Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the

Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M. Sc. Fanny Gatzmann

born in Bonn, Germany

Oral examination: 18.01.2019

DNA methylation in the marbled crayfish *Procambarus virginalis*

Referees

Prof. Dr. Frank Lyko

Prof. Dr. Henrik Kaessmann

Abstract

The all-female marbled crayfish *Procambarus virginalis* is a freshwater crayfish which is the only known obligatory parthenogen among the decapod crustaceans. Marbled crayfish are recent descendants of the sexually reproducing slough crayfish *Procambarus fallax* and have most likely emerged through a recent evolutionary macromutation event in *P. fallax*. Marbled crayfish reproduce by apomictic parthenogenesis, where oocytes do not undergo meiosis and all offspring are genetically identical clones of the mother. Nevertheless, marbled crayfish show a high degree of phenotypic variation and are a highly invasive species, where (through parthenogenesis) a single animal can establish a whole population. Moreover, they have been distributed via the pet trade and anthropogenic releases, and have formed stable populations in a variety of ecological habitats. Earlier this year, our group performed whole-genome sequencing for 11 marbled crayfish animals from different populations and countries, and found only four non-synonymous single nucleotide variances in coding regions. Since the marbled crayfish's remarkable adaptability is not due to genetic variability, it is crucial to investigate epigenetic programming in this organism.

I present here a comprehensive analysis of DNA methylation in marbled crayfish. Whole-genome bisulfite sequencing data was used to directly compare methylation patterns from multiple replicates in different tissues and from different marbled crayfish and *Procambarus fallax* animals. These methylation maps were integrated with RNA-seq and ATAC-seq data to comprehensively analyse the interplay between DNA methylation, chromatin accessibility, and gene expression. I found 18% of CpGs in marbled crayfish to be methylated. Repeats showed overall low methylation levels, with the exception of a single class of DNA transposons, which was ubiquitously methylated. DNA methylation was mainly targeted to the coding regions of housekeeping genes in marbled crayfish. In contrast to paradigmatic mammalian methylomes, I only observed very moderate methylation differences between tissues for both gene bodies and promoters. I did, however, identify a set of approximately 700 genes that showed a high variance in their methylation across tissues and animals. Gene body methylation was significantly inversely correlated with gene expression variability. Interestingly, the marbled crayfish shows overall lower methylation levels and higher gene expression variability than its parent species *P. fallax*. Since plasticity in gene expression can be a beneficial trait for adapting to new environments, this trait might contribute to the marbled crayfish's adaptive and invasive success. The integrative analysis of DNA methylation, chromatin accessibility, and gene expression revealed that genes with highly methylated gene bodies were located in regions of poorly accessible chromatin and showed stable expression patterns. In contrast, lowly methylated genes were found in more accessible chromatin when stably expressed, and in more condensed chromatin when variably expressed. In this context, gene body methylation might function to stabilise gene expression in regions of limited chromatin accessibility.

These findings broaden our knowledge of evolutionary conservation of DNA methylation patterns in invertebrates and provide novel insights on the interplay between gene body methylation, chromatin accessibility, and gene expression.

Kurzzusammenfassung

Der Marmorkrebs *Procambarus virginalis* ist ein Süßwasserkrebs, der sich durch obligatorische Parthenogenese fortpflanzt und aus der Familie der Decapoda (Zehnfußkrebse) stammt, die zum Subphylum Crustacea (Krebstiere) gehört. Er ist vermutlich vor etwa 30 Jahren durch eine Makromutation aus dem Everglades-Sumpfkrebs *Procambarus fallax* hervorgegangen, der sich geschlechtlich fortpflanzt. Bei der Entstehung der Eizellen im Marmorkrebs findet keine Meiose statt, sodass der rein weibliche Nachwuchs genetisch identisch mit der Mutter ist. Trotzdem zeigen Marmorkrebse eine auffällige phenotypische Variabilität und sind sehr invasiv, da durch die parthenogenetische Reproduktion ein einzelnes Tier eine neue Population gründen kann. Zusätzlich wurden sie durch den Tierhandel verbreitet und von Menschen ausgesetzt, und haben weltweite stabile Populationen gebildet. Unsere Arbeitsgruppe hat dieses Jahr die Genome von 11 Marmorkrebsen verschiedenen Ursprungs miteinander verglichen und nur vier nicht-synonyme Einzelnukleotid-Varianten in kodierenden Regionen identifiziert. Da die Anpassungsfähigkeit des Marmorkrebses folgendermaßen nicht durch genetische Variation zu begründen ist, untersuche ich in dieser Arbeit epigenetische Regulation im Marmorkrebs.
Ich präsentiere eine umfassende Analyse der DNA Methylierung im Marmorkrebs. Für verschiedene Gewebe und Tiere, sowohl vom Marmorkrebs als auch von *Procambarus fallax*, wurde Bisulfit-Sequenzierung des ganzen Genomes durchgeführt. Diese Datensätze wurden durch RNA-Sequenzierung und ATAC-Sequenzierungsdaten ergänzt, um das Zusammenspiel dieser drei Faktoren vollständig zu analysieren. 18% aller CpGs im Marmorkrebs waren methyliert. Repetetive DNA war, mit Ausnahme einer stark methylierten Transposonklasse, schwach methyliert. DNA Methylierung fand sich primär in kodierenden Regionen von Genen, vor allem in konstitutiv exprimierten Genen. Im Gegensatz zu oft analysierten Säugetier-Methylomen habe ich nur schwache Unterschiede der Methylierung zwischen Geweben beobachtet. Ich habe jedoch knapp 700 Gene identifiziert, die eine starke Variabilität in ihrer Methylierung zwischen den analysierten Proben zeigten. Die Methylierung in kodierenden Regionen korreliert statistisch signifikant invers mit der Genexpressionsvariabilität. Interessanterweise ist der Marmorkrebs im Vergleich mit seiner Elternspezies signifikant schwächer methyliert und zeigt auch eine signifikant erhöhte Genexpressionsvariabilität. Da Plastizität in der Genexpression von Vorteil sein kann wenn sich ein Organismus an ein neues Habitat anpassen muss, besteht die Möglichkeit dass Genexpressionsvariabilität durch Hypomethylierung dazu beiträgt dass der Marmorkrebs so erfolgreich verschiedene Habitate besetzt. Die integrative Analyse von DNA Methylierung, Chromatinzugänglichkeit und Genexpression zeigte dass stark methylierte Gene sich generell in kondensiertem Chromatin befanden und gleichmäßig exprimiert wurden, während schwach methylierte Gene sich in zugänglicherem Chromatin befanden wenn sie stabil exprimiert waren, und in weniger zugänglichem Chromatin wenn sie variabel exprimiert wurden. In diesem Zusammenhang könnte Methylierung der kodierenden Regionen dafür sorgen dass Gene stabil exprimiert werden, auch wenn sie in kondensiertem Chromatin liegen.
Die Ergebnise dieser Arbeit erweitern unser Wissen über die evolutionäre Konservierung von Methylierungsmustern in Invertebraten, und verschaffen neue Eindrücke über das Zusammenspiel zwischen DNA Methylierung in kodierenden Regionen, Chromatinzugänglichkeit und Genexpressionsmustern.

Contents

1 Introduction

1.1 The marbled crayfish *Procambarus virginalis*

The marbled crayfish *Procambarus virginalis* is a triploid organism which is the only known obligatory parthenogen among the decapod crustaceans (Martin et al., 2015; Scholtz et al., 2003). It represents a novel freshwater crayfish species that emerged in the German aquarium trade in 1995 (Lyko, 2017a). It is assumed that marbled crayfish are descendants of the sexually reproducing slough crayfish *Procambarus fallax* that originated through an evolutionary recent macromutation event (Martin et al., 2010; Vogt et al., 2015). The reproduction mode of the all-female marbled crayfish is by apomictic parthenogenesis, i.e., oocytes do not undergo meiosis, females lay unfertilised, triploid eggs, and all progeny are genetically identical clones of the mother (Martin et al., 2007). In the course of its lifetime, a marbled crayfish can reproduce up to seven times, with an average of 400 genetically identical offspring per clutch (Seitz et al., 2005).

Marbled crayfish animals show a high degree of phenotypic variation, where even offspring of the same clutch display differences in their growth, coloration pattern, lifespan, reproduction and behavior despite growing up in the same environment (Vogt, 2008). Figure 1A shows siblings from the same clutch of eggs, raised in the same environment. The parthenogenetic mode of reproduction and high fecundity of marbled crayfish make them a highly invasive species, which can establish large populations from a single animal. Moreover, marbled crayfish have been distributed via the pet trade and anthropogenic releases, which resulted in populations in a wide range of countries and different ecological habitats (Jones et al., 2009; Chucholl et al., 2012). This includes confirmed reports in Madagascar, Germany, Slovakia, Czech Republic, Romania, Hungary, Ukraine and Malta (Gutekunst et al., 2018; Jones et al., 2009; Chucholl et al., 2012; Lipták et al., 2016; Patoka et al., 2016; Pârvulescu et al., 2017; Lőkkös et al., 2016; Deidun et al., 2018) (Figure 1B). Especially in Madagascar, marbled crayfish have rapidly spread in the last 15 years, with 33 different positive confirmed sites that span more than 100,000 km^2.

Earlier this year, our group published a study where we performed whole-genome sequencing for 11 marbled crayfish animals from a variety of sources and countries (Gutekunst et al., 2018). We found that amongst all these animals, only four non-synonymous single nucleotide variances occurred in coding regions. This is striking, considering the wide variety of environments these animals were found in. Since genetic variation can be disregarded as the mechanism for the marbled crayfish's remarkable adaptability, it is crucial to investigate epigenetic regulation in this organism to explore the possibility that its invasive capability is driven by epigenetic mechanisms. The genetic homogeneity of the marbled crayfish, with very few confounding mutations, makes it a particularly promising model organism for epigenetics research.

1.2 Epigenetics

The genomic sequence of a single fertilised cell contains all the information needed to develop into a complex organism of various cell types. The term epigenetics originates from

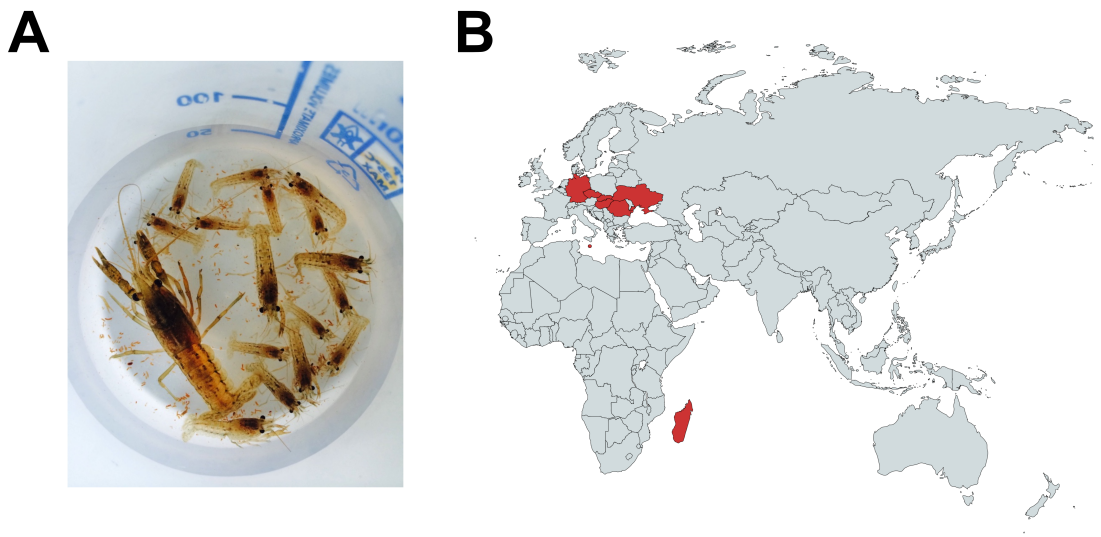

Figure 1: **Marbled crayfish siblings and global marbled crayfish populations.** (A) Marbled crayfish siblings from the same clutch of eggs displaying different coloration patterns and size. (B) Confirmed populations have been reported in Madagascar, Germany, Slovakia, Czech Republic, Romania, Hungary, Ukraine and Malta (Gutekunst et al., 2018; Jones et al., 2009; Chucholl et al., 2012; Lipták et al., 2016; Patoka et al., 2016; Pârvulescu et al., 2017; Lőkkös et al., 2016; Deidun et al., 2018).

the term epigenotype, and initially only described all the developmental events that lead from the zygote to a fully developed organism (Waddington, 1942). Today, the definition of epigenetics has evolved and is understood as the study of heritable changes in expression profiles and thereby, eventually, phenotypes, independent of alterations in the underlying Watson-Crick base-pairing face (Riggs et al., 1996). Epigenetics can establish cellular identity through lineage-specific expression patterns, which are maintained through replication and cell division (Fisher, 2002). The mechanisms inducing these expression profiles are reversible (Ramchandani et al., 1999) and can act much faster than genetic mutations (Rando and Verstrepen, 2007), which means cells are, to a degree, plastic in their expression patterns and differentiation. Since the epigenome can be altered by the environment (Duncan et al., 2014; Dolinoy et al., 2007; Lyko et al., 2010), epigenetic mechanisms may show a dynamic response to the environment, or a phenotypic plasticity, by changing expression patterns accordingly.

While epigenetic marks consist of a number of modifications and components, the majority of epigenetics studies have been focused on covalent modifications of nucleic acids (mainly DNA methylation), post-translational histone modifications, as well as non-coding RNAs and RNA editing (Goldberg et al., 2007; Bernstein et al., 2007; Peschansky and Wahlestedt, 2014; Eisenberg and Levanon, 2018).

1.3 DNA methylation

In eukaryotic genomes, DNA methylation is a biochemical process that predominantly involves the addition of a methyl group to the carbon 5 of cytosine (to form 5mC) through a covalent bond. This modification was first described in 1948 (Hotchkiss, 1948) and has early on been proposed as an epigenetic mark that might be involved in X-chromosome inactivation (Riggs, 1975) and gene regulation during development (Holliday and Pugh, 1975). It is an ancient modification which is present in all three domains of life, and is the best studied epigenetic mark today. It is essential for mammalian embryonic development, as has been shown by the lethality in mice where gene targeting was used on the enzyme that catalyses this mark, leading to drastically reduced methylation levels (Li et al., 1992; Okano et al., 1999). In animals, DNA methylation is mainly targeted to CpG dinucleotides (which means a cytosine and guanine pair ordered in 5' to 3' direction, linked with a phosphate) (Jones, 2012) and has been shown to play a vital role in a variety of biological processes such as development, gene regulation, chromatin remodeling and the suppression of transposable elements (for functions of DNA methylation, see section 1.3.4).

The family of DNA methyltransferases (DNMTs) establishes and maintains the methylation of genomic cytosines in a wide range of organisms (see section below). Other major DNA modifications include the oxidised derivative of 5mC, 5-hydroxymethylcytosine (5hmC), as well as N6-methyladenine (6mA) (Breiling and Lyko, 2015), which is the predominant DNA modification in prokaryotes (Wion and Casadesús, 2006).

1.3.1 The DNA methyltransferase family

The family of DNMTs catalises the transfer of a methyl group from S-adenyl methionine (SAM) to the fifth carbon of cytosine. DNMT3 is usually described as the *de novo* methyltransferase which establishes an initial methylation of unmethylated DNA. DNMT3 in animal genomes methylates primarily CpG dinucleotides, while DRMs, the DNMT3 homolog in plants, can methylate cytosine in any context (called Domains Rearranged Methyltransferases since their catalytic domains are rearranged with respect to DNMT3) (Lyko, 2017b). DNMT1, on the other hand, is considered the maintenance methyltransferase, which propagates methylation patterns through DNA replication: during replication, in a newly synthesised double strand of DNA, only the original strand will carry the epigenetic mark of methylation. DNMT1 recognises this hemimethylated DNA and mediates methylation of the unmethylated daughter strand (Lyko, 2017b; Goll and Bestor, 2005; Law and Jacobsen, 2010).

Both DNMT1 and DNMT3 show strong conservation in the context of eukaryotic methylation: all plants and animals that display cytosine methylation in their genome possess at least one copy of either DNMT1 or DNMT3 (Zemach and Zilberman, 2010; Lyko, 2017b). DNMT3 appears to be somewhat more dispensable than DNMT1, having been lost in a number of organisms such as algae and the silk moth *Bombyx mori* (Zemach and Zilberman, 2010; Lyko, 2017b).

DNMT2 does not methylate DNA, but is actually a tRNA methyltransferase which methylates cytosine 38 in some tRNAs, primarily tRNAs carrying the amino acid aspar-

tic acid, but also amino acids glycine and valine (Goll, 2006; Legrand et al., 2017). Still, it is highly conserved in eukaryotes and often mentioned alongside the traditional DNA methyltransferases DNMT1 and DNMT3.

Enzymes of the Ten–eleven translocation (TET) family, on the other hand, usually catalyse the stepwise oxidation of 5-methylcytosine to 5-hydroxymethylcytosine and further oxidation products (Lyko, 2017b). They are defined by a catalytic dioxygenase domain that performs the oxidisation step. They initiate demethylation, and thereby prevent hypermethylation in the genome, antagonising the DNMT enzymes.

Figure 2A shows a schematic overview of a eukaryotic DNA methylation system.

With the first discovery of a mammalian DNA methyltransferase (DNMT1 in mouse cells), it has been demonstrated that the C-terminal catalytic domain of DNMT1 shows strong conservation and could also be found in bacteria (Bestor et al., 1988). Indeed, five signature catalytic domains of DNMTs are strongly conserved across species, while another five also show some degree of conservation (Pósfai et al., 1989). Animal DNMT enzymes are normally grouped into a regulatory domain at the N-terminus, and a catalytic domain at the C-terminus (Figure 2B) (Lyko, 2017b).

While both DNMT1 and DNMT3 are evolutionary highly conserved and are present in a wide range of animal genomes, each has experienced gains and losses of their number of paralogues in different organisms (Lyko, 2017b) (Figure 2C). Similarly to DNMT1 and DNMT3, DNMT2 shows a high evolutionary conservation across the animal kingdom (Lyko, 2017b) (Figure 2C).

1.3.2 The DNA methylation landscape in animals and plants

In animals, DNA methylation predominantly occurs at CpG dinucleotides and is symmetrical, i.e., it occurs on both strands of the DNA for a paired CpG dinucleotide (Bird, 1980; Zemach et al., 2010; Feng et al., 2010). The frequency of CpG dinucleotides in methylated genomes is lower than would be statistically expected based on the frequency of Cs and Gs (Bird, 1980). This is most likely due to the spontaneous deamination of methylated cytosines to thymines, resulting in a C$>$T mutation, which leads to a CpG depletion over time. In contrast, unmethylated cytosines deaminate to uracil, which is recognised as a DNA-foreign base and excised by the uracil-DNA glycosylase (Coulondre et al., 1978).

While DNA methylation is present in all domains of life, the methylation levels and landscapes in a genome can show substantial diversity (Breiling and Lyko, 2015; Feng et al., 2010; Zemach and Zilberman, 2010). Mammalian genomes, for example, are highly methylated (with 70-80% of all CpG dinucleotides showing methylation) (Li and Zhang, 2014), where the modification has been termed the "fifth base" of the genome. Major exceptions to these global methylation levels are CpG islands, i.e., regions of high CpG density ($>$50%) that are mostly devoid of methylation (Bird et al., 1985). The CpG density in these regions is due to the fact that these cytosines are unmethylated: since they do not deaminate to thymine, but to uracil, the mismatch repair system can accurately re-

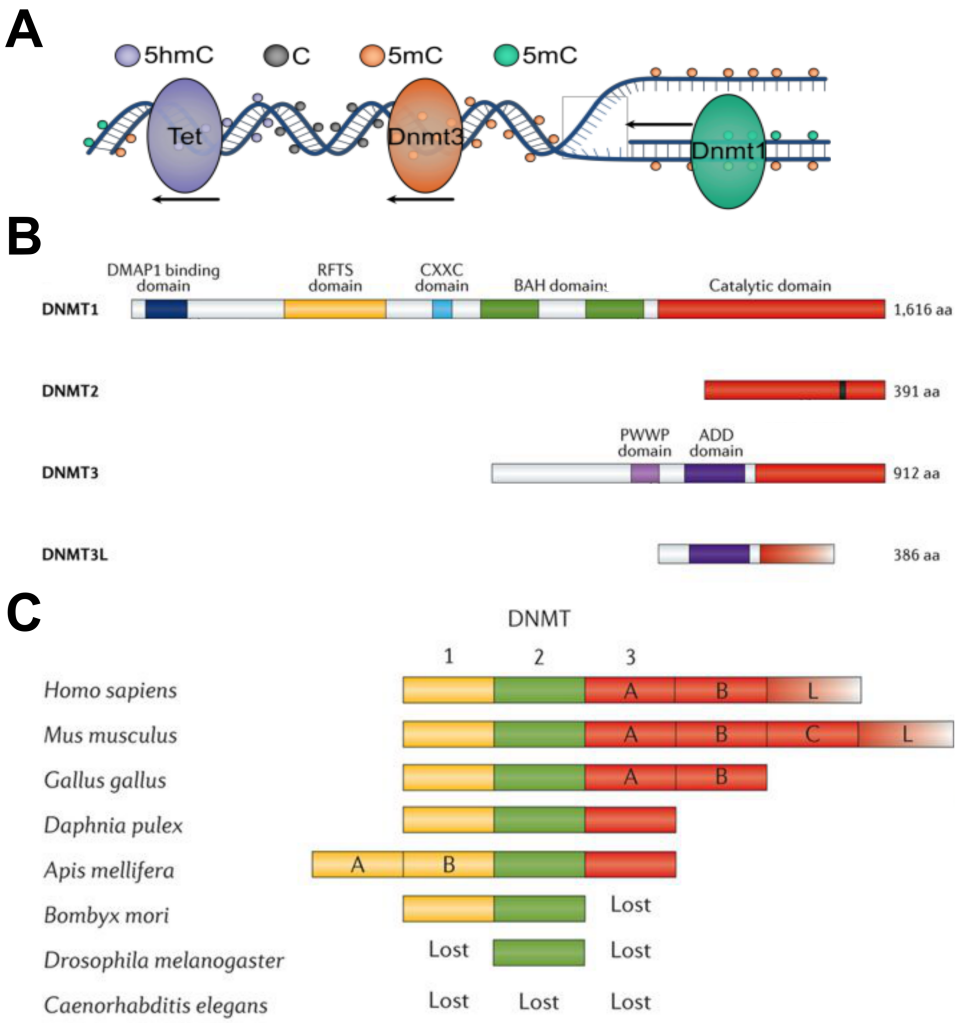

Figure 2: **Conserved catalytic domains, catalytic mechanism and paralogs in different organisms for DNMTs.** (A) Schematic illustration of a DNA methylation system. Enzymes (front) and base modifications (back) are displayed on the DNA strand. DNMT1 (green, on the right) copies the methylation mark from the maternal strand (orange) to the nascent daugter strand during replication (green). DNMT3 (orange, middle) establishes *de novo* methylation patterns. TET (purple) oxidises 5-methylcytosine to 5-hydroxymethylcytosine (purple). (B) Catalytic domains in DNMTs. Each conserved domain is shown in different colors. The catalytic domain (red) is conserved in all DNMTs. The number of amino acids (aa) indicated is representative of the human homologue. DNMT3A is shown representatively for DNMT3. DNMT3L is a catalytically inactive DNMT3 variant that lacks the N-terminal part of the regulatory domain (including the Pro-Trp-Trp-Pro (PWWP) domain) and the C-terminal part of the catalytic domain. (C) Copy number variation of DNMT paralogues in different organisms. (A) adopted from Falckenhayn (2016), (B) and (C) adopted from Lyko (2017b).

cognise and correct mismatches that occur due to deamination (Coulondre et al., 1978). About 60% of promoters in the human genome are associated with CpG islands (Bernstein

et al., 2007; Li and Zhang, 2014), suggesting a functional relevance (see section 1.3.4).

While the majority of studies still focus on mammalian and vertebrate methylation, an ever-increasing number of single-base resolution methylation maps for invertebrate species and plants is being published. In contrast to the ubiquitously methylated vertebrate genomes, methylation levels in invertebrates show a high degree of diversity in their methylation levels (Feng et al., 2010; Zemach et al., 2010; Bewick et al., 2016).

The sea squirt *Ciona intestinalis*, for example, displays intermediate CpG methylation levels of 23.6%, arranged in a so-called "mosaic" methylation pattern (Suzuki et al., 2013). The termite *Zootermopsis nevadensis* with 12% methylated CpGs (Glastad et al., 2016a) is an example for moderate methylation levels. Furthermore, there are sparsely methylated genomes like that of the ant *Dinoponera quadriceps* (3% of all CpGs methylated), or the honeybee (0.7% of all CpGs methylated), where patterns would be termed "sporadic" along the genome (Patalano et al., 2015; Lyko et al., 2010). Finally, some invertebrate species have lost 5mC methylation, like *Drosophila melanogaster*, *Schistosoma mansoni* or *Caenorhabditis elegans* (Raddatz et al., 2013; Simpson et al., 1986).

Plants, again, also show diverse methylation levels that can be intermediate, an example being *Arabidopsis thaliana* with approximately 22% methylation, or high, like *Oryza sativa* with approximately 59% methylation levels (Feng et al., 2010). Plant species have also been shown to exhibit methylation at non-CpG cytosines, in the CHG and CHH context (where H denotes A, C or T) (Law and Jacobsen, 2010).

Methylation in organisms that are not ubiquitously methylated is usually targeted to specific genomic features. Most notably, these include gene bodies and repetitive elements (Suzuki and Bird, 2008; Feng et al., 2010; Zemach et al., 2010). Plants mostly methylate both their gene bodies and their repeats. Methylated invertebrate genomes, on the other hand, show a preference for gene body methylation over repeat methylation (Zemach et al., 2010; Feng et al., 2010).

Gene body methylation in plants, animals, and fungi

Except for most fungi species, all eukaryotic methylomes exhibit methylation at the bodies of protein-coding genes (Zemach et al., 2010; Feng et al., 2010). Figure 3 shows methylation patterns around coding regions for two land plants species (A and B, first row), two invertebrates (C and D, second row), two vertebrate genomes (E and F, third row) and two fungal species (G and H, bottom row).

Specific for the analysed land plant gene body methylation species is a drop in methylation around the transcription start site and the transcription termination site that extends generously into the transcribed region (Figure 3A and B). It should also be noted that plants display gene body methylation only at CpG dinucleotides (Zemach et al., 2010; Feng et al., 2010). The methylation levels in the center of the gene body and the genome upstream and downstream are comparable, with the largest increase in the gene body found in *Arabidopsis thaliana*. Interestingly, a water plant, *Chlamydomonas reinhardtii*, which is a unicellular freshwater green algae, exhibits overall much lower methylation levels than the

land plants, and methylates its gene bodies with only a slight increase in methylation compared to the surrounding genome (Feng et al., 2010).

