. Three replicates were analysed for each control and hypoxia groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean.
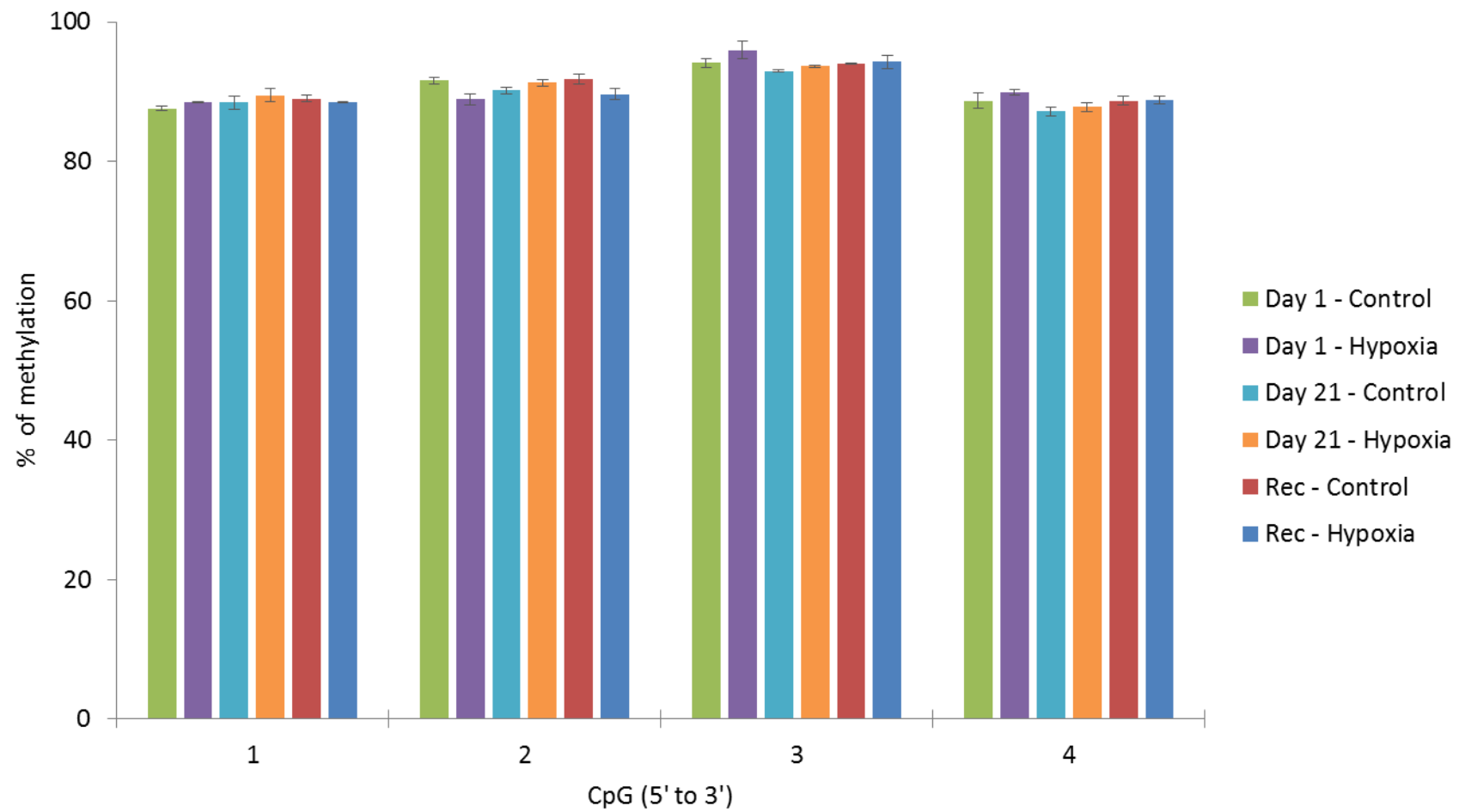
Figure 5.9 Direct bisulfite sequencing PCR data for region 2398. Three replicates were analysed for control and hypoxia groups. Methylation percentage is shown in the y axis and CpG sites in the x axis. Error bars indicate standard error of the mean.

5.3.2 Phenotypic alterations caused by the stressors

The impacts of stressors on phenotypic endpoints were assessed by measuring: (1) Body length and (2) Haemoglobin concentration. Body length was measured for all the treatment and control groups and haemoglobin concentration was measured for the groups exposed to hypoxic conditions after 21 days and after a recovery in clean media.

5.3.2.1 Body length

The measurement of body length is a useful endpoint to determine fitness of the organism and physiological effects of the exposures that could be affecting development and growth (Lampert and Trubetskova, 1996).

Body length was measured using pictures taken from the animals at the different time points, at the beginning of exposure, after the exposure period and after recovery. The pictures were analysed with the Software Image Measurement (KLONK, Denmark). Body length was measured from the base of the spine to the top of the head and was expressed in millimetres. The values were compared for each group using non-parametric statistical analyses (Mann-Whitney) in SPSS.

No change in body length was observed for the animals exposed to arsenic at any of the time points compared to their corresponding controls. However, for the animals exposed to hypoxic conditions a statistically significant decrease in size was observed when compared to animals exposed to normal oxygen conditions. This decrease in body length was observed for neonates (1 day-old), adults (21 days-old) and was maintained even after the recovery period.

For 5-azacytidine the time of exposure was shorter than for the other treatments. No effects were observed for the neonates and 5 days-old *Daphnia* exposed to this chemical. However, the animals after the recovery period (12 days-old) had smaller body length compared to the control group. This result indicates a possible long lasting effect or a delayed effect for this chemical.
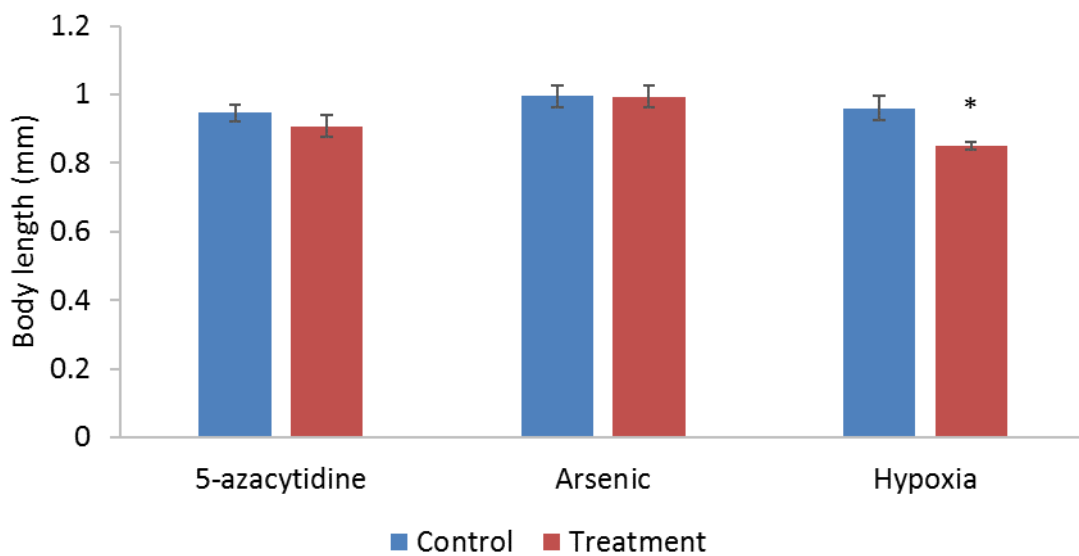
Figure 5.10 Body length of *Daphnia* neonates exposed to stressors during development. * Statistically significant difference between neonates exposed to normoxia and hypoxia (p<0.05). For 5-azacytidine and arsenic group, 15 animals were measure, while 10 animals were measured for hypoxia groups. Values were compared using non-parametric test Mann-Whitney between control and treatment groups. Error bars show standard error of the mean.
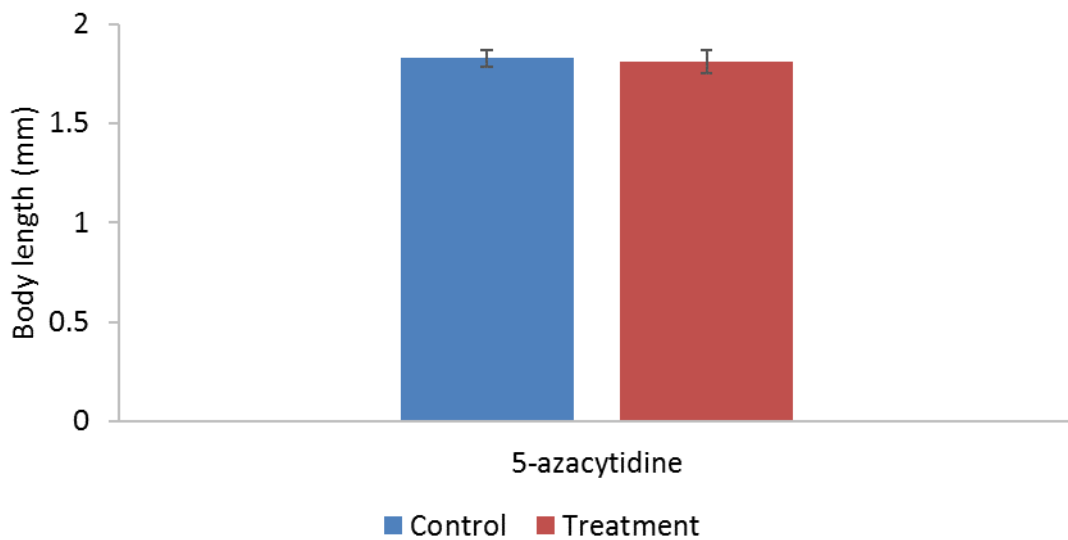
Figure 5.11 Body length of *Daphnia* exposed for 5 days to 5-azacytidine and control. No statistically significant difference between control and treatment (p > 0.05). For each group, 30 animals were measured. Values were compared using non-parametric test Mann-Whitney between control and treatment groups. Error bars show standard error of the mean.
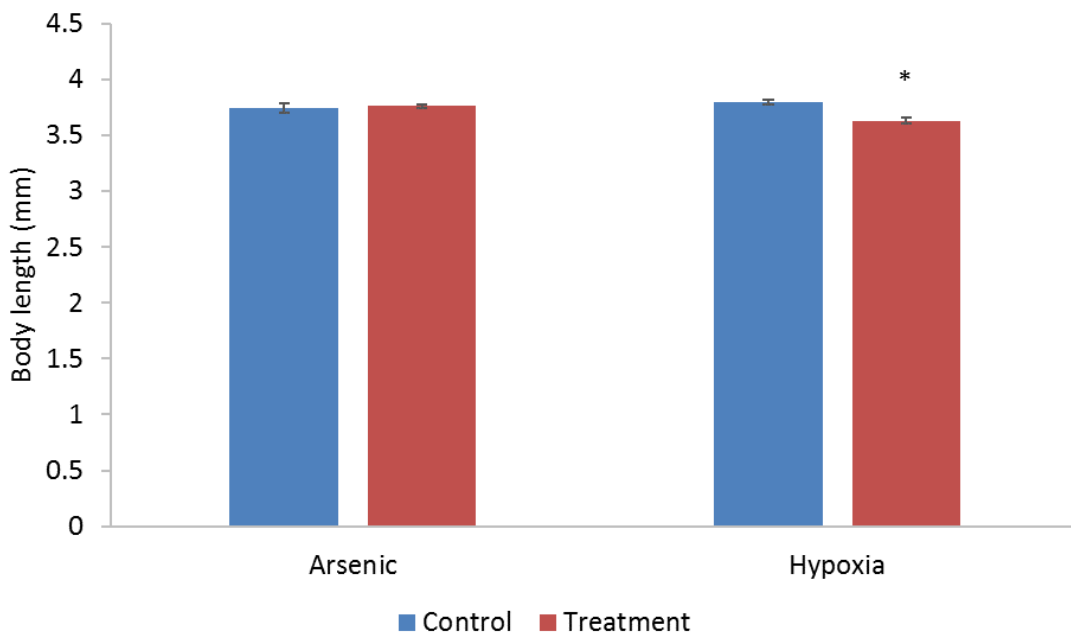
Figure 5.12 Body length of *Daphnia* exposed to arsenic, hypoxia and respective controls for 21 days. *Statistically significant different from respective control (p<0.05). For each group, 30 animals were measured. Values were compared using non-parametric test Mann-Whitney between control and treatment groups. Error bars show standard error of the mean.
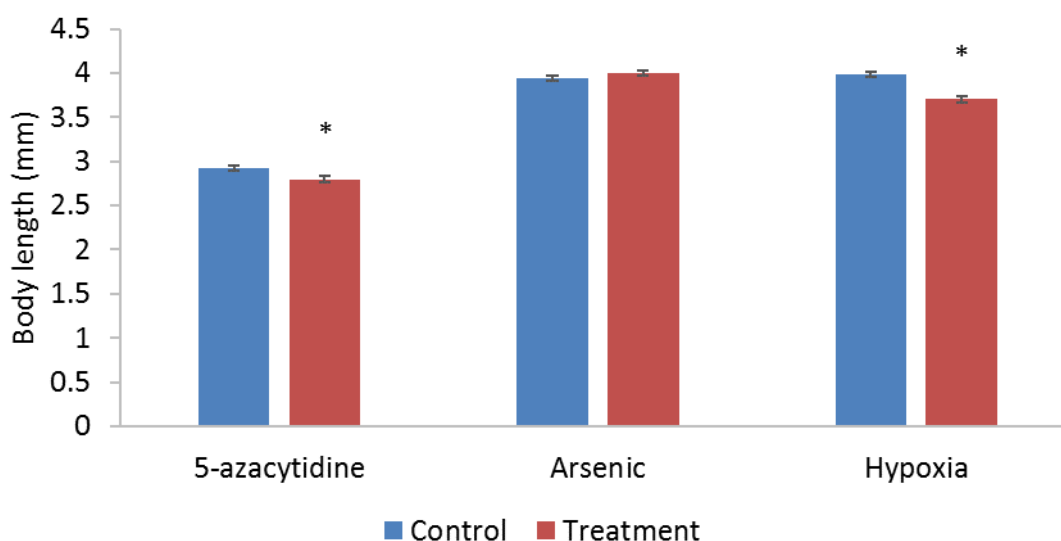
Figure 5.13 Body length of *Daphnia* exposed to 5-azacytidine, arsenic and hypoxia after a recovery period in clean media. *Statistically significant different from respective control (p<0.05). For each group, 30 animals were measured. Values were compared using non-parametric test Mann-Whitney between control and treatment groups. Error bars show standard error of the mean.

5.3.2.2   Haemoglobin quantitation

The exposure of *Daphnia* to hypoxic conditions is known to induce the production of haemoglobin (Gerke *et al.*, 2011; Paul *et al.*, 2004; Pirow *et al.*, 2001; Zeis *et al.*, 2013). Haemoglobin is an extracellular protein that supports transportation and distribution of oxygen in the *Daphnia* circulatory system (Ebert, 2005). Haemoglobin genes are located in a tandem-duplicated gene cluster in *Daphnia*. The composition of protein can vary according to the different subunits expressed and post translational modifications, altering the oxygen affinity in an oxygen-dependent manner (Colbourne *et al.*, 2011; Gerke *et al.*, 2011; Trotter *et al.*, 2015).

A quick and accessible method was used for haemoglobin quantification of the adult animals exposed to hypoxia and normoxia conditions for 21 days and after the recovery period in normoxia for 7 days. The results were obtained using single adult *Daphnia* and six replicates in each group. The protocol followed the method described in Yampolsky *et al.* (2014), with modifications described in section 2.5.2. The results are expressed using the average value for absorbance at 576 nm, normalized by the total protein content. The values were compared for each group using ANOVA with post hoc Tukey test in SPSS.

After the 21 days of treatment the animals exposed to low oxygen levels were visibly red (Figure 5.14). The hypoxia group presented a statistically significant 2-fold increase in the haemoglobin concentration when compared to normoxia group (Figure 5.15).
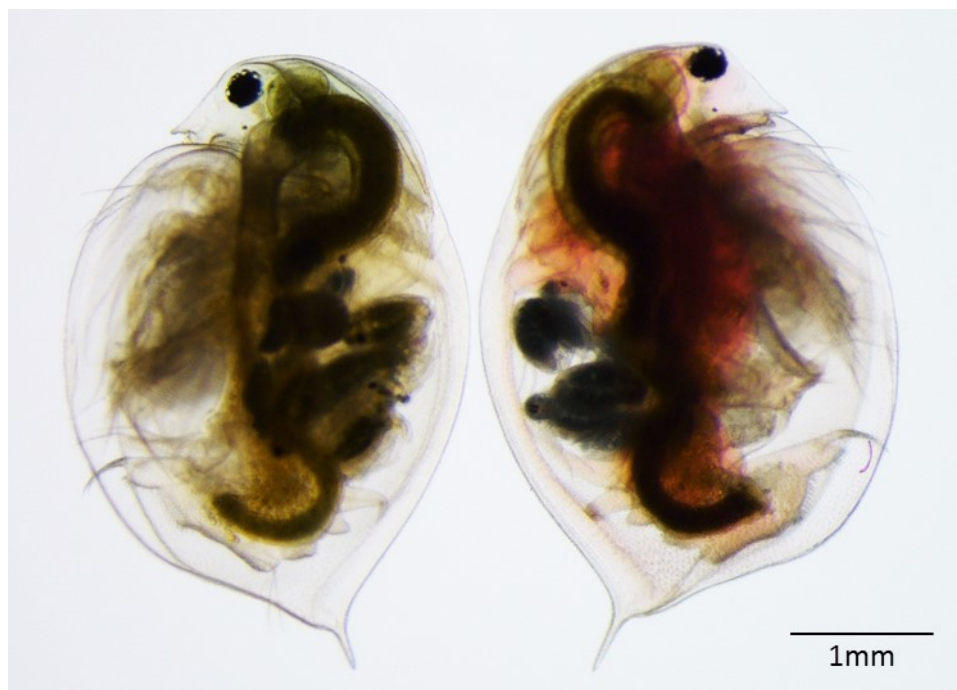
Figure 5.14 Adult *Daphnia* exposed to normoxic (left) and hypoxic (right) conditions.

After 7 days of recovery the haemoglobin levels were decreased by 1.3 fold compared to hypoxic group. However, the levels of haemoglobin, although reduced, were still significantly higher in the recovery group compared to normoxia group by 1.7 fold.
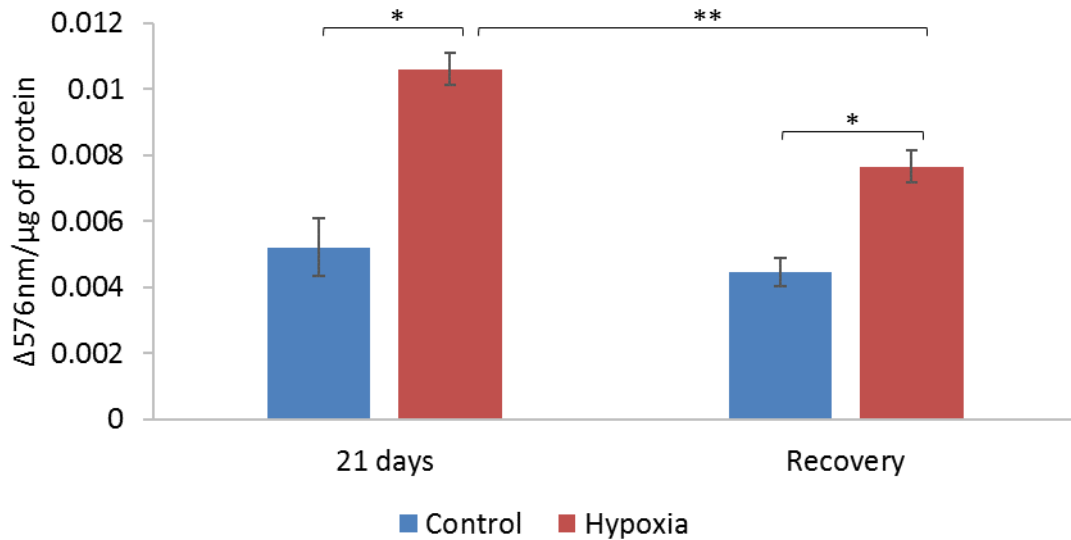
Figure 5.15 Haemoglobin concentration in adult *Daphnia* exposed to hypoxia for 21 days and after 7 days of recovery in normoxic conditions and respective controls. Following homogeneity of variance and normality test the groups were compared using ANOVA with Tukey post hoc test * Statistically significant difference between control and treatment (p<0.01). ** Statistically significant difference between hypoxia treatment at day 21 and after recovery (p<0.05).

5.3.3 Gene expression

The expression levels of genes involved in the one-carbon cycle and demethylation pathway were investigated using RT-PCR. The primers used are described in section 2.9.1. The data were normalised to the geometric average of the two reference genes as described by (Vandesompele *et al.*, 2002). Results are expressed as log 2 fold changes between control and treatment groups for each time point.

As demonstrated in Figure 5.16, exposure to arsenic did not induced any statistically significant changes in the expression levels of the genes involved in one carbon and de-methylation pathways at any of the investigated time points. Although not statistically significant, some trends were observed. For example, *DNMT1*, *DNMT2* and *DNMT3* expression levels were slightly increased at day 1, while *DNMT3* was increased at day 21 (Figure 5.16; t-test, p>0.05).

On the other hand, hypoxic conditions induced changes in the expression levels of several genes. *DNMT1* was upregulated at day 1 but downregulated by 4-fold at day 21 and after recovery. *DNMT2* was upregulated for all time points with 6-fold increase at day 21 and more than 4-fold increase after recovery. *DNMT3* was upregulated by almost 4-fold after 21 days of exposure but it was downregulated after recovery. *TET_*1 was upregulated at day 21 by 2-fold while *TET_*2 was slightly increased at day 1. *MS*, *SAHH* and *GNMT* expression levels were statistically significantly altered when compared to controls. *MS* was upregulated at day 1 and downregulated at day 21. *SAHH* expression was increased after recovery. *GNMT* was upregulated at days 1 and 21.