The two invertebrate species, *Apis mellifera* and *Ciona intestinalis*, display a distinct in-

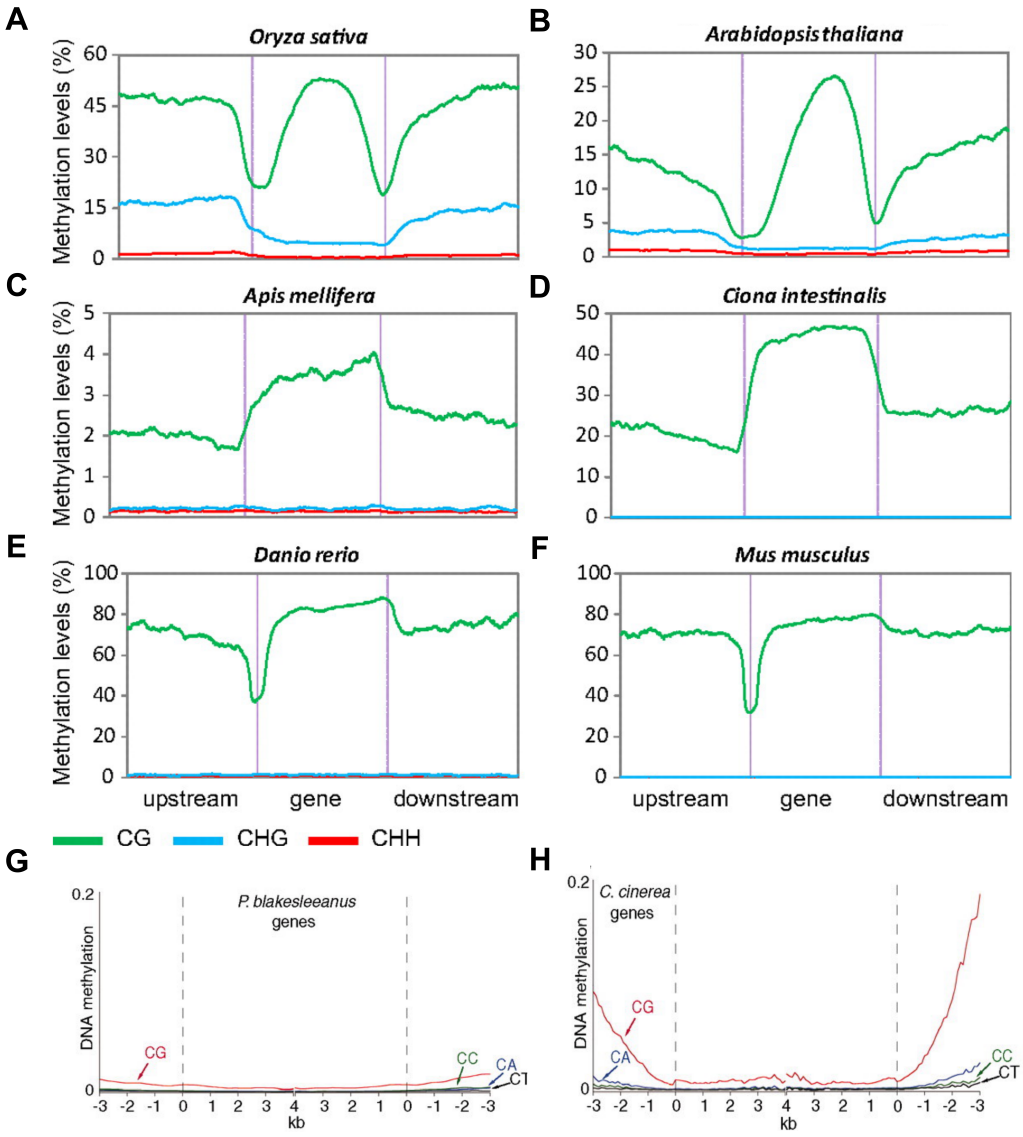

Figure 3: **Gene body methylation patterns for eight animals, plants and fungi.** (A)-(F), metagene plot showing methylation along the gene body where colours denote methylation context (CpG, CHG, or CHH) as indicated. (G)-(F), similar analysis for fungi, context again as indicated by colour (CC, CA, CG). Animals and plants adopted from Feng et al. (2010), Fungi adopted from Zemach et al. (2010).

crease of methylation in their gene bodies compared to the surrounding genome background (Figure 3, second row). Then again, vertebrates, with the highest amount of methylation, show comparable levels of methylation in the surrounding genome and the gene body, with a short, sharp drop of methylation at the transcription start and end site (Figure 3, third row).

Figure 3G and H shows gene body methylation for two fungi species, which do not exhibit gene body methylation. However, a moderate amount of body methylation has been observed in the fungal species *Uncinocarpus reesii* (Zemach et al., 2010).

Methylation of repetitive elements in plants, animals, and fungi

Repeat methylation shows different conservation patterns in eukaryotes compared to gene bodies: vertebrates, plants, and fungi generally methylate their repetitive elements, while this targeting is not as conserved in invertebrates, whose repeats are often unmethylated (Feng et al., 2010; Zemach et al., 2010).

Figure 4 shows repeat methylation for the same species as described above: two land plant species (top row), two invertebrates (second row), two vertebrates (third row), and two fungal species (bottom row). For land plants, methylation levels in repeats displays a distinct plateau compared to the surrounding genome (Figure 4C and D). This plateau is not quite as pronounced in water plants (Feng et al., 2010). For plants, methylation in repeats occurs predominantly in a CpG and CHG context, with only a few occurrences in a CHH context (Feng et al., 2010; Zemach et al., 2010).

Invertebrates, on the other hand, experience either barely any change in methylation levels for repeats (*Ciona intestinalis*), or even display a slight decrease (*Apis mellifera*) compared to the surrounding genome. In other invertebrate methylomes, a slight methylation of repeats was observed, like the desert locust *Schistocerca gregaria* (Falckenhayn et al., 2013), the sand flea *Parhyale hawaiensis* (Kao et al., 2016), or the Pacific oyster *Crassostrea gigas* (Wang et al., 2014b).

In vertebrates, repeats are similarly methylated in comparison to the surrounding genome (Figure 4, third row).

The bottom row of Figure 4 shows repeat methylation in two fungal species, which both display repeat methylation. Another fungal species, *Uncinocarpus reesii* has been reported very little CpG methylation in repeats, since repeats in this species are depleted of CpG dinucleotides (Zemach et al., 2010). Methylation in repeats this species occurs preferably at CC, CT and CA dinucleotides (Zemach et al., 2010).

1.3.3 Functions of gene body DNA methylation

The functionality of gene body methylation remains to be fully understood (Zilberman, 2017; Bewick and Schmitz, 2017). A wide variety of roles in the regulation of genes have been proposed for gene body methylation, including an involvement in chromatin remodelling (Lorincz et al., 2004) and in modulating alternative splicing of mRNA (Lyko et al., 2010). Further functions that are closer in connection with this thesis are described below.

Active transcription of gene body-methylated genes

Actively transcribed genes are usually methylated in their gene bodies. The association between methylation of coding regions and expression has been shown in plants (Zilberman et al., 2007), as well as humans (Ball et al., 2009) and a number of invertebrates (Zemach et al., 2010; Glastad et al., 2016a; Bonasio et al., 2012; Suzuki et al., 2013). It should be noted that, for both plants and invertebrates, genes with the highest methyla-

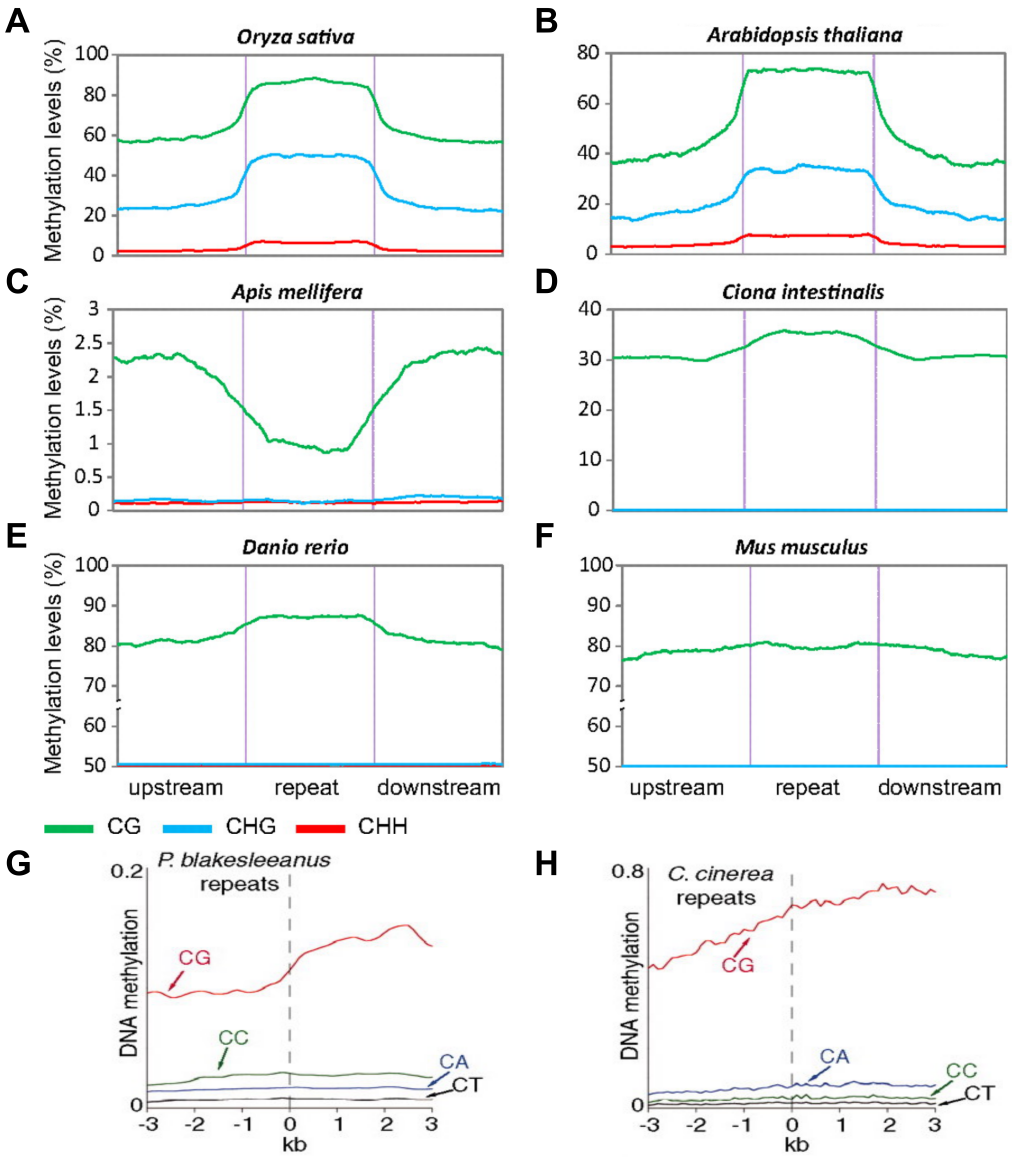

Figure 4: **Repeat methylation patterns for eight animals. plants and fungi.** (A)-(F), metagene plot showing methylation along the gene body where colours denote methylation context (CpG, CHG, or CHH) as indicated. (G)-(F), similar analysis for fungi, context again as indicated by colour (CC, CA, CG). Animals and plants adopted from Feng et al. (2010), Fungi adopted from Zemach et al. (2010).

tion levels in their gene bodies are often moderately expressed, while the most and the least expressed genes are usually undermethylated (Zilberman et al., 2007; Zemach et al., 2010). Zilberman et al. (2007) proposed that aberrant transcripts can result from cryptic intragenic promoters that are exposed as the polymerase travels along the gene body and disrupts the chromatin. They suggested that these short, aberant transcripts can lead to methylation of homologous DNA through the short interfering RNA (siRNA) pathway (Chan et al., 2005).

Methylation targets housekeeping genes and reduces transcriptional noise
In invertebrates and plants, methylation is preferentially targeted to groups of genes that are highly conserved and constitutively expressed. These groups most likely consist mainly of housekeeping genes, while presumably tissue-specific genes are less methylated (Takuno and Gaut, 2012; Sarda et al., 2012; Suzuki and Bird, 2008). This has been shown for a wide range of organisms including invertebrate and insect species, notable examples being *Apis mellifera*, *Bombyx mori*, *Ciona intestinalis*, the sea anemone *Nematostella vectensis* (Sarda et al., 2012), the termite *Zootermopsis nevadensis* (Glastad et al., 2016a)), and *Arabidopsis thaliana* (Takuno and Gaut, 2012). Methylation of coding regions has also been associated with the suppression of cryptic intragenic promoters or transcriptional noise (Neri et al., 2017), which makes sense, considering that housekeeping genes should produce accurate, not aberrant transcripts. In this context, it is also interesting that an inverse correlation between gene body methylation and gene expression variation has been demonstrated for a number of species. These include invertebrates *Z. nevadensis* (Glastad et al., 2016a), *C. intestinalis* (Suzuki et al., 2013), but also humans (Huh et al., 2013). However, a mechanistic explanation of how gene body methylation stabilises the number of transcripts per gene remains lacking.

1.3.4 Functions of DNA methylation outside of coding regions

DNA methylation has been shown and proposed to carry out a variety of biochemical functions. Early on, DNA methylation has been suggested to be the responsible epigenetic mark for X-chromosome inactivation (Riggs, 1975), gene regulation during development (Holliday and Pugh, 1975) and cell differentiation (Compere and Palmiter, 1981). Later on, with improving techniques to analyse DNA methylation, further functions have been added to the list, including imprinting, transcriptional regulation, chromatin remodelling, as well as silencing of transposable elements (Bird, 2002).

Gene regulation and embryonic development
Since the discovery that *in vitro* methylated DNA is not expressed when transfected into *Xenopus laevis* oocytes, methylation been associated with the silencing of genes when it is present near gene regulatory regions (Vardimon et al., 1982; Breiling and Lyko, 2015). In such promoters, methylation may either inhibit transcription factors from binding to these regions, or attract methyl-CpG binding protein complexes, which function as active repressors (Li and Zhang, 2014). While DNA methylation has been shown to be essential for mammalian embryonic development (Li et al., 1992; Okano et al., 1999; Smith and Meissner, 2013), it is not generally essential to organismal development, since it is absent in many organisms like *Drosophila melanogaster*, *Schistosoma mansoni* or *Caenorhabditis elegans* (Raddatz et al., 2013; Simpson et al., 1986).

Transposon silencing
In vertebrates, plants, and fungi, methylation of transposable elements (TEs) is associated with the suppression of their transpositioning activities. It is therefore considered a

genome defense mechanism (Walsh1998, Zemach2010). In invertebrates, however, the data is inconclusive. Kao et al. (2016), for example, reported moderate amounts of transposon methylation in the sand flea *Parhyale hawaiensis*, as did Glastad et al. (2016a) for the termite *Zootermopsis nevadensis* and Falckenhayn et al. (2013) for the desert locust *Schistocerca gregaria*. On the other hand, neither *Apis mellifera* nor *Ciona intestinalis* displayed significant enrichment of DNA methylation in repeats (Lyko et al., 2010; Simmen et al., 1999). It follows that repeats in invertebrates must be, at least partially, silenced by other mechanisms (Zemach and Zilberman, 2010).

DNA methylation as a quick response to a changing environment
In recent years, DNA methylation has repeatedly been suggested as an mechanism that allows the organism to respond to environmental cues through gene expression changes (Jaenisch and Bird, 2003; Verhoeven et al., 2016; Duncan et al., 2014), in addition to transcription factors, translational modification elements, and post-transcriptional modification factors that adapt to environmental factors. However, experimental data to determine the importance of methylation in this context has remained sparse. DNA methylation has been shown to rapidly respond to environmental stress in mice (Radford et al., 2014), plants (Dowen et al., 2012) and the crustacean *Daphnia pulex* (Asselman et al., 2017). Such changes in DNA methylation patterning may induce plasticity in organisms (Vogt et al., 2008), possibly by altering phenotypic traits through gene expression changes. Furthermore, it has been suggested that certain genetic variants could induce variable phenotypes that are epigenetically mediated. In this case, the stochasticity of methylation patterns would produce different phenotypes, which would be subjected to natural selection of the environment (Feinberg and Irizarry, 2010).

1.4 Computational analysis of DNA methylation using bisulfite sequencing data

1.4.1 Bisulfite sequencing to determine the methylation status

In 1980, Wang et al. reported that, when subjecting 5-methylcytosine to bisulfite treatment, it deaminates to uracil much more slowly than unmethylated cytosines. Based on this observation, Frommer et al. (1992) proposed that this differing reaction rate could be exploited to analyse DNA methylation patterns in genomic DNA. With this method, not a specific chemical modification would have to be detected, but a base exchange, which was much easier since sequencing methods had already been established. This concept for methylation analysis, i.e., subjecting DNA to bisulfite treatment, selectively deaminating unmethylated cytosines to uracil while methylated cytosines stay unchanged, would lay the foundation for a number of methods to come in the following years. The most notable techniques to use this mechanism for methylation analysis include 454 sequencing (Taylor et al., 2007), SOLiD sequencing (Sequencing by Oligonucleotide Ligation and Detection) (Pandey et al.), RRBS (Reduced Representation Bisulfite Sequencing) (Meissner et al., 2005), and, finally, whole-genome bisulfite sequencing.

Today, whole-genome bisulfite sequencing (Cokus et al., 2008; Lister and Ecker, 2009)

(WGBS) represents the gold-standard methodology to analyse genome-wide DNA methylation patterns at single-base resolution. After its isolation, the DNA is fragmented and treated with sodium bisulfite. This leads to the deamination of unmethylated cytosines into uracil, while methylated and hydroxymethylated cytosines will stay the same (Huang et al., 2010). The DNA is then subjected to standard whole-genome sequencing protocols including library preparation and a PCR, which substitutes uracils for thymines. The libraries are sequenced on second-generation Illumina platforms.

1.4.2 Computational analyses of whole-genome bisulfite sequencing data

After sequencing, whole-genome bisulfite sequencing reads are computationally processed and mapped to the reference genome of the respective organism. The basic data processing steps include the quality control of reads, trimming of reads for quality, trimming adapters, aligning reads to the reference genome and methylation calling. The trimming step is crucial since the failure to do so may result in low mapping efficiencies, misalignments and errors in methylation calling since adapters can appear methylated if they were added after bisulfite conversion. Mapping of bisulfite-converted reads is challenging due to the reduced sequence complexity of reads (few cytosines), and because cytosines could be converted to thymines, so a thymine in the sequenced reads could be mapped against either cytosine or thymine in the reference genome. This drastically increases the search space for mapping and makes the matching process more complicated.

The bisulfite alignment software BSMAP (Xi and Li, 2009) has been developed to address these issues. It converts thymines in the bisulfite reads to cytosines only for positions where there is a cytosine in the reference, while keeping all other thymines in the bisulfite reads unchanged. Then, it maps the read directly to the reference. This conversion during mapping is performed through position-specific bitwise masking of the read, and combined with hashing of the genome, making it sensitive and efficient. After mapping, DNA methylation at cytosine positions in the genome is assessed by quantifying the number of methylated and unmethylated reads mapping to a cytosine position.

To assess the quality of the data, the success of the bisulfite conversion should be evaluated. This can be done, for example, by analysing the post-mapping methylation status of mitochondrial DNA, which should be completely unmethylated. Alternatively, unmethylated spike-in DNA (viral DNA of the lambda phage is usually used for this purpose) can be added before sequencing, and its cytosines should have been largely converted to thymines. An appropriate coverage of the data should be reached for meaningful analyses (10x is currently considered a good target coverage).

Average methylation levels can then be calculated, general genome-wide profiling and targeting can be analysed, and methylation differences between biological groups can be assessed. Especially when studying differential methylation, it is advisable to generate biological replicates to account for natural within-group-variation, since otherwise, differences will be over-interpreted and can result in a high false-positive rate of methylation differences. Methylation data can be analysed visually, for examples in Genome Browser

tracks, or can be submitted to customised computational downstream analyses. A more precise description and protocol to overcome the challenges of processing and analysing bisulfite sequencing data is described in Gatzmann and Lyko (2018).

1.5 Chromatin structure in the context of DNA methylation

Chromatin is a dynamic structure that helps to package DNA into a dense and compact shape, but also regulates DNA accessibility for replication, DNA repair and gene expression. The nucleosome is the smallest structural unit of chromatin. It contains 147 base pairs of DNA which are wrapped a little less than two times around a histone octamer. This histone octamer is composed of two copies each of histone proteins H2A, H2B, H3 and H4 (Tessarz and Kouzarides, 2014). Modifications of histones, such as methylation, acetylation, and ubiquitylation, or methylation of the DNA can fundamentally alter the structure of chromatin, making it more compressed (usually termed heterochromatin) or open (euchromatin). Generally, euchromatin contains most of the active genes in a cell, while heterochromatin is more commonly found at centromeres, telomeres and inactive genes (Bannister and Kouzarides, 2011). However, this is a simplified classification of chromatin states, since there are intermediate configurations (Bannister and Kouzarides, 2011).

DNA methylation does not influence chromatin accessibility states on its own. Instead, there are a number of complex interactions between histone modifications and DNA methylation that eventually model chromatin states. Histone lysine methylation on position 27 of H3 (H3K27me), for example, is a marker of open chromatin that is mostly found in unmethylated stretches of DNA like CpG islands (Rose and Klose, 2014). The same goes for the modification H3K4me3, which is associated with states of open chromatin and which blocks the *de novo* methyltransferase DNMT3 from binding the H3 tail and keeps it from methylating the underlying DNA (Rose and Klose, 2014). On the other hand, H3K36 trimethylation (H3K36me3) is targeted to gene bodies of actively transcribed genes and correlates with an enrichment of DNA methylation. Then again, di- and trimethylated lysine 9 on histone H3 (H3K9me2/3) is usually found at inactive genes, and it has been revealed that DNA methylation in such regions is dependent on the presence of these marks in *Neurospora crassa* (Tamaru et al., 2003).

1.5.1 Computational analysis of genome-wide chromatin structure using ATAC-seq

In 2013, Buenrostro et al. (2013) introduced an assay for transposase-accessible chromatin using sequencing (ATAC-seq) as a rapid method to profile regions of open chromatin without enriching for specific histone modifications. The prokaryotic Tn5 transposase is loaded with adaptors for high-throughput DNA sequencing and added to isolated DNA. Here, it fragments DNA in regions of accessible chromatin, while this is less likely to happen in more condensed chromatin due to steric hindrances. It then tags the DNA with the integrated adaptors. The resulting amplifiable fragments can be used for standard high-throughput sequencing.

This data can be aligned to a reference sequence by standard genome sequencing mapping tools, and an enrichment of read coverage in a genomic region is interpreted as a region where chromatin was accessible for the transposase fragmentation. ATAC-seq data can also be used to identify nucleosome positioning: by grouping reads into reads that are shorter than the canonical length that is usually protected by a nucleosome (147 bp), and by reads that are consistent with that length, their specific alignment start and end points provide information about the position of a nucleosome, as well as nucleosome-free regions and potential transcription factor binding regions.

1.6 Previous work on methylation in the marbled crayfish

In the recently published a draft genome assembly of the marbled crayfish (Gutekunst et al., 2018), my colleagues identified homologs of the DNA methyltransferases DNMT1 and DNMT3, as well as a single copy of a TET hydroxymethylase (see Figure 5A, B and C). To confirm the expression of these enzymes, qRT-PCR was performed for developmental stages and adult tissues, revealing low mRNA levels for all three enzymes in early embryonic stages, which increased during embryonic development (Figure 5D). In adult tissues, DNMT1 and DNMT3 were mostly moderately but stably expressed. TET was highly expressed in all adult tissues except for ovaries (Figure 5E).

The presence of methylation in the marbled crayfish was confirmed by previous studies in our lab using whole-genome bisulfite sequencing (Falckenhayn, 2016). Hereby, methylation patterns in the marbled crayfish could be characterised at single-base resolution. Initial analyses were performed for the hepatopancreas tissues of an animal from our lab strain (Vogt et al., 2015). Mapping the trimmed data against the marbled crayfish assembly (Gutekunst et al., 2018) (see 4.3 for details of processing the data) yielded a mapping rate of approximately 76% at 9.8x genome coverage and 78% covered CpGs. Mitochondrial DNA was used to determine bisulfite conversion efficiency at approximately 99.77% (see Table 1 in section 2.1.1), ensuring the trustworthiness of the data.

Basic patterns

These first analyses have shown that methylation in the marbled crayfish is specific to CpG dinucleotides (Figure 6A), and bimodally distributed, i.e., within a tissue, single CpGs are observed to be either completely methylated or completely unmethylated (Figure 6B). It is also symmetric on both strands (Figure 6C), (Falckenhayn, 2016).

Analysing 100 scaffolds from the draft genome assembly for their methylation landscape revealed that most scaffolds show a mosaic methylation pattern that is typical for many invertebrate methylomes (Figure 6D).

Additionally, it has been shown that gene body methylation is also bimodally distributed, with two distinct populations of genes that are either lowly or highly methylated (6E). Moreover, highly methylated genes tend to be long, evolutionary conserved, CpG-poor and moderately expressed (Falckenhayn, 2016). Since these are common features of house-

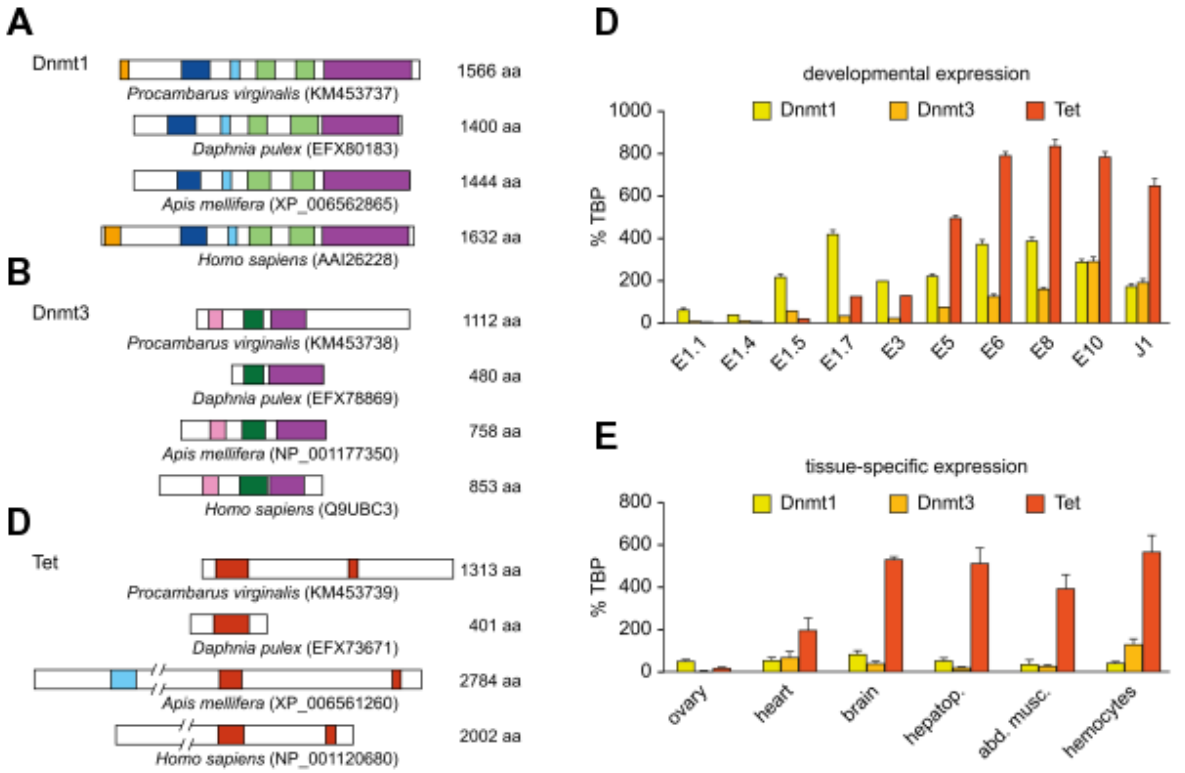

Figure 5: **Previous work: Conservation of Dnmt1, Dnmt3 and Tet in marbled crayfish** Paper: Genome annotation revealed the presence of a DNA methylation system consisting of single homologs for Dnmt1 (a), Dnmt3 (b) and Tet (c), respectively. Shown are comparisons of virtually translated protein sequences with three reference organisms: *Daphnia pulex*, *Apis mellifera* and *Homo sapiens*. Numbers in brackets represent accession numbers. Conserved domains are shown as colored boxes. mRNA expression levels are indicated relative to the TBP (TATA-box-binding protein) housekeeping gene. Bars indicate standard deviations from at least three independent measurements. E: embryonic stages; J: juvenile stages. e mRNA levels of Dnmt1, Dnmt3 and Tet in various adult marbled crayfish tissues (hepatop. hepatopancreas, abd. musc. abdominal musculature). Adopted from Gatzmann et al. (2018).

keeping genes, it was confirmed that DNA methylation in the marbled crayfish is enriched at gene bodies of housekeeping genes (Figure 6F).

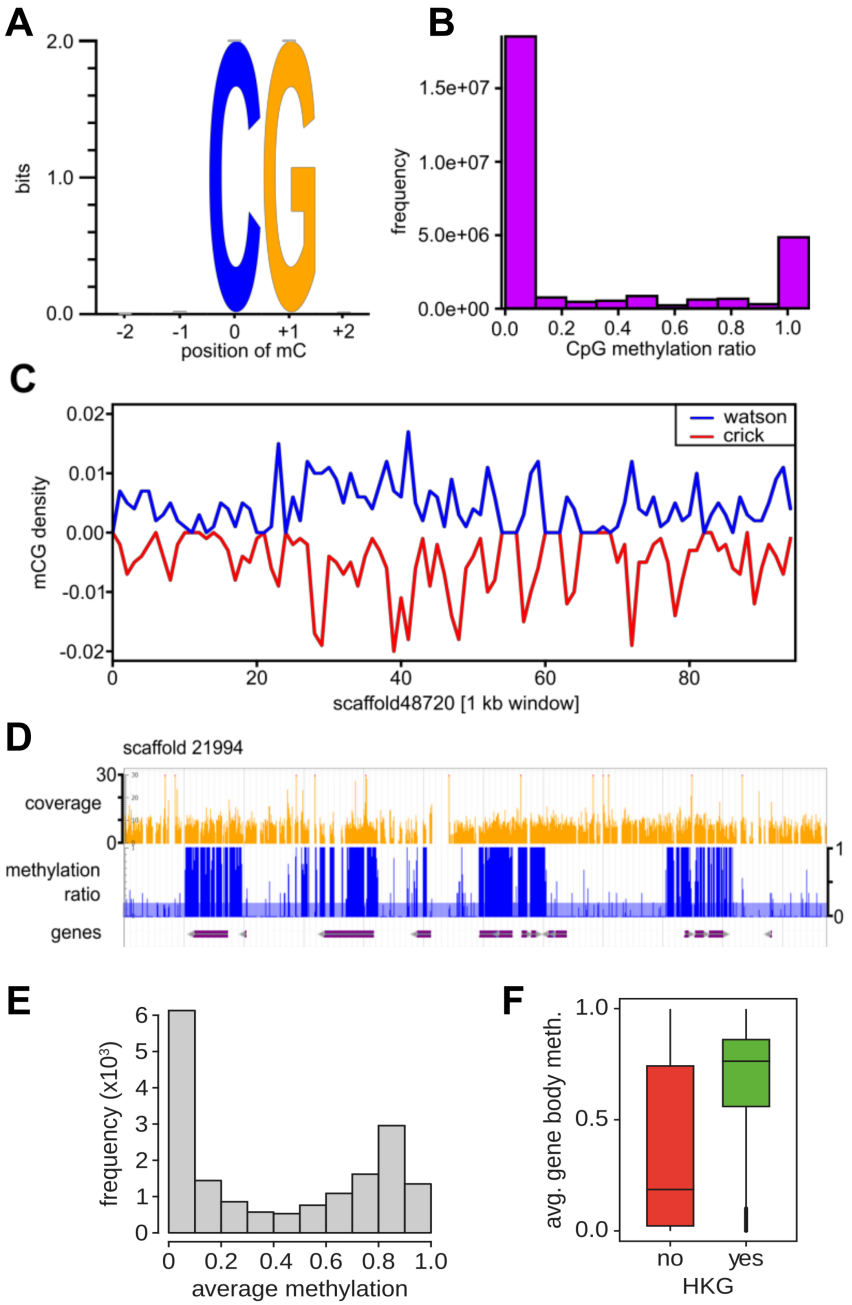

Figure 6: **Previous work: characterisation of the marbled crayfish methylome** (A) Logoplot of the two bases at methylated cytosine residues showing CpG-specific methylation. (B) Histogram showing the average methylation ratio for each CpG. (C) Strand specific density of methylated CpGs along the scaffold 48,720 (Watson strand: blue, Crick strand: red). (D) representative scaffold is shown to demonstrate the mosaic methylation landscape in the marbled crayfish. (E) Histogram showing the distribution of gene body methylation levels. (F) Boxplots showing the distribution of methylation ratios for non-housekeeping genes (red) compared to housekeeping genes (green).

1.7 Aims of this PhD thesis

The marbled crayfish is an emerging invasive freshwater crayfish with a high genetic homogeneity that shows a remarkable adaptability to different environments and a strong degree of phenotypic variability. Since genetic variation is likely not the cause for its invasive capacity, the aim of this thesis is to analyse epigenetic mechanisms in the marbled crayfish to provide insights into the potential environmental adaptation through epigenetic regulation.

I present in this thesis a detailed analysis of DNA methylation in the marbled crayfish using whole-genome bisulfite sequencing data. This information was integrated with RNA-seq and ATAC-seq data to integratively analyse the interplay between the three epigenetic layers of DNA methylation, gene expression, and chromatin accessibility. The genetic homogeneity of the marbled crayfish, with very few confounding mutations, makes it a particularly promising model organism for epigenetics research.

These analyses will

- establish the methylome of the marbled crayfish
- provide insights into the evolutionary conservation of methylation patterns in invertebrates and crustaceans
- broaden the knowledge on the interplay of gene body methylation, chromatin accessibility and gene expression.

2 Results

2.1 Data

2.1.1 Whole-genome bisulfite sequencing data

Whole-genome bisulfite sequencing was used on nine marbled crayfish DNA samples to generate high-resolution DNA methylation maps. Samples originated from five different tissues and 6 different animals. Table 1 summarises the results of sequencing and basic processing. In short, we generated methylation data for four distinct adult tissues (hepatopancreas, abdominal musculature, hemocytes, and gills) and a sample from early embryos (embryonic stage 1.7, Grimmer (2015)). For hepatopancreas and abdominal musculature, we sequenced three biological replicates from three different animals each, and single replicates for the hemocytes, embryos, and gills. Genome coverages for all samples ranged from 9x to 23x (Table 1), and the total percentage of covered CpGs in the genome ranged from 76% to 87%, averaging at 80.5%. Mapping rates were mostly around 50%, with the exception of three samples. One of these samples had been kept at around 30 degrees in ethanol for several weeks which could have lead to DNA degradation, while the other two showed some degree of bacterial contamination when blasting reads against the NCBI Reference Sequence Project database (O'Leary et al., 2016). Bisulfite conversion efficiencies, estimated from reads mapping to the mitochondrial DNA, averaged around 99.5%.

Table 1: **Whole-genome bisulfite sequencing results for marbled crayfish animals.** Abbreviations: Map.: mapping rate, Cover.: coverage, Conv.: bisulfite conversion rate, % CpGs: % CpGs covered in the genome, hep.: hepatopancreas, musc.: abdominal musculature, E1.7: embryonic stage 1.7, hemo.: hemolymph, bp: base pairs, PE: paired-end.

ID	Origin	Tissue	Map. %	Cover.	Conv. %	% CpGs	read length
Pvir2	lab stock	hep.	55	9.8x	99.7	78.5	100 bp PE
Pvir3	lake Moosweiher	hep.	53	8.8x	99.8	75.6	100 bp PE
Pvir6	lab stock	hep.	52	11.4x	99.2	76.9	100 bp PE
Pvir2	lab stock	musc.	45	23.4x	99.6	86.9	150 bp PE
Pvir3	lake Moosweiher	musc.	50	20.4x	99.5	84.9	150 bp PE
Mora	Moramanga	musc.	26	13.8x	99.3	79.6	150 bp PE
E1.7	lab stock	embryos	33	15.2x	99.7	80.5	150 bp PE
hem	lake Reilingen	hemo.	27	17.9x	99.1	81.2	150 bp PE
Pvir3	lake Moosweiher	gills	rate	19.4x	99.4	84.8	150 bp PE

We also generated whole-genome bisulfite sequencing for the parent species, *Procambarus fallax*. Sequencing was performed for three DNA samples from two female animals

Table 2: **Whole-genome bisulfite sequencing results for *P. fallax* animals.** Abbreviations: Pff: indicates female animals, Map.: mapping rate, Cover.: coverage, Conv.: bisulfite conversion rate, % CpGs: % CpGs covered in the genome, hep.: hepatopancreas, musc.: abdominal musculature, bp: base pairs, PE: paired-end.

ID	Origin	Tissue	Map. %	Cover.	conv. %	% CpGs	read length
Pff3	aquarium supply	hep.	50	10.8x	99.9	68.2	100 bp PE
Pff4	aquarium supply	hep.	51	10.2x	99.5	67.1	100 bp PE
Pff4	aquarium supply	musc.	52	10.8x	99.5	67.2	100 bp PE

and two distinct tissues, namely hepatopancreas and abdominal musculature. Mapping rates ranged between 50-52%, and genome coverages averaged around 10.6x. Conversion efficiencies ranged from 99.5-99.9%, and approximately 67.5% of all CpGs were covered. This data is shown in Table 2.

2.1.2 RNA-seq data

To address the relationship between methylation and gene expression, we also generated RNA-seq data for the marbled crayfish and its parent species. For the marbled crayfish, we sequenced three biological replicates each for adult tissues hepatopancreas, abdominal musculature, and hemocytes. Yields per sample ranged from 3,200 to 42,000 mega base pairs (Mbp), and mapping rates ranged from 86-92% (see Table 3). Yields vary so strongy because samples were sequenced at different time points in this project, on different machines, and some samples were pooled on a single lane while others were sequenced alone on a single lane.

Table 3: **RNA sequencing results for marbled crayfish animals.** Abbreviations: hep.: hepatopancreas, musc.: abdominal musculature, hemo.: hemolymph, Mbp: mega basepairs, Map.: mapping rate, bp: base pairs, PE: paired-end.

ID	Origin	Tissue	Yield[Mbp]	Map. %	read type
Pvir2	lab stock	hep.	42,032	88	100 bp PE
Pvir6	lab stock	hep.	10,656	90	100 bp PE
Pvir7	lab stock	hep.	11,047	90	125 bp PE
Pvir2	lab stock	musc.	9,144	89	150 bp PE
Pvir6	lab stock	musc.	8,752	92	150 bp PE
Pvir7	lab stock	musc.	4,719	90	150 bp PE
hem37	lake Reilingen	hemo.	3,488	86	50 bp SE
hem39	lake Reilingen	hemo.	3,231	89	50 bp SE
hem40	lake Reilingen	hemo.	3,920	87	50 bp SE

We also sequenced three biological replicates for abdominal musculature from the

Table 4: **RNA sequencing results for *P. fallax* animals.** Abbreviations: Pff: indicates female animals, musc.: abdominal musculature, Mbp: mega basepairs, Map.: mapping rate, bp: base pairs, PE: paired-end.

ID	Origin	Tissue	Yield[Mbp]	Map. %	read type
Pff1	aquarium supply	musc.	9,495	90	125 bp PE
Pff3	aquarium supply	musc.	9,161	90	125 bp PE
Pff4	aquarium supply	musc.	8,716	90	125 bp PE

parent species *Procambarus fallax*. Yields averaged around 9,000 Mbp per sample, and mapping rates were at 90% per sample (data shown in Table 4).

2.1.3 ATAC-seq data

For further insight into the regulatory functions of DNA methylation, I also investigated the interplay between methylation, chromatin accessibility, and gene expression. For this purpose, we used the Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013) to analyse genome-wide chromatin accessibility patterns. ATAC-seq was performed on marbled crayfish hemocytes, which could be analysed together with WGBS and RNA-seq data from hemocytes, where all hemocyte samples came from the same pool of cells collected by different animals. We sequenced three biological replicates, with yields ranging from 6,400-12,700 Mbp (see Table 5).

Table 5: **ATAC sequencing results.** Abbreviations: hemo.: hemocytes, Mbp.: mega base-pairs, Map.: mapping rate, bp: base pairs, PE: paired-end.

ID	Origin	Tissue	Yield[Mbp]	Map. %	read type
hem1	lake Reilingen	hemo.	12,717	78	125 bp PE
hem2	lake Reilingen	hemo.	10,849	74	125 bp PE
hem3	lake Reilingen	hemo.	6,423	62	125 bp PE

2.2 The marbled crayfish methylome

2.2.1 Continued basic characterisation of the marbled crayfish methylome

In addition to the basic characterisations of the marbled crayfish methylome done in our lab before (see section 1.6), I conducted a comparative analysis of average methylation levels in 2kb windows for the marbled crayfish and other animals with methylated genomes. It revealed that methylation levels in the marbled crayfish were distinctly higher than those of other crustaceans, namely *Daphnia pulex* (Asselman et al., 2016) and the sand flea *Parhyale hawaiensis* (Kao et al., 2016))(Figure 7A). Genome-wide, 18% of CpGs in the marbled crayfish were methylated. Figure S1 shows a violin plot of a vertebrate species (mouse), which is, as expected, highly methylated in more than 75% of all 2kb windows.

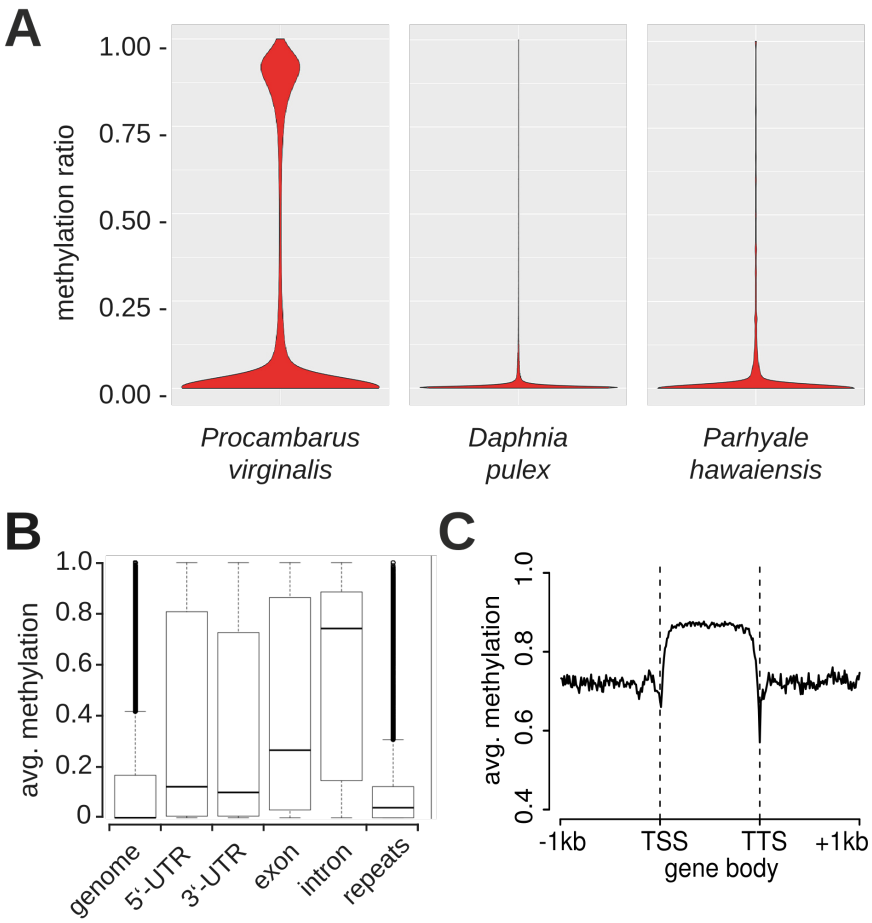

Figure 7: **Characterisation of the marbled crayfish methylome** (A) Comparative analysis of the three currently known crustacean methylomes. Violin plots show DNA methylation levels for 2kb-windows. (B) Methylation levels of the genome and of predicted gene features. (C) Metagene plot showing methylation levels for gene bodies and surrounding 1 kb upstream and downstream.

Methylation targets

Methylation in the marbled crayfish genome appears mostly targeted to gene bodies, with introns showing higher average methylation levels than exons (Figure 7B). The overall genome, repeats and 5' and 3' UTRs (defined as the 1 kb upstream of the transcription start site and downstream of the transcription termination site, respectively) of genes appear mostly hypomethylated. Methylation levels in the 5' UTR appear somewhat elevated in comparison to the 3' UTR. This can also be observed in the metagene plot in 7C, where in the 3' UTR, just before the transcription start site, I saw a slight elevation of methylation levels. At the same time, a distinct drop in methylation levels can be detected right after the transcription termination site. Clearly, methylation levels in the genome are lower than in the gene body. This is in accordance with studies from Zemach et al. (2010); Feng et al. (2010). As noted before, previous analyses have shown that genes that are targeted by methylation are primarily housekeeping genes (see section 1.6, Figure 6F).

These results show that the marbled crayfish methylome shares the basic features of Dnmt1-Dnmt3-dependent animal methylomes (Zemach et al., 2010).

2.2.2 The marbled crayfish methylome only shows subtle tissue-specificity

DNA methylation of gene bodies

To further characterise methylation patterns in the marbled crayfish, I used a primary set of samples comprising DNA samples from abdominal musculature, hepatopancreas, hemocytes and embryos. For these samples, I generated methylation heatmaps using hierarchical clustering on genes. All genes with sufficient coverage were included, and the average methylation of each gene was calculated and used in the analysis. A heatmap for all eight samples in the primary sample set and all genes is shown in Figure 8A. Strikingly, the heatmap gave the impression that methylation is not tissue-specific or specific to developmental stages. This impression was also confirmed by a Wilcoxon-rank-sum test applied to the two tissues for which replicates were available (hepatopancreas and abdominal musculature). Using this test, no gene was identified to be significantly differentially methylated between the two tissues at a p-value cutoff of 0.1. One hepatopancreas shows distinctly lower methylation levels than the other two, and hemocytes appear to have somewhat lower methylation levels, too. However, for hemocytes, it is difficult to draw a clear conclusion since only a single replicate was available for analysis.

To further test whether tissue-specific methylation might occur in gene bodies, I generated another heatmap and applied hierarchical clustering not only to the genes, but also to the samples. This is shown in Figure 8B. This approach was able to group the three abdominal musculature samples into the same clade on the tree. The lowly-methylated hepatopancreas sample clustered together with the hemocytes sample, which also had overall lower gene body methylation levels.

Promoter methylation (defined as the 5' UTR) showed similar methylation patterns, with no distinctly visible tissue-specificity (Figure S2A). For hierarchical clustering on samples for promoter methylation, again, the three abdominal musculature samples clustered in one clade, and two of the hepatopancreas samples did, too (Figure S2B). Again, the Wilcoxon-rank-sum-test did not classify any promoter as significantly differentially methylated.

Separating samples in methylation space

In addition to heatmaps and statistical testing, I used principal components analysis (PCA) and metric multidimensional scaling (MDS) to see if tissues could be separated from each other based on their average methylation ratio. Pairwise examination of the first three components of a PCA showed that abdominal musculature samples tend to group together closely in two out of three cases (circles, Figure 9A and C). When plotting components 2 and 3 against each other, hepatopancreas samples could be separated from all other tissues (Figure 9C), even though they themselves were quite widely spread out across the analysed space. Hemocytes and the embryonic stage have a tendency to be placed

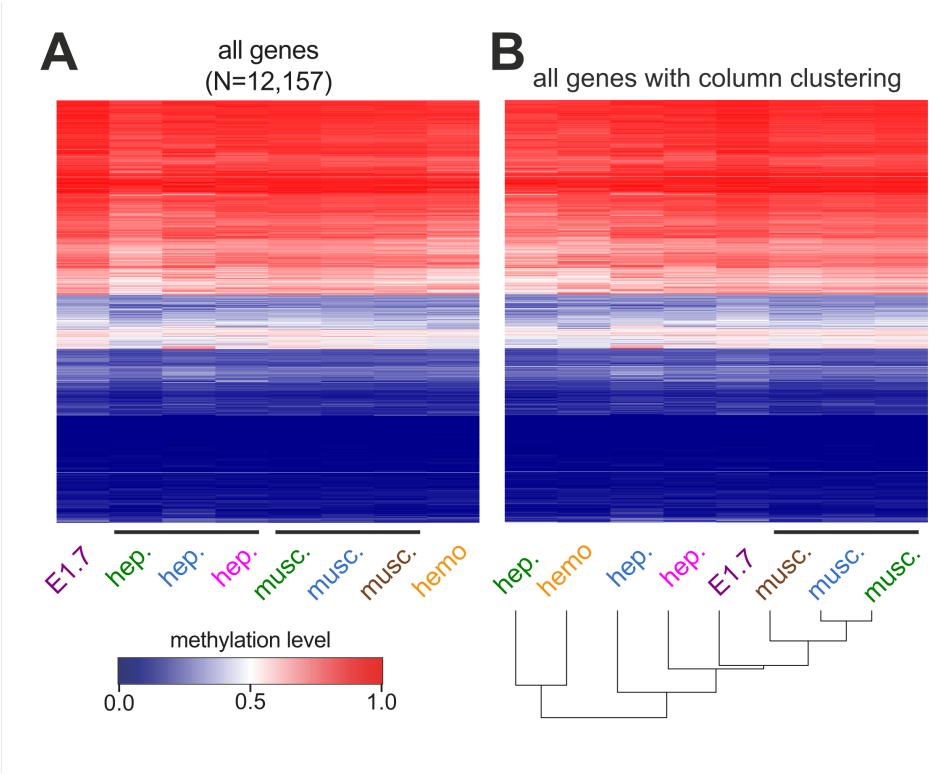

Figure 8: **Comparison of gene body methylation comprising a set of different developmental stages, tissues and animals.** (A) Comparative analysis of gene body methylation patterns shown in a heatmap where hierarchical clustering for rows (genes) was used. The heatmap shows average gene body methylation levels for each gene for the set of 8 independent samples (columns). Colors indicate individual animals. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). Only genes containing at least 10 CpGs with a strand-specific coverage of 3x in all 8 samples are shown. E1.7: stage 1.7 embryos, hep.: hepatopancreas, musc.: abdominal musculature, hemo: hemocytes. (B) Similar heatmap, but including clustering for individual samples and tissues.

slightly apart from other samples, except for components 2 and 3 (Figure 9C), where they fall in between the abdominal musculature and hepatopancreas samples.

For MDS, again, abdominal musculature samples showed a tendency to group together (Figure 9D, E and F). A combination of coordinates 2 and 3 was again able to separate hepatopancreas samples from the other tissues (Figure 9F). This was observed even though again, hepatopancreas samples appear to vary in their methylation patterns since they were quite widely distributed across the analysed space. This could partially be attributed to the hypomethylated state of one of the hepatopancreas samples. However, the other two also did not group closely together in any of the analysed plots. Again, hemocytes and the embryonic stage are found more towards the outside of the grouped samples in two out of three cases, but in coordinates 2 and 3 (Figure 9F), they group closely together with abdominal musculature samples.