Similar to hypoxic conditions, exposure to 5-azacytidine affected the expression levels of several genes. The expression level of *DNMT3* was decreased and increased after 1 day and 5 days of exposure, respectively. Similar to *DNMT3*, the expression level of *DNMT1* was increased after 5 days of chronic exposure. *TET*_1 and *TET*_2 genes were both downregulated after one day of exposure with expression levels of *TET*_1 increasing after 5 days of exposure to 5-azacytidine. In the one-carbon pathways, opposite to *MTRR* expression, the expression level of *MAT* gene was increased and decreased after one day and 5 days of exposure, respectively. *MS* expression, similar to *MAT* gene, was downregulated after 5 days of exposure. Interestingly, the expression level of *GNMT* was down-regulated compared to control and previous time points after the recovery period.
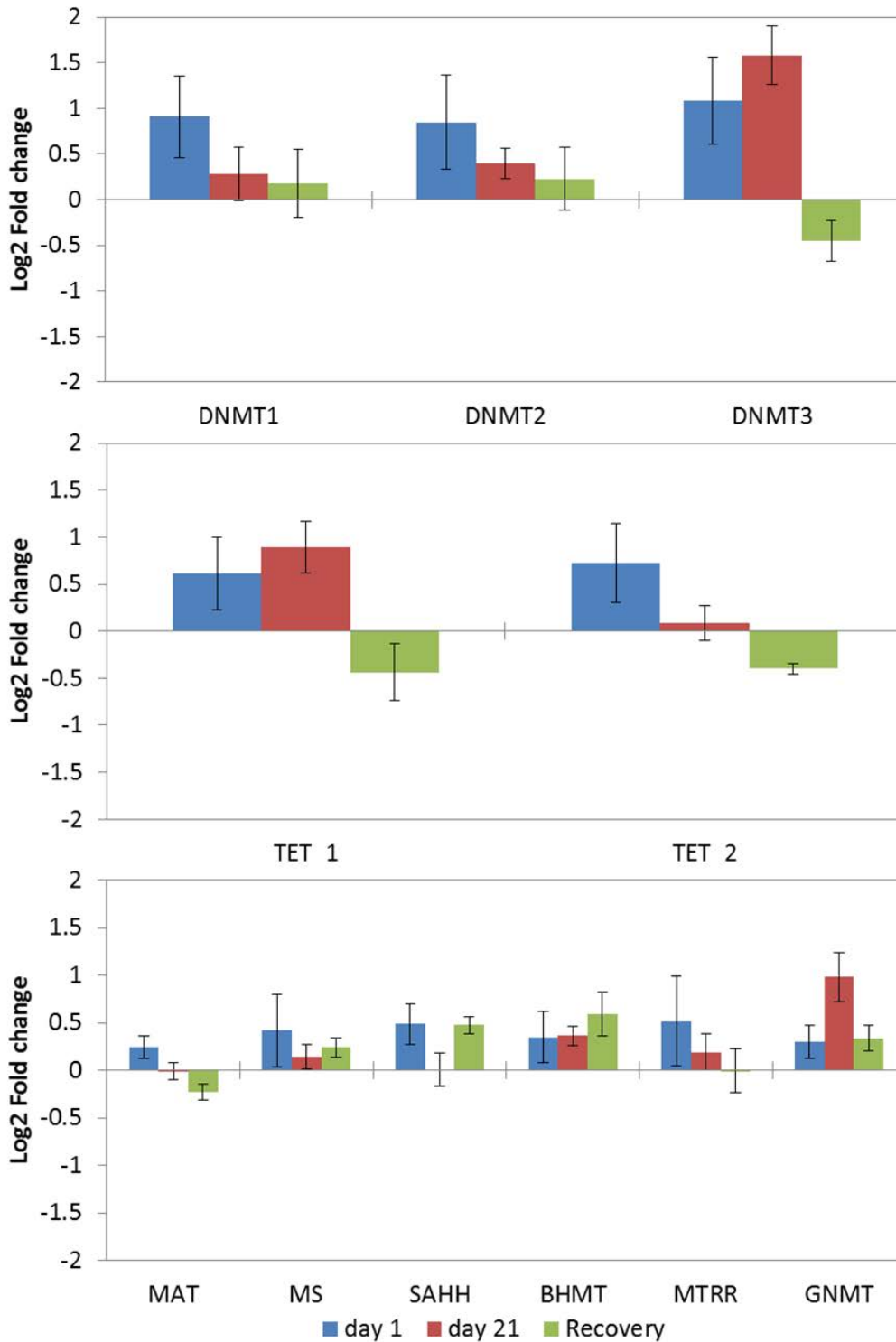
Figure 5.16 Gene expression results for the groups exposed to arsenic. Genes involved in the one-carbon pathway and demethylation pathway. Relative log2 fold change to control group expression. No statistical significantly different from control (t-test; p<0.05). Error bars indicate standard error of the mean. Three biological replicates analysed with two technical replicates.
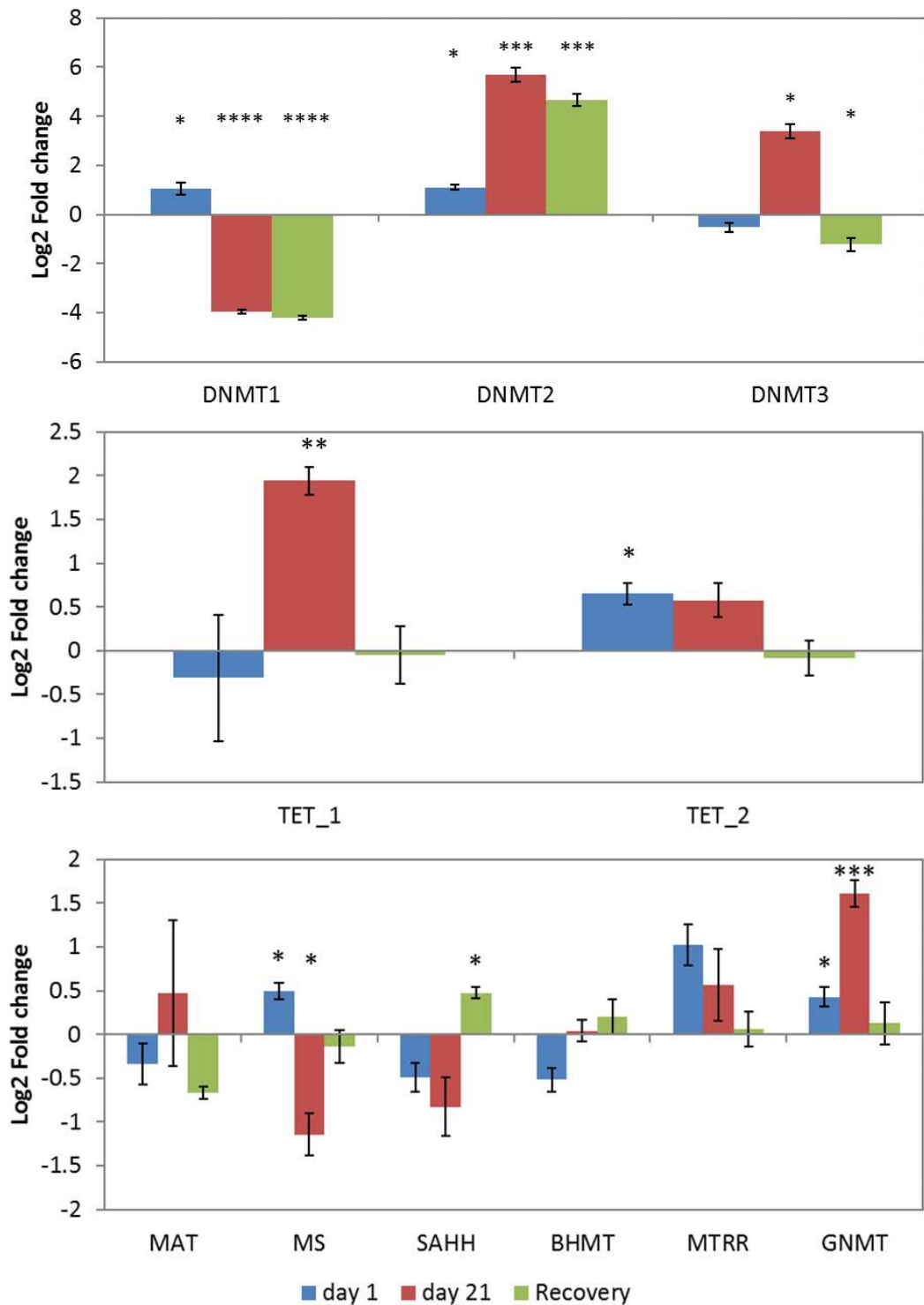
Figure 5.17 Gene expression results for the groups exposed to hypoxia. Genes involved in the one-carbon pathway and demethylation pathway. Relative log2 fold change to control group expression. T-test was used as statistical test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Error bars indicate standard error of the mean. Three biological replicates analysed with two technical replicates.
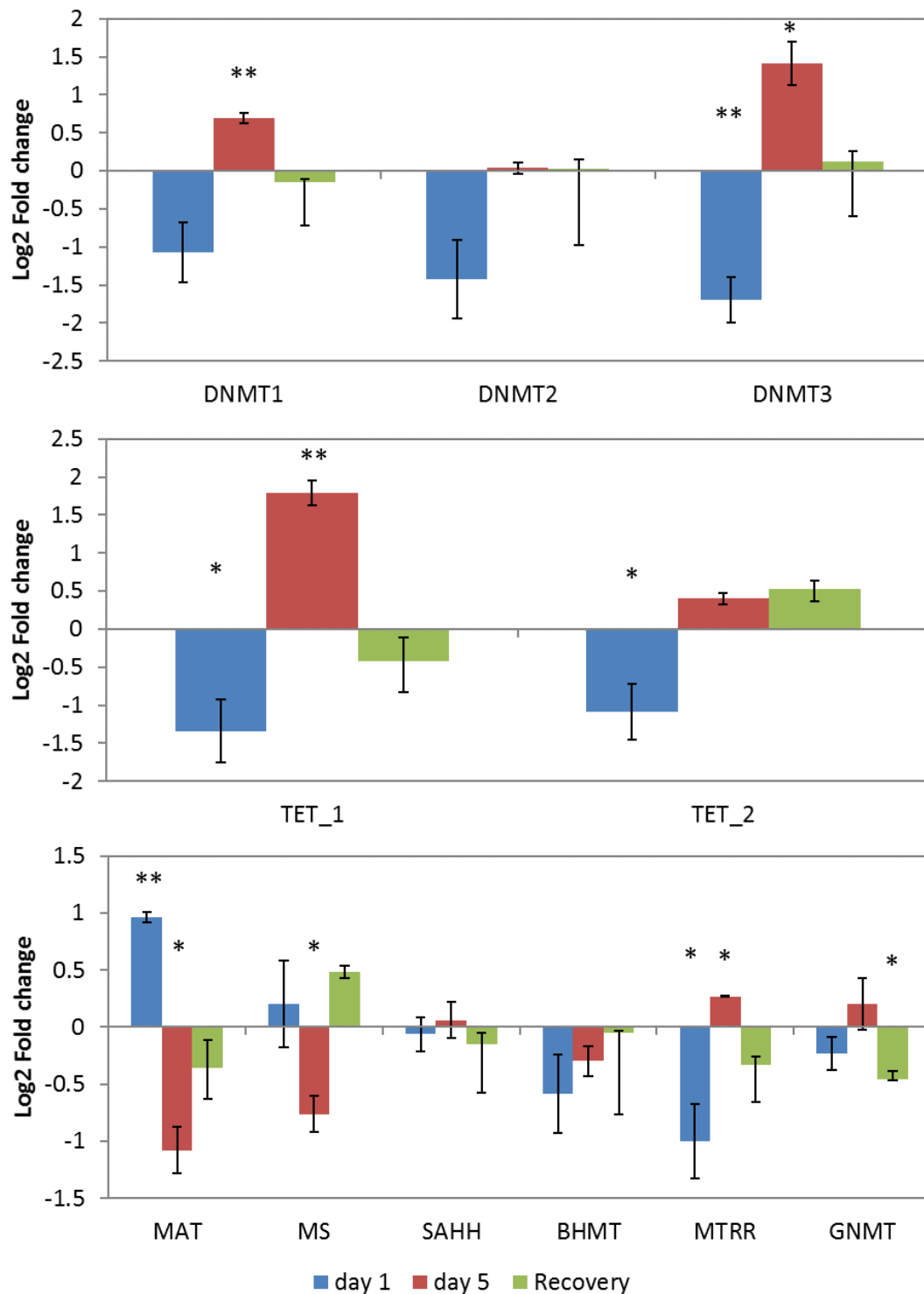
Figure 5.18 Gene expression results for the groups exposed to 5-azacytidine. Genes involved in the one-carbon pathway and demethylation pathway. Relative log2 fold change to control group expression. T-test was used as statistical test. * p<0.05; ** p<0.01. Error bars indicate standard error of the mean. Three biological replicates analysed with two technical replicates.

5.3.4 Target quantitation of metabolites involved in the one-carbon cycle

The levels of 10 metabolites involved in the one-carbon cycle were quantified using liquid chromatography coupled to mass spectrometry as described in section 2.10 for all three treatment groups and all time points (acute and chronic exposures, and recovery). The metabolites quantified were: methionine, choline, adenosine, betaine, sarcosine, stachydrine, glycine, dimethylglycine (DMG), S- adenosylhomocysteine (SAH) and S-adenosyl-L-methionine (SAM).

Statistically significant changes in the concentration of metabolites were observed, mostly for the samples exposed to hypoxia and 5-azacytidine. For arsenic, adenosine was the only metabolite affected by the treatment, demonstrating increased concentration after 21 days of exposure (Figure 5.20).

Hypoxic conditions induced changes in the concentration of several metabolites with the concentration of some of the metabolites not restored to the control level even after the recovery period. For example, DMG, methionine, sarcosine and SAH levels were elevated in the recovery groups compared to controls. Betaine and methionine concentrations were increased after chronic exposure to hypoxia while SAH levels were decreased. The only metabolites that were affected after acute exposure to hypoxia were choline and methionine, with both metabolites decreasing in the treatment group compared to control.