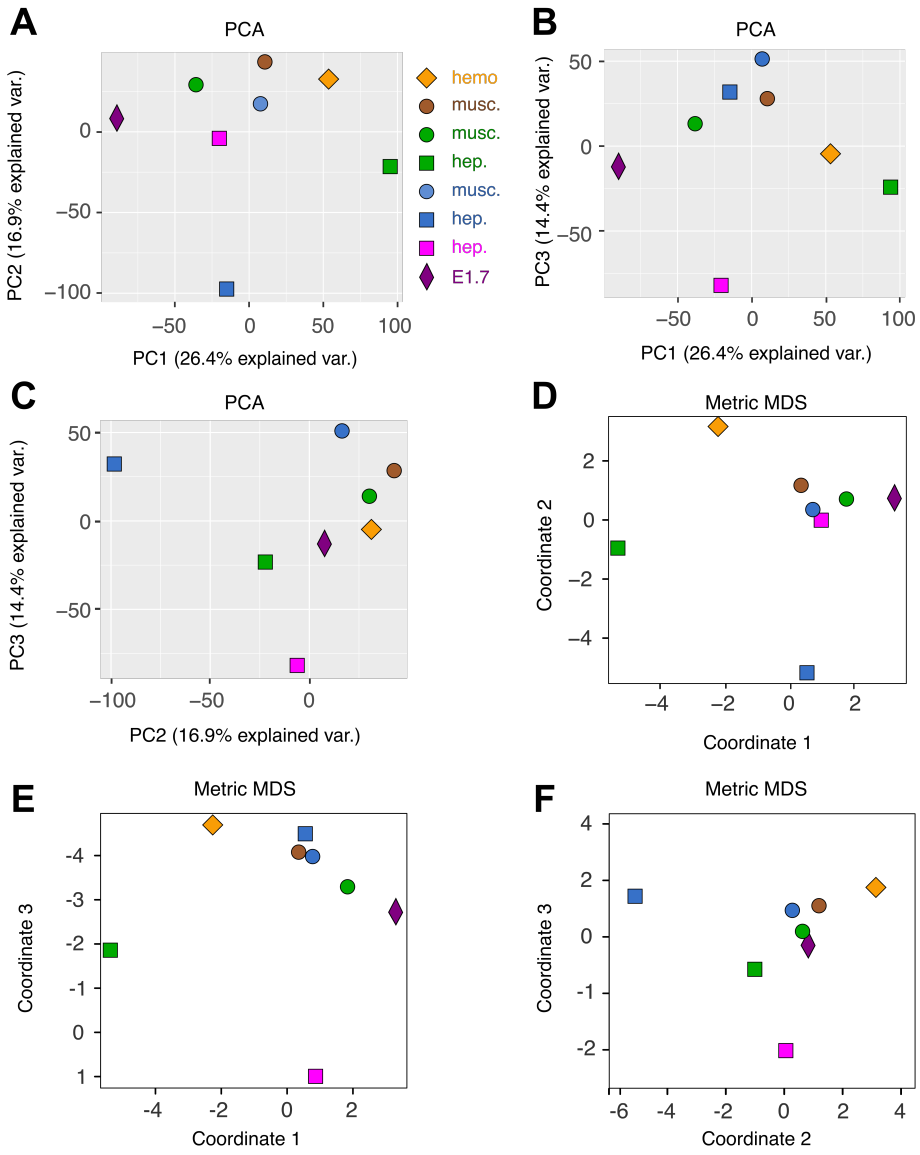

Figure 9: **Principal components analysis and metric multidimensional scaling for gene body methylation.** (A)-(C): Principal components analysis for gene body methylation. (D)-(F) Metric multidimensional scaling for gene body metylation. Symbols represent sampes as indicated (top row), colors denote animals as shown before.

DNA methylation of housekeeping genes

Next, I examined gene body methylation patterns by grouping genes into housekeeping genes and tissue-specific genes. This is shown in Figure 10A. This analysis confirmed the high methylation levels of housekeeping genes shown before, while consequently, non-housekeeping genes showed much lower levels of methylation (Figure 10B). Again, by simple visual examination, no cluster of genes appears to show elevated levels of tissue-specific methylation. Since the Wilcoxon-rank-sum-test had been applied to each gene individually, it was not applied again here. Table 6 shows a gene set enrichment analysis

for methylated genes, where enriched functions include protein folding and functions related to cell division.

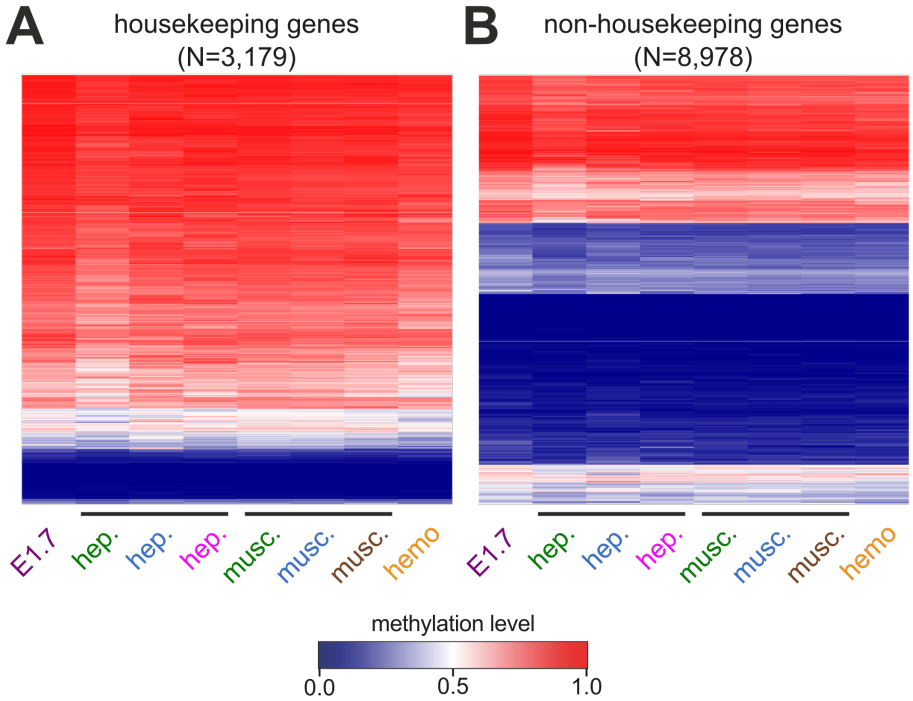

Figure 10: **Comparison of gene body methylation in different sets of genes.** (A) Comparative analysis of gene body methylation patterns in housekeeping genes shown in a heatmap where hierarchical clustering for rows (genes) was used. The heatmap shows average gene body methylation levels for each gene for the set of 8 independent samples (columns). Colors indicate individual animals. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). Only genes containing at least 10 CpGs with a strand-specific coverage of 3x in all 8 samples are shown. E1.7: stage 1.7 embryos, hep.: hepatopancreas, musc.: abdominal musculature, hemo: hemocytes. (B) Similar heatmap, but only for non-housekeeping genes.

I also examined promoter methylation in housekeeping and non-housekeeping genes. Promoter methylation patterns did not show such a pronounced difference between the two gene groups, even though again, a slight decrease in methylation for tissue-specific promoters could be observed (Figure S3).

In conclusion, it seems methylation in and around genes is, for most parts, invariant to tissue or developmental stages in the marbled crayfish. However, there appear to be some genes that show a limited amount of tissue-specificity, since both hierarchical clustering, as well as PCA and MDS, showed some capacity to group tissues together under certain

Table 6: **Gene set enrichment analysis for methylated genes.** Gene set enrichment for methylated genes showing biological process, enrichment and p-value for the eight most enriched processes.

biological process	enrichment	p-value
assembly of the pre-replicative complex	8.7	0.00074
protein folding	7.9	0.035
establishment of sister chromatid cohesion	7.5	0.0038
mitotic prometaphase	5.8	0.0042
interconversion of 2-oxoglutarate and 2-hydroxyglutarate	5.1	0.0072
ethanol oxidation	5.0	0.008
synthesis of ketone bodies	4.3	0.012

conditions. So it might be smaller sets of genes, or regions in genes, that exhibit some degree of tissue-specific methylation patterns.

2.2.3 Repeats are sparsely methylation in the marbled crayfish

Figure 11A shows repeat methylation for the 5 most frequent repeat classes and their frequency in the marbled crayfish genome. The majority of repeats was unmethylated, while every class had at least a small percentage of repeats that showed elevated methylation levels. DNA transposons had an especially high number of highly methylated repeats, with approximately 50% of repeats being methylated.

Repeat methylation generally did not seem to be associated with the age of the repeats, with the correlation coefficient between repeat divergence and its methylation only being 0.012 (p-value 2.26e-3, Figure 11B). However, repeat methylation was strongly associated with the location of the repeat in the genome: repeats within genes had substantially higher levels of methylation than repeats outside of genes (Figure 11C).

Of all repeat types within DNA transposons, TcMar-Tiggers in particular showed substantial elevation of methylation levels, as is shown in Figure 11D. For this class, I tested whether the high methylation level might be due to the age of this repeat class in the marbled crayfish, or whether it was found more often within gene bodies. First, I tested whether TcMar-Tiggers with lower divergence in the marbled crayfish had different methylation levels than older TcMar-Tiggers. However, Figure S4A shows that the correlation of TcMar-Tigger methylation levels and their age did not appear to be particularly strong. Next, I tested whether DNA transposons or TcMar-Tiggers were particularly diverged or conserved in evolutionary age in the marbled crayfish, and whether this might be associated with their high methyation levels. Firstly, DNA transposons generally did not appear to be significantly more or less diverged than other repeat classes (Figure S4B). Neither did TcMar-Tiggers appear significantly old or young within DNA transposons (Figure S4C). They were also not primarily located within genes, as is shown in Figure S4D.

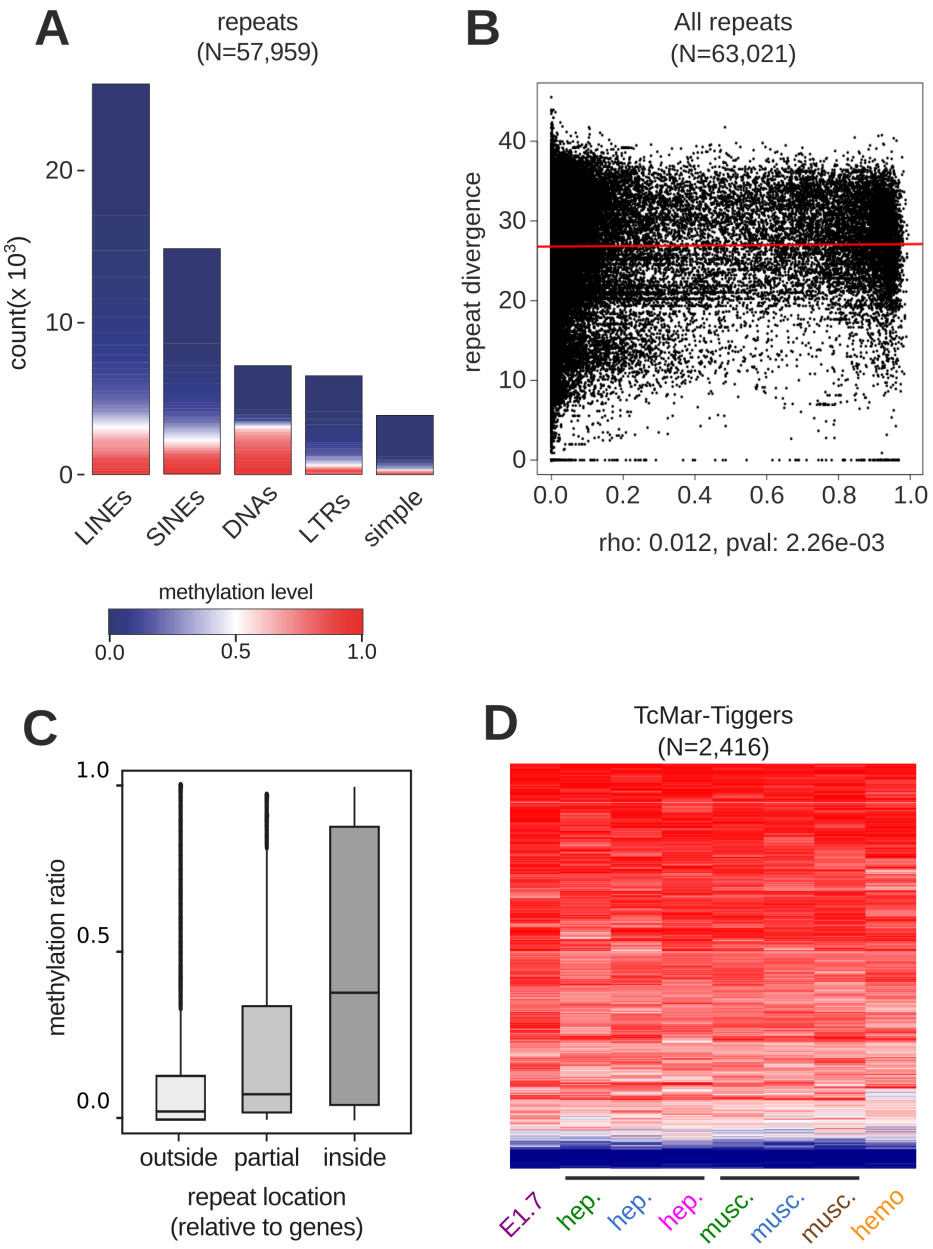

Figure 11: **Repeat methylation in the marbled crayfish.** (A) Methylation patterns of the 5 most frequent repeat classes annotated in the marbled crayfish. LINEs (long interspersed nuclear elements): N=25,622, SINEs (short interspersed nuclear elements): N=14,821, DNAs (DNA transposons): N=7,144, LTRs (long terminal repeats): N=6,483, simple (simple repeats): N=3,889. (B) Repeat methylation plotted against repeat age (divergence as taken from the repeatmasker pipeline). High divergence values imply older repeats. (C) Location-dependent methylation of repeats. (D) TcMar-Tigger repeats as the repeat type with the highest levels of methylation.

Figure 12 shows methylation heatmaps for the five most frequent repeat classes. I could

see more clearly that only small sets of repeats that are methylated, with only approximately 10% of repeats methylated. The exceptions are DNA transposons, where TcMar-Tiggers are mostly responsible for the high number of methylated repeats. Repeats encoding ribosomal RNA (rRNAs) are included as an example for essential, non-transposon repeats, and show similar proportions of methylated elements compared to the major repeat classes. Consistent with the observation for gene bodies and promoters, repeat methylation again appeared largely tissue-invariant in these heatmaps. One exception is a small cluster of genes in SINEs, where the three hepatopancreas samples appear somewhat hypomethylated.

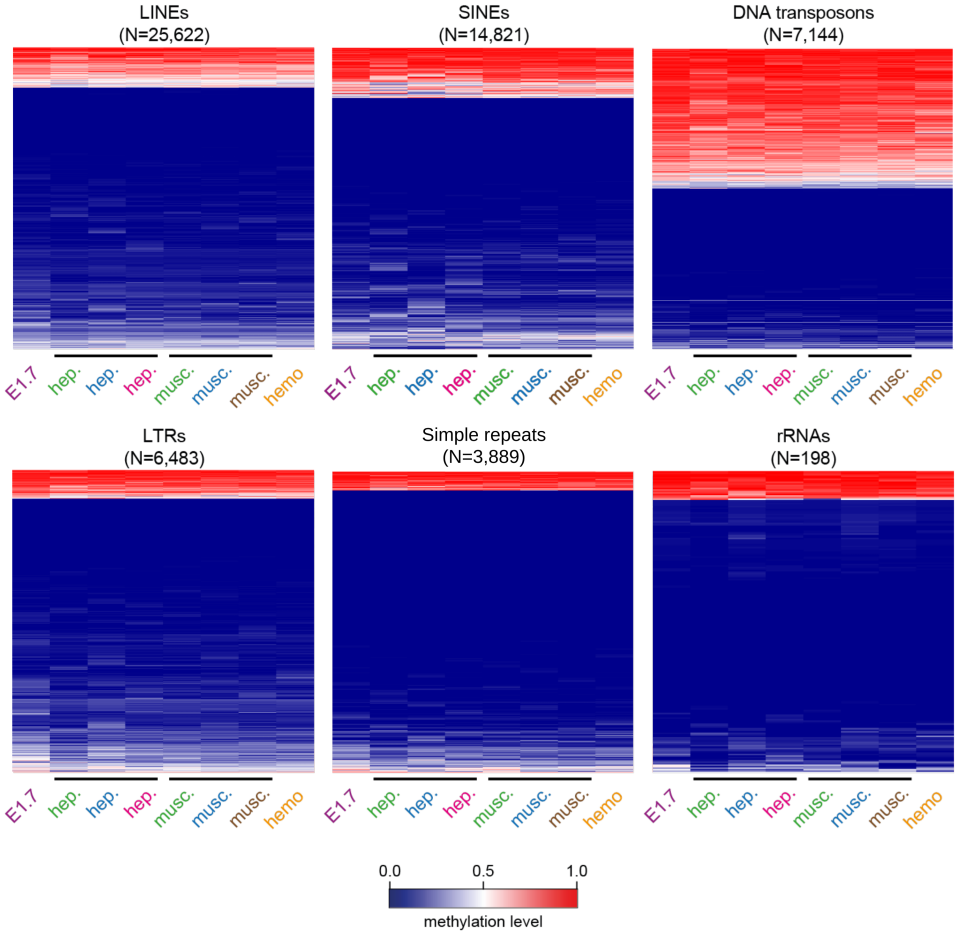

Figure 12: **Methylation heatmaps for major repeat classes.** The heatmaps show average methylation levels of selected major repeat classes in eight independent samples (columns). Only repeats with sufficient coverage in all 8 samples are shown. The five most frequent repeat classes are shown (LINEs, SINEs, DNA transposons, LTRs, Simple repeats), as well as rRNAs as an example for a non-transposon repeat class. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). E1.7: stage 1.7 embryos, hep.: hepatopancreas, musc.: abdominal musculature, hem.: hemocytes. Colors denote individual animals.

2.2.4 Variable gene body methylation in the marbled crayfish

In a parallel analysis with a slightly different set of samples, I investigated whether for certain genes, methylation might not be significantly elevated or lowered in different tissues, but simply variable between samples and tissues. For this analysis, the hemocyte methylation maps were replaced by those of gills. When examining the methylation variance (Figure 13A), it is clear that the majority of genes are stably methylated, while a smaller set shows high variance. When choosing a variance cutoff of 0.006, 846 genes are identified as highly variable. Since some of these were consistently lowly methylated or highly methylated ($0.2 <$ average methylation ratio < 0.8), I excluded them from further analyses. This defined 697 genes as highly variable in their methylation levels across samples. A heatmap for these genes is shown in Figure 13B.

To investigate whether these samples were defined by a common pathway, I used PANTHER Gene list analysis (Mi et al., 2012) to enrich for Gene Ontology pathway analysis terms. This approach identified several biological processes as enriched in this gene set, including processes related to cellular biosynthesis and metabolism (Figure 13C). Metric multidimensional scaling of variably methylated genes was able to clearly separate hepatopancreas from all other samples, as opposed to the slimmer margin observed earlier for all genes (Figure 13D). Again, the three abdominal musculature grouped closely together, this time with the samples from gills. This could suggest a moderate amount of tissue-specific methylation in some of these genes. A capture array has been generated for the set of 697 variably methylated genes to allow large-scale comparison of more than one hundred animals from different tissues and environments. This investigation is ongoing and could show whether methylation changes can be contributed to tissues or the environmental origin of a sample.

2.3 DNA methylation and expression in the marbled crayfish

2.3.1 Moderate correlations of gene body methylation and expression levels

RNA-seq datasets were generated in addition to WGBS to investigate whether DNA methylation in the marbled crayfish plays a role in gene regulation. For these analyses, I integrated three independent biological replicates from each hepatopancreas and abdominal musculature samples with the methylation data from these tissues. Details on this data are described in section 2.1.2 and in tables 3 and 4.

Analysing gene expression levels and the according methylation levels revealed a weak correlation between both promoter methylation and expression and gene body methylation and expression (rho=0.183 and rho=0.189, respectively) (Figure 14A). While the correlations are weak, they appeared statistically significant (p-value=1.38 e-89 and p-value=4.95 e-84). For housekeeping genes, which generally show higher methylation levels, these correlations were weaker (rho=-0.017 and rho=0.15, respectively) and the statistical significance dropped, too (p-value=3.4e-01 and p-value=3.93e-01) (Figure 14B).

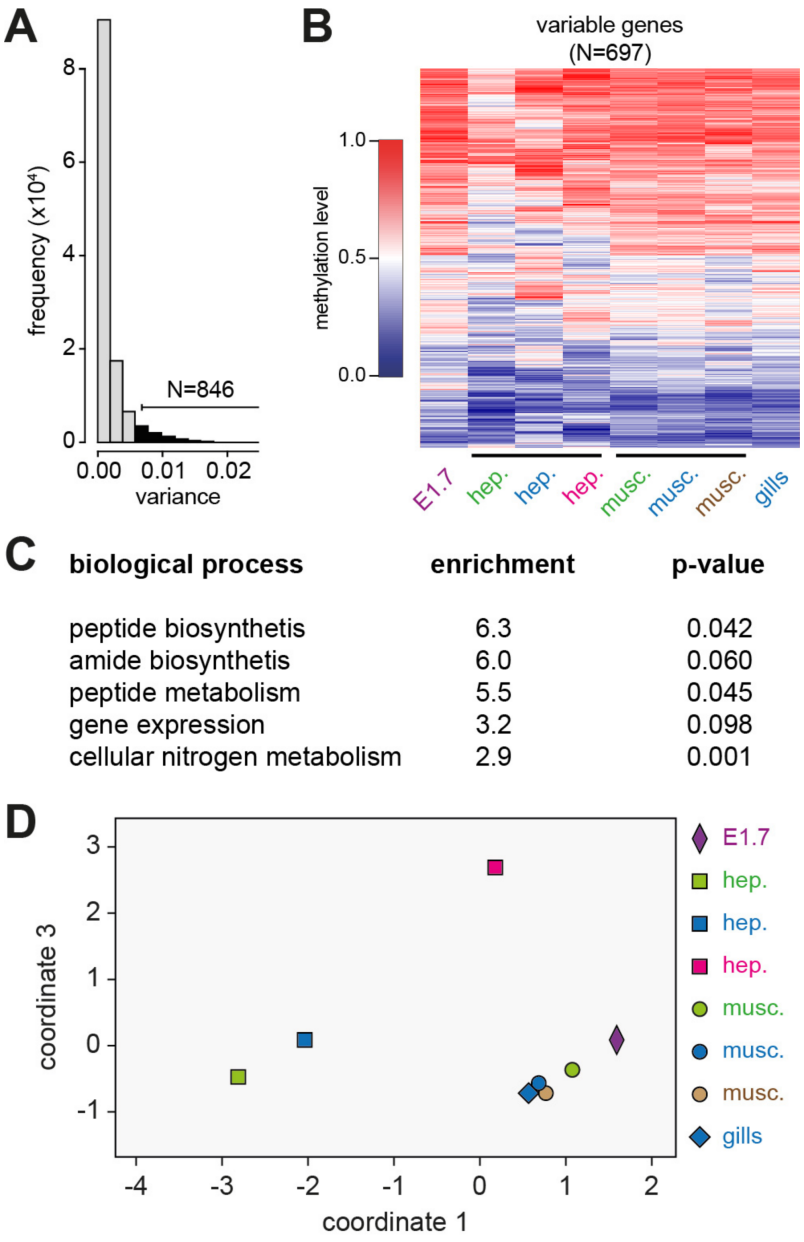

Figure 13: **Identification and characterization of variably methylated genes.** (A) Histogram of methylation variance for 12,244 genes with sufficient coverage in all 8 samples. 846 of these genes had a methylation variance $>$0.006. (B) Comparative analysis of variably methylated genes. The heatmap shows average gene body methylation levels in 8 independent samples (columns) for the 697 variably methylated genes with a mean ratio $>$0.2 and $<$0.8. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). (C) Gene ontology analysis. The five most strongly enriched biological processes with a p-vale <0.1 are shown. (D) Metric multidimensional scaling analysis based on the methylation levels of the 697 variably methylated genes.

When binning genes into 8 distinct expression ranks, the results showed a parabolic correlation between gene body methylation and gene expression levels, that is, genes with

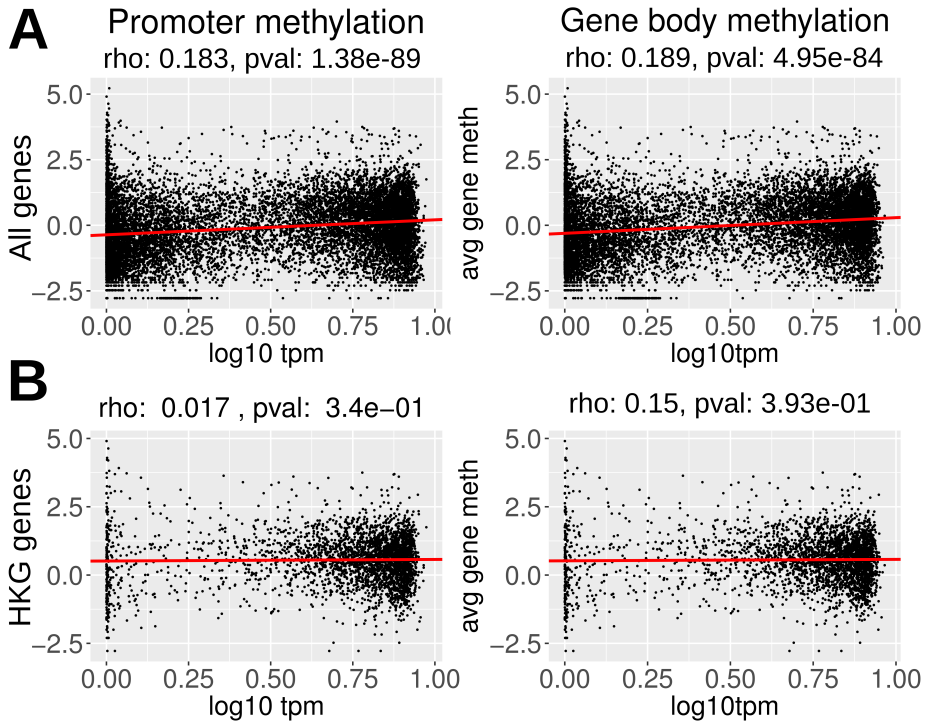

Figure 14: **Methylation and levels of expression.** (A) Correlation of promoter methylation and expression, and gene body methylation and expression for all genes. Correlation lines are shown in red, with the correlation coefficient rho and the according p-value shown above. (B) Parallel analysis for housekeeping genes.

moderate expression levels showed the highest degree of methylation. This is shown in Figure 15A. This parabolic association is observed more strongly in housekeeping genes (Figure 15B).

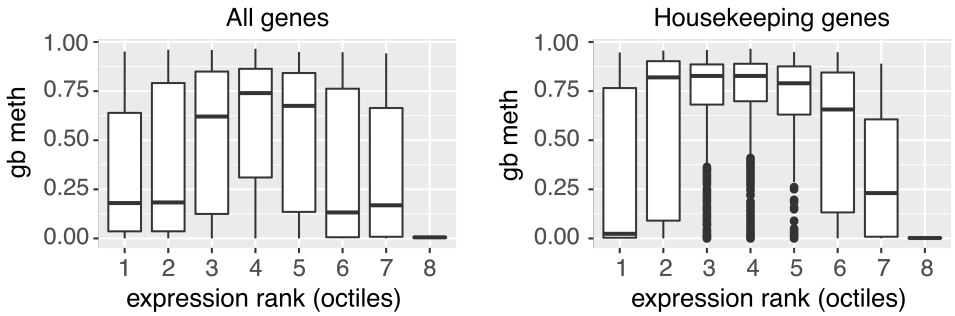

Figure 15: **Highly methylated genes are moderately expressed.** Boxplots showing the relationship between expression rank of genes (x-axis) and their associated methylation levels. Left panel, all genes. Right panel, housekeeping genes. Abbreviations: gb meth: gene body methylation.