Interestingly, 5-azacytidine treatment only caused a decrease in the concentration of metabolites (adenosine, stachydrine, sarcosine, glycine, SAM and SAH). This effect was only observed after 5 days of exposure. However, the concentration of all the metabolites was restored to the control level after the recovery period.
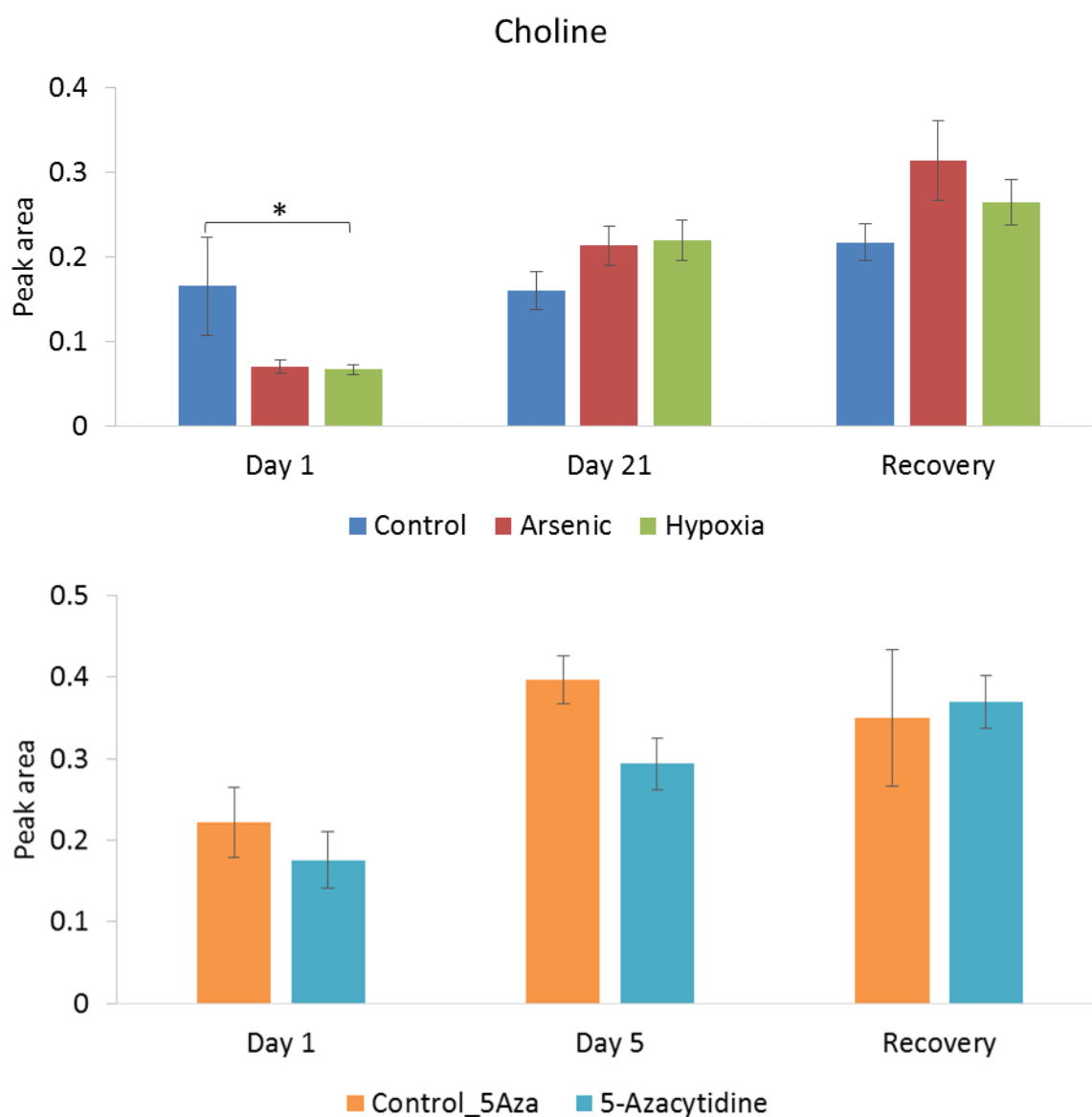
Figure 5.19 Choline concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *(p < 0.05).
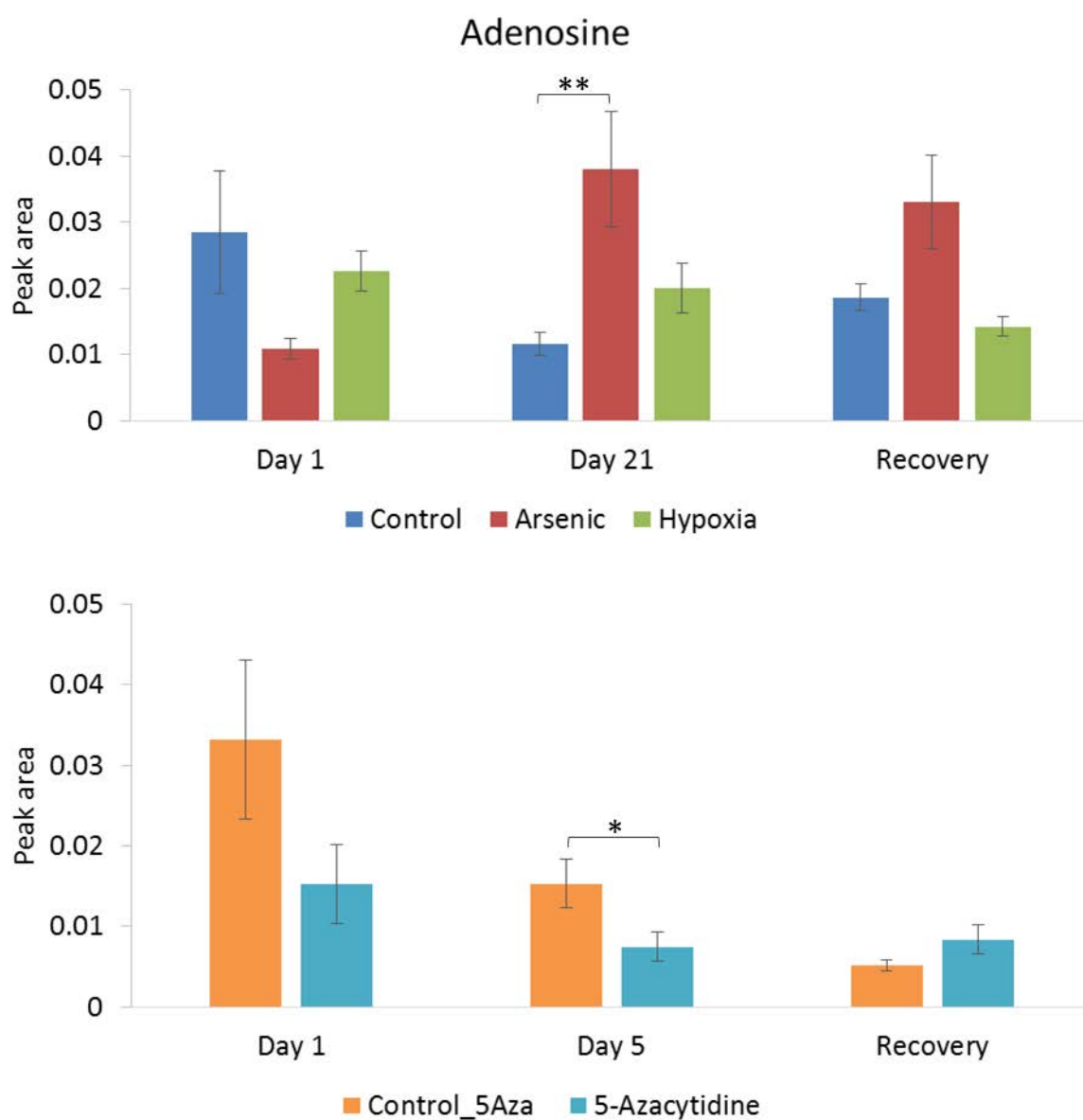
Figure 5.20 Adenosine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *(p < 0.05); **(p < 0.01).
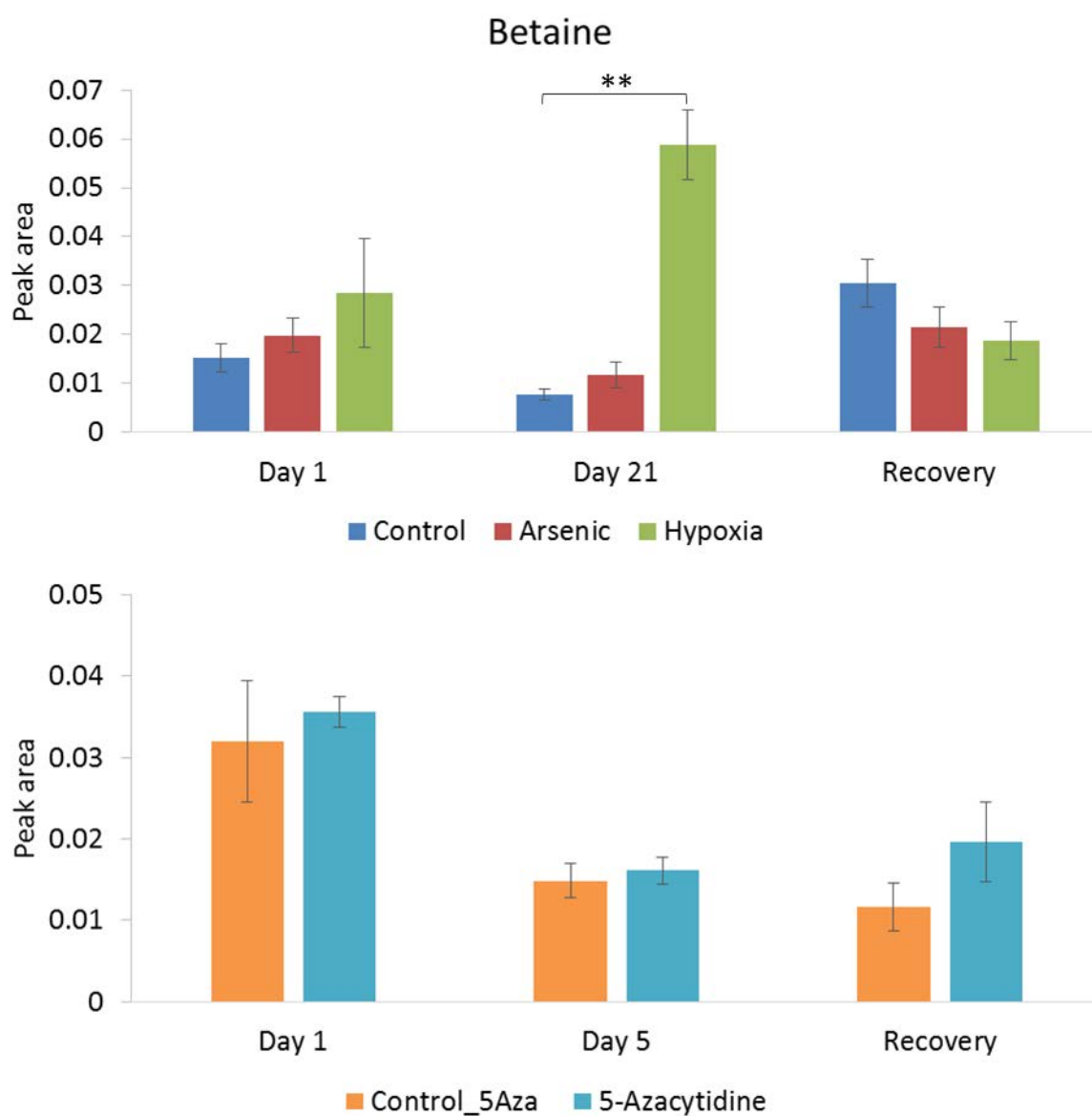
Figure 5.21 Betaine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. **(p < 0.01).
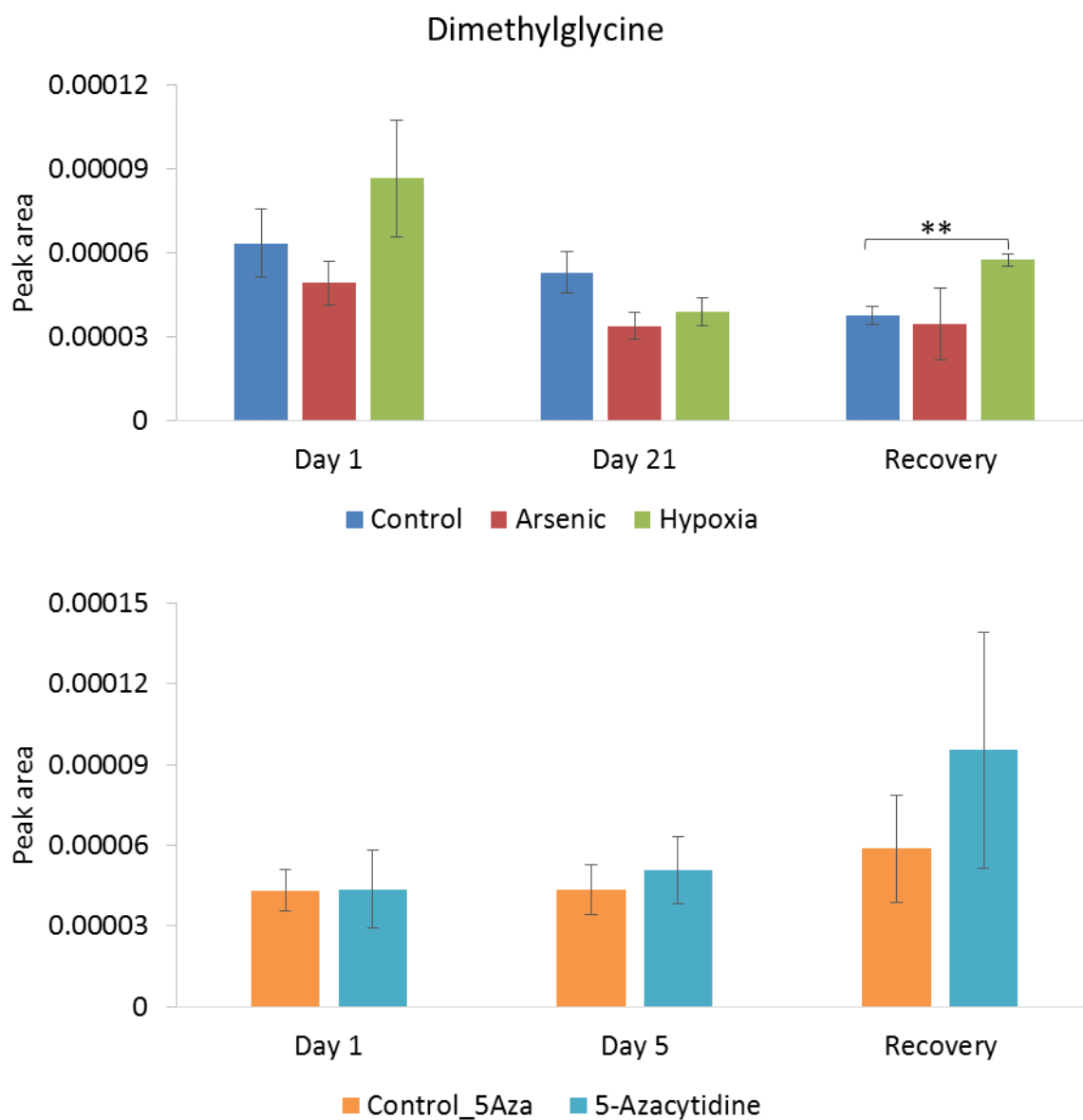
Figure 5.22 Dimethylglycine (DMG) concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. **(p < 0.01).
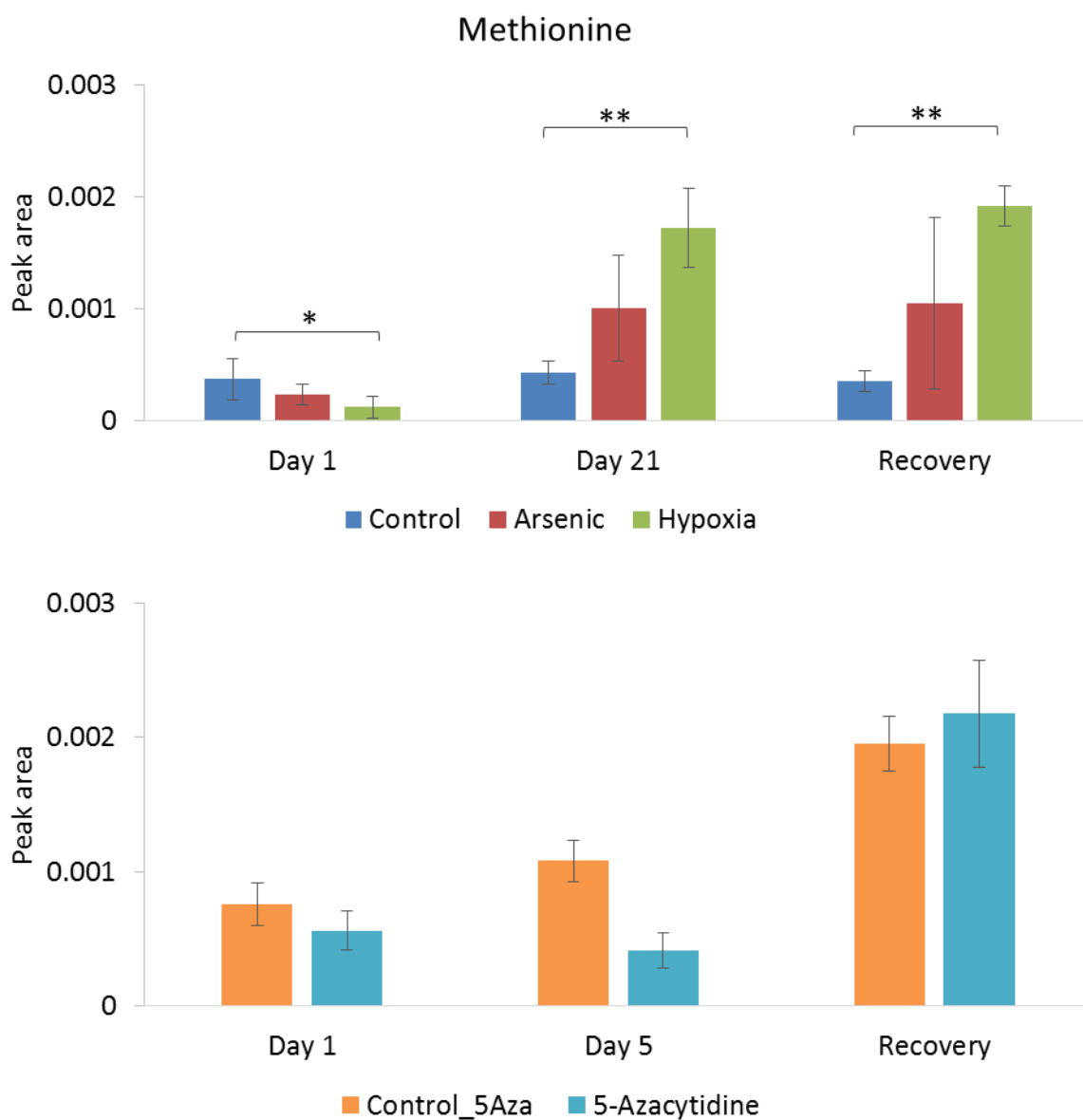
Figure 5.23 Methionine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *(p < 0.05); **(p < 0.01).
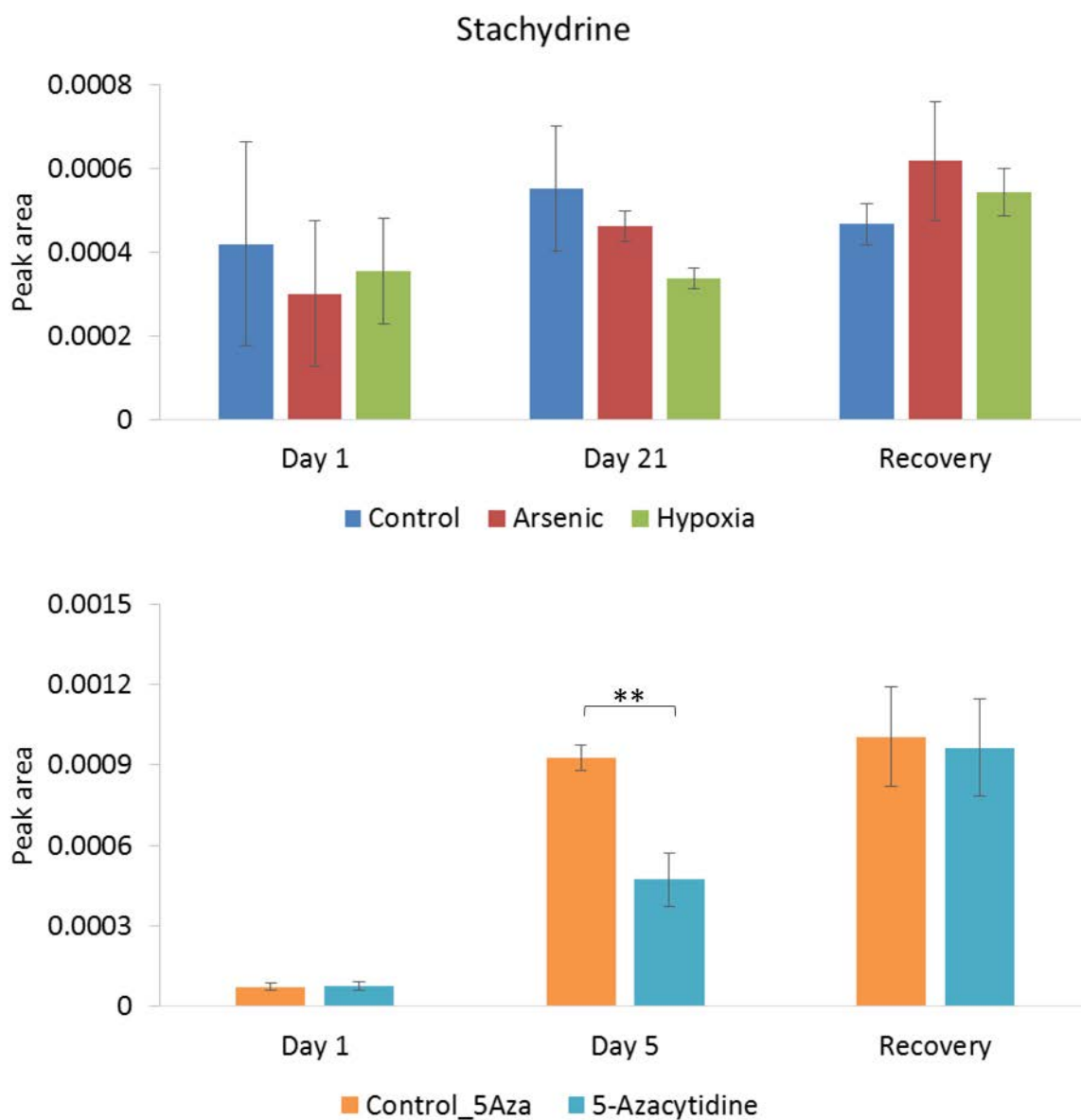
Figure 5.24 Stachydrine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. **(p < 0.01).
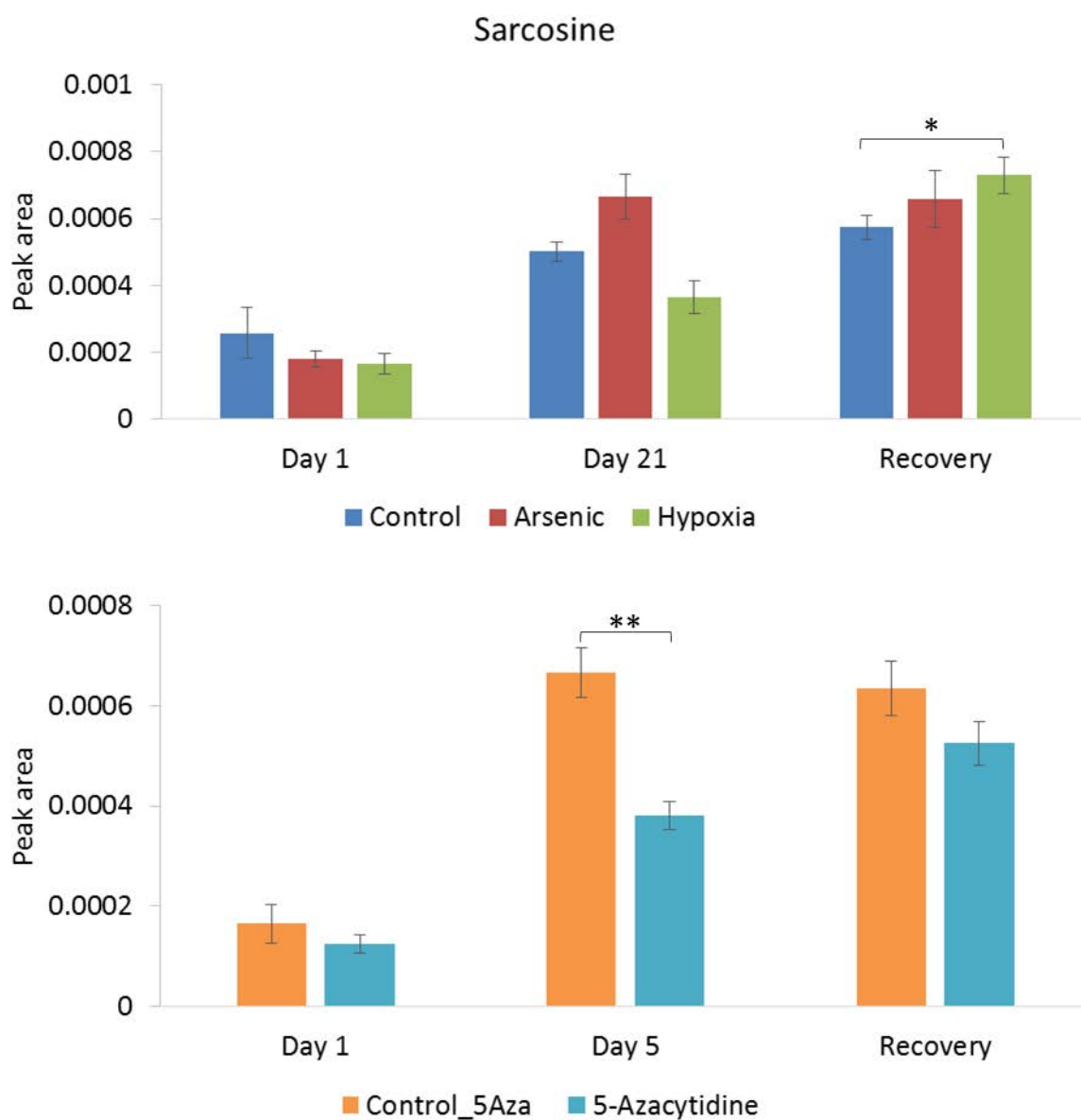
Figure 5.25 Sarcosine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *(p < 0.05); **(p < 0.01).
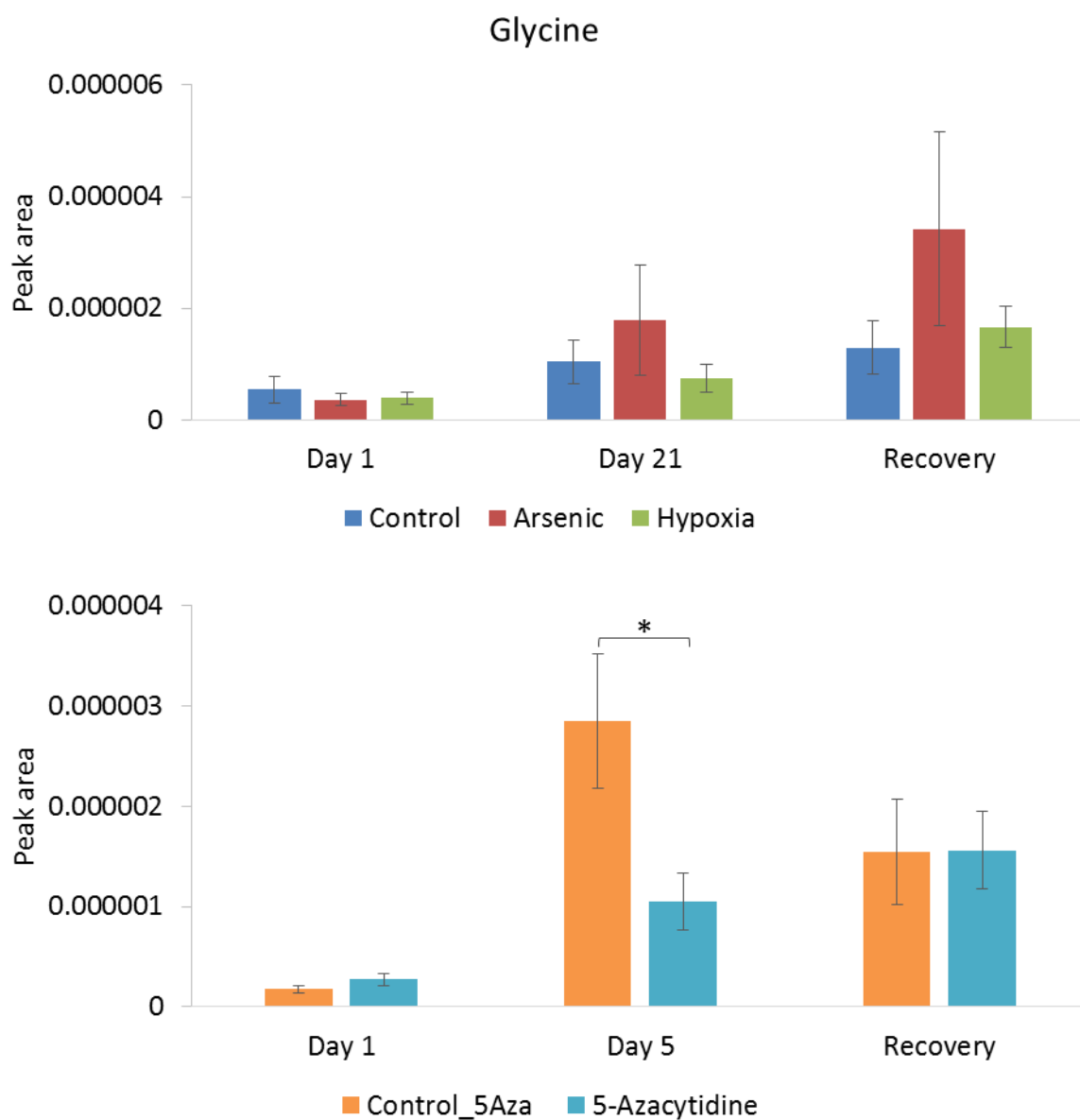
Figure 5.26 Glycine concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *(p < 0.05).
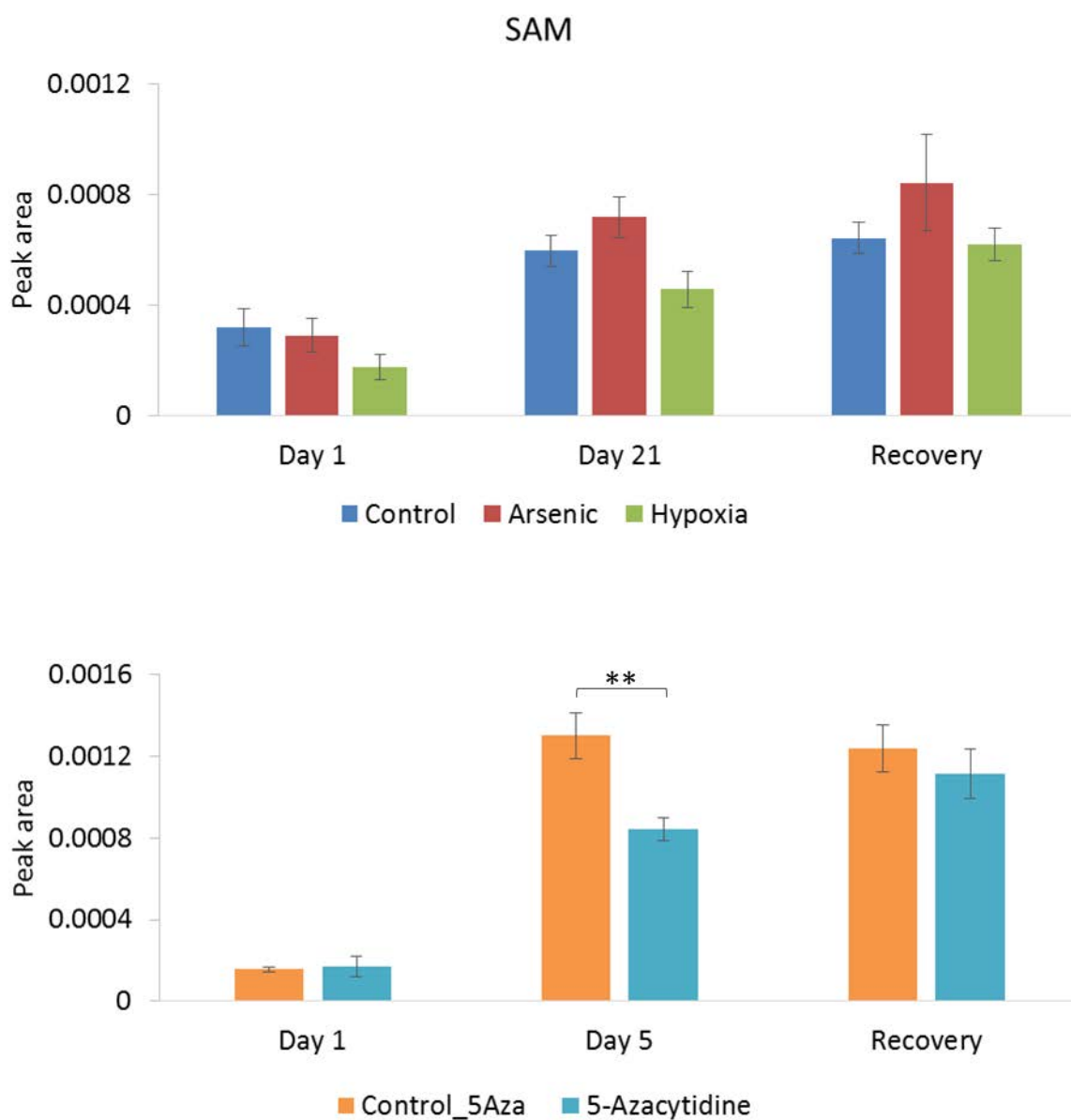
Figure 5.27 SAM concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. **(p < 0.01).
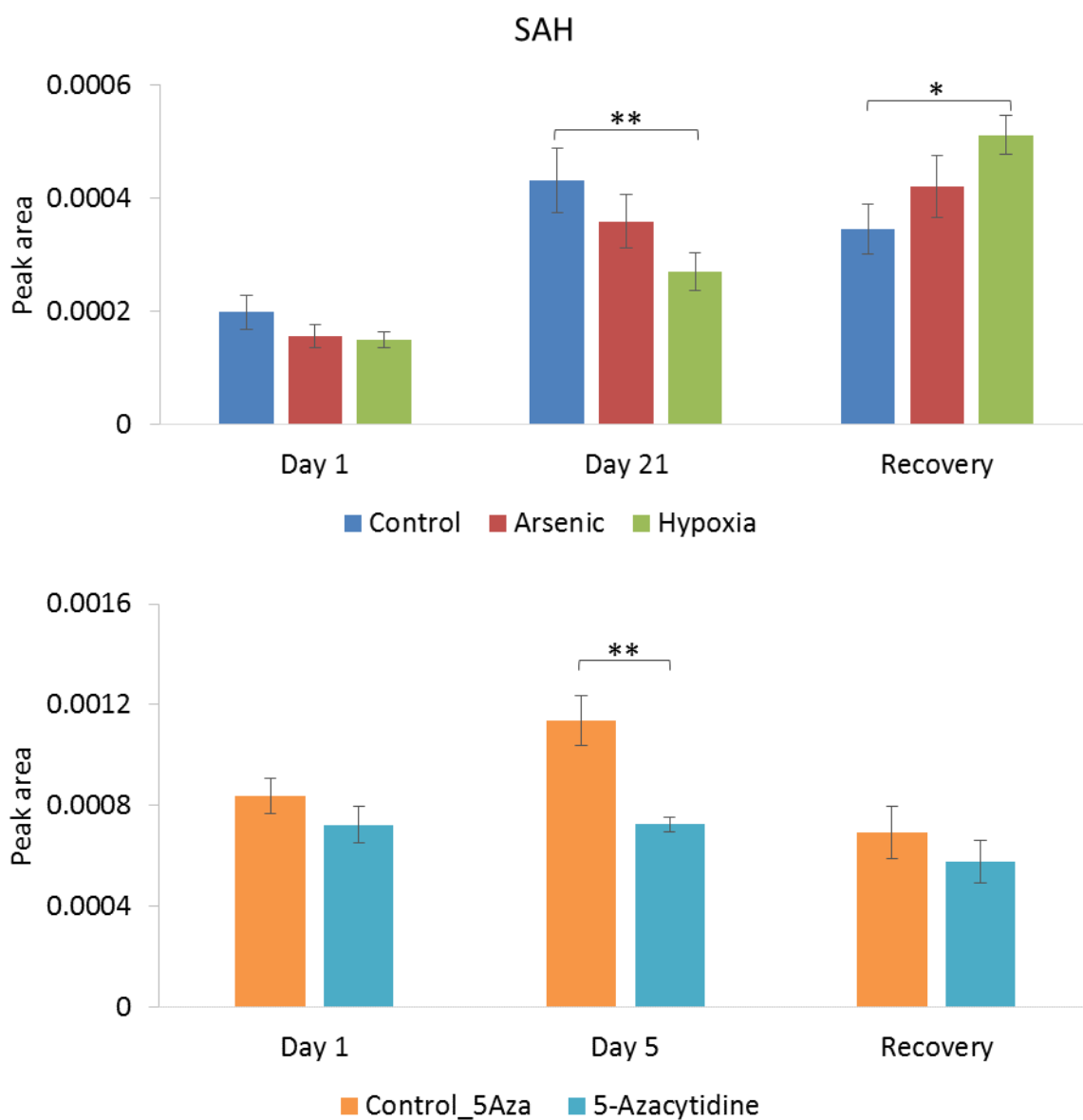
Figure 5.28 SAH concentration for samples exposed to arsenic, hypoxia and 5-azacytidine. Peak area was normalised to the peak are for the internal standard (S-adenosyl-L-methionine-d3). Six replicates were used for each group. Groups were compared using Mann-Whitney statistical test. *($p < 0.05$); **($p < 0.01$).