2.3.2 Changes in methylation are not correlated with changes in expression

After analysing whether methylation is tissue-specific or modulates expression levels, I investigated more specifically whether expression level changes between tissues could be attributed to changes in methylation patterns. To this end, I first performed differential gene expression analysis, identifying 3,131 differentially expressed genes between hepatopancreas and abdominal musculature (Figure 16).

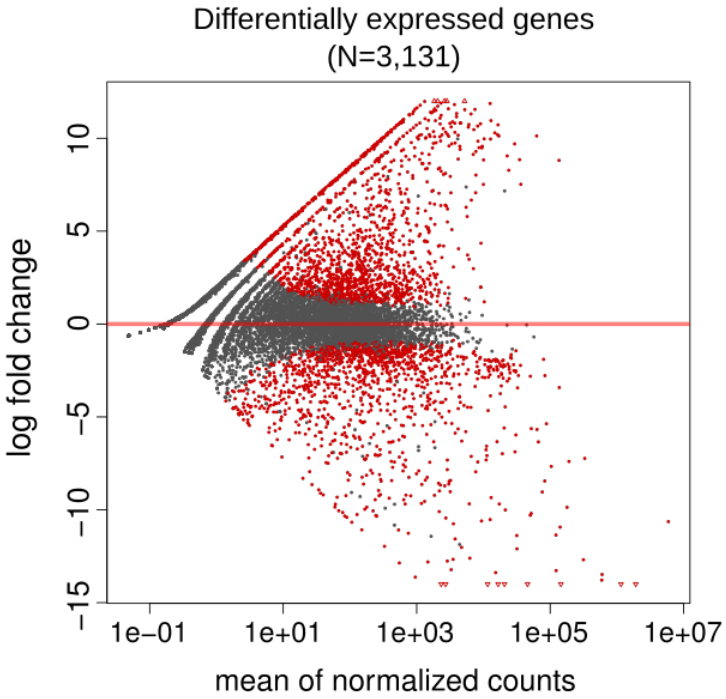

Figure 16: **Differential expression in the marbled crayfish.** MA plot showing the log-fold change and the log average expression level. Statistically significantly differentially expressed genes are shown in red (N=3,131, p-value cutoff 0.1).

Next, I calculated differential methylation between tissues for promoters (that is, 5'UTRs) and gene bodies. This was done by using the average methylation for each gene for one tissue, and subtracting it by the average of the other tissue. Figure 17A shows this differential methylation for gene bodies and promoters of all genes, plotted against the log2-fold change of expression. Methylation changes and expression changes do not appear correlated for either gene bodies or promoters. This impression persists when examining different groups of genes, like housekeeping or non-housekeeping genes (Figure 17B), or gene age (Figure 17C).

To try a different visualisation approach, I plotted the average expression levels of the two tissues against each other, and examined which genes exhibited a particular high difference in methylation between the tissues. Figure 18A shows this relationship for gene body methylation differences, where I could confirm the previous impression that expres-

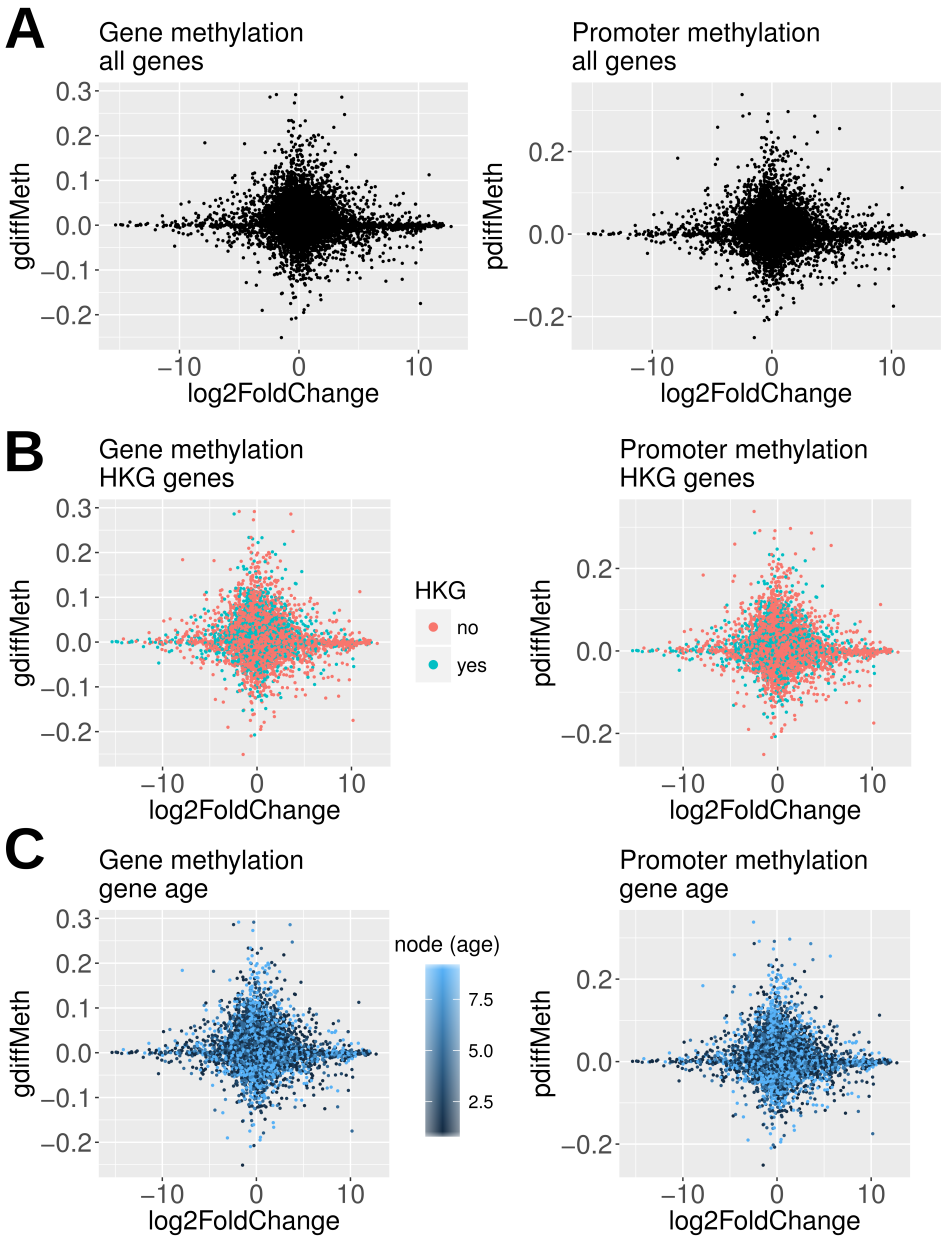

Figure 17: **Differential methylation and differential expression patterns.** Scatter plots showing the log-fold change between tissues abdominal musculature and hepatopancreas and the methylation differences between these tissues. (A) All genes (B) All genes, colors denoting housekeeping genes (red) and non-housekeeping genes (turquoise) (C) All genes, colors denoting gene age (light blue=youngest, dark blue=oldest).

sion differences do not seem to be associated with methylation differences. For better visibility, the right panel of Figure 18A shows only genes with an average methylation ratio difference between tissues of at least $\pm$ 0.18. Large differences in methylation did not necessarily deviate from the similarly expressed genes on the diagonal. Figure 18B shows this relationship for promoter methylation differences. In this case, it is interesting to note that a slightly larger number of differently methylated gene promoters appears to be un-

expressed in abdominal musculature. The majority of these promoters are more strongly methylated in abdominal musculature. This makes sense, since methylation in promoter regions has been associated with silencing of these genes.

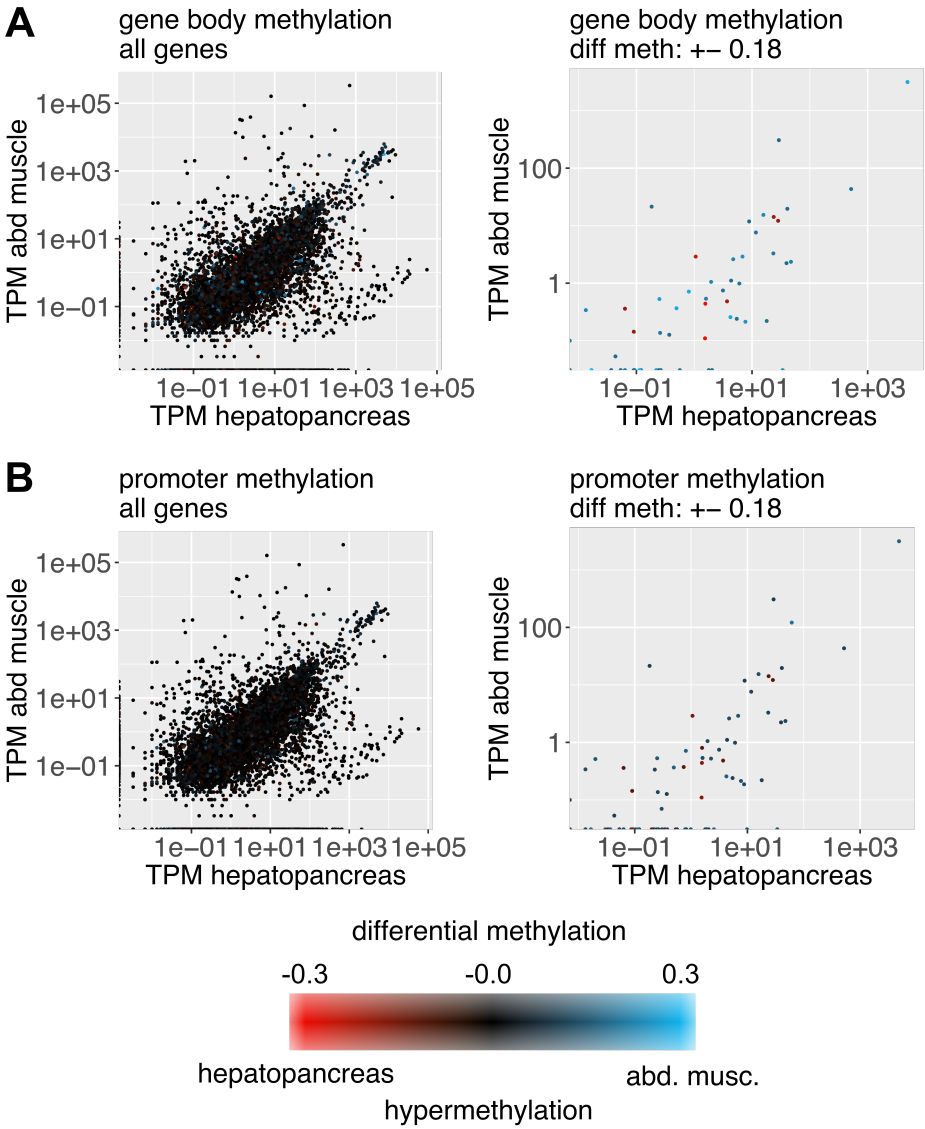

Figure 18: **Correlation of expression between tissues under consideration of differential methylation.** (A) Correlation of expression of hepatopancreas and abdominal musculature, where color indicates gene body methylation. For all genes (left panel), and only genes with an absolute differential methylation of at least 0.18. (B) Parallel analysis for promoter methylation.

2.3.3 Methylation stabilises gene expression

It has been shown in other organisms that elevated gene body methylation levels are associated with stable expression levels, while low-methylated genes usually show higher variation in their expression levels. To find out whether this relationship holds in the marbled

crayfish, I calculated expression variation as the coefficient of variation per gene from the three replicates per tissue. Consistent with previous reports, I observe a statistically significant inverse correlation between gene body methylation and gene expression variability with a correlation coefficient of -0.314 and -0.28 for hepatopancreas and abdominal musculature, respectively (Figure 19A). The p-values suggested a high significance of this correlation (p-values 2.9e-235 and 6.4e-144, respectively). Figure 19B shows the same relationship in boxplots, where the negative association between the two is more visible.

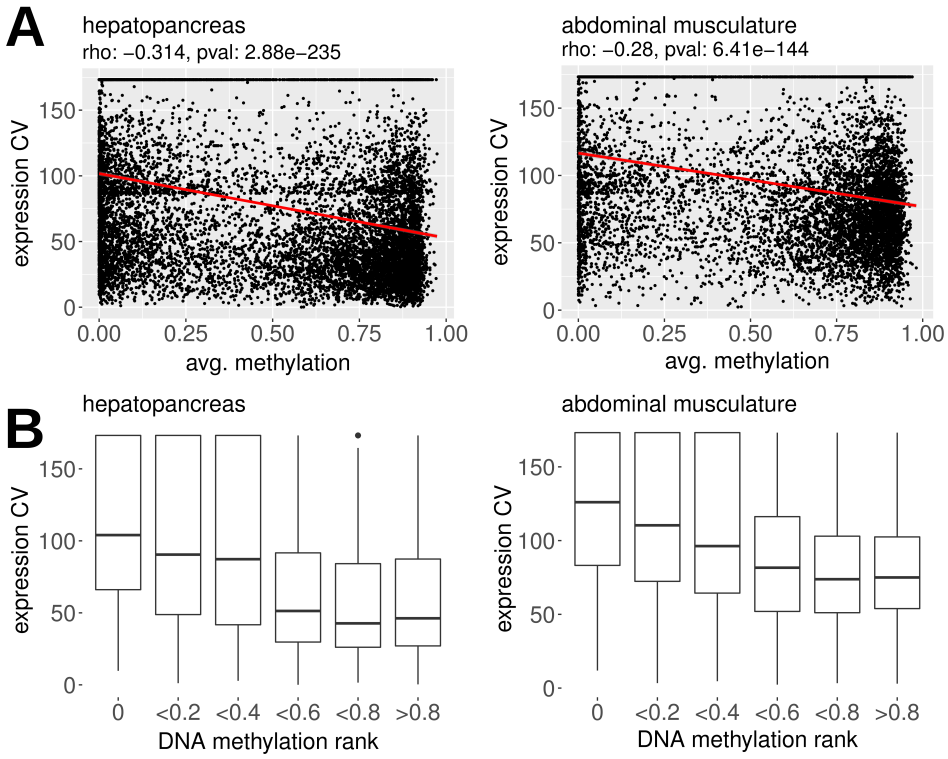

Figure 19: **Correlation between DNA methylation and gene expression variation in hepatopancreas and abdominal musculature.** (A) Gene body methylation shows an inverse correlation with gene expression variability in hepatopancreas and abdominal muscle. The correlation coefficients and p-values for the significance of the correlation are indicated. (B) Boxplot showing the same relationship with regression lines. Methylation rank zero represents completely unmethylated genes.

To give a representative example, a housekeeping gene in the marbled crayfish showed high methylation levels across the gene body in both tissues analysed, and had similar expression levels in all samples. It encodes a homolog of the eukaryotic translation initiation factor 3 subunit E (EIF3E) (Figure 20, top panel). In contrast, a tissue-specific gene that encodes a homolog of vitelline membrane outer layer protein 1 (VMO1) was mostly unmethylated across the gene body and showed substantial levels of gene expression variability in all samples (Figure 20, bottom panel). Additional examples are provided in Figure S5.

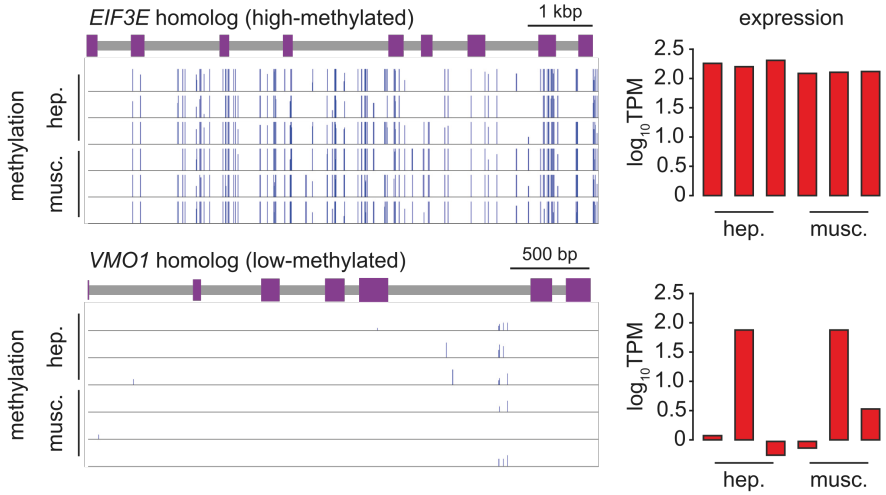

Figure 20: **Example genes for the inverse relationship of gene body methylation and expression variability.** (A) Representative example for a highly methylated gene with low gene expression variability. Genome browser tracks (left panel) show high methylation throughout the gene body. Barplot shows a stable expression across all replicates and both tissues (right panel). (B) Representative example for a lowly methylated gene with high gene expression variability. Genome browser tracks (left panel) show low methylation levels in the gene body, while the barplot (right panel) shows high variation in gene expression levels.

2.4 Increased gene body methylation and reduced gene expression variability in *Procambarus fallax*

Interestingly, the parent species of the marbled crayfish, *Procambarus fallax*, shows no evidence for invasiveness like the marbled crayfish does, but can be found only in defined habitats in Florida and southern Georgia (Hendrix and Loftus, 2000; Hobbs, 1981). To investigate whether methylation might facilitate some degree of environmental adaptability in the marbled crayfish, we also generated whole-genome bisulfite sequencing data for *P. fallax*. These comprised three samples, namely 2x hepatopancreas and 1x abdominal musculature (see Table 2). For a comparative analysis of gene body methylation, two marbled crayfish hepatopancreas and one abdominal musculature samples were chosen.

While statistical testing was not possible with such a limited number of samples, Figure 21A shows that the methylation pattern of *P. fallax* appears in wide parts similar to that of the marbled crayfish. However, when calculating methylation differences for each gene between the species, I identified 2,357 genes with a difference of at least 0.1 for the mean methylation ratio per gene (Figure 21B). The majority of these genes appears hypomethylated in the marbled crayfish, with a few exceptions of hypermethylation in marbled crayfish.

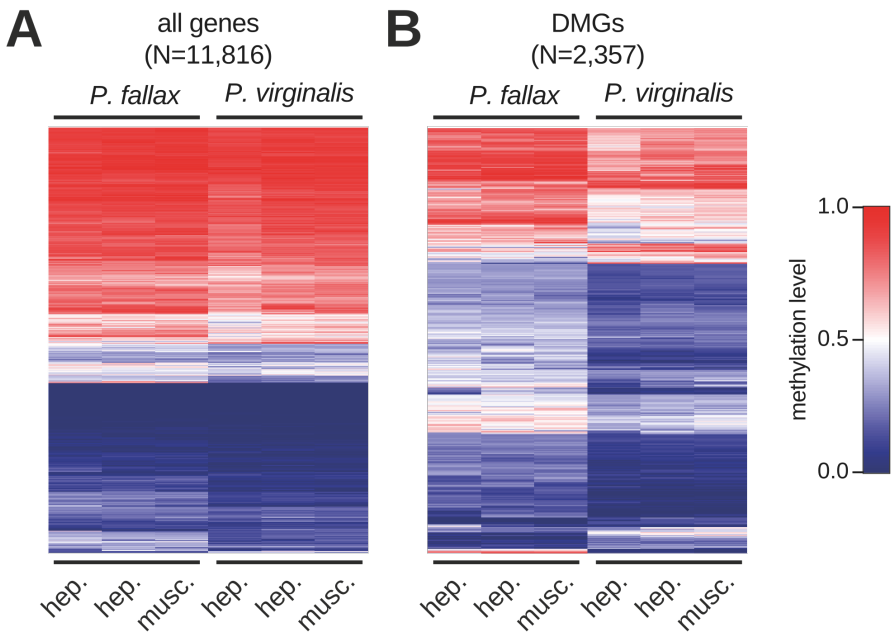

Figure 21: **Gene body methylation in the marbled crayfish and *P. fallax*.** (A) Comparative analysis of gene body methylation patterns in *P. virginalis* and *P. fallax*. (B) Species-specific differential gene body methylation. The heatmap shows differentially methylated genes (DMGs) with an average methylation ratio difference $>$ 0.1.

It can be shown that generally, gene body methylation levels were significantly reduced in the marbled crayfish (Figure 22A, p-value$<$2.2e-16). Strikingly, gene expression variability was significantly elevated in marbled crayfish (Figure 22B, p-value$<$5.58e-13). This is particularly interesting because these findings suggest that gene body hypomethylation in the marbled crayfish leads to increased gene expression variability levels when compared to its parent species.

2.5 Integrating methylation, chromatin accessibility, and expression

Finally, I explored the relationship between DNA methylation, chromatin accessibility, and gene expression patterns. ATAC-seq data was used to sensitively characterise high-resolution chromatin accessibility patterns for marbled crayfish hemocytes. These data were integrated with WGBS and RNA-seq data, which gives a comprehensive insight into the interplay between these three epigenetic layers. All three datasets were generated from the same pool of marbled crayfish hemocytes. The three ATAC-seq replicates were confirmed to have a high correlation among each other (data not shown), and were then pooled for increased coverage. In the pooled data, 89,156 high-confidence accessible peaks were identified. Of these, 4,558 overlapped with promoter regions.

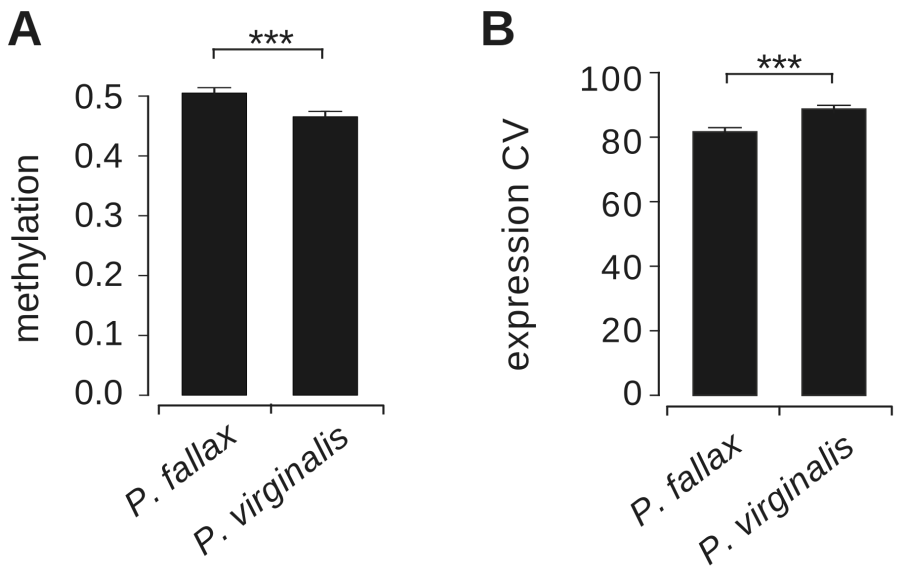

Figure 22: **Gene body hypomethylation and elevated gene expression variability in *P. virginalis*.** (A) *P. fallax* has significantly higher methylation than *P. virginalis* ($p<2.2e-16$, two-tailed t-test). (B) Comparison of gene expression variation between marbled crayfish and *P. fallax*. Coefficients of expression variation are indicated for triplicate RNA-seq datasets from the abdominal musculature. The difference between the two species is highly significant (p<5.58e-13, two-tailed t-test).

2.5.1 Accessibility, gene body methylation, and expression levels

To shed light on the association between chromatin accessibility and gene body methylation first, I generated heatmaps for low-methylated and high-methylated genes and used raw mapped read counts per chromosomal position as a measure of chromatin accessibility. For genes with lowly methylated coding regions, I found that chromatin around the transcription start site is distinctly more accessible than for highly methylated genes (Figure 23).

Analysing chromatin accessibility in metagene plots, I observed that ATAC signals (the raw mapped read counts) are most strongly in the region around 190 bp downstream of the transcription start site (Figure 24A). Unsurprisingly and consistent with observations from mouse embryonic stem cells (Clark et al., 2018), genes in very open chromatin were also the most expressed (Figure 24B). It is interesting to note that all other expression levels appear highly similar in their more closed chromatin state.

Visualising accessibility in the context of methylation, the metagene plot shows more clearly what I already observed in the heatmaps: genes with low to moderate gene body methylation are usually found in more open chromatin states than those with higher methylation levels (Figure 23C and Supplementary Figure S6). Interestingly, genes with low to intermediate levels of methylation are more accessible than completely unmethylated

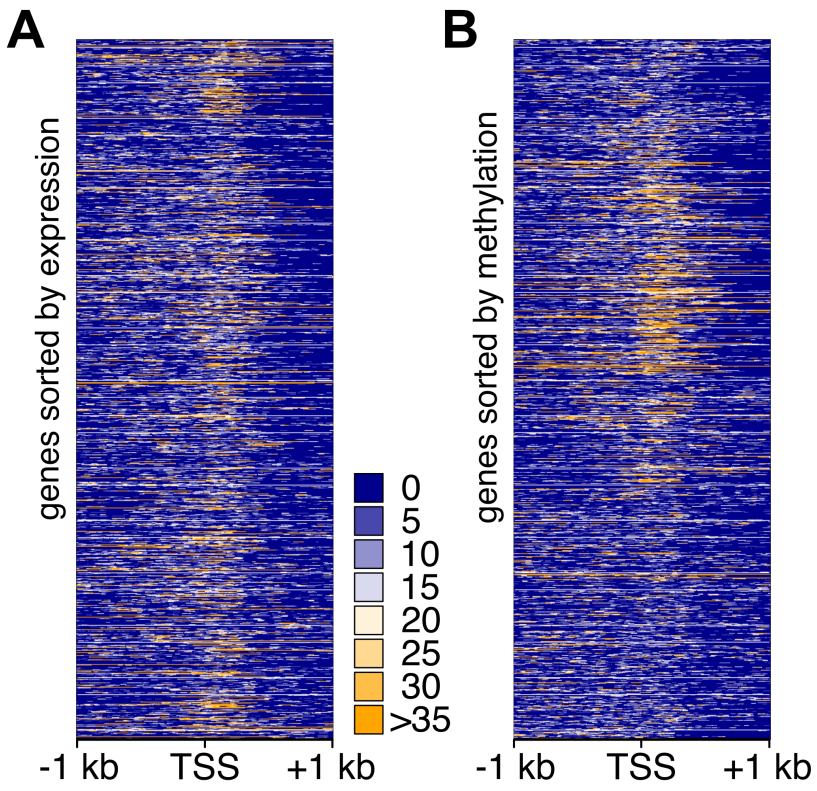

Figure 23: **Heatmaps of chromatin accessibility around transcription start sites (TSS).** Chromatin accessibility is shown for high-methylated (average methylation ratio > 0.5, left) and low-methylated (methylation level < 0.5, right) genes. Colors indicate raw mapped read counts per position, i.e., yellow indicates more accessible chromatin and blue indicates more closed chromatin.

genes. Consistent with the overall trend of highly methylated genes being found in more closed chromatin, housekeeping genes, which are often highly methylated, were found in more closed chromatin than tissue-specific genes (Figure 24D).

2.5.2 Gene body methylation might promote stable expression of poorly accessible genes

Finally, I analysed the interplay between gene body methylation, chromatin accessibility, and gene expression variability. First, the observation that hypomethyated genes show greater gene expression variability is also conserved in hemocytes (Figure S7).

Looking at gene expression variability in a metagene plot, I observed that stably expressed genes are generally found in more open chromatin states than variably expressed genes (Figure 25, left panel). This is consistent with findings in mouse embryonic stem

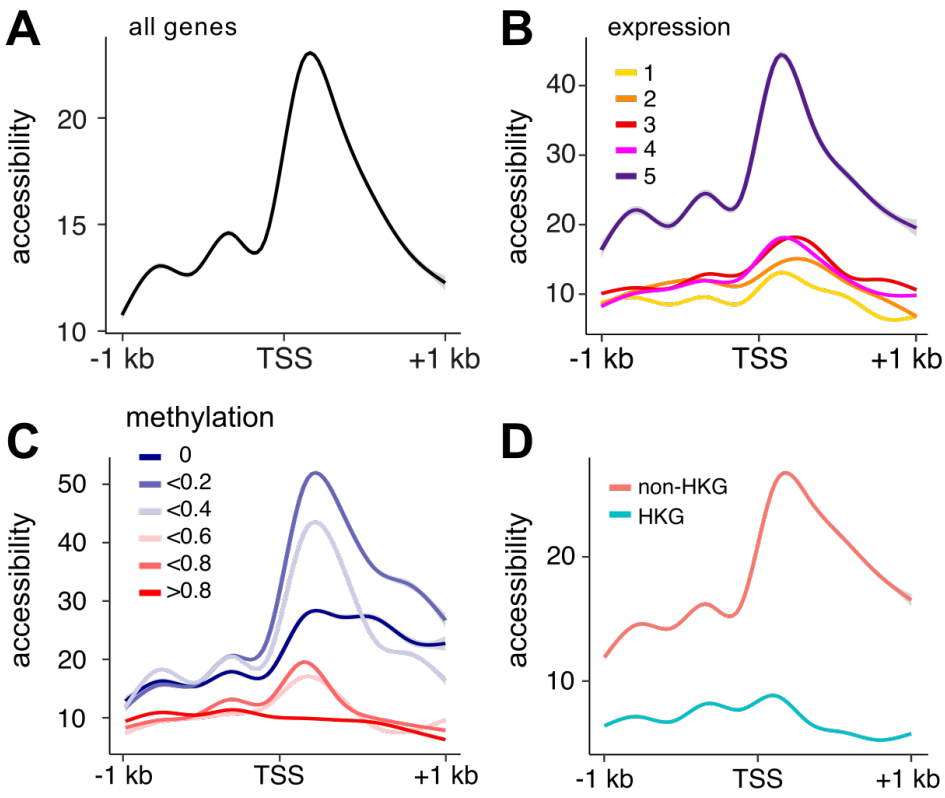

Figure 24: **Metagene plots of chromatin accessibility around transcription start sites (TSS).** (A) ATAC signals for all genes. (B) ATAC signals for all genes grouped by expression quintiles. (C) ATAC signals for all genes rouped by methylation quintiles. (D) ATAC signals for genes classified as housekeeping gene or tissue-specific gene.

cells, where conflicting chromatin marks of both open and closed chromatins were found to be associated with gene expression noise (Faure et al., 2017). Still, this is somewhat surprising since housekeeping genes are generally highly methylated and should, therefore, be found in less accessible chromatin, but are also stably expressed, and could therefore preferably lie in more accessible chromatin.

To understand this phenomenon, I divided genes into hypomethylated (average methylation ratio <0.4) and hypermethylated (average methylation ratio >0.4). I observed that lowly methylated genes with stable expression were distinctly more accessible than lowly methylated genes with variable expression (Figure 25, middle panel). In contrast, high-methylated genes are generally less variably expressed, and are consistently associated with more closed chromatin states (Figure 25, right panel), independent of their (relative) gene expression variability. These findings suggest a mechanism where gene body methylation could be involved in promoting the stable expression levels of poorly accessible genes, which would usually be variably expressed. This would hold, for example, for housekeeping genes, where variable gene expression is not a desirable trait.

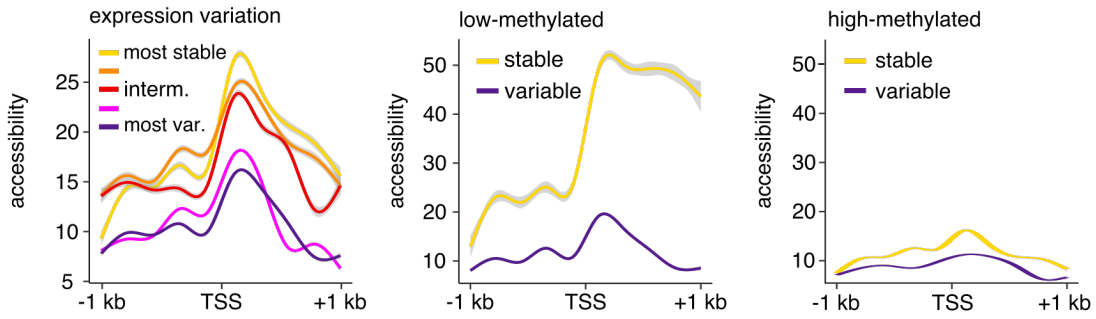

Figure 25: **Metagene plots of chromatin accessibility for expression variation.** Left panel, all genes grouped by their level of variable expression. Middle panel, lowly methylated genes grouped by their level of variable expression. Right panel, highly methylated genes grouped by their level of variable expression.

3 Discussion

DNA methylation is an ancient modification which is present in all three domains of life. Its patterns between vertebrates and invertebrates show substantial differences (Breiling and Lyko, 2015; Feng et al., 2010; Zemach et al., 2010), the first and major one being that vertebrate genomes are generally ubiquitously methylated, while most invertebrates show greater diversity in their methylation levels and targets. Despite these differences, the majority of studies on epigenetic regulation have been conducted for mammalian genomes. And while an increasing number of invertebrate methylomes has been published in the last few years (Rošić et al., 2018; Bewick et al., 2016; Glastad et al., 2016a; Kao et al., 2016; Libbrecht et al., 2016; Cunningham et al., 2015; Patalano et al., 2015; Lyko et al., 2010; Feng et al., 2010; Zemach and Zilberman, 2010), the majority of these have analysed methylation patterns from DNA that was isolated either from whole animals or from a single tissue.

Of all the published invertebrate methylomes, only two crustacean species have been analysed to date, namely *Daphnia pulex* (Asselman et al., 2016) and the sand flea *Parhyale hawaiensis* (Kao et al., 2016). This is despite the fact that crustaceans are dominant components of aquatic ecosystems which comprise more than 66,000 species and are the major constituent of zooplankton, making many of them keystone species for aquatic environments (LeBlanc, 2007).

In this thesis, I present an in-depth analysis of DNA methylation in the marbled crayfish, an emerging model organism that belongs to the order of decapod crustaceans and shows a great invasive capacity despite its global genetic homogeneity (Gutekunst et al., 2018). Whole-genome bisulfite sequencing data was used to directly compare methylation patterns from different tissues and from different marbled crayfish animals. These methylation maps were integrated with RNA-seq and ATAC-seq data to collectively analyse the interactions and correlations between DNA methylation, chromatin accessibility, and gene expression. The genetic homogeneity of the marbled crayfish, with very few confounding mutations, makes it a particularly promising model organism for epigenetics research. These analyses will broaden the knowledge on the conservation patterns in invertebrates, and provide insights into the effect of gene body methylation on gene expression and chromatin accessibility.

3.1 A highly methylated crustacean genome with specific methylation targets

Global methylation levels

An initial analysis of genome-wide CpG methylation levels was performed for a single hepatopancreas sample. It revealed that approximately 18% of all CpGs in the marbled crayfish genome are methylated. In comparison, the other two published crustacean methylomes, *Daphnia pulex* and *Parhyale hawaiensis* (Asselman et al., 2016; Kao et al., 2016), displayed lower methylation levels of 1.1 and 12.3%, respectively, when analysed by the same pipeline. The marbled crayfish genome is also more strongly methylated compared

to most insect species (Bewick et al., 2016), but lower than, for example, the sea tunicate *Ciona intestinalis* with 23.1% methylated CpGs (Suzuki et al., 2013).

Methylation is compartmentalised in the marbled crayfish genome
An analysis of 2kb windows revealed that large parts of 2 kb regions in the genome were unmethylated, while a distinct cluster was largely methylated. This indicates that methylation of CpGs in the marbled crayfish is not randomly or evenly distributed in the genome, but is more compartmentalised and enriched in specific regions. While this is a frequently observed pattern in invertebrate methylomes (Feng et al., 2010; Zemach et al., 2010), it is not always conserved: in 2015, Patalano et al. (2015) showed that the genomes of two eusocial insect species, *Polistes canadensis* (a wasp), and *Dinoponera quadriceps* (an ant), were more evenly methylated along the entire genome when compared to *Apis mellifera*, in which they found few, highly methylated regions, while the majority of the genome remained unmethylated.

In vertebrates, on the other hand, comparing 2 kb regions of methylation in the mouse genome showed, as expected, that most of these regions were methylated at more than 75% of the contained CpG sites, which is consistent with the ubiquitous methylation commonly reported for mammals and vertebrates (Feng et al., 2010; Zemach et al., 2010).

The marbled crayfish is sparsely methylated in repeats and methylated in gene bodies
Methylation in the marbled crayfish is primarily targeted to coding regions, and interestingly, introns show higher methylation levels than exons. While invertebrates generally have a strong preference for methylation enrichment in gene bodies (Feng et al., 2010; Zemach et al., 2010; Sarda et al., 2012), this observed pattern is unlike most animal and invertebrate species or plants, where exon methylation is usually higher than intron methylation (Glastad et al., 2016a; Cunningham et al., 2015; Patalano et al., 2015; Wang et al., 2013; Feng et al., 2010). However, a similar targeting preference of introns over exons has also been reported in the draft methylome of *Locusta migratoria* (Wang et al., 2014a). Section 3.3 discusses gene body methylation in the marbled crayfish in more detail.

5' UTRs displayed a moderate degree of methylation, while methylation levels in 3' UTRs were somewhat lower. While the overall methylation levels in these regions were comparable to other invertebrate species, for most species where UTRs were analysed, usually a more elevated level of methylation was reported for the 3' UTRs as opposed to the 5' UTRs (Patalano et al., 2015; Cunningham et al., 2015; Wang et al., 2013; Bonasio et al., 2012; Xiang et al., 2010).

A subset of repeats appears methylated in the marbled crayfish. Generally, invertebrates tend to be unmethylated or lowly methylated at repeats (Feng et al., 2010; Zemach et al., 2010; Zemach and Zilberman, 2010). However, for some species, it was reported that few, specific subsets of repeats are methylated (Rošić et al., 2018; Glastad et al., 2016a; Keller et al., 2016). This is discussed below.

3.2 Most repeats in the marbled crayfish are methylated in the context of gene bodies

For most major repeat classes, only a fraction of repeats in the marbled crayfish genome are methylated. This includes LINEs, SINEs, LTRs, and simple repeats. In contrast, approximately half of the annotated DNA transposons in the marbled crayfish are methylated. The majority of these are TcMar-Tiggers, an ancient group of repeats that move by a cut-and-paste mechanism (Smit and Riggs, 1996).

Methylation of repeats, or of TcMar-Tiggers in particular, did not show a correlation with the age or divergence of the repeat. However, high methylation levels of repeats strongly coincided with a location within a gene body. Introns are frequently made up of DNA transposons, and since gene bodies (and especially introns) are targeted by methylation in the marbled crayfish, it is possible that repeats are just incidentally methylated when they reside in an intron. Alternatively, these repeats could be specifically targeted by DNA methylation, potentially explaining the enhanced intron methylation compared to exons in the marbled crayfish.

In many animal and plant genomes, methylation of repeats is a functionally important defense mechanism against the activity of DNA transposons (Walsh et al., 1998; Zemach and Zilberman, 2010). Repeat methylation in invertebrates is often moderate (Glastad et al., 2016a; Kao et al., 2016; Wang et al., 2014b; Falckenhayn et al., 2013), and in some of these cases, the location of methylated repeats also coincides with introns (Glastad et al., 2016a). Other invertebrates like *Apis mellifera* or *Ciona intestinalis* show extremely low methylation levels at repeats, and in the case of the latter, these few, very slightly methylated repeats are also located within introns, so they might not constitute active methylation targets (Lyko et al., 2010; Simmen et al., 1999).

On the other hand, the nematode *Caenorhabditis elegans* has lost methylation completely, but more basal nematodes have reported DNA methylation, targeted exclusively to repetitive sequences (Rošić et al., 2018). Zemach and Zilberman (2010) suggested that, in invertebrates, repeat methylation is not a defense against transposon activity, but that this lineage silences transposable elements by other mechanisms like interfering RNAs (Aravin et al., 2007). However, the highly specific targeting of DNA methylation to a single transposon class, independent of a location within a gene body, raises the question whether repeat methylation does sometimes function as a silencing mechanism in invertebrates.

3.3 Distribution and targets of gene body methylation

3.3.1 Subtle tissue methylation differences in a largely tissue-invariant methylome

I analysed gene body methylation for a set of eight marbled crayfish DNA samples comprising different tissues and embryonic stages, using for each gene the average methylation ratio of all well-covered CpGs in that gene body. While direct statistical testing did

not identify differentially methylated genes between abdominal musculature and hepatopancreas, hiearchical clustering applied to the samples placed the abdominal musculature replicates into a single clade on the tree (Figure 8). Similarly, principal components analysis and multi-dimensional scaling narrowly separated the three hepatopancreas samples from all other tissues. Hence, there must be subtle differences in smaller sets of genes, or regions in genes, that exhibit a limited degree of tissue-specific methylation patterns while the majority of gene bodies displays tissue-invariant methylation.

These conclusions are different from paradigmatic mammalian methylomes, where statistically significant differences in methylation between adult tissues are frequently observed (Rakyan et al., 2008; Ziller et al., 2013). However, many of these differences in mammals are manifested in annotated promoter regions. I analysed 5'UTRs in the marbled crayfish for tissue-specific methylation and again, found no statistically significant differences, and only slight differences for hierarchical clustering. However, the fact that precise promoter locations are still unknown in the marbled crayfish draft genome assembly, as well as the limited number of samples, limits the significance of these analyses.

The question of how strong tissue-specific methylation is in invertebrate genomes has not been often addressed. In *Ciona intestinalis*, it has been observed that the same gene groups are methylated or unmethylated in two distinct tissues (Suzuki et al., 2013), whereas slight methylation differences in tissues were reported for *Bombyx mori* (Wu et al., 2017). However, Lea et al. discussed in 2017 that, to make a significant statement about differential methylation patterns in different samples, tissues or embryonic stages, much larger numbers of samples would have to be compared.

The methylation landscape in mammals undergoes drastic changes during development from the zygote through to post-implantation (Smith et al., 2012). While the single replicate of an embryonic stage in the marbled crayfish dataset did not display substantial differences to differentiated tissues when subjected to a principal components analysis or multi-dimensional scaling, it stems from a developmental stage obtained 7 days after egg deposition, where more than 1,000 cells are present in an egg (Alwes and Scholtz, 2006; Grimmer, 2015). It is therefore possible that reprogramming in the marbled crayfish occurs before this stage, and that major adult methylation patterns have already been established at this point. On the other hand, changes in the methylation landscape during development in mice still show methylation differences between post-implantation embryos on day 7.5 and adult tissues (Smith et al., 2012). It would therefore be an interesting question for the future to investigate high-resolution epigenetic reprogramming during marbled crayfish development, especially since is a question that has not yet been addressed in invertebrates.

3.3.2 Gene body methylation is targeted at housekeeping genes and weakly associated with gene expression

Previous comparisons of gene body methylation in the marbled crayfish revealed a bimodal distribution, where methylation seems enriched in a specific set genes, while other genes were largely unmethylated. Further analysis showed that methylation is preferentially tar-

geted to housekeeping genes (Falckenhayn, 2016). This is most consistent with reports from invertebrates and plants, where it is frequently reported that evolutionary conserved and moderately expressed genes are the main targets of gene body methylation (Sarda et al., 2012; Wang et al., 2014b, 2013; Bonasio et al., 2012; Zilberman, 2017; Takuno and Gaut, 2012). In the ubiquitously methylated vertebrate genomes, nearly all gene bodies are heavily methylated.

Gene body methylation has been associated with the active transcription of these genes in a variety of plants and animals (Zilberman et al., 2007; Jones, 2012; Ball et al., 2009; Zemach et al., 2010; Glastad et al., 2016a; Bonasio et al., 2012; Suzuki et al., 2013; Xiang et al., 2010). For the marbled crayfish methylome, I only observed a moderate, but statistically significant correlation of gene body methylation levels and expression. Genes with moderate expression levels show the highest amount of methylation in the marbled crayfish, while the most and least expressed genes are less likely to be methylated. This observation is conserved in plants and invertebrates (Zilberman et al., 2007; Zemach et al., 2010).

It has been shown in plants and insects that DNA methylation in the gene body inversely correlates with Polymerase II occupancy (Zilberman et al., 2007; Glastad et al., 2016b), suggesting that methylation actually interferes with the transcription or elongation despite its association with actively transcribed genes. For mammals, Lorincz et al. (2004) have reported a similar depletion of Polymerase II in methylated transgenes downstream of the promoter. Zilberman et al. (2007) proposed that aberant transcripts that are generated from cryptic intragenic promoters could actively methylate the according homologous DNA in the gene body through the short interfering (siRNA) pathway (Chan et al., 2005). Furthermore, they suggested that methylation of gene bodies is therefore likely beneficial, since it could inhibit transcriptional initiation at such cryptic intragenic promoters, preventing the generation of aberrant transcripts. This notion has been demonstrated by Neri et al. in 2017. However, this benefit carries the cost of reduced elongation efficiency.

On a different note, while it is widely agreed that methylation of promoters leads to silencing of genes, I even observe a similarly weak positive correlation between promoters (or 3'UTRs) and expression levels of the respective genes. I did, however, observe slight indications that 5'UTRs that were hypermethylated in abdominal musculature when compared to hepatopancreas, resulted in unexpressed genes in abdominal musculature (Figure 18). However, before promoters have been properly annotated in the marbled crayfish genome, it is difficult to draw conclusions from these observations.

3.3.3 A set of genes is variably methylated in the marbled crayfish

The high stability of methylation patterns in gene bodies notwithstanding, I also identified nearly 700 genes that were defined by their variable methylation across the analysed samples. These genes showed an enrichment for functions that are related to cellular biosynthesis and metabolism. Interestingly, when subjecting these to a principal components analysis, hepatopancreas samples were more easily separated from other tissues than

they could be when using all genes. These results could suggest a more important role in tissue-specific regulation for this set of genes. Their variability could even reflect local adaptation, considering that the samples come from different animals that were raised in different conditions such as lab stock and different wild origins. However, the set of samples that I used would again be too small for significant analyses in this respect (Lea et al., 2017).

DNA methylation has been repeatedly suggested as a quick response to sudden environmental changes that could otherwise not be answered quickly enough by genetic mutations (Jaenisch and Bird, 2003; Verhoeven et al., 2016; Duncan et al., 2014). In 2015, Dubin et al. (2015) reported that gene body methylation in *Arabidopsis thaliana* displayed higher methylation levels for plants that originated from colder regions compared to those from warmer regions. This was associated with changes in transcription levels for the respective genes, possibly reflecting a local adaptation mechanism (Dubin et al., 2015). Similar associations have been observed in the 1001 Genomes collection of *Arabidopsis thaliana*, where the geographic origin of a plant was reported to be a major predictor of methylation levels as well as altered gene expression patterns (Kawakatsu et al., 2016). Rapid methylation changes in response to direct biotic stress have been shown in mice, *Daphnia pulex*, and again *Arabidopsis thaliana* (Radford et al., 2014; Asselman et al., 2017; Dowen et al., 2012).

In all these cases, the hypothesis is that a specific change in methylation implies a specific expression change as a response to the environment, possibly resulting in novel phenotypes. Furthermore, Feinberg and Irizarry (2010) suggested that genetic variants could exist that induce variable phenotypes through epigenetic stochasticity. In this case, the variability of methylation patterns would produce different phenotypes, which would be naturally selected by the environment.
And while I did not observe gene body methylation changes to be associated with differential gene expression between tissues, a larger number of samples from different animals, origins, and tissues could probably address this question in a more appropriate depth.

3.4 Gene body methylation as a stabilising mechanism for gene expression levels

3.4.1 Gene body methylation is inversely correlated with gene expression variability

In the marbled crayfish, I observed a strong and statistically significant inverse correlation of gene body methylation levels and the variability of their expression. This is a conserved mechanism in some insects (Glastad et al., 2016a; Wang et al., 2013) and humans (Huh et al., 2013). In line with this observation, Neri et al. (2017) have shown that gene body methylation serves as a mechanism to suppress cryptic intragenic promoters in transcribed genes. Since gene body methylation is generally targeted at housekeeping genes, it makes sense that such genes should be both stably expressed, and not produce aberrant transcripts.

Gene expression variability as a mechanism for environmental adaptability
It has long been shown that gene expression levels are inherently variable, even in clonal cell populations and homogeneous environments (Elowitz, 2002). This phenomenon has been linked to intrinsic noise, extrinsic noise and phenotypic plasticity (Elowitz, 2002; Newman et al., 2006). The stochastic switching between phenotypic traits can be beneficial, since it can confer plasticity as a quick response to a changing environment (Beaumont et al., 2009).
For example, Kenkel and Matz (2016) found that inshore corals from a thermally variable environment display a greater capacity for gene expression plasticity than corals from a more stable habitat: when re-located to a new environment, these inshore corals were more capable at adopting gene expression profiles of native corals than those from a stable habitat. However, if variability of gene expression increases too much, it could lead to the disruption of gene regulatory networks, suggesting that it is a trait which should be controlled.

Notably, I found that the gene bodies of marbled crayfish were significantly less methylated than those of its parent species, *Procambarus fallax*. Furthermore, gene expression variability was significantly increased in the marbled crayfish in comparison to *P. fallax*. In contrast to the marbled crayfish, *P. fallax* does not display evidence for an invasiveness like that of the marbled crayfish and has, so far, only been reported in defined habitats southern Georgia and Florida (Hendrix and Loftus, 2000; Hobbs, 1981).
And while the parthenogenetic reproduction mode of the marbled crayfish is most likely a major contributor to its invasive potential, there is a possibility that hypomethylation of gene bodies in marbled crayfish increases its gene expression variability, and this might provide an additional mechanism for its remarkable adaptability.

3.4.2 DNA methylation as a mechanism to stabilise gene expression variability in poorly accessible genes

Finally, I investigated the relationship between DNA gene body methylation, gene expression patterns and chromatin accessibility around the transcription start site. To our knowledge, this is the first study to simultaneously use whole-genome bisulfite sequencing, RNA-seq, and ATAC-seq data for comprehensive analyses of the interplay of DNA methylation, gene expression and chromatin accessibility, which can deepen our understanding of the gene regulatory potential of gene body methylation.

Consistent with observations from mouse embryonic stem cells (mESCs) and the malaria parasite *Plasmodium falciparum* (Clark et al., 2018; Toenhake et al., 2018), genes in highly accessible chromatin were also the most expressed. Also consistent with mESCs, the transcription start sites of highly body-methylated genes and housekeeping genes were found in more condensed chromatin states than unmethylated and tissue-specific genes. Interestingly, low to moderate methylation levels were associated with higher chromatin accessibility than completely unmethylated genes. In this context, it is notable that Yin et al. (2017) reported that certain transcription factors bind specifically to methylated DNA, which were enriched for functions with roles in embryonic and organismal development. This is also interesting when considering that moderately methylated genes show

higher chromatin accessibility than completely unmethylated genes.

I also observed that genes with high gene expression variability are generally found in more poorly accessible chromatin states than stably expressed genes. Faure et al. (2017) made a similar report for mESCs: most mESC promoters (>90%) possess the H3K4me3 mark, which is associated with active promoters. Furthermore, they found the generally repressive histone modification H3K27me3 primarily at genes with high expression noise. This simultaneous occurrence of apparently conflicting histone modifications is called bivalency and has been associated with genes that are rapidly up- or downregulated during organismal development. This effect could allow a rapid response to internal and external environmental cues through an according increase or decrease of gene expression levels.

These findings are somewhat surprising when considering that housekeeping genes are usually highly methylated, which is associated with more condensed chromatin, but are stably expressed, which should make them located in open chromatin, according to the results discussed before. To build a more comprehensive picture and integrate information from my three datasets, I grouped genes into lowly methylated and highly methylated sets of genes and could show the following:

- highly methylated genes have overall lower gene expression variability
- highly methylated genes are generally associated with more condensed chromatin, regardless of their relative gene expression variability
- lowly methylated genes are located in open chromatin when stably expressed, and in more condensed chromatin when variably expressed.

Variable gene expression can be beneficial, as discussed before when rapid up- or downregulation is a desirable trait (e.g., during embryonic development), or when it can provide a quick response to a changing environment. However, in the case of housekeeping genes, for example, this would not be advantageous.

Taken together, we can state that less accessible chromatin, potentially involving conflicting histone modifications, appears to facilitate the variable expression of genes. In this context, my studies suggest that gene body methylation helps to promote stable expression of poorly accessible genes like housekeeping genes, where variability in expression is not a desirable trait (see Figure 26).