5.4   Discussion

5.4.1 Stress-specific effects

The effects of several stressors were investigated on the expression and methylation of selected genes and levels of metabolites involved in the one-carbon pathway. In general, exposure to an environmentally relevant concentration of arsenic induced little response in all the investigated end-points. Hypoxia significantly affected the different endpoints analysed (gene expression, metabolites concentration, phenotypic measurements and DNA methylation). In contrast to hypoxia, 5-azacytidine induced changes mainly in DNA methylation with some changes observed for the gene expression and metabolites.

Arsenic has been reported to induce alterations at DNA methylation level (Reichard and Puga, 2010). In most mammalian species, the mechanism of arsenic detoxification is centred on the methylation and reduction of inorganic arsenic generating monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA). Both metabolites are less reactive and less toxic than its inorganic form. S-adenosylmethionine (SAM) is the main methyl donor for this reaction (Vahter and Concha, 2008; Vahter, 2002). Thus, both DNA methylation and arsenic detoxification pathways compete for the same methyl donor, therefore the presence of arsenic can indirectly impact DNA methylation. Furthermore, is reported that arsenic can disrupt the activity of DNMTs possibly leading to global hypomethylation (Reichard and Puga, 2010; Reichard *et al.*, 2007; Zhao *et al.*, 1997).

Despite several studies suggesting effects on DNA methylation following arsenic exposure, mostly performed in cell cultures, no effects were observed for *Daphnia* in this study. No differentially methylated regions were identified with whole genome bisulfite sequencing and

no effects on the expression levels of the genes involved in one-carbon metabolism were detected. Only adenosine concentration was increased, which could not be unambiguously related to alterations in the DNA methylation pathway.

Hypoxia affects a wide range of biological processes. It is often caused by increase in anthropogenic input of organic matter and nutrients into water bodies that leads to a growth in primary production, reducing aquatic oxygen concentrations (Wu, 2002). Oxygen depletion can impair growth, disturb reproduction and even cause death of aquatic populations (Long *et al.*, 2015).

Due to the importance of the biological pathways that can be affected by hypoxia, aquatic organisms have developed mechanisms of acclimation and tolerance to oxygen depletion (Long *et al.*, 2015). Most of these mechanisms are activated and regulated by hypoxia inducible factors (HIFs). However, some studies have shown regulation by epigenetic mechanisms (Brown and Rupert, 2014; Hattori *et al.*, 2015; Lachance *et al.*, 2014; Tsai and Wu, 2014; Tudisco *et al.*, 2014).

The organisms exposed to hypoxia presented several changes in metabolites concentration and in gene expression. A summary of the results for gene expression and changes in the levels of metabolites for the groups exposed to hypoxia are presented in Figure 5.29.

In general, the effects were spread along the pathway, affecting it at different points. The effects could either be observed on essential nutrients, like choline, or directly on transcription of the enzymes involved in DNA methylation reaction. The changes in the levels of metabolites and gene expression were affected both by acute and chronic exposure. However, it is interesting to note that the effects were mostly divergent. For example, 1 day of exposure to hypoxia led to a decrease in methionine levels, however, chronic exposure (21

days) caused an increase in the concentration of this metabolite when compared to respective controls. Methionine concentration was also increased after the recovery period. Methionine and ATP are the precursors of SAM in a reaction catalysed by methionine adenosyltransferase (*MAT*). No statistically significant changes were seen for *MAT* gene expression; however, SAM concentration was slightly decreased for all timepoints. This could indicate an inefficiency of *MAT* enzyme under hypoxic conditions or an effect due to limitation in ATP due to oxygen deprivation.

The same effect was observed on the expression of *DNMT1*, being upregulated at day 1 and downregulated at day 21. These examples highlight differences in the effects of acute and chronic exposure. The physiological responses are most likely divergent when comparing acute and chronic responses to stressors, and compensatory changes might happen with time. Acute responses to oxygen depletion are controlled by the HIF pathway. Studies have shown that HIF-1 biding is directly controlled by the presence of methylation of specific CpGs within hypoxia responsive elements (HRE) (Mariani *et al.*, 2014; Watson *et al.*, 2010). Besides that, the maintenance of HIF regulation for long periods of time could be costly to the organism. Epigenetic mechanisms could work as a long term regulator of hypoxia responsive genes.

Despite the possible effects, no changes in region-specific methylation were observed for the samples exposed to hypoxia. However, it is important to remember that the analyses were conducted using a draft genome, therefore, some genes involved in hypoxia response pathway were missing and were not covered with WGBS.
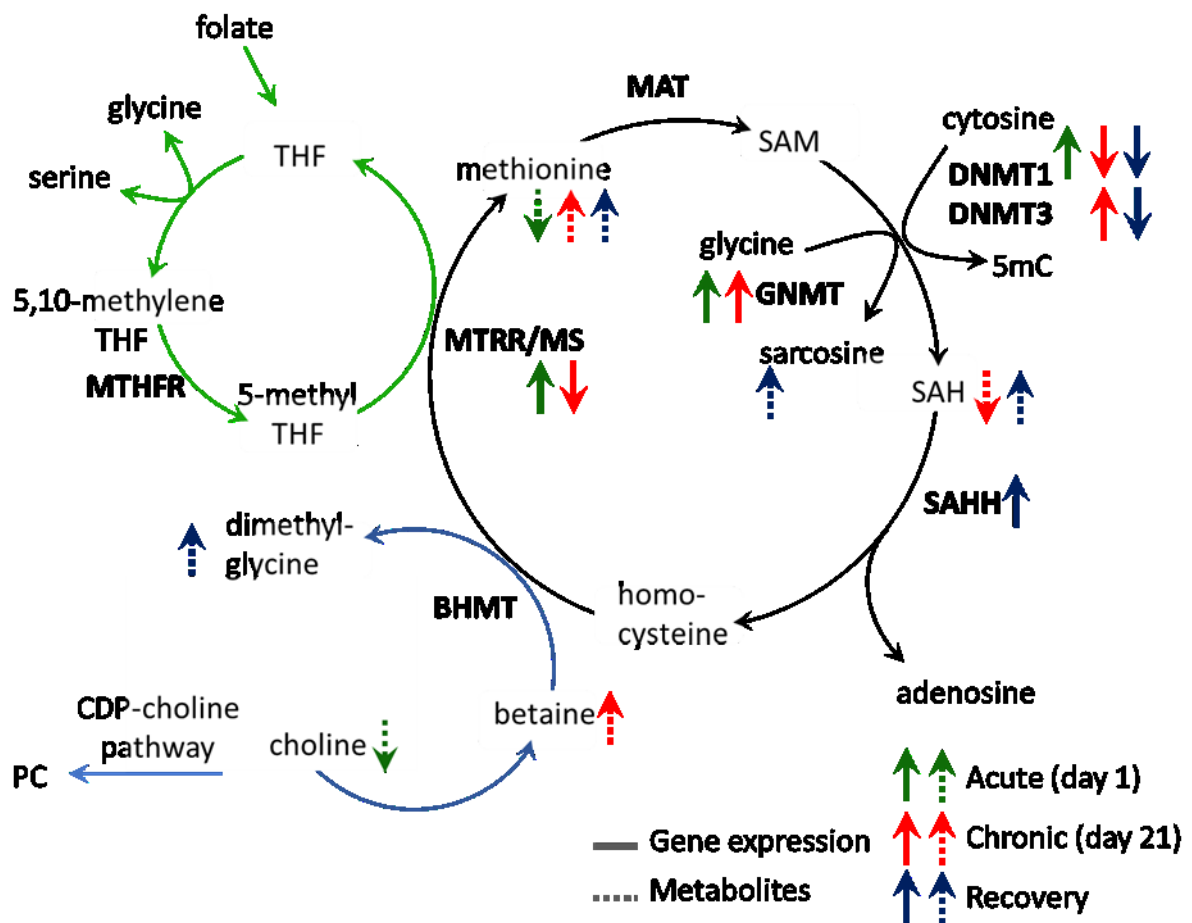
Figure 5.29 Overview of the changes observed for hypoxia after acute exposure (day 1), chronic exposure (day 21) and recovery (day 28). The significant changes on gene expression and levels of metabolites in the one-carbon pathway when comparing control and hypoxia exposure are presented in this figure. Arrows: decrease or increase. Continuous lines show results for gene expression while intermittent lines show changes in the levels of metabolites. Different colours are used for different time points: green – acute exposure (day 1); red – chronic exposure (day 21); blue – recovery (day 28).

Abbreviations: BMHT: betaine- homocysteine methyltransferase, CDP-choline: cytidine diphosphate-choline, DNMT: DNA methyltransferase, GNMT: glycine N-methyltransferase, MAT: methionine adenosyltransferases, MS: methionine synthase, MTHFR: methylenetetrahydrofolate reductase, MTRR: Methionine synthase reductase, PC: phosphatidylcholine; SAH: S-adenosylhomocysteine, SAHH: S- adenosylhomocysteine hydrolase, SAM dependent MT: S-adenosylmethionine dependent methyltransferase, THF: tetrahydrofolate.

DNMT2 was originally assigned as a DNA methyltransferase due to its highly conserved DNA methylase domain. Goll *et al.* (2006) have demonstrated that in fact this DNMT methylates small tRNAs rather than DNA, although the function of RNA methylation is not yet clear.

For samples exposed to hypoxia, *DNMT2* was upregulated for all time points. Two studies have indicated the relationship between *DNMT2* overexpression and stress tolerance (Lin *et al.*, 2005; Schaefer *et al.*, 2010).

TET enzymes are involved in the demethylation pathway through the conversion of 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine (Cadet and Wagner, 2014; Ito *et al.*, 2011; Kohli and Zhang, 2013). *TET1* expression is known to be altered in neuroblastoma cells exposed to hypoxia, leading to an increase in 5-hmC and demethylation of HIF biding sites (Mariani *et al.*, 2014). From the two *TET* homologs found in *Daphnia* genome, *TET*_1 presented the same expression pattern, being upregulated after day 21, returning to normal level of expression after recovery.

In general, the hypoxia effects demonstrated in *Daphnia* are related to changes in gene expression and metabolites concentration rather than alteration in DNA methylation profile. Nevertheless, according to the observed results, possible changes targeting DNA methylation cannot be discarded.

Contrary to the physiological effects expected for hypoxia, 5-azacytidine's mode of action is targeted to DNA methylation. It is an analogue of cytosine that can be incorporated to the DNA during replication. DNMTs are then sequestrated by 5-azacytidine and remain attached to the DNA being unavailable for further methylation (Stresemann and Lyko, 2008).

Global methylation reduction has already been cited to occur in *Daphnia* exposed to 5-azacytidine (Vandegehuchte *et al.*, 2010b). Several regions of the genome were shown to be

altered by 5-azacytidine exposure (in Section 4.3.5). Also, direct bisulfite sequencing PCR showed changes in DNA methylation for target regions. Despite the broad effect of this chemical, some regions did not present altered DNA methylation indicating that 5-azacytidine may have different effects on different parts of the genome. This could be related to chromatin structure, accessibility of DNA or CpG density.

Figure 5.30 shows a summary of the results for gene expression and changes in the levels of metabolites on the one-carbon pathway for the groups exposed to 5-azacytidine. In general, only gene expression showed alterations with acute exposure. However, after 5 days of exposure, the effects could be observed for the expression of the enzymes of the one-carbon pathway, especially those responsible for the methylation of DNA. Effects on metabolite levels were only observed after chronic exposure.