3.5 Summary

This thesis established the methylome the marbled crayfish and provided an in-depth analysis of DNA methylation patterns and its targets, as well as the interplay between gene body methylation, gene expression, and chromatin accessibility. By directly comparing methylation in different tissues, I could show that the marbled crayfish displays only very moderate methylation differences between tissues, which is in contrast to well-studied mammalian methylomes. Not many comparisons of methylation between tissues have been conducted in invertebrates, so this finding helps to broaden our knowledge

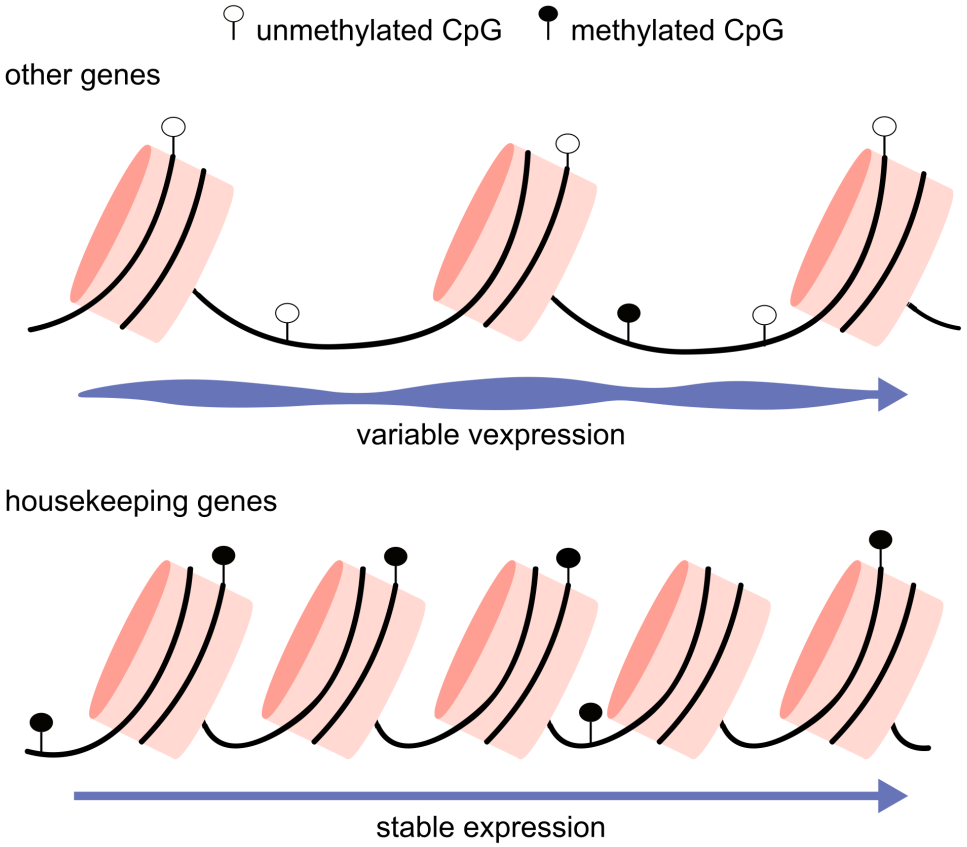

Figure 26: **Gene body methylation stabilises gene expression of poorly accessible genes in the marbled crayfish.** Housekeeping genes are usually located in chromatin states with limited accessibility and commonly show high methylation levels (lower panel). This stabilises their expression compared with other non-housekeeping genes, which are lowly methylated and usually found in more accessible chromatin (upper panel). Black filled circle, methylated CpG. Empty circle, unmethylated CpG

about the conservation of methylation patterns in invertebrate tissues.

While most repeat classes in the marbled crayfish were sparsely methylated, a distinct class of transposable elements was heavily targeted by methylation. Most invertebrate genomes show sparse repeat methylation, and make use of other ways to silence repeats. The observation made here could imply that repeat methylation still sometimes functions as a silencing mechanism in invertebrates.

While the correlation between gene body methylation and expression level is only moderate in the marbled crayfish, it displayed a strong inverse correlation between gene body methylation and gene expression variability. Notably, the marbled crayfish is significantly hypomethylated in comparison to its less invasive parent species, and its genes are also significantly more variably expressed. Since gene expression variability can be a beneficial trait in an often-changing environment, it is possible that hypomethylation of gene

bodies provides a mechanism for the adaptability of the marbled crayfish by increasing gene expression variability.

Finally, genes with high gene expression variability are frequently observed in more condensed chromatin. This is in accordance with previous reports in mouse embryonic stem cells, where conflicting chromatin states with both active and inactive histone marks are associated with gene expression noise and are attributed to genes where rapid up- or downregulation of expression is a desirable trait. Simultaneously, I found that highly methylated genes, or housekeeping genes also often reside in poorly accessible chromatin. I could show that methylated genes are less variably expressed, despite their preferential location in more condensed chromatin. In this context, gene body methylation could function to stabilise gene expression of genes that lie in condensed chromatin, but where variable expression is not beneficial.

3.6 Outlook

Gene regulation is a complex interplay involving a variety of mechanisms like DNA methylation, histone modifications, and chromatin remodelling enzymes, non-coding RNAs and transcription factors. While, to our knowledge, this is the first comprehensive study making use of whole-genome bisulfite sequencing, RNA-seq and ATAC-seq at the same time, more studies and more types of data would be needed to understand the exact process during which gene body methylation promotes stable gene expression. Studies of histone modifications and RNA editing, for example, are other epigenetic layers that would be worth examining. In fact, RNA editing is currently under investigation in our group. While a larger variety of data types for simultaneous analyses will of course deepen our understanding for such mechanisms, it will be challenging to integrate all these at the same time to form a comprehensive picture.

Many of the analyses in this thesis were limited by the quality and annotation of the draft genome assembly, which was generated from short-read sequencing data. Using short-read sequencing data for assemblies of complex genomes is especially challenging in a triploid genome like the marbled crayfish, where different alleles impose an additional confounding factor for the assembly pipeline. Third-generation-sequencing technologies like PacBio single molecule real time sequencing (SMRT) generate long sequencing reads of up to > 30 kbp lengths, which can be used either individually, or simultaneously with short-read sequencing data to create a hybrid assembly. Other approaches like Dovetail Hi-C sequencing use cross-linking of chromatin to detect long-range chromatin interactions in the genome, and can drastically improve existing assemblies. Both Dovetail Hi-C, as well as PacBio sequencing data have been generated for the marbled crayfish, and an improved reference genome version is currently being assembled. This will be followed by an updated annotation of the new genome.

For example, precise promoter annotations would allow much more precise analyses of gene regulation through DNA methylation and chromatin accessibility in the marbled crayfish. I observed slight indications that methylated 5'UTRs in abdominal musculature result in unexpressed genes, and true promoter annotations would provide deeper insight

and significance for such analyses. In this context, it would also be exciting to analyse whether marbled crayfish embryo genomes undergo reprogramming of methylation in promoters or gene bodies during early embryonic development. Additionally, only 8-9% of the genome is annotated as repeats. This number is likely much higher, given the difficulty of performing correct assemblies for repetitive regions with short-read sequencing data. Moreover, new repeat annotation tools designed for invertebrate genomes could be used to more correctly and comprehensively annotate repeats and transposons in the marbled crayfish.

Furthermore, it would be exciting to be able to pinpoint exact methylation pattern and expression profile differences to different environmental factors. While the significance of my analyses regarding this topic was also limited by the number of samples, for the set of approximately seven hundred variably methylated genes, a capture array has been generated to perform bisulfite-sequencing just for these genes. This permits a large number of animals to be processed, and more than one hundred samples from a variety of environmental sources and different tissues are currently being analysed. This could allow the identification of epigenetic ecotypes, where a certain methylation modification corresponds to a specific environment, and deepen our knowledge of environmental epigenetics.

4 Materials and Methods

This chapter describes the technical infrastructure (hardware and software) used for data analysis, as well as sample origins and preparation methods. Additionally, details on downstream bioinformatic analyses are given.

4.1 Computing system

4.1.1 Hardware

Basic data processing (trimming, mapping and database generation) was run on the DKFZ's high-performance computing cluster (HPC) or on the group's local server. The cluster has a total of 352 cores on 21 nodes, comprising 2 TB RAM. Methylation data was stored in MySQL databases on the group's local server. Downstream analyses were performed on a local desktop computer that holds a total of eight CPUs in four cores and 32 GB of RAM.

4.1.2 Software packages

Table 7 gives an overview of all major software packages used for this project, along with their versions and sources. Different software versions for one tool were either used because specific tools require specific versions of other softwares, or were used on different platforms, or at different stages in the project. Exact analyses and minor packages or methods will be discussed in section 4.3.

Table 7: **Major software packages.** Overview of major software used for this project. Program version and their sources are provided. The exact usage and minor packages or functions are described in section 4.3.

Software	Version	Source
FastQC	0.11.3	Andrews (2010)
Trimmomatic	0.35	Bolger et al. (2014)
TrimGalore	0.4.4	Martin (2011); Krueger (2012)
Bsmap	2.73	Xi and Li (2009)
picard	1.137	Broadinstitute
SAMtools	1.3, 1.8	Li et al. (2009)
hisat2	2.0.4	Kim et al. (2015)
htseq-count	0.60.	Anders et al. (2015)
DESeq2	1.14.1	Anders and Huber (2010)
BLAST	2.2.28+	Altschul et al. (1990)
PANTHER enrichment analysis	13.1	Mi et al. (2017)
bowtie2	2.2.6	Langmead and Salzberg (2012)
python	2.7, 3.4	van Rossum (1995)
R	3.3.0	R Development Core Team (2013)

4.2 Sample acquisition, preparation and sequencing

4.2.1 Origin of samples and animal culture

Animals used for the analyses in this thesis came from various sources. These are described in detail in table 8. Laboratory animals come from the lineage in our division (Vogt et al., 2015). Wild animals were collected from lakes in Germany and Madagascar in accordance with local fishery regulations.

Table 8: ***Procambarus* animals used for sequencing.** Species, specimen and the origin of the sample are indicated. Samples from lake Moosweiher, Moramanga and lake Reilingen were collected by previous and current lab members or cooperation partners. Eggs from several animals were pooled for an embryonic stage, just as hemolymph was pooled for several animals.

Species	Specimen	Origin
P. virginalis	pooled eggs	lab stock
P. virginalis	Pvir#2	lab stock
P. virginalis	Pvir#3	lake Moosweiher, Germany
P. virginalis	Pvir#6	lab stock
P. virginalis	Pvir#2	lab stock
P. virginalis	Pvir#3	lake Moosweiher, Germany
P. virginalis	Mora	Moramanga, Madagascar
P. virginalis	pooled hemolymph	lake Reilingen, Germany
P. fallax	Pfal#1	lab stock
P. fallax	Pfal#3	lab stock
P. fallax	Pfal#4	lab stock
P. fallax	Pfal#4	lab stock

4.2.2 Sample preparations and DNA and RNA extractions

For lab stock animals and animals collected in Germany, the samples of hepatopancreas and abdominal musculature from adult crayfish were taken from adult animals under a dissection microscope. They were snap-frozen in liquid nitrogen and stored at -80 degree C until extraction of nucleic acids. Embryonic eggs were snap-frozen in liquid nitrogen after collecting them from the mother. The sample form Moramanga, Madagascar, was stored in ethanol until extraction of DNA. Genomic DNA was isolated using the Blood & Cell Culture DNA Kit (Qiagen, Hilden,Germany), and total RNA was purified with Trizol (Invitrogen, Darmstadt, Germany).
For ATAC-seq, three independent biological samples were collected by bleeding individual marbled crayfish. Approximately 500 mykroL of hemolymph was collected using a 23G needle inserted in the abdomen of the a cold-anesthetized crayfish. Extractions were performed by Katharina Hanna and Dr. Vitor Coutinho.
For RNA-seq, total RNA was isolated from 20-60 mg frozen tissues. Thawed tissues were homogenized in 1 ml Trizol (Ambion), precipitated with isopropanol and resuspended in

20-100 mikroliter RNase-free water (Gibco Life Technologies). Total RNA was treated with DNase using the RNeasy Mini Kit (Qiagen) following the manufacturer's RNeasy Mini Protocol for RNA Cleanup in combination with the On-Column DNase Digestion Protocol.

4.2.3 Library preparation and high-throuput sequencing

The DKFZ Genomics and Proteomics Core Facility (Heidelberg, Germany) prepared libraries for WGBS and RNA-seq. In short, the TruSeq PCR-Free Library Prep Kit (LT) (Illumina, San Diego, U.S.) was used for library preparation, and the Epitect Kit (Qiagen) for bisulfite conversion. Library amplification was performed using the Kapa HiFi HotStart Uracil+ ReadyMix (2X) (Kapa Biosystems, Wilmington, U.S.), and libraries were sequenced on an Illumina HiSeq platform.
ATAC-seq libraries were generated by Dr. Vitor Coutinho. After collecting hemolypmh, one volume of anti-coagulant solution (0.14M and NaCl, 0.1M glucose, 30mM Na3Citrate.2 H20, 26mM citric acid, 0.5M EDTA) was added prior centrifugation of 300 x g, 5 minutes at 4°C. After washing the cell pellet twice with sterile and cold PBS 1X, 50.000 hemocytes were immediately used for the ATAC library preparation. The transposase reaction was optimized and a 20 minutes reaction was used to avoid DNA overdigestion. The subsequent steps were followed strictly to the original protocol (Buenrostro et al., 2013). Samples were then sequenced on different Illumina HiSeq platforms (see Tables 1 and 2).
For RNA-seq, sequencing libraries were prepared using 1 mikrogram of DNAse-treated total RNA in the first step of the TruSeq RNA Sample Preparation v2 protocol (Illumina, Part 15026495 Rev. A) as recommended by the manufacturer.

4.3 Bioinformatic analyses

4.3.1 Quality assurance and basic processing of the data

As a first measure, quality of data was assured using the FastQC (Andrews, 2010) report. Raw reads were trimmed for adapters and low quality bases or reads using Trimmomatic (Bolger et al., 2014) with the following parameters:

- ILLUMINACLIP: Remove Illumina adapters provided in a separate file. Trimmomatic looks for seed matches (16 bases) and allows a maximum of 2 mismatches. Seeds will be extended and clipped if a score of 30 is reached (paired-end reads), or of 10 (single-end reads)
- LEADING, TRAILING: remove leading and trailing bases with a quality less than 3
- SLIDINGWINDOW: for sliding 4-base windows, cut the read if an average quality of a window drops to less than 15
- MINLEN: only keep reads with a minimum length of 36

This was done for WGBS and RNA-seq. ATAC-seq data was trimmed using TrimGalore (Martin, 2011; Krueger, 2012) using default parameters except for an expected minimum Phred score of 20 required.

4.3.2 Alignment and analyses of whole-genome bisulfite sequencing data

Mapping and storing the data
Trimmed WGBS reads from both marbled crayfish and *P. fallax* were mapped against the marbled crayfish draft genome assembly (Gutekunst et al., 2018). *P. fallax* reads could also be mapped to the marbled crayfish genome, since the two species are closely related and genetically very similar (Gutekunst et al., 2018). BSMAP (Xi and Li, 2009) was used for mapping because it addresses a number of complications that arise in the process of mapping WGBS data (see section 1.4.2. The software was used with default parameters, i.e., a seed of length 16 and no mismatches allowed in the seed. Only read pairs that mapped onto the same scaffold with the appropriate orientation and distance to each other were used downstream for methylation calling. Also, only reads pairs mapping uniquely in one genomic regions were considered for further analyses. Methylation calling was performed using the python script provided by the BSMAP package. The script was modified slightly to improve the quality of data:

- a minimum Phred score of 30 was required for the analysed cytosine
- a minimum Phred score of 20 was required for the four surrounding bases

Only cytosines with a strand-specific coverage of 3X were used for analyses. Methylation ratio was determined as formula: ratio = # methylated reads / # total reads The data was saved in MySQL databases containing only CpG dinucleotides, their scaffold, position, methylation and coverage.
Data for other methylomes (*Parhyale hawaiensis*, *Daphnia pulex* and *Mus musculus*) was processed in the same way, using their respective reference genomes.

Global methylation levels
Violin plots were generated for 2 kb running windows for individual methylomes using R's ggplot2.violinplot function.

Targets and metagene plot
Barplots for methylation targets in the marbled crayfish methylome was defined as the average methylation ratio for all CpGs in the given regions (all genes, all repeats, intergenic regions). Annotations were used as published (Gutekunst et al., 2018). The metagene plot was generated as the average methylation per position across all annotated genes, scaling for gene lengths accordingly.

Heatmaps and repeat analyses
Heatmaps were generated for genomic regions(promoters, genes and repeats) as follows: for each genomic region, the average methylation ratio for all CpGs in that genomic region was calculated. The minimum number of CpGs required to consider a gene was set at ten, and for repeats and promoters at 3. R's heatmap.2 function was used to generate heatmaps, using hierarchical clustering to group rows (genomic regions) and columns (samples).
For the comparison of gene body methylation between *P. virginalis* and *P. fallax*, the same cutoffs for the number of CpGs and coverage were used, and differentially methylated

genes between species were defined as having an difference in methylation ratio larger than 0.1.
Repeat locations with respect to coding regions were defined as follows:

- outside: repeat has no overlap to gene body
- partial: repeat has an incomplete overlap with gene body of at least 10 bp
- within: repeat lies within the gene body

Repeats were grouped, and their methylation plotted as boxplots with R's ggplot2.
Repeat age was obtained from the repeatmasker output from the maker pipeline (Holt and Yandell, 2011), and plotted as a scatterplot with a regression line using R's abline and lm functions. The correlation coefficient was assessed using the cor function, and tested for significance using cor.test.

Variably methylated genes and grouping/separating the samples
A set of variably methylated genes was defined as follows: After plotting a histogram of the variance of the methylation ratio for all genes across the set of 8 samples, a variance cutoff of 0.008 was chosen, i.e., all genes with a variance of greater than 0.008 were used for further consideration. In addition, only genes with a mean methylation ratio of larger than 0.2 and less than 0.8 were chosen to only include with a greater spread of methylation. A heatmap was generated as described before for these samples.
Both a principal components analysis (PCA) as well as multi-dimensional scaling (MDS) were performed for the samples. This was done on the set of all genes as well as for all genes. PCA was computed using R's prcomp() function. Metric MDS was computed with R's cmdscale function, and a non-metric MDS with R's isoMDS function.

Visualizing methylation in the Apollo Genome Browser
Methylation was examined in the Apollo Genome Browser set up at marmorkrebs.dkfz.de (Gutekunst et al., 2018).

Differential methylation between tissues to be correlated with expression changes
Methylation differences between tissues were calculated for each gene as the average gene body methylation between three samples of one tissue, subtracted by the average gene body methylation between three samples of another tissue. Differential promoter methylation was calculated in the same manner.

Testing for significantly different levels of methylation between species
For the two species, a two-sided t-test was used in R to determine whether differences in methylation were significant.

Gene set enrichment analyses
Sets of genes, in this case the variably methylated genes, were blasted against the *Drosophila melanogaster* uniprot protein set downloaded from
https://www.uniprot.org/proteomes/UP000000803
on August 23, 2017. Hits with an e-value cutoff of 0.1 were submitted to PANTHER gene list analysis (Mi et al., 2017) using a corrected p-value cutoff of 0.1.

4.3.3 RNA-seq analyses

Mapping and units
Trimmed RNA-seq data was mapped with the splice-aware RNA alignment software hisat2 (Kim et al., 2015). Gene expression levels for correlation analyses were calculated using rsem (Li and Dewey, 2011), which computes a number of units to measure gene expression levels, including FPKM (Fragments Per Kilobase Million) and TPM (Transcripts Per Kilobase Million). TPMs were used since they allow for better comparison of expression levels between samples. They are calculated as follows:

1. Read counts per gene are normalized by the length of the gene (in kilobases). This gives the reads per kilobase (RPK)

2. The sum of all RPK values in a sample is divided by 1,000,000. This gives the "per million" scaling factor

3. RPK values are divided by the "per million" scaling factor, resulting in the TPM values.

Differential gene expression
For differential gene expression between marbled crayfish tissues hepatopancreas and between marbled crayfish and *P. fallax* abdominal musculature tissue, htseq-count (Anders et al., 2015) was used to calculate gene expression counts. These were used as an input for DESeq2 (Anders and Huber, 2010), which was applied with a p-value cutoff of 0.1 to identify differentially expressed genes. MA plots were generated using plotMA() from the DESeq2 package. For heatmaps, the DESeq2 matrix of counts was log2 transformed for normalization, and R's dist() function was then applied to generate a distance matrix. This matrix was plotted using heatmap.2. For both analyses, two sets of three biological replicates each were used and differentially expressed genes were enriched as described in the last paragraph of section 4.3.2.

Correlation of methylation and expression levels
Methylation ratios and gene expression levels (TPMs) were plotted in R as a scatterplot with a regression line and tested for significance of correlation as described in 4.3.2. Additionally, genes were binned into expression ranks by their log10 (TPM) value and plotted against methylation as boxplots.
Differences in expression between groups were determined as the average log2 fold change as calculated by DESeq2. Differences in methylation were computed by taking the average methylation ratio per gene (or promoter) in each replicate group, and subtracting the two from each other. The log2 fold change was then plotted against the difference in methylation in a scatterplot. Ages of genes and housekeeping gene definitions were taken as defined by Falckenhayn (2016).
Differences in expression between tissues were examined by plotting average TPM values of tissues against each other in R and coloring them by differential methylation calculated as described above.

Expression variation
Variation in expression was computed from the TPM values of three biological replicates.

The coefficient of variation was used since it normalises the squared standard deviation by the mean. Genes were binned into methylation ranks and plotted as boxplots against the expression coefficient of variation in R. Additionally, scatterplots and regression lines / correlation testing was performed as described before.
A two-sided t-test was used in R to determine whether differences in the expression coefficient of variation were significant between the two species.

4.3.4 ATAC-seq

Mapping and units
ATAC-seq data was trimmed using TrimGalore (Martin, 2011; Krueger, 2012) using default parameters. TrimGalore was used in this case to clip Nextera adapters used for sequencing. Trimmed reads were mapped against the marbled crayfish assembly using bowtie2 (Langmead and Salzberg, 2012) using default parameters. This was done by my colleague Julian Gutekunst. From the BAM alignment files, pure read counts (or coverage) per position were then extracted and used as a measure for accessibility at a given site.

Assessing accessibility for different sets of genes
Heatmaps were generated for chromatin accessibility around the transcription start sites of genes. For 1000 bp upstream and downstream of transcription start sites, ATAC read coverage was plotted for different sets of genes using R's image function. Genes were ordered in ascending order by their average accessibility.
Metagene plots were generated for different sets of genes using R's ggplot2 geom_smooth function. Genes were grouped by methylation level, expression level or expression variability level.

A Appendix

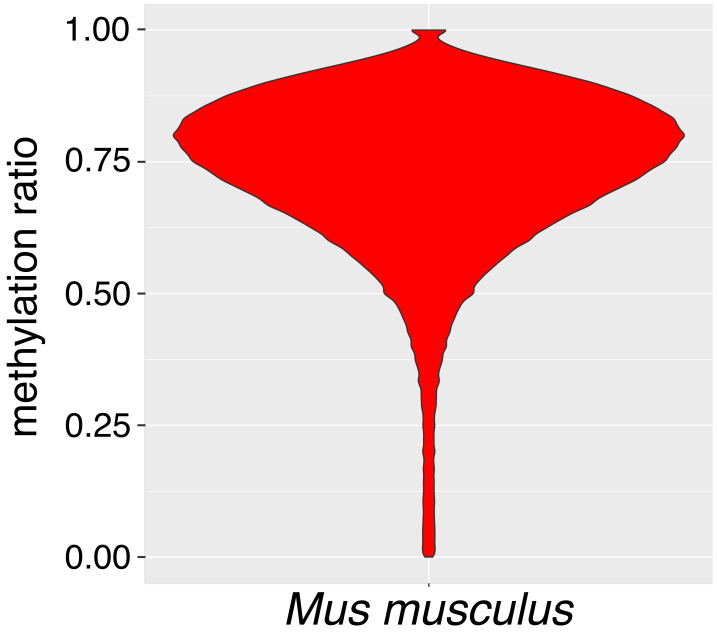

Figure S1: **Violin plot for the mouse showing DNA methylation levels for 2kb-windows.** The violin plot illustrates the ubiquitous methylation levels that are typical for mammalian genomes.

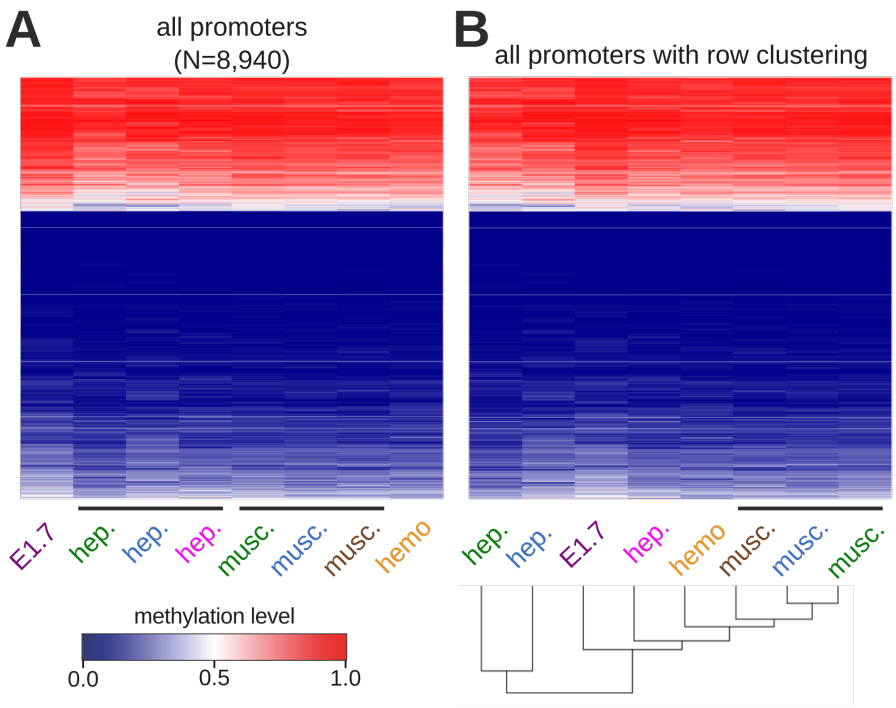

Figure S2: **Comparison of promoter methylation comprising a set of different developmental stages, tissues and animals.** (A) Comparative analysis of promoter patterns shown in a heatmap where hierarchical clustering for rows (promoters) was used. The heatmap shows average promoter levels for each gene for the set of 8 independent samples (columns). Colors indicate individual animals. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). Only genes containing at least 10 CpGs with a strand-specific coverage of 3x in all 8 samples are shown. E1.7: stage 1.7 embryos, hep.: hepatopancreas, musc.: abdominal musculature, hemo: hemocytes. (B) Similar heatmap, but including clustering for individual samples and tissues.