DNMT3 expression was firstly downregulated at day 1, however, at day 5 both *DNMT1* and *DNMT3* transcripts were upregulated. In this context, both SAM and SAH concentrations were decreased following 5-azacytidine exposure. According to James *et al.* (2002), SAH is known to act as a regulator of *DNMTs* expression. Often, high levels of SAH are known to repress expression of *DNMTs*. In this case, the lower levels of SAH, caused by lower rates of cytosine methylation due to *DNMTs* inefficiency, could be acting as a stimulus for *DNMT* expression. The lower levels of SAM could be due to the downregulated expression of *MAT* as well as the reduction of methionine concentration, although not statistically significant different.
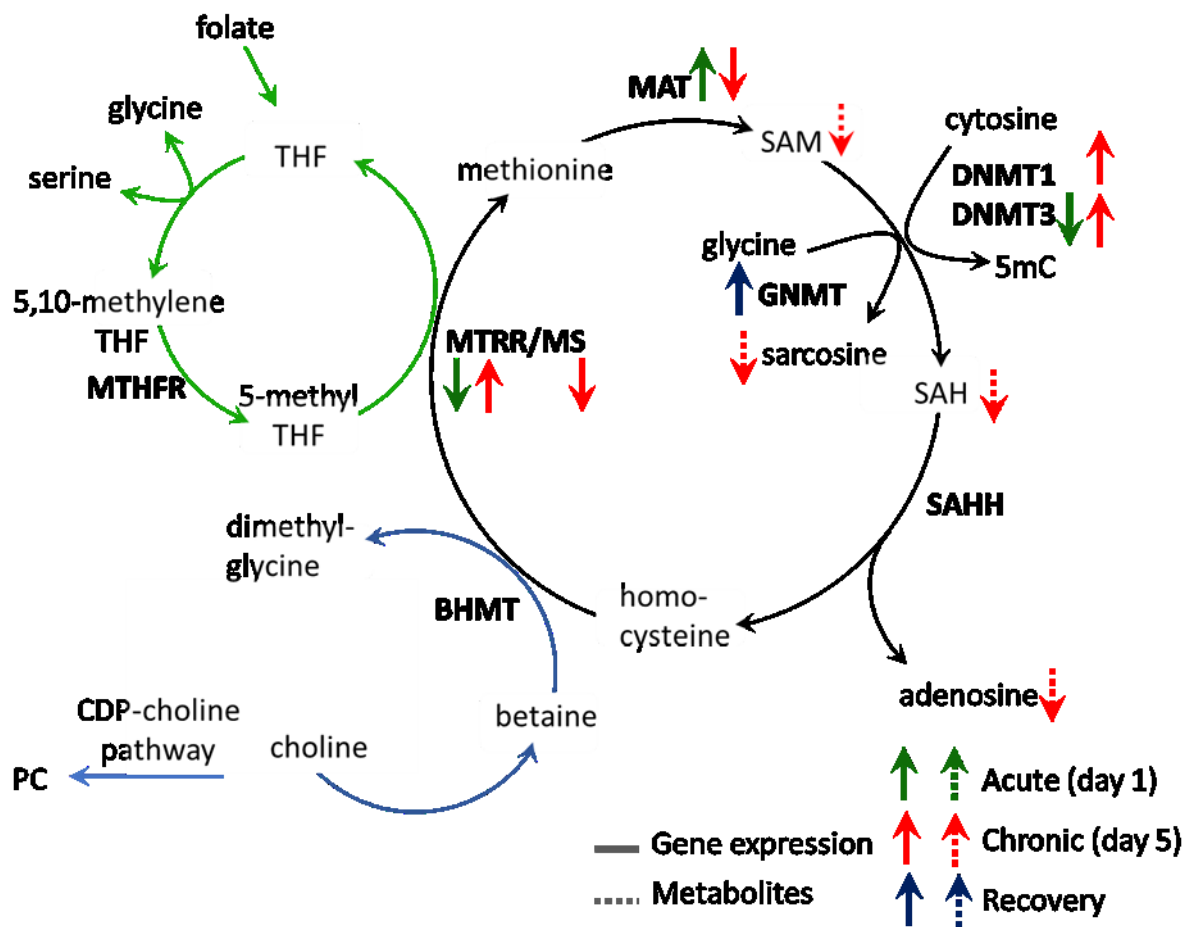
Figure 5.30 Overview of the changes observed for 5-azacytidine after acute exposure (day 1), chronic exposure (day 5) and recovery (day 12). The significant changes on gene expression and levels of metabolites in the one-carbon pathway when comparing control and 5-azacytidine exposure are presented in this figure. Arrows: decrease or increase. Continuous lines show results for gene expression while intermittent lines show changes in the levels of metabolites. Different colours are used for different time points: green – acute exposure (day 1); red – chronic exposure (day 5); blue – recovery (day 12).

Abbreviations: BMHT: betaine- homocysteine methyltransferase, CDP-choline: cytidine diphosphate-choline, DNMT: DNA methyltransferase, GNMT: glycine N-methyltransferase, MAT: methionine adenosyltransferases, MS: methionine synthase, MTHFR: methylenetetrahydrofolate reductase, MTRR: Methionine synthase reductase, PC: phosphatidylcholine; SAH: S-adenosylhomocysteine, SAHH: S- adenosylhomocysteine hydrolase, SAM dependent MT: S-adenosylmethionine dependent methyltransferase, THF: tetrahydrofolate.

5.4.2 Cumulative and persistent changes in animals exposed to 5-azacytidine and hypoxia

Observing the different time points of exposure, two main types of information can be assessed. Firstly, comparing the data for day 1 and day 5 or 21 the cumulative effects of exposures can be identified. Looking at the recovery timepoint, the persistence of the alterations can be analysed.

Focusing on the results for gene expression and metabolites quantitation, both hypoxia and 5-azacytidine appear to have a cumulative effect, since mostly divergent effects were observed for acute and chronic exposure. For gene expression, a shift in response was observed, where genes that were upregulated at day 1 were downregulated at the end of exposure and the ones downregulated at day 1 are upregulated at the end. These results indicate that the effects of exposures are divergent with time, and are different from the ones identified for acute exposures.

For hypoxia exposures, no effects were observed in DNA methylation. However, the effects on gene expression and levels of metabolites still remained after the normalisation of oxygen levels, suggesting a long lasting outcome on the organisms' biology.

It is known that the regulation of haemoglobin expression through the HIF pathway can only be maintained in low oxygen environment since in normoxic conditions the HIF-1α is degraded (Semenza, 2007). Since the metabolism was still altered and haemoglobin concentration was elevated even when moved to normal oxygen conditions (recovery), these findings suggest a possible long term regulation of haemoglobin expression that could be due to an epigenetic mechanism, however more studies need to be conduct to explore this hypothesis. As an

example, the time necessary for haemoglobin degradation in the *Daphnia*'s haemolymph needs could be measured to assess the protein stability in this animal.

Regarding the possible persistence of the effect, groups exposed to 5-azacytidine did not present altered gene expression and metabolite levels after the end of exposure and period of recovery. This indicates that the organisms were able to recover to normal levels of gene expression and metabolites even after a period of chronic exposure to 5-azacytidine.

One the other hand, BSP data for groups exposed to 5-azacytidine indicate both a cumulative effect on region-specific methylation and a persistent effect. For DMRs that presented altered methylation level, a tendency for lower methylation levels at day 5 when compared to day 1 was observed. This is consistent with the mechanisms of action of 5-azacytidine, where during replication the cytosine analogue is incorporated to the DNA causing a cumulative effect on the DNA methylation levels at specific CpG sites. However, it is already known that the effects of 5-azacytidine exposure can be blocked after extended exposure, due to depletion of nucleotides, arrest of the cell cycle and cytotoxicity related to high doses (Choi *et al.*, 2007). Therefore, it is impossible to reach complete demethylation after exposure to this chemical.

Looking at the levels of DNA methylation after a recovery period, some regions analysed recovered to the normal levels. Nonetheless, some have still less methylation at specific sites that the control samples.

Therefore, an important highlight of these findings is that although no effects were observed for gene expression and metabolite concentrations after the removal of the exposure, DNA methylation was still decreased in target regions of the genome. Consequently, based on these results, it is suggested that some effects of 5-azacytidine exposure on CpG methylation are reversible, but not all. This is consistent with the mechanism of "epigenetic memory"

where it is proposed that chemicals or stressors can cause specific changes in DNA methylation profile that can be maintained during life time (Bird, 2002; Mirbahai and Chipman, 2014).

The "epigenetic memory" can also play a role in the transmission of disrupted epigenetic information to following generations. The possibility of transgenerational inheritance across multiple generations has been suggested for several organisms, including mammals, plants and invertebrates (Hauser *et al.*, 2011; Skinner, 2014; Skinner *et al.*, 2010). The identification of DNA methylation disruption that is not reversed after the removal of the stressors is another strong indication that those changes can act later in the organism's life.

5.5 Conclusions

In this study, it was demonstrated that most of the stressors can affect the organisms in a stress-specific manner. Arsenic concentration was selected based on reported values that were found in different environments (Smedley and Kinniburgh, 2005) and no effects were observed for the endpoints analysed.

Differing from arsenic, hypoxia affected the expression of genes and the levels of metabolites of the one-carbon pathway. Oxygen depletion also caused a decrease in body length showing that growth and energy metabolism were impaired. Thus, the results suggested physiological responses to hypoxia and effects that could target the DNA methylation machinery, although no alterations in region specific DNA methylation could be identified.

It was also possible to identify long-lasting changes in DNA methylation caused by 5-azacytidine. This in consistent with the proposed concept of epigenetic memory and highlights the need for further studies to understand the implications of long-lasting effects of altered

DNA methylation profiles, especially because gene expression and metabolite concentrations were restored after the removal of the stressor.

These results are very relevant when thinking about the current approaches for chemical risk assessment and environmental monitoring. Usually, no test is employed to evaluate the effects of chemicals on the transcriptome and metabolome, although efforts have been to develop such tests.

Chapter 6

General discussion

The overall aim of this thesis was to describe the methylome of *Daphnia magna* as well as to investigate if the methylome of *Daphnia* is responsive to environmentally relevant exposure conditions. To achieve this, several specific aims were outlined (section 1.6). Initially, the overall pattern of DNA methylation as well as the DNA methylation machinery were characterised for *D. magna* (Chapter 3). Following this, different methods of differentially methylated regions (DMRs) identification were tested and compared after treating *Daphnia* with 5-azacytidine, a chemical that is known to alter DNA methylation. Subsequently, the chosen method of analysis of DMRs was applied to the samples obtained from chronic exposures to environmentally relevant concentration of two selected stressors, arsenic and hypoxia as well as the positive control, 5-azacytidine (Chapter 4). The effects of the stressors were also evaluated after acute and chronic exposure, on DNA methylation and the epigenetic machinery through gene expression and the concentration of metabolites of one-carbon pathway. The persistence of the identified alterations was also analysed to test the concept of epigenetic memory (Chapter 5).

6.1 Evolution of DNA methylation and the potential benefit of using the model organism *Daphnia magna* in epigenetics research

The first aim of this research was to characterise the overall distribution pattern of DNA methylation and the second aim was to identify the mechanisms involved in DNA methylation in the crustacean *Daphnia magna*, an important model organism for ecotoxicology studies. As reviewed in section 1.3 and discussed in chapter 3 (section 3.4), the DNA methylation patterns in vertebrates and invertebrates vary dramatically (Breiling and Lyko, 2015; Feng *et al.*, 2010a; Jiang *et al.*, 2014). Nevertheless, gene body methylation is a conserved feature of

these evolutionary distant organisms, including plants, fungi and invertebrate and vertebrate animals, although it has been nearly lost in some organism in a lineage-specific manner (Feng *et al.*, 2010a; Glastad *et al.*, 2011).

Although the occurrence of gene body methylation is well conserved during evolution, the function of gene body methylation is not completely understood (Feng *et al.*, 2010a; Tweedie *et al.*, 1997). It is hypothesised that it is often linked with transcriptionally active chromatin regions. The organisms' genes can also be classified into two distinct groups, which are high and low methylated genes (Dixon *et al.*, 2016).

DNA methylation has been suggested to increase mutation rates, reducing the numbers of CpG dinucleotides in DNA sequences (Goll and Bestor, 2005; Zemach *et al.*, 2010). This leads to the idea that methylated genes should show reduced sequence conservation among different taxa. However, studies conducted in insects and plants have contradicted this hypothesis. In fact, highly methylated genes have shown high sequence conservation in invertebrates (Dixon *et al.*, 2016; Sarda *et al.*, 2012).

Daphnia magna presents a methylation pattern very similar to those found for other invertebrates, showing increased DNA methylation in gene bodies, particularly on 5' ends, rarely occurring in transposable elements and repetitive sequences and with CGIs largely unmethylated. Also, DNA methylation is restricted to the CpG dinucleotide context. The restriction of methylation to sites in gene bodies corroborates the low occurrence of methylated cytosines evaluated by global measurements (Chapter 3, section 3.3.5.2).

The strain used in this study presented the lowest level of DNA methylation for this species measured so far (0.14%). It is important to cite that both *D. magna* strains analysed before (Xinb3: 0.52% and Iinb1: 0.49%) are inbred strains. Potentially this could have influenced the

normal methylation pattern and caused a higher level of cytosine methylation of their genome (Asselman *et al.*, 2015). Therefore, it is proposed that, in contrast to quantitative genetic studies that rely heavily on using inbred strains (Xinb3, Iinb1), epigenomic studies should be conducted using non-inbred strains (i.e. Bham2) with a normal DNA methylation pattern. The possible deleterious effects of inbreeding (Charlesworth and Willis, 2009), and the differences in global DNA methylation detected between the two inbred strains (up to 0.52%) and the strain used in this study (0.14%), indicates that it is preferable to use non-inbred strains in epigenetics studies, at least until the exact effect of inbreeding is investigated on DNA methylation profile.

The identified DNA methylation profile in *Daphnia* (Figure 3.11) and other invertebrates is different from those of vertebrates (Chapter 1, section 1.3). Vertebrate animals present a global distribution of methylation, with ummethylated regions, often located in promoter sequences of different genes. This feature, considering the species analysed so far, seems to be specific to vertebrata, since it was not found in lower chordate taxa (*e.i. Ciona* sp.) (Breiling and Lyko, 2015; Tweedie *et al.*, 1997). It has been proposed that the global distribution of methylated cytosine and the higher levels of methylated cytosine arose with the need for silencing of transposable elements and repetitive regions that were transferred through sexual reproduction (Zemach *et al.*, 2010).

Despite the current knowledge on differences in DNA methylation distribution and function between vertebrates and invertebrates, often epigenetic studies exclusively focus on analysing methylation changes around the promoter regions of the genes and CGIs, without considering the taxa of their model system. This could be due to lack of information on DNA methylation in invertebrate species, since studies on DNA methylation in invertebrates were

for a long time neglected. This lack of investment in understanding the role of DNA methylation in invertebrate models stems from the fact that DNA methylation is nearly completely absent in the two main invertebrate model organisms for genetic studies, *D. melanogaster* and *C. elegans* (Capuano *et al.*, 2014; Park *et al.*, 2011; Simpson *et al.*, 1986). This resulted in generalising the finding to other invertebrate organisms and abandoning research on DNA methylation in invertebrates until recently. As shown in other invertebrate species, such as *C. floridanus*, *N. vitripennis* and *A. mellifera*, DNA methylation is present in invertebrates and plays a significant role in regulating various aspects of their life (Beeler *et al.*, 2014; Drewell *et al.*, 2014; Glastad *et al.*, 2015; Pegoraro *et al.*, 2016; Rasmussen and Amdam, 2015; Wang *et al.*, 2013).

Daphnia magna has the potential to revolutionise our understanding of the function of epigenetic factors in invertebrates. As presented in section 1.5.1, there are great advantages of using *D. magna* in epigenetic studies, as they are easy to maintain in large numbers under laboratory conditions, they have low genetic diversity due to their parthenogenetic reproduction, their genome is responsive to a wide range of stimuli, they are phenotypically plastic and a wide range of genomics resources is being developed for them. Furthermore, there is extensive knowledge regarding the ecology of this taxon, which can be used to help address environmentally relevant questions, including exposure to environmentally relevant stressors, trophic level interactions and evolution biology.

Therefore, *D. magna* is a valuable addition to the model organisms used for epigenetic studies and can contribute to the knowledge on DNA methylation in arthropods and in invertebrates in general. It can also be used to investigate the interactions between various epigenetic mechanisms, as already described (Harris *et al.*, 2012; Robichaud *et al.*, 2012).

6.2 Pipeline of analysis of WGBS data and applicability for non-model organisms

Whole genome bisulphite sequencing has been applied for the analysis of DNA methylation modifications, detecting methylated cytosines at single nucleotide resolution throughout the entire genome. An important step for accurate detection of methylated cytosines is to use an appropriate software for mapping and visualisation of WGBS data (Lee *et al.*, 2015). Several programs have been used for methylation mapping. In this thesis, Bismark software was chosen for mapping of the sequenced reads (Krueger and Andrews, 2011).

After the initial visualisation of the global DNA methylation profiling, often the next step is to identify the regions with differential methylation across a set of samples. Differentially methylated regions (DMRs) can be single cytosines or large regions, such as entire genes. The definition of the region length is totally arbitrary and will depend on the biological question that needs to be answered and the resources available (Bock, 2012).

Different approaches can be used to identify DMRs. Numerous statistical methods and software packages have been developed to compare and detect regions with altered DNA methylation (Rackham *et al.*, 2015). Consequently, it is extremely important to select the most appropriate statistical approach based on the experimental design and the model organism in order to identify and analyse DMRs.

In this thesis, DMR identification was performed using the software SeqMonk. It allows the visualisation of the data using the direct output from Bismark analyses and can be used for further analysis of the data. It is possible to set custom genomes and annotation tracks,

allowing analyses of non-annotated, or incomplete, genomes for DNA methylation studies (Chatterjee *et al.*, 2012).

This software allowed the use of different methods of DMRs detection. These were classified as "biased", when the methylation levels were compared using a predefined region, such as genes or CGIs, or "unbiased", when the comparison was performed without any predefined region and covered the entire genome (details in Chapter 4, section 4.3.3).

It is important to note that the majority of pipelines for WGBS analyses were developed to be used in studies with mammalian species, such as human and mouse. An advantage of using these species is the availability of comprehensive genome sequence, assembly and annotation, meaning that not only the genes are annotated but several regulatory regions are also identified. In consequence, conclusions can be draw easily. As discussed in section 4.4.1, for the studies in species such as human and mouse, the use of "biased" analyses has great advantages. It limits the number of regions being tested, improving statistical power and targeting the analyses to specific features (i.e. genes, CGIs, promoters, exons/intron, enhancers) (Baumann and Doerge, 2014). However, for species, such as *D. magna* with genomes under construction, targeting the analyses to annotated regions could be less informative than analysing the whole genome. As demonstrated in section 4.3.3, the use of annotated genes caused a large averaging effect on the methylation levels in *Daphnia*. Additionally, the use of a biased analysis targeted to already known regions for detection of differential methylation limits the discovery of novel regions presenting altered methylation profiles (Robinson *et al.*, 2014). Consequently, as the genome of *D. magna* was smaller than anticipated and not well annotated, the DMRs identification in this study was performed using an unbiased approach (Chapter 4, section 4.3.3.2.3). Additionally, of the proposed unbiased

methods, the best approach for DMR analyses for the conditions used in this study was to use the level of methylation of single cytosines and compare the groups using a windowed replicate test. This allowed the search for DMRs in the entire genome, without bias to annotated regions, and the identification of consistent effects on DNA methylation along short sequences (100bp) (Zhong *et al.*, 2013).

The use of short windows for the statistical tests was arbitrary. This was to achieve greater resolution of the observed effects and to easily target the identified region, providing greater biological relevance to the findings (Beissinger *et al.*, 2015; Robinson *et al.*, 2014; Wang *et al.*, 2015). Importantly, due to the large number of statistical tests performed for the comparison of genome loci, it is crucial to correct for multiple testing. In this study, the Benjamini and Hochberg multiple test correction was used. With the comparison of large numbers of CpGs and multiple test correction, there is a tendency for a large number of false negative values (Bock, 2012). The use of a windowed test allowed the identification of neighbouring CpGs that presented similar differences in DNA methylation.

In conclusion, there are different approaches that can be used for the identification of DMRs. The choice of software and parameters used is of extreme importance when analysing DNA methylation data and need to be appropriate to the organism used and the availability of its genomic resources. For organisms with incomplete genome sequences, the analysis is not straight forward and the pipelines of analysis often need to be modified and adapted to each circumstance.

The use of software that allowed input of a custom genome track was an important factor in the software selection step. It did not limit the analysis only to genomes publically available or to those entirely sequenced and annotated. The methylation level distribution could be

visualised over the entire genome and could be tested for DMRs, even in organisms with a non-sequenced or non-annotated genome.

6.3   Exposure approaches and the implications for toxicity testing

Daphnia magna and other *Daphnia* species are currently used in ecotoxicological testing. The guidelines provided by the Organisation for Economic Co-operation and Development (OECD) are the main literature that guides the laboratory procedures. The acute immobilisation test and the reproduction test are standardised by OECD test number 202 and 211, respectively (OECD, 2012, 2004).

The acute immobilisation test assesses the effects of chemicals and stressors on the mortality/immobilisation of the animals after an acute exposure, usually for 24 or 48 hours. For this, *Daphnia* neonates are exposed to different concentrations of the test substance within 24 hours of their release from their mothers. Then, immobilised animals are recorded at the end of the period of exposure. Results are expressed as the concentration that causes immobilisation of 50% of the organisms (EC50) (OECD, 2004). The *Daphnia* reproduction test records first the mortality of the animals, then the number of offspring produced and the effect caused by chronic exposure to the tested stressor (OECD, 2012).

The first point that should be considered to advance the current protocols for toxicological testing is the design of the exposures. Laboratory exposures should aim to better represent of what is happening in the natural environment. Current exposure procedures are based on using *Daphnia* that are less than 24hrs old. This means that the exposures miss the sensitive period of embryogenesis. In natural environments animals are usually exposed to compounds in their environment throughout their life, including during embryogenesis. Current

toxicological tests do not take this into consideration. In this study the animals were exposed form the beginning of their development, when still in the mothers' brood pouch, to cover any possible interval of susceptibility to the epigenetic mechanisms that the animals could present during this period. Currently, the closest approach to this are the multigenerational studies that are being conducted in *Daphnia*. Several pollutants, such as propranolol and carbendazim have shown deleterious effects in *Daphnia* exposed for multiple generations. Indeed, some adverse effects of chemicals, especially compounds that are known to bioaccumulate, can become more severe following several generations of exposure. Furthermore, sometimes subtle adverse effects of chemicals are only recognised after persistence into the following generations (Jeong *et al.*, 2015; Silva *et al.*, 2016). Therefore, this method of exposure is much more representative of what is encountered in the environment and should improve our understanding of the real effects of environmental exposures.

Secondly, it is important to highlight that the usual endpoints analysed in ecotoxicological tests using *Daphnia*, such as mortality and reproduction impairment, are very crude endpoints that are usually observed following exposure to high concentrations of a single compound, which are usually not environmentally relevant. It is also recognised that mode of action (MoA) of certain compounds are concentration dependent and the MoA of a compound can vary based on the concentration and duration of exposure (Nendza and Wenzel, 2006). In most circumstances animals in the environment are usually exposed to low concentrations of mixtures of compounds either for an extended or short period of time. Thus the observed effects are usually different (i.e. more subtle) from the ones detected in the laboratory setting and under high concentrations. Furthermore, some subtle effects can cause an adverse effect

at a later stage in an organism's life rather than having an immediate effect (Jirtle and Skinner, 2007). This highlights the need for introducing and considering a wider range of chemical testing guidelines and procedures.

For that reason, in recent years, the application of omics technologies in toxicology studies has increased dramatically. The use of transcriptomics, proteomics and metabolomics can help to study how chemicals affect aquatic and terrestrial organisms in the environment and to screen for the molecular events that occur in response to exposure to different stressors (Martyniuk and Simmons, 2016). In addition to these omics technologies, epigenomics can be used to identify effects of the stressors on the level of epigenetic mechanisms. It is known that effects on the epigenetic marks, such as DNA methylation, can potentially act later in the organism's life or be transmitted to further generations (Bird, 2002).

Therefore, it is important to add to the usual strategies applied in toxicological studies, tests that analyse the effects on genetic and epigenetic mechanisms to allow further investigations of the effects of the stressors on the organisms. This would add a new layer of information to current strategies in risk assessment, being used for the screening of effects and possibly helping to extrapolate and relate the effects observed in different organisms.

6.4 Stress-specific changes in DNA methylation and the concept of epigenetic memory

In this thesis, the effects of three stressors were investigated on the expression and methylation of selected genes and levels of metabolites involved in the one-carbon pathway (details in Chapter 5). In general, exposure to an environmentally relevant concentration of arsenic induced little response in the investigated end-points. Hypoxia significantly affected

all the different endpoints analysed (gene expression, metabolites concentration, phenotypic measurements and DNA methylation). In contrast to hypoxia, 5-azacytidine mainly induced changes in DNA methylation and the enzymes and metabolites closely related to the DNA methylation reaction.

Despite several studies linking arsenic exposure to changes in DNA methylation, no effects were observed for *Daphnia* in this study. Arsenic could affect DNA methylation due to the mechanism of detoxification that compete with DNA methylation for the same methyl donor (Reichard and Puga, 2010; Reichard *et al.*, 2007; Zhao *et al.*, 1997). However, no differentially methylated regions were identified with WGBS and no effects on the expression levels of the genes involved in one-carbon metabolism were detected. The reported EC50 value for arsenic is 7400 µg L-1 (Biesinger and Christensen, 1972). Therefore, it is possible that higher levels of arsenic are required to affect DNA methylation since the proposed mechanisms of toxicity requires the competition for the methyl donors.

Hypoxia in water bodies is often caused by increase in anthropogenic input of organic matter and nutrients that leads to a growth in primary production (i.e. algae bloom) leading to even higher levels of organic matter. The decomposition of the organic matter reduces the dissolved oxygen concentrations in the water (Wu, 2002). Oxygen depletion can affect a wide range of biological processes, causing the impairment of growth, disturb reproduction and even cause death of aquatic populations (Long *et al.*, 2015).

In general, the effects of hypoxia exposure were spread along the one-carbon pathway, affecting it at different points. The effects could be observed on essential nutrients, like choline, and directly on transcription of the enzymes involved in DNA methylation reactions.

Due to the importance of the biological pathways that can be affected by hypoxia, aquatic organisms have developed mechanisms of acclimation and tolerance to oxygen depletion (Long *et al.*, 2015). Some studies have shown hypoxic responsive elements regulated by epigenetic mechanisms (Brown and Rupert, 2014; Hattori *et al.*, 2015; Lachance *et al.*, 2014; Tsai and Wu, 2014; Tudisco *et al.*, 2014).

Daphnia has developed strategies to cope with low levels of oxygen. This was one of the main reasons for selecting hypoxia as a stressor and investigating the effects of low oxygen levels on the epigenetic mechanisms. Despite no direct effects on DNA methylation profile, hypoxia affects the one-carbon pathway. Therefore, a relationship between hypoxia and DNA methylation cannot be excluded. Additionally, several questions arose that can be the subject of future studies: 1) Are epigenetic mechanisms controlling and regulating the response to hypoxia or are the epigenetic changes secondary to hypoxia? 2) If epigenetics is the main mechanisms of response to hypoxia, can pollutants affect the epigenetic machinery and subsequently make the organisms susceptible to hypoxia effects by altering the main regulatory mechanism? 3) Can exposure to hypoxia cause persistent changes in the epigenome of an organism and what would be the consequences at the population level?

Contrary to the physiological effects identified for hypoxia, 5-azacytidine acts by targeting DNA methylation. It is an analogue of cytosine that can be incorporated into the DNA during replication. DNMTs are then sequestrated by 5-azacytidine and remain attached to the DNA being unavailable for further methylation reactions (Stresemann and Lyko, 2008).

Besides the global methylation reduction already cited to occur in *Daphnia* exposed to 5-azacytidine (Vandegehuchte *et al.*, 2010b), several regions of the genome were shown to be altered by 5-azacytidine exposure by WGBS and confirmed with direct bisulphite sequencing

PCR (in Section 4.3.6). Despite the broad effect of this chemical, altered DNA methylation was not observed in the entire genome. This indicated that 5-azacytidine may have different effects on different parts of the genome. Such differences could be related to chromatin structure, CpG density or accessibility of DNA.

The important finding related to 5-azacytidine exposure is the maintenance of the observed effects on DNA methylation after the removal of the stressor. A decrease in methylation level was observed for CpGs in the target regions, as demonstrated in section 5.3.1. Although the methylation level of some CpG sites were restored to the original level after removal of the stressor, the methylation levels of some CpG sites did not recover. This indicates that certain modifications to the methylome, in contrast to changes at transcriptome and metabolome levels, are more persistent and thus have the potential to alter the response of an organism even in the absence of a stressor. Thus, these findings provide more evidence in support of the concept of "epigenetic memory". Epigenetic memory means that chemicals or stressors can cause specific changes in DNA methylation profile that can be maintained during an individual's life time (Bird, 2002; Mirbahai and Chipman, 2014). Additionally, it can play a role in the transmission of disrupted epigenetic information to following generations. The identification of DNA methylation disruption that is not reversed after the removal of the stressors is another strong indication that those changes can act later in the organism's life or be transmitted to further generations. Therefore, in this study it was demonstrated that the stressors can affect the organisms in a stressor-specific manner and that it is consistent with the proposed concept of epigenetic memory.

The results of this study indicate the need for working towards developing a standard set of epigenetic assays for incorporation into current chemical risk assessment procedures as well

as refinement of the current OECD guidelines for chemical risk assessment. Using multiple 'omic approaches, such as transcriptomics, proteomics, metabolomics and epigenomics, will aid the process of understanding the MoA of chemicals and that of developing a set of biomarkers, which can be used to predict the effects of compounds both at individual and population levels.

6.5   Gaps in knowledge and future work

As demonstrated in chapter 3 and discussed in section 6.1, it was possible to identify the overall distribution of methylated cytosines in *D. magna*. However, due to the lack of genome resources, the description of the methylome and altered profiles is still incomplete. The genome generated in this thesis was not able to overcome the problem of a poor constructed genome. Therefore, further major advances with the *D. magna* methylome and knowledge of other epigenetic mechanisms, depend on the construction of a complete and fully annotated genome.

In this study whole organisms were used to obtain DNA samples for WGBS and direct bisulphite PCR analyses. It was useful to identify the normal DNA methylation distribution and the effects caused by stressors in *D. magna*. The further step would be to investigate the tissue-specific DNA methylation profile and possible effects of stressors. Single cell DNA methylation analyses and laser-capture microdissection could be alternative methods for tissue-specific analyses. However, optimisation of these techniques are likely necessary, since until now they have not been applied for studies in *Daphnia*.

The identification of persistent effects on DNA methylation profile indicates that the mechanisms of repair might not be efficient, at least within the same generation. Due to the

need to incorporate different exposure procedures to classical ecotoxicological assays and the recent findings regarding increased effects after multigenerational exposures, the investigation of epigenetic reprogramming events is essential, especially to determine the potential of transgenerational effects in this species. It will allow the identification of possible susceptible periods during *Daphnia*'s development that should be either considered or deliberately excluded from the assays.

Finally, with the development of genomic resources for *D. magna*, the construction of a detailed DNA methylation profile and the investigation of reprogramming events, future steps would be to work towards development of standard epigenetic assays for chemical risk assessment. However, it might be premature to suggest the use of epigenetic profiles as biomarkers of exposures, since there is still a lot to be investigated, the findings in this thesis have contributed to the development of the field and have highlighted the potential of DNA methylation studies in *D. magna* and the application of some concepts in studies with other species. Therefore, ideally in the next few years scientists will be able to construct a comprehensive database profiling the responses of various epigenetic markers to a range of chemicals and stressors and under multiple conditions using *Daphnia* as one of the key environmentally relevant species.

References

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., Walter, J., 2012. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genetics* 8, e1002750. doi:10.1371/journal.pgen.1002750

Argos, M., Chen, L., Jasmine, F., Tong, L., Pierce, B.L., Roy, S., Paul-Brutus, R., Gamble, M. V, Harper, K.N., Parvez, F., Rahman, M., Rakibuz-Zaman, M., Slavkovich, V., Baron, J.A., Graziano, J.H., Kibriya, M.G., Ahsan, H., 2015. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in bangladesh. *Environmental Health Perspectives* 123(1):64-71. doi:10.1289/ehp.1307884

Asselman, J., De Coninck, D.I., Pfrender, M.E., De Schamphelaere, K.A., 2016. Gene body methylation patterns in *Daphnia* are associated with gene family size. *Genome Biology and Evolution* 8 (4): 1185-1196. doi:10.1093/gbe/evw069

Asselman, J., De Coninck, D.I., Vandegehuchte, M.B., Jansen, M., Decaestecker, E., De Meester, L., Bussche, J. V, Vanhaecke, L., Janssen, C.R., De Schamphelaere, K.A., 2015. Global cytosine methylation in *Daphnia magna* depends on genotype, environment and their interaction. *Environmental Toxicology and Chemistry* 34(5):1056-1061. doi:10.1002/etc.2887

Athanasio, C.G., Chipman, J.K., Viant, M.R., Mirbahai, L., 2016. Optimisation of DNA extraction from the crustacean *Daphnia*. *PeerJ* 4, e2004. doi:10.7717/peerj.2004

Baccarelli, A., Bollati, V., 2009. Epigenetics and environmental chemicals. *Current Opinion in Pediatrics* 21(2):243–251. doi:10.1097/MOP.0b013e32832925cc

Baer, K.N., Goulden, C.E., 1998. Evaluation of a high-hardness COMBO medium and frozen algae for *Daphnia magna*. *Ecotoxicology and Environmental Safety* 39(3):201–206. doi:10.1006/eesa.1997.1627

Bastos Sales, L., Kamstra, J.H., Cenijn, P.H., van Rijt, L.S., Hamers, T., Legler, J., 2013. Effects of endocrine disrupting chemicals on *in vitro* global DNA methylation and adipocyte differentiation. *Toxicology in Vitro* 27(6):1634–1643. doi:10.1016/j.tiv.2013.04.005

Baumann, D.D., Doerge, R.W., 2014. MAGI: Methylation analysis using genome information. *Epigenetics* 9(5):698–703. doi:10.4161/epi.28322

Beeler, S.M., Wong, G.T., Zheng, J.M., Bush, E.C., Remnant, E.J., Oldroyd, B.P., Drewell, R.A., 2014. Whole-genome DNA methylation profile of the jewel wasp (*Nasonia vitripennis*). *G3* 4(3):383–388. doi:10.1534/g3.113.008953

Beissinger, T.M., Rosa, G.J., Kaeppler, S.M., Gianola, D., de Leon, N., 2015. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics Selection Evolution* 47, 30. doi:10.1186/s12711-015-0105-9

Bestor, T.H., 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics* 9(16):2395–2402. doi:10.1093/hmg/9.16.2395

Bhutani, N., Burns, D.M., Blau, H.M., 2011. DNA demethylation dynamics. *Cell* 146(6):866–872. doi:10.1016/j.cell.2011.08.042

Biesinger, K.E., Christensen, G.M., 1972. Effects of Various Metals on Survival, Growth, Reproduction, and Metabolism of *Daphnia magna*. *Journal of the Fisheries Research Board of* Canada 29(12):1691–1700. doi:10.1139/f72-269

Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes and Development* 16:6–21. doi:10.1101/gad.947102

Bock, C., 2012. Analysing and interpreting DNA methylation data. *Nature Reviews Genetics* 13:705–719. doi:10.1038/nrg3273

Boetzer, M., Henkel, C. V, Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579. doi:10.1093/bioinformatics/btq683

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., Balasubramanian, S., 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336(6083):934–937. doi:10.1126/science.1220671

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72(1-2):248–254. doi:10.1016/0003-2697(76)90527-3

Breiling, A., Lyko, F., 2015. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin* 8(24). doi:10.1186/s13072-015-0016-6

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A.J., Socci, N.D., Scandura, J.M., 2011. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* 6, e14524. doi:10.1371/journal.pone.0014524

Brown, C.J., Rupert, J.L., 2014. Hypoxia and Environmental Epigenetics. *High Altitude Medicine and Biology* 15(3):323–330. doi:10.1089/ham.2014.1016

Cadet, J., Wagner, J.R., 2014. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutatation Research. Genetic Toxicology and Environmental Mutagenesis* 764-765:18–35. doi:10.1016/j.mrgentox.2013.09.001

Capuano, F., Mülleder, M., Kok, R., Blom, H.J., Ralser, M., 2014. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Analytical Chemistry* 86(8):3697–3702. doi:10.1021/ac500447w

Castonguay, E., Angers, B., 2012. The key role of epigenetics in the persistence of asexual lineages. *Genetics Research International* 2012,534289. doi:10.1155/2012/534289

Charlesworth, D., Willis, J.H., 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* 10:783–96. doi:10.1038/nrg2664

Chatterjee, A., Stockwell, P.A., Rodger, E.J., Morison, I.M., 2012. Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Research 40(10):e79. doi:10.1093/nar/gks150

Chen, C.-C., Wang, K.-Y., Shen, C.-K.J., 2013. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *The Journal of Biological Chemistry* 288:9084–9091. doi:10.1074/jbc.M112.445585

Chen, C.-C., Wang, K.-Y., Shen, C.-K.J., 2012. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *The Journal of Biological Chemistry* 287:33116–33121. doi:10.1074/jbc.C112.406975

Chinnusamy, V., Zhu, J.K., 2009. Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology* 12(2):133-139. doi:10.1016/j.pbi.2008.12.006

Choi, S.H., Byun, H.-M., Kwan, J.M., Issa, J.-P.J., Yang, A.S., 2007. Hydroxycarbamide in combination with azacitidine or decitabine is antagonistic on DNA methylation inhibition. *British Journal of Haematology* 138(5):616–623. doi:10.1111/j.1365-2141.2007.06707.x

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., Jacobsen, S.E., 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452:215–219. doi:10.1038/nature06745

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Cáceres, C.E., Carmel, L., Casola, C., Choi, J.-H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Fröhlich, T., Geiler-Samerotte, K. a, Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E. V, Kültz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews, J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E. V, Zdobnov, E.M., Grigoriev, I. V, Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331(6017):555–561. doi:10.1126/science.1197761

Collier, J., 2009. Epigenetic regulation of the bacterial cell cycle. *Current Opinion in Microbiology* 12(6):722-729. doi:10.1016/j.mib.2009.08.005

Crider, K.S., Yang, T.P., Berry, R.J., Bailey, L.B., 2012. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Advances in Nutrition* 3(1):21–38. doi:10.3945/an.111.000992

Cunningham, C.B., Ji, L., Wiberg, R.A.W., Shelton, J., McKinney, E.C., Parker, D.J., Meagher, R.B., Benowitz, K.M., Roy-Zokan, E.M., Ritchie, M.G., Brown, S.J., Schmitz, R.J., Moore, A.J., 2015. The Genome and Methylome of a Beetle with Complex Social Behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biology and Evolution* 7(12):3383–3396. doi:10.1093/gbe/evv194

Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J.P., 2011. Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein hormones. *Journal of Proteome Research* 10:4478-4504. doi:10.1021/pr200284e

Dixon, G., Bay, L.K., Matz, M. V, 2016. Evolutionary consequences of DNA methylation in a basal metazoan. *Molecular Biology and Evolution* 33(9):2285-2293. doi:10.1093/molbev/msw100

Doyle, J.J., Doyle, J.D., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11 – 15.

Drewell, R.A., Bush, E.C., Remnant, E.J., Wong, G.T., Beeler, S.M., Stringham, J.L., Lim, J., Oldroyd, B.P., 2014. The dynamic DNA methylation cycle from egg to sperm in the honey bee *Apis mellifera*. Development 141:2702–2711. doi:10.1242/dev.110163

Eads, B.D., Andrews, J., Colbourne, J.K., 2008. Ecological genomics in *Daphnia*: stress responses and environmental sex determination. *Heredity* 100:184-190 doi:10.1038/sj.hdy.6800999

Ebert D. 2005. Introduction to *Daphnia* Biology. In: Ecology, epidemiology, and evolution of parasitism in daphnia. Bethesda: National Center for Biotechnology Information.

Elliott, E.N., Sheaffer, K.L., Kaestner, K.H., 2016. The "*de novo*" DNA methyltransferase Dnmt3b compensates the Dnmt1-deficient intestinal epithelium. *Elife* 5:e12975 doi:10.7554/eLife.12975.001

Erdmann, R.M., Souza, A.L., Clish, C.B., Gehring, M., 2014. 5-hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA. *G3* 5(1):1–8. doi:10.1534/g3.114.014670

Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., Tress, M.L., 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Human Molecular Genetics* 23(22):5866–5878. doi:10.1093/hmg/ddu309

Fang, X., Thornton, C., Scheffler, B.E., Willett, K.L., 2013. Benzo[a]pyrene decreases global and gene specific DNA methylation during zebrafish development. *Environmental Toxicology and Pharmacology* 36(1):40–50. doi:10.1016/j.etap.2013.02.014

Feil, R., Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics* 13:97–109. doi:10.1038/nrg3142

Feliciello, I., Parazajder, J., Akrap, I., Ugarković, Đ., 2013. First evidence of DNA methylation in insect *Tribolium castaneum*. *Epigenetics* 8(5):534–541. doi:10.4161/epi.24507

Felsenfeld, G., 2014. A brief history of epigenetics. *Cold Spring Harbor Perspectives in Biology* 6:a018200. doi:10.1101/cshperspect.a018200

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M., Jacobsen, S.E., 2010a. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy os Sciences of the United States of America* 107(19):8689–8694. doi:10.1073/pnas.1002720107

Feng, S., Jacobsen, S.E., Reik, W., 2010b. Epigenetic reprogramming in plant and animal development. *Science* 330(6004):622–627. doi:10.1126/science.1190614

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2015. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44:D279–D285. doi:10.1093/nar/gkv1344

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.-Z., Plass, C., Esteller, M., 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy os Sciences of the United States of America* 102(30):10604–10609. doi:10.1073/pnas.0500398102

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy os Sciences of the United States of America* 89(5):1827–1831. doi:10.1073/pnas.89.5.1827

Fuchs, S.M., Quasem, I., 2014. Budding yeast as a model to study epigenetics. *Drug Discovery Today Disease Model* 12:1–6. doi:10.1016/j.ddmod.2014.04.004

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. doi:10.1186/1471-2164-11-483

Geerts, A.N., Vanoverbeke, J., Vanschoenwinkel, B., Van Doorslaer, W., Feuchtmayr, H., Atkinson, D., Moss, B., Davidson, T.A., Sayer, C.D., De Meester, L., 2015. Rapid evolution of thermal tolerance in the water flea *Daphnia*. *Nature Climate Change* 5:665-668. doi:10.1038/nclimate2628

Gerke, P., Börding, C., Zeis, B., Paul, R.J., 2011. Adaptive haemoglobin gene control in *Daphnia pulex* at different oxygen and temperature conditions. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology* 159(1):56–65. doi:10.1016/j.cbpa.2011.01.017

Giessler, S., Mader, E., Schwenk, K.. 1999. Morphological evolution and genetic differentiation in *Daphnia* species complexes. *Journal of Evolutionary Biology* 12:710-723. doi:10.1046/j.1420-9101.1999.00065.x

Glastad, K.M., Hunt, B.G., Goodisman, M.A.D., 2015. DNA methylation and chromatin organization in insects: Insights from the ant *Camponotus floridanus*. *Genome Biology and Evolution* 7(4):931–942. doi:10.1093/gbe/evv039

Glastad, K.M., Hunt, B.G., Yi, S. V, Goodisman, M.A.D., 2011. DNA methylation in insects: on the brink of the epigenomic era. *Insect Molecular Biology* 20(5):553–565. doi:10.1111/j.1365-2583.2011.01092.x

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry* 74:481–514. doi:10.1146/annurev.biochem.74.010904.153721

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* 311(5759):395–398. doi:10.1126/science.1120976

Gong, Z., Zhu, J.-K., 2011. Active DNA demethylation by oxidation and repair. *Cell Research* 21:1649–1651. doi:10.1038/cr.2011.140

Gowher, H., Jeltsch, A., 2001. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. *Journal of Molecular Biology* 309(5):1201–1208. doi:10.1006/jmbi.2001.4710

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.H., Aravind, L., He, C., Shi, Y., 2015. DNA methylation on N6-Adenine in *C. elegans*. *Cell* 161(4):868–878. doi:10.1016/j.cell.2015.04.005

Grunstein, M., Gasser, S.M., 2013. Epigenetics in *Saccharomyces cerevisiae*. *Cold Spring Harbor Perspectives in Biology* 5:a017491. doi:10.1101/cshperspect.a017491

Harris, K.D.M., Bartlett, N.J., Lloyd, V.K., 2012. *Daphnia* as an emerging epigenetic model organism. *Genetic Research International* 2012:147892. doi:10.1155/2012/147892

Hattori, M., Yokoyama, Y., Hattori, T., Motegi, S.-I., Amano, H., Hatada, I., Ishikawa, O., 2015. Global DNA hypomethylation and hypoxia-induced expression of the ten eleven translocation (TET) family, TET1, in scleroderma fibroblasts. *Experimental Dermatology* 24(11):841-846. doi:10.1111/exd.12767

Hauser, M.T., Aufsatz, W., Jonak, C., Luschnig, C., 2011. Transgenerational epigenetic inheritance in plants. *Biochimica et Biophysica Acta* 1809(8):459–68. doi:10.1016/j.bbagrm.2011.03.007

Hayakawa, T., Kobayashi, Y., Cui, X., Hirano, S., 2005. A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. *Archives of Toxicology* 79(4):183–191. doi:10.1007/s00204-004-0620-x

Head, J.A., Dolinoy, D.C., Basu, N., 2012. Epigenetics for ecotoxicologists. *Environmental Toxicology and Chemistry* 31(2):221–227. doi:10.1002/etc.1707

Head, J.A., 2014. Patterns of DNA Methylation in Animals: An Ecotoxicological Perspective. *Integrative and Comparative Biology* 54(1):77–86. doi:10.1093/icb/icu025

Heard, E., Martienssen, R.A., 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157(1):95-109. doi:10.1016/j.cell.2014.02.045

Hendrich, B., Tweedie, S., 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* 19(5):269–277. doi:10.1016/S0168-9525(03)00080-5

Herceg, Z., Vaissière, T., 2011. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 6(7):804–819. doi:10.4161/epi.6.7.16262

Hochmuth, J.D., De Meester, L., Pereira, C.M.S., Janssen, C.R., De Schamphelaere, K.A.C., 2015. Rapid adaptation of a *Daphnia magna* population to metal stress is associated with heterozygote excess. *Environmental Science and Technology* 49(15):9298-9307. doi:10.1021/acs.est.5b00724

Hu, C.-W., Chen, J.-L., Hsu, Y.-W., Yen, C.-C., Chao, M.-R., 2015. Trace analysis of methylated and hydroxymethylated cytosines in DNA by isotope-dilution LC-MS/MS: first evidence of DNA methylation in *Caenorhabditis elegans*. *The Biochemical Journal* 465(1):39–47. doi:10.1042/BJ20140844

Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431(7011):931–945. doi:10.1038/nature03001

Iguchi-Ariga, S.M., Schaffner, W., 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes and Development* 3:612–619. doi:10.1101/gad.3.5.612

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S., Feinberg, A.P., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics* 41:178–86. doi:10.1038/ng.298

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* 333(6047):1300–1303. doi:10.1126/science.1210597

Itoh, H., Iwasaki, M., Kasuga, Y., Yokoyama, S., Onuma, H., Nishimura, H., Kusama, R., Yoshida, T., Yokoyama, K., Tsugane, S., 2014. Association between serum organochlorines and global methylation level of leukocyte DNA among Japanese women: a cross-sectional study. *Science of the Total Environment* 490:603–609. doi:10.1016/j.scitotenv.2014.05.035

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* 33:245–254. doi:10.1038/ng1089

James, S.J., Melnyk, S., Pogribna, M., Pogribny, I.P., Caudill, M.A., 2002. Elevation in S-Adenosylhomocysteine and DNA Hypomethylation: Potential Epigenetic Mechanism for Homocysteine-Related Pathology. *Journal of Nutrition* 132(8 Suppl):2361S–2366S.

Jeong, T.Y., Kim, H.Y., Kim, S.D., 2015. Multi-generational effects of propranolol on *Daphnia magna* at different environmental concentrations. *Environmental Pollution* 206:188–194. doi:10.1016/j.envpol.2015.07.003

Jiang, N., Wang, L., Chen, J., Wang, L., Leach, L., Luo, Z., 2014. Conserved and divergent patterns of DNA methylation in higher vertebrates. *Genome Biology and Evolution* 6(11):2998–3014. doi:10.1093/gbe/evu238

Jin, B., Robertson, K.D., 2013. DNA methyltransferases, DNA damage repair, and cancer. In: Epigenetic Alterations in Oncogenesis. ed. by Karpf, R.A. New York, NY: Springer New York, 3–29. doi:10.1007/978-1-4419-9967-2

Jirtle, R.L., Skinner, M.K., 2007. Environmental epigenomics and disease susceptibility. *Nature Reviews Genetics* 8:253–62. doi:10.1038/nrg2045

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* 13:484–92. doi:10.1038/nrg3230

Jones, P.A., Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics* 3:415–428. doi:10.1038/nrg816

Jones, P.A., 1999. The DNA methylation paradox. *Trends in Genetics* 15(1):34–37. doi:10.1016/S0168-9525(98)01636-9

Jones, P.A., Liang, G., 2009. Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics* 10:805–11. doi:10.1038/nrg2651

Jones, P.A., Takai, D., 2001. The role of DNA methylation in mammalian epigenetics. *Science* 293(5532):1068–70. doi:10.1126/science.1063852

Kamstra, J.H., Aleström, P., Kooter, J.M., Legler, J., 2015a. Zebrafish as a model to study the role of DNA methylation in environmental toxicology. *Environmental Science and Pollution Research International* 22(21):16262–16276. doi:10.1007/s11356-014-3466-7

Kamstra, J.H., Løken, M., Aleström, P., Legler, J., 2015b. Dynamics of DNA hydroxymethylation in zebrafish. *Zebrafish* 12(3):230–237. doi:10.1089/zeb.2014.1033

Kass, S.U., Landsberger, N., Wolffe, A.P., 1997. DNA methylation directs a time-dependent repression of transcription initiation. *Current Biology* 7(3):157–165. doi:10.1016/S0960-9822(97)70086-1

Keating, K.I., Dagbusan, B.C., 1984. Effect of selenium deficiency on cuticle integrity in the Cladocera (Crustacea). *Proceedings of the National Academy of Sciences of the United States of America* 81(11):3433–3437.

Keil, K.P., Lein, P.J., 2016. DNA methylation: a mechanism linking environmental chemical exposures to risk of autism spectrum disorders? *Environmental Epigenetics* 2(1). doi:10.1093/eep/dvv012

Kelsey, G., Feil, R., 2013. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philosophical Transactions of the Royal Society B* 368(1609):20110336. doi:10.1098/rstb.2011.0336

Kilham, S.S., Kreeger, D.A., Lynn, S.G., Goulden, C.E., Herrera, L., 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* 377:147–159. doi:10.1023/A:1003231628456

Kim, K.Y., Kim, D.S., Lee, S.K., Lee, I.K., Kang, J.H., Chang, Y.S., Jacobs, D.R., Steffes, M., Lee, D.H., 2010. Association of low-dose exposure to persistent organic pollutants with global DNA hypomethylation in healthy Koreans. *Environmental Health Perspectives* 118(3):370–374. doi:10.1289/ehp.0901131

Klose, R.J., Bird, A.P., 2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* 31(2):89–97. doi:10.1016/j.tibs.2005.12.008

Kohli, R.M., Zhang, Y., 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502(7472):472–9. doi:10.1038/nature12750

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572. doi:10.1093/bioinformatics/btr167

Lachance, G., Uniacke, J., Audas, T.E., Holterman, C.E., Franovic, A., Payette, J., Lee, S., 2014. DNMT3a epigenetic program regulates the HIF-2α oxygen-sensing pathway and the cellular response to hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* 111(21):7783–7788. doi:10.1073/pnas.1322909111

Laing, L. V., Viana, J., Dempster, E.L., Trznadel, M., Trunkfield, L.A., Webster, T.M.U., Aerle, R. van, Paull, G.C., Wilson, R.J., Mill, J., Santos, E.M., 2016. Bisphenol A causes reproductive toxicity, decreases dnmt1 transcription, and reduces global DNA methylation in breeding zebrafish (*Danio rerio*). *Epigenetics* 11(7):526-538. doi:10.1080/15592294.2016.1182272.

Lampert, W., Kinne, O., 2011. *Daphnia*: Development of a Model Organism in Ecology and Evolution. International Ecology Institute.

Lampert, W., Trubetskova, I., 1996. Juvenile Growth Rate as a Measure of Fitness in *Daphnia*. *Functional Ecology* 10(5):631–635. doi:10.2307/2390173

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25. doi:10.1186/gb-2009-10-3-r25

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* 11(3):204–220. doi:10.1038/nrg2719

Lee, J.-H., Park, S.-J., Kenta, N., 2015. An integrative approach for efficient analysis of whole genome bisulfite sequencing data. *BMC Genomics* 16(Suppl 12):S14. doi:10.1186/1471-2164-16-S12-S14

Lees-Murdock, D.J., Walsh, C.P., 2008. DNA methylation reprogramming in the germ line. *Epigenetics* 3(1):5–13. doi:10.4161/epi.3.1.5553.

Lev Maor, G., Yearim, A., Ast, G., 2015. The alternative role of DNA methylation in splicing regulation. *Trends in Genetics* 31(5):274-280. doi:10.1016/j.tig.2015.03.002

Li, D., Lu, C., Wang, J., Hu, W., Cao, Z., Sun, D., Xia, H., Ma, X., 2009. Developmental mechanisms of arsenite toxicity in zebrafish (*Danio rerio*) embryos. *Aquatic Toxicology* 91(3):229–237. doi:10.1016/j.aquatox.2008.11.007

Li, L.-C., Dahiya, R., 2002. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18(11):1427–1431.

Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., Zhang, H., Chen, Q., Luo, R., Chen, M., He, Y., Jin, X., Zhang, Q., Yu, C., Zhou, G., Sun, J., Huang, Y., Zheng, H., Cao, H., Zhou, X., Guo, S., Hu, X., Li, X., Kristiansen, K., Bolund, L., Xu, J., Wang, W., Yang, H., Wang, J., Li, R., Beck, S., Wang, J., Zhang, X., 2010. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol*ogy 8:e1000533. doi:10.1371/journal.pbio.1000533

Lin, M.-J., Tang, L.-Y., Reddy, M.N., Shen, C.-K.J., 2005. DNA methyltransferase gene dDnmt2 and longevity of *Drosophila. The Journal of Biological Chemistry* 280:861–864. doi:10.1074/jbc.C400477200

Lind, L., Penell, J., Luttropp, K., Nordfors, L., Syvänen, A.C., Axelsson, T., Salihovic, S., van Bavel, B., Fall, T., Ingelsson, E., Lind, P.M., 2013. Global DNA hypermethylation is associated with high serum levels of persistent organic pollutants in an elderly population. *Environment International* 59:456–461. doi:10.1016/j.envint.2013.07.008

Lindberg, A.-L., Kumar, R., Goessler, W., Thirumaran, R., Gurzau, E., Koppova, K., Rudnai, P., Leonardi, G., Fletcher, T., Vahter, M., 2007. Metabolism of low-dose inorganic arsenic in a central European population: influence of sex and genetic polymorphisms. *Environmental Health Perspectives* 115(7):1081–1086. doi:10.1289/ehp.10026

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402–408. doi:10.1006/meth.2001.1262

Locasale, J.W., 2013. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nature Reviews Cancer* 13(8):572–583. doi:10.1038/nrc3557

Long, Y., Yan, J., Song, G., Li, X., Li, X., Li, Q., Cui, Z., 2015. Transcriptional events co-regulated by hypoxia and cold stresses in Zebrafish larvae. *BMC Genomics* 16:385. doi:10.1186/s12864-015-1560-y

Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., Weissman, A.M., 1999. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* 96(20):11364–11369. doi:10.1073/pnas.96.20.11364

Lu, S.C., 2000. S-Adenosylmethionine. *The International Journal of Biochemistry and Cell Biology* 32(4):391–395.

Luo, G.-Z., Blanco, M.A., Greer, E.L., He, C., Shi, Y., 2015. DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? *Nature Reviews Molecular Cell Biology* 16:705–10. doi:10.1038/nrm4076

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi:10.1186/2047-217X-1-18

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. doi:10.1371/journal.pbio.1000506

Lyko, F., Maleszka, R., 2011. Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics* 27(4):127–131. doi:10.1016/j.tig.2011.01.003

Lyko, F., Ramsahoye, B.H., Jaenisch, R., 2000. DNA methylation in *Drosophila melanogaster*. *Nature* 408:538–540. doi:10.1038/35046205

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research* 43(D1):D222–226. doi:10.1093/nar/gku1221

Mariani, C.J., Vasanthakumar, A., Madzo, J., Yesilkanal, A., Bhagat, T., Yu, Y., Bhattacharyya, S., Wenger, R.H., Cohn, S.L., Nanduri, J., Verma, A., Prabhakar, N.R., Godley, L.A., 2014. TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. *Cell Reports* 7(5):1343–1352. doi:10.1016/j.celrep.2014.04.040

Martyniuk, C.J., Simmons, D.B., 2016. Spotlight on environmental omics and toxicology: a long way in a short time. *Comparative Biochemistry and Physiology Part D Genomics Proteomics* 19:97–101. doi:10.1016/j.cbd.2016.06.010

Maunakea, A.K., Chepelev, I., Cui, K.R., Zhao, K.J., 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research* 23:1256–1269. doi:10.1038/cr.2013.110

Menzel, S., Bouchnak, R., Menzel, R., Steinberg, C.E.W., 2011. Dissolved humic substances initiate DNA-methylation in cladocerans. *Aquatic Toxicology* 105(3-4):640–642. doi:10.1016/j.aquatox.2011.08.025

Messiaen, M., De Schamphelaere, K.A.C., Muyssen, B.T.A., Janssen, C.R., 2010. The micro-evolutionary potential of *Daphnia magna* population exposed to temperature and cadmium stress. *Ecotoxicology and Environmental Safety* 73(6):1114-1122. doi:10.1016/j.ecoenv.2010.05.006

Messiaen, M., Janssen, C.R, De Meester, L., De Schamphelaere, K.A.C., 2013. The initial tolerance to sub-lethal Cd exposure is the same among ten naïve pond populations of *Daphnia magna*, but their micro-evolutionary potential to develop resistance is very different. *Aquatic Toxicology* 144–145:322-331. doi:10.1016/j.aquatox.2013.10.016

Mikeska, T., Craig, J.M., 2014. DNA methylation biomarkers: cancer and beyond. *Genes* 5(3):821–864. doi:10.3390/genes5030821

Miner, B.E., De Meester, L., Pfrender, M.E., Lampert, W., Hairston, N.G., 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings of the Royal Society B Biological Sciences* 279(1735):1873-1882. doi:10.1098/rspb.2011.2404

Mirbahai, L., Chipman, J.K., 2014. Epigenetic memory of environmental organisms: a reflection of lifetime stressor exposures. *Mutation Research Genetic Toxicology and Environmental Mutagenesis* 764-765:10–7. doi:10.1016/j.mrgentox.2013.10.003

Mirbahai, L., Southam, A.D., Sommer, U., Williams, T.D., Bignell, J.P., Lyons, B.P., Viant, M.R., Chipman, J.K., 2013. Disruption of DNA Methylation via S-Adenosylhomocysteine Is a Key Process in High Incidence Liver Carcinogenesis in Fish. *Journal of Proteome Research* 12(6):2895–2904. doi:10.1021/pr400195u

Mirbahai, L., Yin, G., Bignell, J.P., Li, N., Williams, T.D., Chipman, J.K., 2011. DNA methylation in liver tumorigenesis in fish from the environment. *Epigenetics* 6(11):1319–1333. doi:10.4161/epi.6.11.17890

Mitsudome, T., Mon, H., Xu, J., Li, Z., Lee, J.M., Patil, A.A., Masuda, A., Iiyama, K., Morokuma, D., Kusakabe, T., 2015. Biochemical characterization of maintenance DNA methyltransferase DNMT-1 from silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology* 58:55–65. doi:10.1016/j.ibmb.2015.01.008

Mukherjee, K., Twyman, R.M., Vilcinskas, A., 2015. Insects as models to study the epigenetic basis of disease. *Progress in Biophysics and Molecular Biology* 118(1-2):69–78. doi:10.1016/j.pbiomolbio.2015.02.009

Nendza, M., Wenzel, A., 2006. Discriminating toxicant classes by mode of action. 1. (Eco)toxicity profiles. *Environmental Science and Pollution Research* 13(3):192–203. doi:10.1065/espr2006.01.013

Noble, W.S., 2009. How does multiple testing correction work? *Nature Biotechnology* 27:1135–1137. doi:10.1038/nbt1209-1135

OECD, 2012. OECD Guidelines for the Testing of Chemicals, Section 2, Test No. 211: Daphnia magna Reproduction Test.

OECD, 2004. OECD Guidelines for the Testing of Chemicals, Section 2, Test No. 202: Daphnia sp. Acute Immobilisation Test.

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(3):247–257. doi:10.1016/S0092-8674(00)81656-6

Okano, M., Xie, S., Li, E., 1998. Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research* 26(11):2536–2540. doi: 10.1093/nar/26.11.2536

Olsvik, P.A., Williams, T.D., Tung, H.S., Mirbahai, L., Sanden, M., Skjaerven, K.H., Ellingsen, S., 2014. Impacts of TCDD and MeHg on DNA methylation in zebrafish (*Danio rerio*) across two generations. *Comparative Biochemistry and Physiology Part - C Toxicology and Pharmacology* 165:17–27. doi:10.1016/j.cbpc.2014.05.004

Omilian, A.R., Lynch, M., 2009. Patterns of intraspecific DNA variation in the *Daphnia* nuclear genome. *Genetics* 182:325-336. doi:10.1534/genetics.108.099549

Orsini, L., Gilbert, D., Podicheti, R., Jansen, M., Brown, J.B., Solari, O.S., Spanier, K.I., Colbourne, J.K., Rush, D., Decaestecker, E., Asselman, J., De Schamphelaere, K.A.C., Ebert, D., Haag, C.R., Kvist, J., Laforsch, C., Petrusek, A., Beckerman, A.P., Little, T.J., Chaturvedi, A., Pfrender, M.E., De Meester, L., Frilander, M.J., 2016. *Daphnia magna* transcriptome by RNA-Seq across 12 environmental stressors. *Scientific Data* 3:160030. doi:10.1038/sdata.2016.30

Orsini, L., Spanier, K.I., DE Meester, L., 2012. Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: validation in space, time and experimental evolution. *Molecular Ecology* 21(9):2160–2175. doi:10.1111/j.1365-294X.2011.05429.x

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(12):3345–54. doi:10.1093/molbev/msr168

Patalano, S., Hore, T.A., Reik, W., Sumner, S., 2012. Shifting behaviour: Epigenetic reprogramming in eusocial insects. *Current Opinion in Cell Biology* 24(3):367–373. doi:10.1016/j.ceb.2012.02.005

Paul, R.J., Zeis, B., Lamkemeyer, T., Seidl, M., Pirow, R., 2004. Control of oxygen transport in the microcrustacean *Daphnia*: regulation of haemoglobin expression as central mechanism of adaptation to different oxygen and temperature conditions. *Acta Physiologica* 182(3):259–275. doi:10.1111/j.1365-201X.2004.01362.x

Pegoraro, M., Bafna, A., Davies, N.J., Shuker, D.M., Tauber, E., 2016. DNA methylation changes induced by long and short photoperiods in *Nasonia*. *Genome Research* 26:203–10. doi:10.1101/gr.196204.115

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(4):781–790.

Pfrender, M.E., Spitze, K., Lehman, N., 2000. Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Molecular Ecology* 9:1717-1735. doi:10.1046/j.1365-294x.2000.01062.x

Piccolo, F.M., Fisher, A.G., 2014. Getting rid of DNA methylation. *Trends in Cell Biology* 24(2):136–143. doi:10.1016/j.tcb.2013.09.001

Pirow, R., Bäumer, C., Paul, R.J., 2001. Benefits of haemoglobin in the cladoceran crustacean *Daphnia magna*. *The Journal of Experimental Biology* 204:3425–3441.

Ponger, L., Li, W.-H., 2005. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution* 22(4):1119–1128. doi:10.1093/molbev/msi098

Pradhan, S., Bacolla, A., Wells, R.D., Roberts, R.J., 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of *de novo* and maintenance methylation. *The Jornal of Biological Chemistry* 274:33002–33010. doi:10.1074/jbc.274.46.33002

Probst, A. V, Dunleavy, E., Almouzni, G., 2009. Epigenetic inheritance during the cell cycle. *Nature Reviews Molecular Cell Biology* 10:192–206. doi:10.1038/nrm2640

Putiri, E.L., Robertson, K.D., 2011. Epigenetic mechanisms and genome stability. *Clinical Epigenetics* 2:299. doi:10.1007/s13148-010-0017-z

Qi, W., Nong, G., Preston, J.F., Ben-Ami, F., Ebert, D., 2009. Comparative metagenomics of *Daphnia* symbionts. *BMC Genomics* 10:172. doi:10.1186/1471-2164-10-172

Quinlivan, E.P., Gregory, J.F., 2008. DNA digestion to deoxyribonucleoside: a simplified one-step procedure. *Analytical Biochemistry* 373(2):383–385. doi:10.1016/j.ab.2007.09.031

Rackham, O.J.L., Dellaportas, P., Petretto, E., Bottolo, L., 2015. WGBSSuite: Simulating Whole Genome Bisulphite Sequencing data and benchmarking differential DNA methylation analysis tools. *Bioinformatics* 31(14):2371-2373. doi:10.1093/bioinformatics/btv114

Raddatz, G., Guzzardo, P.M., Olova, N., Fantappié, 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 Stated of America.* 110(21):8627–8631. doi:10.1073/pnas.1306723110

Rasmussen, E.M.K., Amdam, G. V., 2015. Cytosine modifications in the honey bee (*Apis mellifera*) worker genome. *Frontiers in Genetics* 5:8. doi:10.3389/fgene.2015.00008

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(7):880–891.

Reichard, J.F., Puga, A., 2010. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. *Epigenomics* 2(1):87–104. doi:10.2217/epi.09.45

Reichard, J.F., Schnekenburger, M., Puga, A., 2007. Long term low-dose arsenic exposure induces loss of DNA methylation. *Biochemical and Biophysical Research Communications* 352(1):188–192. doi:10.1016/j.bbrc.2006.11.001

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-W., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., Baylin, S.B., Vogelstein, B., 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416:552–556. doi:10.1038/416552a

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16(6):276–277. doi:10.1016/S0168-9525(00)02024-2

Rijlaarsdam, M.A., van der Zwan, Y.G., Dorssers, L.C.J., Looijenga, L.H.J., 2014. DMRforPairs: identifying differentially methylated regions between unique samples using array based methylation profiles. *BMC Bioinformatics* 15:141. doi:10.1186/1471-2105-15-141

Rivière, G., 2014. Epigenetic features in the oyster *Crassostrea gigas* suggestive of functionally relevant promoter DNA methylation in invertebrates. *Frontiers in Physiology* 5:129. doi:10.3389/fphys.2014.00129

Robertson, A.B., Klungland, A., Rognes, T., Leiros, I., 2009. DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cellular and Molecular Life Sciences* 66(6):981–993. doi:10.1007/s00018-009-8736-z

Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A., Jones, P.A., 1999. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Research* 27(11):2291–2298. doi:10.1093/nar/27.11.2291

Robichaud, N.F., Sassine, J., Beaton, M.J., Lloyd, V.K., 2012. The epigenetic repertoire of *Daphnia magna* includes modified histones. *Genetics Research International* 2012:174860. doi:10.1155/2012/174860

Robinson, M.D., Kahraman, A., Law, C.W., Lindsay, H., Nowicka, M., Weber, L.M., Zhou, X., 2014. Statistical methods for detecting differentially methylated loci and regions. *Frontiers in Genetics* 5:324. doi:10.3389/fgene.2014.00324

Rössler, J., Stolze, I., Frede, S., Freitag, P., Schweigerer, L., Havers, W., Fandrey, J., 2004. Hypoxia-induced erythropoietin expression in human neuroblastoma requires a methylation free HIF-1 binding site. *Journal or Cellular Biochemistry* 93(1):153–161. doi:10.1002/jcb.20133

Rountree, M.R., Bachman, K.E., Baylin, S.B., 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genetics* 25:269–277. doi:10.1038/77023

Routtu, J., Hall, M.D., Albere, B., Beisel, C., Bergeron, R.D., Chaturvedi, A., Choi, J.-H., Colbourne, J., De Meester, L., Stephens, M.T., Stelzer, C.-P., Solorzano, E., Thomas, W.K., Pfrender, M.E., Ebert, D., 2014. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. *BMC Genomics* 15:1033. doi:10.1186/1471-2164-15-1033

Rusiecki, J.A., Baccarelli, A., Bollati, V., Tarantini, L., Moore, L.E., Bonefeld-Jorgensen, E.C., 2008. Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic inuit. *Environmental Health Perspectives* 116(11):1547–1552. doi:10.1289/ehp.11338

Santangeli, S., Maradonna, F., Gioacchini, G., Cobellis, G., Piccinetti, C.C., Dalla Valle, L., Carnevali, O., 2016. BPA-Induced Deregulation Of Epigenetic Patterns: Effects On Female Zebrafish Reproduction. *Scientific Reports* 6:21982. doi:10.1038/srep21982

Santi, D. V, Norment, A., Garrett, C.E., 1984. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proceedings of the National Academy of Sciences of the United States of America* 81(22):6993–6997.

Sarda, S., Zeng, J., Hunt, B.G., Yi, S. V, 2012. The evolution of invertebrate gene body methylation. *Molecular Biology and Evolution* 29(8):1907–1916. doi:10.1093/molbev/mss062

Schaefer, M., Lyko, F., 2010. Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation. *Nature Genetics* 42:920–921. doi:10.1038/ng1110-920

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., Lyko, F., 2010. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development* 24(15):1590–1595. doi:10.1101/gad.586710

Schübeler, D., 2015. Function and information content of DNA methylation. *Nature* 517:321–326. doi:10.1038/nature14192

Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., Reik, W., 2013. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical Transactions of the Royal Society B* 368(1609):20110330. doi:10.1098/rstb.2011.0330

Semenza, G.L., 2007. Life with Oxygen. *Science* 318(5847):62–64. doi:10.1126/science.1147949

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371):74–79. doi:10.1038/nature10442

Silva, A.R.R., Cardoso, D.N., Cruz, A., Pestana, J.L.T., Mendo, S., Soares, A.M.V.M., Loureiro, S., 2016. Multigenerational effects of carbendazim in *Daphnia magna*. Environmental Toxicology and Chemistry. doi:10.1002/etc.3541

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* 283(5405):1164–1167. doi:10.1126/science.283.5405.1164

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research* 19:1117–1123. doi:10.1101/gr.089532.108

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(16):6711–6719. doi:10.1093/nar/14.16.6711

Skinner, M.K., 2014. Environmental stress and epigenetic transgenerational inheritance. *BMC Medicine* 12:153. doi:10.1186/s12916-014-0153-y

Skinner, M.K., 2008. What is an epigenetic transgenerational phenotype? F3 or F2. *Reproductive Toxicology* 25(1):2–6. doi:10.1016/j.reprotox.2007.09.001

Skinner, M.K., Manikkam, M., Guerrero-Bosagna, C., 2010. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends in Endocrinology and Metabolism* 21(4):214–222. doi:10.1016/j.tem.2009.12.007

Slieker, R.C., Bos, S.D., Goeman, J.J., Bovée, J.V., Talens, R.P., van der Breggen, R., Suchiman, H.E.D., Lameijer, E.-W., Putter, H., van den Akker, E.B., Zhang, Y., Jukema, J.W., Slagboom, P.E., Meulenbelt, I., Heijmans, B.T., 2013. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics and Chromatin* 6:26. doi:10.1186/1756-8935-6-26

Smedley, P.L., Kinniburgh, D.G., 2005. Source and behaviour of arsenic in natural waters Importance of arsenic in drinking water. Chapter 1 in United Nations Synthesis Report on Arsenic in Drinking Water.

Smith, F.E., 1963. Population Dynamics in *Daphnia magna* and a New Model for Population Growth. *Ecology* 44(4):651–663. doi:10.2307/1933011

Smith, T.H.L., Collins, T.M., McGowan, R.A., 2011. Expression of the dnmt3 genes in zebrafish development: Similarity to Dnmt3a and Dnmt3b. *Development Genes and Evolution* 220(11-12):347–353. doi:10.1007/s00427-010-0347-z

Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., Gao, J., Liu, P., Li, L., Xu, G.-L., Jin, P., He, C., 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153(3):678–691. doi:10.1016/j.cell.2013.04.001

Song, F., Mahmood, S., Ghosh, S., Liang, P., Smiraglia, D.J., Nagase, H., Held, W.A., 2009. Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. *Genomics* 93(2):130–139. doi:10.1016/j.ygeno.2008.09.003

Spruijt, C.G., Vermeulen, M., 2014. DNA methylation: old dog, new tricks? *Nature Structural and Molecular Biology* 21:949–954. doi:10.1038/nsmb.2910

Standage, D.S., Berens, A.J., Glastad, K.M., Severin, A.J., Brendel, V.P., Toth, A.L., 2016. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology* 25(8):1769–1784. doi:10.1111/mec.13578

Stollewerk, A., 2010. The water flea *Daphnia*-a "new" model system for ecology and evolution? *Journal of Biology* 9:21. doi:10.1186/jbiol212

Strepetkaitė, D., Alzbutas, G., Astromskas, E., Lagunavičius, A., Sabaliauskaitė, R., Arbačiauskas, K., Lazutka, J., 2015. Analysis of DNA Methylation and Hydroxymethylation in the Genome of Crustacean *Daphnia pulex*. *Genes* 7(1):1. doi:10.3390/genes7010001

Stresemann, C., Lyko, F., 2008. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer* 123:8–13. doi:10.1002/ijc.23607

Strömqvist, M., Tooke, N., Brunström, B., 2010. DNA methylation levels in the 5' flanking region of the vitellogenin I gene in liver and brain of adult zebrafish (*Danio rerio*)—Sex and tissue differences and effects of 17α-ethinylestradiol exposure. *Aquatic Toxicology* 98(3):275–281. doi:10.1016/j.aquatox.2010.02.023

Suzuki, M.M., Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* 9:465–76. doi:10.1038/nrg2341

Suzuki, M.M., Kerr, A.R.W., De Sousa, D., Bird, A., 2007. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Research* 17:625–631. doi:10.1101/gr.6163007

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* 324(5929):930–935. doi:10.1126/science.1170116

Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S.J., Pachter, L., Martin, D.I.K., 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. doi:10.1101/gr.162412.113

Takiguchi, M., Achanzar, W.E., Qu, W., Li, G., Waalkes, M.P., 2003. Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. *Experimental Cell Research* 286(2):355–365. doi:10.1016/S0014-4827(03)00062-4

Tan, L., Shi, Y.G., 2012. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* 139(11):1895–902. doi:10.1242/dev.070771

Taylor, N.S., Weber, R.J.M., Southam, A.D., Payne, T.G., Hrydziuszko, O., Arvanitis, T.N., Viant M.R., 2008. A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* 5:44-58. doi:10.1007/s11306-008-0133-3

Trotter, B., Otte, K. a, Schoppmann, K., Hemmersbach, R., Fröhlich, T., Arnold, G.J., Laforsch, C., 2015. The influence of simulated microgravity on the proteome of *Daphnia magna*. *npj Microgravity* 1:15016. doi:10.1038/npjmgrav.2015.16

Tollefsbol, T.O. (Ed.), 2012. Epigenetics in Human Disease. Academic Press.

Tsai, Y.-P., Wu, K.-J., 2014. Epigenetic regulation of hypoxia-responsive gene expression: focusing on chromatin and DNA modifications. *Internation Journal of Cancer* 134(2):249–256. doi:10.1002/ijc.28190

Tudisco, L., Della Ragione, F., Tarallo, V., Apicella, I., D'Esposito, M., Matarazzo, M.R., De Falco, S., 2014. Epigenetic control of hypoxia inducible factor-1α-dependent expression of placental growth factor in hypoxic conditions. *Epigenetics* 9:600–610. doi:10.4161/epi.27835

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., Frye, M., Helm, M., Stoecklin, G., Lyko, F., 2012. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nature Structural and Molecular Biology 19(9):900–905. doi:10.1038/nsmb.2357

Tweedie, S., Charlton, J., Clark, V., Bird, A., 1997. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology* 17(3):1469–1475. doi:10.1128/MCB.17.3.1469

Ulrey, C.L., Liu, L., Andrews, L.G., Tollefsbol, T.O., 2005. The impact of metabolism on DNA methylation. *Human Molecular Genetics* 14:R139–R147. doi:10.1093/hmg/ddi100

Vahter, M., 2002. Mechanisms of arsenic biotransformation. *Toxicology* 181-182:211–217. doi: 10.1016/S0300-483X(02)00285-8

Vahter, M., Concha, G., 2008. Role of Metabolism in Arsenic Toxicity. *Pharmacology and Toxicology* 89(1):1–5. doi:10.1111/j.1600-0773.2001.890101.x

Vaissière, T., Sawan, C., Herceg, Z., 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutation Research* 659(1-2):40–48. doi:10.1016/j.mrrev.2008.02.004

Vandegehuchte, M.B., De Coninck, D., Vandenbrouck, T., De Coen, W.M., Janssen, C.R., 2010a. Gene transcription profiles, global DNA methylation and potential transgenerational epigenetic effects related to Zn exposure history in *Daphnia magna*. *Environmental Pollution* 158(10):3323–3329. doi:10.1016/j.envpol.2010.07.023

Vandegehuchte, M.B., Janssen, C.R., 2013. Epigenetics in an ecotoxicological context. *Mutation Research Genetic Toxicology and Environmental Mutagenesis* 764-765:36-45 doi:10.1016/j.mrgentox.2013.08.008

Vandegehuchte, M.B., Kyndt, T., Vanholme, B., Haegeman, A., Gheysen, G., Janssen, C.R., 2009a. Occurrence of DNA methylation in *Daphnia magna* and influence of multigeneration Cd exposure. *Environment International* 35(4):700–706. doi:10.1016/j.envint.2009.01.002

Vandegehuchte, M.B., Lemière, F., Janssen, C.R., 2009b. Quantitative DNA-methylation in Daphnia magna and effects of multigeneration Zn exposure. *Comparative Biochemistry and Physiology C Toxicology and Pharmacology* 150(3):343–348. doi:10.1016/j.cbpc.2009.05.014

Vandegehuchte, M.B., Lemière, F., Vanhaecke, L., Vanden Berghe, W., Janssen, C.R., 2010b. Direct and transgenerational impact on Daphnia magna of chemicals with a known effect on DNA methylation. *Comparative Biochemistry and Physiology C Toxicology and Pharmacology* 151(3):278–285. doi:10.1016/j.cbpc.2009.11.007

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3:research0034.1. doi:10.1186/gb-2002-3-7-research0034

Van Doorslaer, W., Stoks, R., Duvivier, C., Bednarska, A., De Meester, L., 2009. Population dynamics determine genetic adaptation to temperature in *Daphnia*. *Evolution* 63(7):1867-1878. doi:10.1111/j.1558-5646.2009.00679.x

Wade, P.A., 2001. Methyl CpG-binding proteins and transcriptional repression. *BioEssays* 23(12):1131-1137. doi:10.1002/bies.10008

Wagner, K.-D., Wagner, N., Wellmann, S., Schley, G., Bondke, A., Theres, H., Scholz, H., 2003. Oxygen-regulated expression of the Wilms' tumor suppressor Wt1 involves hypoxia-inducible factor-1 (HIF-1). *FASEB Journal* 17:1364–1366. doi:10.1096/fj.02-1065fje

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L., Qian, J., 2015. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics* 16:49. doi:10.1186/s12864-015-1271-4

Wang, X., Li, Q., Lian, J., Li, L., Jin, L., Cai, H., Xu, F., Qi, H., Zhang, L., Wu, F., Meng, J., Que, H., Fang, X., Guo, X., Zhang, G., 2014. Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. *BMC Genomics* 15:1119. doi:10.1186/1471-2164-15-1119

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.-H., Colbourne, J.K., Clark, A.G., Werren, J.H., 2013. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genetics* 9:e1003872. doi:10.1371/journal.pgen.1003872

Wang, Z., Li, X., Jiang, Y., Shao, Q., Liu, Q., Chen, B.Y., Huang, D., 2015. SwDMR: A sliding window approach to identify differentially methylated regions based on whole genome bisulfite sequencing. *PLoS One* 10(7):e0132866. doi:10.1371/journal.pone.0132866

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., Clark, S.J., 2002. Identification and resolution of artifacts in bisulfite sequencing. *Methods* 27(2):101–107. doi:10.1016/S1046-2023(02)00060-9

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191. doi:10.1093/bioinformatics/btp033

Watson, J. a, Watson, C.J., McCann, A., Baugh, J., 2010. Epigenetics, the epicenter of the hypoxic response. *Epigenetics* 5(4):293–296. doi:10.4161/epi.5.4.11684

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.P., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C.S.G., 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.G., 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., Gründer, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C., Hurst, G.D.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J.J., Kim, J.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., 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.F., 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., Schlüns, H., Schmitt, T., Schneider, M., Schüler, A., Schurko, A.M., Shuker, D.M., Simões, Z.L.P., 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* 327(5963):343–348. doi:10.1126/science.1178028

Whittle, C.A.W., Otto, S.P.O., Johnston, M.O.J., Krochko, J.E.K., 2009. Adaptive epigenetic memory of ancestral temperature regime in *Arabidopsis thaliana. Botany* 87(6):650–657. doi:10.1139/B09-030

WHO, W.H.O., 2011. Guidelines for Drinking-Water Quality, Fourth edition. WHO 1, 541.

Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M., Maleszka, R., 2014. Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biology* 4:140110. doi:10.1098/rsob.140110

Wu, R.S.S., 2002. Hypoxia: From molecular responses to ecosystem responses. *Marine Pollution Bulletin* 45(1-12):35–45. doi:10.1016/S0025-326X(02)00061-9

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., Tackett, A., Wang, G., Hon, L.S., Fang, G., Swenberg, J.A., Xiao, A.Z., 2016. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* 532:329-333 doi:10.1038/nature17640

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., 2010. Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology* 28:516–520. doi:10.1038/nbt.1626

Xiao, Y., Word, B., Starlard-Davenport, A., Haefele, A., Lyn-Cook, B.D., Hammons, G., 2008. Age and gender affect DNMT3a and DNMT3b expression in human liver. *Cell Biology and Toxicology* 24(3):265–272. doi:10.1007/s10565-007-9035-9

Yampolsky, L.Y., Schaer, T.M.M., Ebert, D., 2014. Adaptive phenotypic plasticity and local adaptation for temperature tolerance in freshwater zooplankton. *Proceedings of the Royal Society B Biological Sciences* 281(1776):20132744. doi:10.1098/rspb.2013.2744

Yang, A.S., Estécio, M.R.H., Doshi, K., Kondo, Y., Tajara, E.H., Issa, J.-P.J., 2004. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Research* 32(3):e38. doi:10.1093/nar/gnh032

Yi, P., Melnyk, S., Pogribna, M., Pogribny, I.P., Hine, R.J., James, S.J., 2000. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. *The Journal of Biological Chemistry* 275:29318–29323. doi:10.1074/jbc.M002725200

Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., Caiafa, P., 2015. Reconfiguration of DNA methylation in aging. *Mechanism of Ageing Development* 151:60–70. doi:10.1016/j.mad.2015.02.002

Zeis, B., Becker, D., Gerke, P., Koch, M., Paul, R.J., 2013. Hypoxia-inducible haemoglobins of *Daphnia pulex* and their role in the response to acute and chronic temperature increase. *Biochimica et Biophysica Acta* 1834(9):1704–1710. doi:10.1016/j.bbapap.2013.01.036

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980):916–919. doi:10.1126/science.1186366

Zhang, C., Edepli, K.S., Hoshida, Y., 2016. Comparative epigenomic profiling of the DNA methylome in mouse and zebrafish uncovers high interspecies divergence. *Frontiers in Genetics* 7:110. doi:10.3389/FGENE.2016.00110

Zhao, C.Q., Young, M.R., Diwan, B.A., Coogan, T.P., Waalkes, M.P., 1997. Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression. *Proceedings of the National Academy of Sciences of the Unites States of America* 94(20):10907–10912. doi:10.1073/pnas.94.20.10907

Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y., Giovannoni, J.J., 2013. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology* 31:154–159. doi:10.1038/nbt.2462

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(1):129–138. doi:10.1111/j.1365-2583.2011.01121.x

Appendix

List of publications and abstracts

Publication

Athanasio, C.G., Chipman, J.K., Viant, M.R., Mirbahai, L., 2016. Optimisation of DNA extraction from the crustacean *Daphnia*. PeerJ 4, e2004. doi:10.7717/peerj.2004

Abstracts

Athanasio, C.G., Mally, A., Viant, M.R., Chipman, J.K., Mirbahai, L., DNA methylation in the model organism *Daphnia magna*. SETAC/iEOS Joint Focused Topic Meeting 2016. Oral presentation. Ghent, Belgium.

Athanasio, C.G., Mirbahai, L., Parameswaran, S., Gopalakrishnan, C., Viant, M.R., Chipman, J.K., Epigenetic memory of environmental stressors in *Daphnia magna*. 25th Annual Meeting SETAC Europe 2015. Poster presentation. Barcelona, Spain.

Athanasio, C.G., Mirbahai, L., Viant, M.R., Chipman, J.K., Epigenetic memory of environmental stressors in *Daphnia magna* - Updates*.* Biosciences Graduate Research School Symposium 2015. Oral presentation. The University of Birmingham, UK.

Athanasio, C.G., Mirbahai, L., Viant, M.R., Chipman, J.K., Epigenetic memory of environmental stressors in *Daphnia magna* – Plans. Biosciences Graduate Research School Symposium 2014. Poster presentation. The University of Birmingham, UK.

List of additional files

Additional files are available on the CD included at the back of this thesis

File 3.1 Genome_B2_Bham_denovo_assembled_genome

File 3.2 Predicted_CpG_Islands_B2

File 3.3 Transcription Start Sites (TSS)_B2

File 3.4 Annotated mRNA_B2

File 3.5 Excluded_scaffolds_Methylation_analyses_B2

File 3.6 List of MethylatedCytosines_14days

File 3.7 List of MethylatedCytosines_5days

File 4.1 List of Probes with Altered Methylation_Age

File 4.2 List of DMRs_Exposure to Stressors