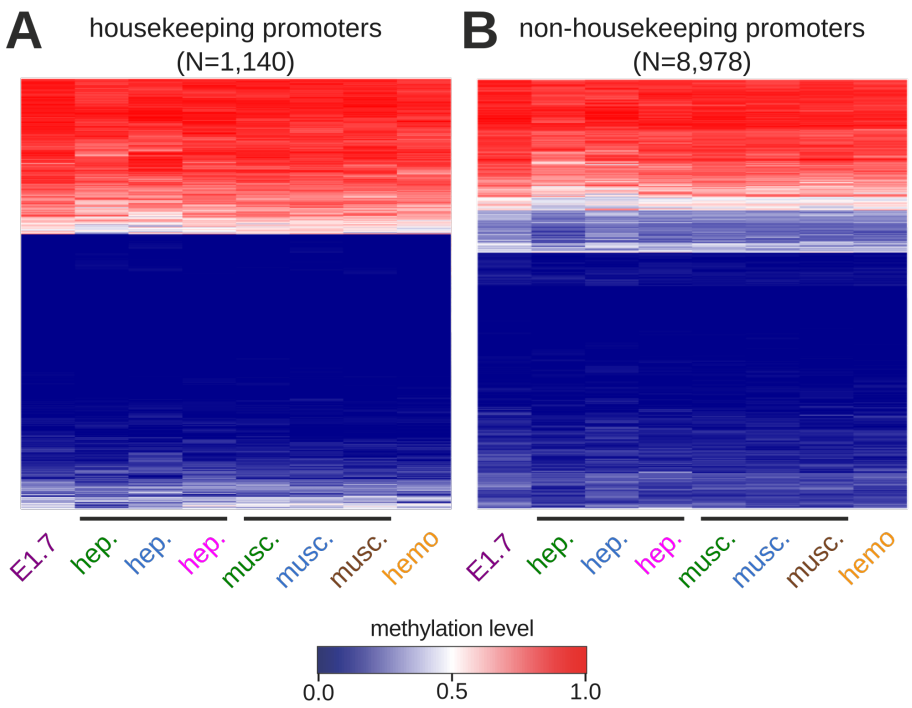

Figure S3: **Comparison of promoter methylation in different sets of genes.** (A) Comparative analysis of promoter methylation patterns in housekeeping genes shown in a heatmap where hierarchical clustering for rows (genes) was used. The heatmap shows average promoter methylation levels for each gene for the set of 8 independent samples (columns). Colors indicate individual animals. Methylation levels are indicated on a scale from 0 (blue) to 1 (red). Only genes containing at least 10 CpGs with a strand-specific coverage of 3x in all 8 samples are shown. E1.7: stage 1.7 embryos, hep.: hepatopancreas, musc.: abdominal musculature, hemo: hemocytes. (B) Similar heatmap, but only for non-housekeeping genes.

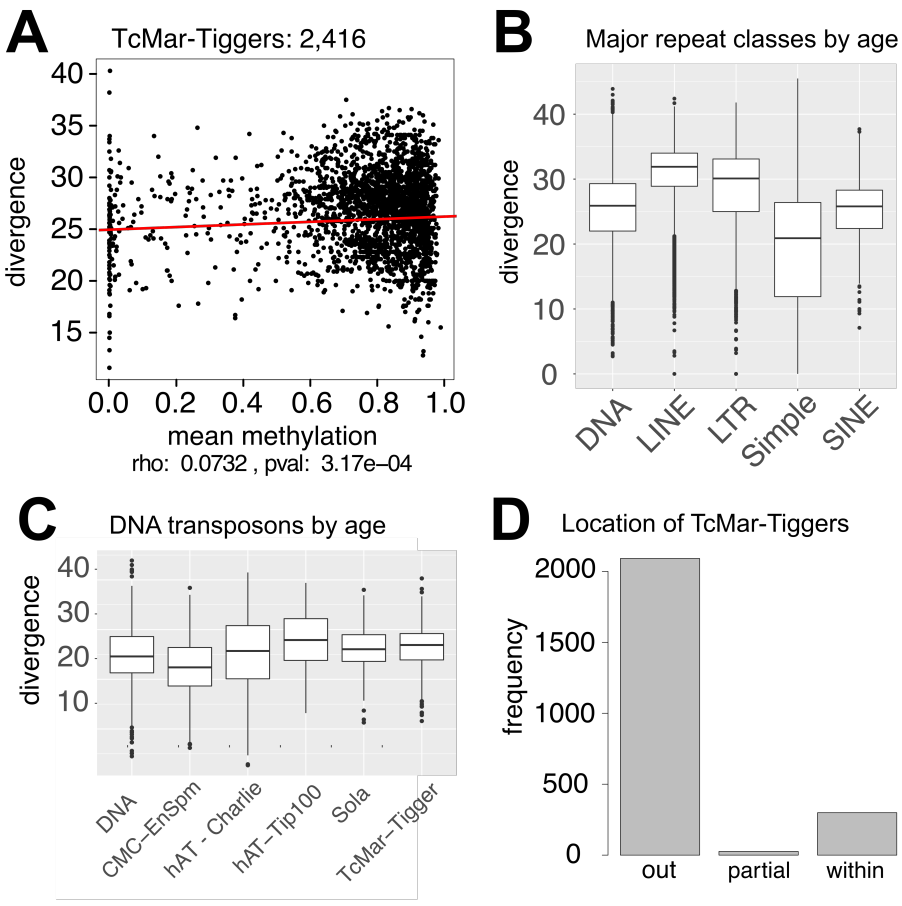

Figure S4: **Analysis of TcMar-Tigger methylation with respect to location and age.** (A) Correlation of methylation of TcMar-Tiggers with respect to their evolutionary divergence as reported by the MAKER pipeline. (B) Major repeat classes and their divergence. (C) Major DNA transposon classes, including TcMar-Tiggers, by divergence. (D) Location of TcMar-Tiggers with respect to genes.

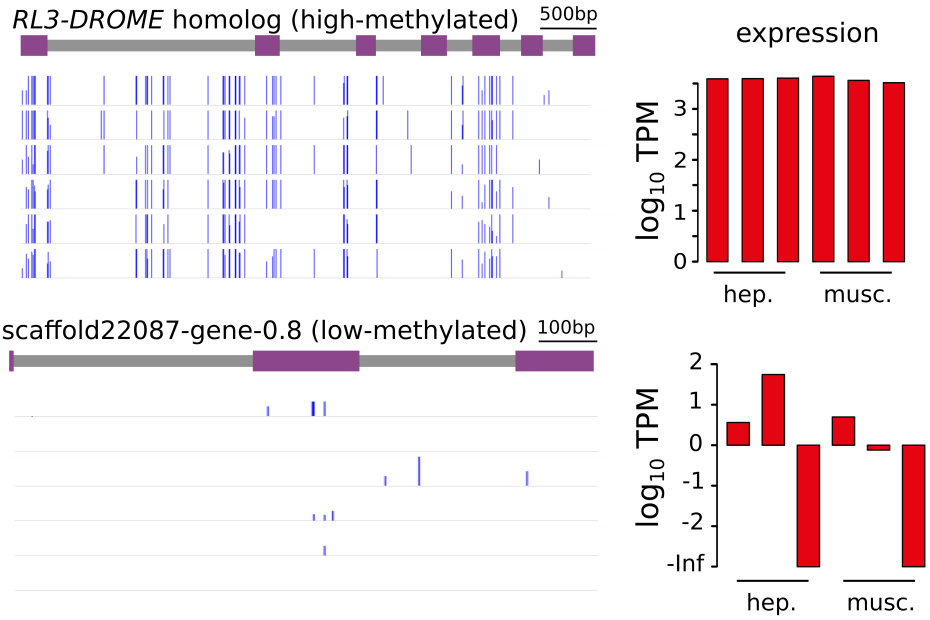

Figure S5: **Additional examples for the relationship between methylation and expression variation.** Top panel: another highly methylated gene (left) with stable expression across samples and tissues (right). Bottom panel: another lowly methylated gene (left) with high gene expression variability. Expression log10 TPM values of -3 indicate zero expression.

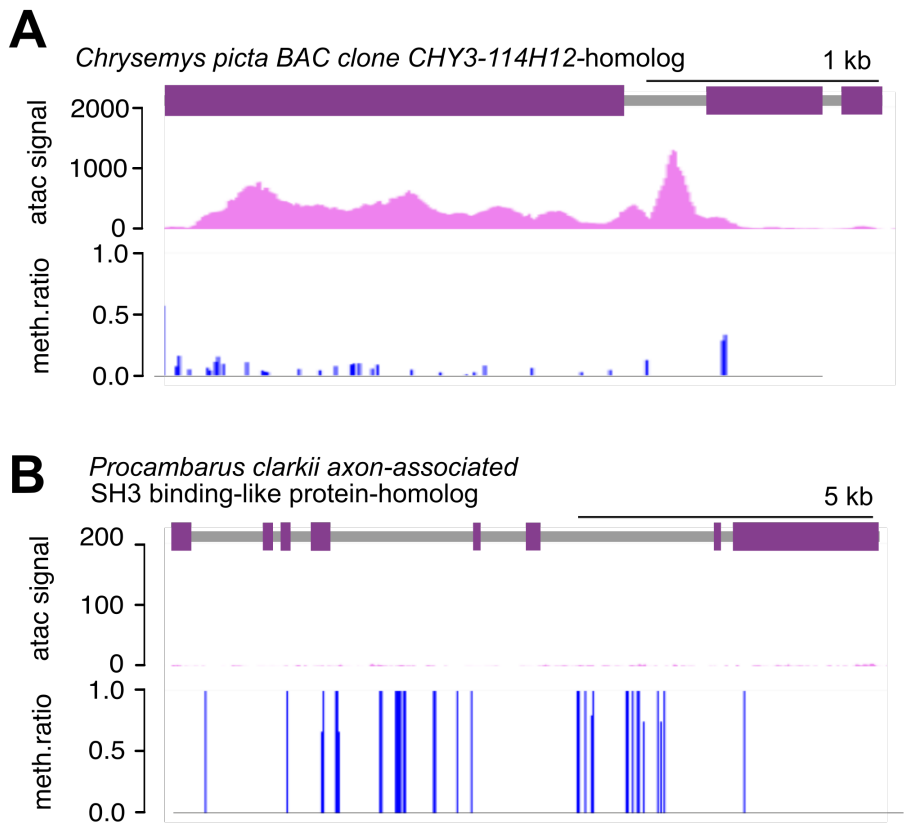

Figure S6: **Genome browser tracks of chromatin accessibility for a low-, and a high-methylated gene.** (A) Gene with low-intermediate meth and high accessibilty. (B) Gene with high gene body methylation levels and limited chromatin accessibility.

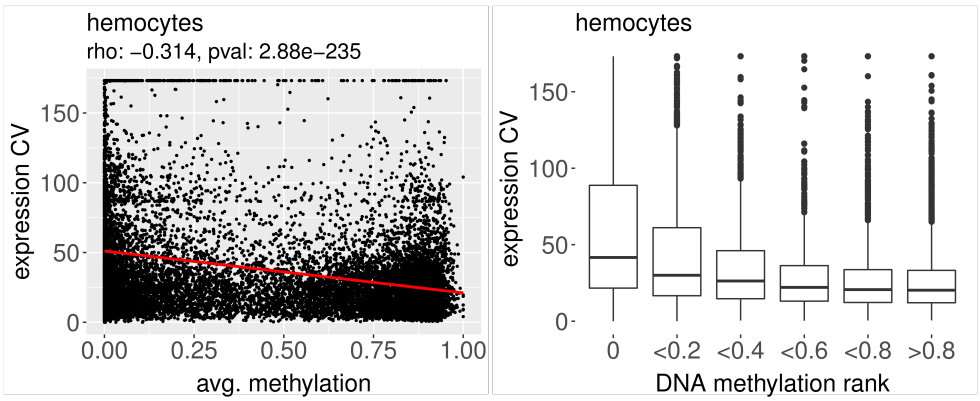

Figure S7: **Correlation of methylation and expression variation in hemocytes.** The negative correlation between gene body methylation and the coefficient of variation of expression levels is conserved in hemocytes. Left panel, scatterplot with regression line for methylation and expression coefficient of variation (expression CV). Correlation coefficient rho and p-value are indicated. Right panel, boxplot showing the same relationship.

List of Figures

List of Tables

List of Abbreviations

5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
6mA	N6-methyladenine
ATAC-seq	assay for transposase-accessible chromatin sequencing
BLAST	Basic Local Alignment Search Tool
CpG	cytosine and guanine in 5' to 3' direction, linked by a phosphate
CPU	Central processing unit
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DRM	domains rearranged methyltransferase
CHG	cytosine-non-guanine-guanine trinucleotide in 5' to 3' direction
CHH	cytosine-non-guanine-non-guanine in 5' to 3' direction
TE	transposable element
H2A, H2B, H3 and H4	histones H2A, 2B, H3 and H4
H3K27me	methylation of lysine on position 27 of histone H3
H3K4me3	trimethylation of lysine on position 4 of histone H3
H3K36me3	trimethylation of lysine on position 36 of histone H3
H3K9me2/3	methylation and dimethylation of lysine 9 on histone H3
HPC	high performance computing
kbp	kilo base pair
LINEs	long interspersed nuclear elements
LTRs	long terminal repeats
mbp	mega base pair
mESC	mouse embryonic stem cell
MDS	multi-dimensional scaling
PCA	principal components analysis
RNA-seq	RNA-sequencing
SAM	S-adenyl methionine
SINE	short interspersed nuclear element
SMRT	single molecule real time (sequencing)
TET	ten-eleven translocation methylcytosine dioxygenase
TPM	Transcripts per Kilobase Million
TSS	transcription start site
TTS	transcription termination site
UTR	untranslated region
WGBS	whole-genome bisulfite sequencing
WGS	whole-genome sequencing

References

S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–10, oct 1990. ISSN 0022-2836. doi: 10.1016/S0022-2836(05)80360-2. URL http://www.ncbi.nlm.nih.gov/pubmed/2231712.

Frederike Alwes and Gerhard Scholtz. Stages and other aspects of the embryology of the parthenogenetic Marmorkrebs (Decapoda, Reptantia, Astacida). *Development Genes and Evolution*, 216(4):169–184, 2006. ISSN 0949944X. doi: 10.1007/s00427-005-0041-8.

Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, 2010. ISSN 1474-760X. doi: 10.1186/gb-2010-11-10-r106. URL http://www.ncbi.nlm.nih.gov/pubmed/20979621http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3218662.

Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2):166–9, jan 2015. ISSN 1367-4811. doi: 10.1093/bioinformatics/btu638. URL http://www.ncbi.nlm.nih.gov/pubmed/25260700http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4287950.

S Andrews. FastQC: : a quality control tool for high throughput sequence data., 2010.

A. A. Aravin, G. J. Hannon, and J. Brennecke. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science*, 318(5851):761–764, nov 2007. ISSN 0036-8075. doi: 10.1126/science.1146484. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1146484.

Jana Asselman, Dieter I. M. De Coninck, Michael E. Pfrender, and Karel A. C. De Schamphelaere. Gene Body Methylation Patterns in Daphnia Are Associated with Gene Family Size. *Genome Biology and Evolution*, 8(4):1185–1196, apr 2016. ISSN 1759-6653. doi: 10.1093/gbe/evw069. URL https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evw069.

Jana Asselman, Dieter IM De Coninck, Eline Beert, Colin R. Janssen, Luisa Orsini, Michael E. Pfrender, Ellen Decaestecker, and Karel AC De Schamphelaere. Bisulfite Sequencing with Daphnia Highlights a Role for Epigenetics in Regulating Stress Response to Microcystis through Preferential Differential Methylation of Serine and Threonine Amino Acids. *Environmental Science & Technology*, 51(2):924–931, jan 2017. ISSN 0013-936X. doi: 10.1021/acs.est.6b03870. URL http://pubs.acs.org/doi/10.1021/acs.est.6b03870.

Madeleine P Ball, Jin Billy Li, Yuan Gao, Je-Hyuk Lee, Emily M LeProust, In-Hyun Park, Bin Xie, George Q Daley, and George M Church. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, 27(4):361–368, apr 2009. ISSN 1087-0156. doi: 10.1038/nbt.1533. URL http://www.nature.com/articles/nbt.1533.

Andrew J Bannister and Tony Kouzarides. Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–95, mar 2011. ISSN 1748-7838. doi: 10.1038/cr.2011.22. URL http://www.ncbi.nlm.nih.gov/pubmed/21321607http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3193420.

Hubertus J. E. Beaumont, Jenna Gallie, Christian Kost, Gayle C. Ferguson, and Paul B. Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–93, nov 2009. ISSN 0028-0836. doi: 10.1038/nature08504. URL http://www.nature.com/articles/nature08504.

Bradley E. Bernstein, Alexander Meissner, and Eric S. Lander. The Mammalian Epigenome. *Cell*, 128(4):669–681, feb 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.01.033. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867407001286.

T Bestor, A Laudano, R Mattaliano, and V Ingram. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology*, 203(4):971–83, oct 1988. ISSN 0022-2836. URL http://www.ncbi.nlm.nih.gov/pubmed/3210246.

Adam J Bewick and Robert J Schmitz. Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, 36:103–110, apr 2017. ISSN 13695266. doi: 10.1016/j.pbi.2016.12.007. URL https://linkinghub.elsevier.com/retrieve/pii/S1369526616301297.

Adam J. Bewick, Kevin J. Vogel, Allen J. Moore, and Robert J. Schmitz. Evolution of DNA Methylation across Insects. *Molecular Biology and Evolution*, page msw264, dec 2016. ISSN 0737-4038. doi: 10.1093/molbev/msw264. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw264.

A. Bird. DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1):6–21, jan 2002. ISSN 08909369. doi: 10.1101/gad.947102. URL http://www.genesdev.org/cgi/doi/10.1101/gad.947102.

A Bird, M Taggart, M Frommer, O J Miller, and D Macleod. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1):91–9, jan 1985. ISSN 0092-8674. URL http://www.ncbi.nlm.nih.gov/pubmed/2981636.

A P Bird. DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research*, 8(7):1499–504, apr 1980. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pubmed/6253938http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC324012.

Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15):2114–20, aug 2014. ISSN 1367-4811. doi: 10.1093/bioinformatics/btu170. URL http://www.ncbi.nlm.nih.gov/pubmed/24695404http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4103590.

Roberto Bonasio, Qiye Li, Jinmin Lian, Navdeep S. Mutti, Lijun Jin, Hongmei Zhao, Pei Zhang, Ping Wen, Hui Xiang, Yun Ding, Zonghui Jin, Steven S. Shen, Zongji Wang, Wen Wang, Jun Wang, Shelley L. Berger, Jürgen Liebig, Guojie Zhang, and Danny Reinberg. Genome-wide and Caste-Specific DNA Methylomes of the Ants Camponotus floridanus and Harpegnathos saltator. *Current Biology*, 22(19):1755–1764, oct 2012. ISSN 09609822. doi: 10.1016/j.cub.2012.07.042. URL http://linkinghub.elsevier.com/retrieve/pii/S0960982212008676.

Achim Breiling and Frank Lyko. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & Chromatin*, 8(1):24, dec 2015. ISSN 1756-8935. doi: 10.1186/s13072-015-0016-6. URL http://www.epigeneticsandchromatin.com/content/8/1/24.

Broadinstitute. Picard tools.

Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12):1213–1218, dec 2013. ISSN 1548-7091. doi: 10.1038/nmeth.2688. URL http://www.nature.com/articles/nmeth.2688.

Simon W.-L. Chan, Ian R. Henderson, and Steven E. Jacobsen. Erratum: Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nature Reviews Genetics*, 6(5):351–360, may 2005. ISSN 1471-0056. doi: 10.1038/nrg1601. URL http://www.nature.com/articles/nrg1601.

Christoph Chucholl, Katharina Morawetz, and Harald Groß. The clones are coming – strong increase in Marmorkrebs [Procambarus fallax (Hagen, 1870) f. virginalis] records from Europe. *Aquatic Invasions*, 7(4):511–519, nov 2012. ISSN 18185487. doi: 10.3391/ai.2012.7.4.008. URL http://www.aquaticinvasions.net/2012/issue4.html.

Stephen J. Clark, Ricard Argelaguet, ChantriolInt-Andreas Kapourani, Thomas M. Stubbs, Heather J. Lee, Celia Alda-Catalinas, Felix Krueger, Guido Sanguinetti, Gavin Kelsey, John C. Marioni, Oliver Stegle, and Wolf Reik. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nature Communications*, 9(1):781, dec 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-03149-4. URL http://www.nature.com/articles/s41467-018-03149-4.

Shawn J. Cokus, Suhua Feng, Xiaoyu Zhang, Zugen Chen, Barry Merriman, Christian D. Haudenschild, Sriharsa Pradhan, Stanley F. Nelson, Matteo Pellegrini, and Steven E. Jacobsen. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452(7184):215–219, mar 2008. ISSN 0028-0836. doi: 10.1038/nature06745. URL http://www.nature.com/articles/nature06745.

S J Compere and R D Palmiter. DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell*, 25(1):233–40, jul 1981. ISSN 0092-8674. URL http://www.ncbi.nlm.nih.gov/pubmed/6168387.

C Coulondre, J H Miller, P J Farabaugh, and W Gilbert. Molecular basis of base substitution hotspots in Escherichia coli. *Nature*, 274(5673):775–80, aug 1978. ISSN 0028-0836. URL http://www.ncbi.nlm.nih.gov/pubmed/355893.

Christopher B. Cunningham, Lexiang Ji, R. Axel W. Wiberg, Jennifer Shelton, Elizabeth C. McKinney, Darren J. Parker, Richard B. Meagher, Kyle M. Benowitz, Eileen M. Roy-Zokan, Michael G. Ritchie, Susan J. Brown, Robert J. Schmitz, and Allen J. Moore. The Genome and Methylome of a Beetle with Complex Social Behavior, Nicrophorus vespilloides (Coleoptera: Silphidae). *Genome Biology and Evolution*, 7(12): 3383–3396, dec 2015. ISSN 1759-6653. doi: 10.1093/gbe/evv194. URL https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evv194.

Alan Deidun, Arnold Sciberras, Justin Formosa, Bruno Zava, Gianni Insacco, Maria Corsini-Foka, and Keith A Crandall. Invasion by non-indigenous freshwater decapods of Malta and Sicily, central Mediterranean Sea. *Journal of Crustacean Biology*, sep 2018. ISSN 0278-0372. doi: 10.1093/jcbiol/ruy076. URL https://academic.oup.com/jcb/advance-article/doi/10.1093/jcbiol/ruy076/5096909.

D. C. Dolinoy, D. Huang, and R. L. Jirtle. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences*, 104(32):13056–13061, aug 2007. ISSN 0027-8424. doi: 10.1073/pnas.0703739104. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0703739104.

R. H. Dowen, M. Pelizzola, R. J. Schmitz, R. Lister, J. M. Dowen, J. R. Nery, J. E. Dixon, and J. R. Ecker. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences*, 109(32):E2183–E2191, aug 2012. ISSN 0027-8424. doi: 10.1073/pnas.1209329109. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1209329109.

Manu J Dubin, Pei Zhang, Dazhe Meng, Marie-Stanislas Remigereau, Edward J Osborne, Francesco Paolo Casale, Philipp Drewe, André Kahles, Geraldine Jean, Bjarni Vilhjálmsson, Joanna Jagoda, Selen Irez, Viktor Voronin, Qiang Song, Quan Long, Gunnar Rätsch, Oliver Stegle, Richard M Clark, and Magnus Nordborg. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife*, 4, may 2015. ISSN 2050-084X. doi: 10.7554/eLife.05255. URL https://elifesciences.org/articles/05255.

Elizabeth J. Duncan, Peter D. Gluckman, and Peter K. Dearden. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 322(4):208–220, jun 2014. ISSN 15525007. doi: 10.1002/jez.b.22571. URL http://doi.wiley.com/10.1002/jez.b.22571.

Eli Eisenberg and Erez Y. Levanon. A-to-I RNA editing — immune protector and transcriptome diversifier. *Nature Reviews Genetics*, 19(8):473–490, aug 2018. ISSN 1471-0056. doi: 10.1038/s41576-018-0006-1. URL http://www.nature.com/articles/s41576-018-0006-1.

M. B. Elowitz. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584): 1183–1186, aug 2002. ISSN 00368075. doi: 10.1126/science.1070919. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1070919.

C. Falckenhayn, B. Boerjan, G. Raddatz, M. Frohme, L. Schoofs, and F. Lyko. Characterization of genome methylation patterns in the desert locust Schistocerca gregaria. *Journal of Experimental Biology*, 216(8):1423–1429, apr 2013. ISSN 0022-0949. doi: 10.1242/jeb.080754. URL http://jeb.biologists.org/cgi/doi/10.1242/jeb.080754.

Cassandra Falckenhayn. *The Methylome of the Marbled Crayfish Procambarus virginalis*. PhD thesis, Ruperto-Carola University of Heidelberg, 2016.

Andre J. Faure, Jörn M. Schmiedel, and Ben Lehner. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Systems*, 5 (5):471–484.e4, nov 2017. ISSN 24054712. doi: 10.1016/j.cels.2017.10.003. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471217304404.

A. P. Feinberg and R. A. Irizarry. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proceedings of the National Academy of Sciences*, 107(suppl_1):1757–1764, jan 2010. ISSN 0027-8424. doi: 10.1073/pnas.0906183107. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0906183107.

S. Feng, S. J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini, and S. E. Jacobsen. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107(19):8689–8694, may 2010. ISSN 0027-8424. doi: 10.1073/pnas.1002720107. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1002720107.

Amanda G. Fisher. Cellular identity and lineage choice. *Nature Reviews Immunology*, 2(12):977–982, dec 2002. ISSN 1474-1733. doi: 10.1038/nri958. URL http://www.nature.com/articles/nri958.

M Frommer, L E McDonald, D S Millar, C M Collis, F Watt, G W Grigg, P L Molloy, and C L Paul. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5):1827–31, mar 1992. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/pubmed/1542678http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC48546.

Fanny Gatzmann and Frank Lyko. Whole-genome bisulfite sequencing for the methylation analysis of insect genomes. *Methods Mol. Biol.*, in press, 2018.

Fanny Gatzmann, Cassandra Falckenhayn, Julian Gutekunst, Katharina Hanna, Günter Raddatz, Vitor Coutinho Carneiro, and Frank Lyko. The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. *Epigenetics & Chromatin*, 11(1):57, dec 2018. ISSN 1756-8935. doi: 10.1186/s13072-018-0229-6. URL https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-018-0229-6.

Karl M Glastad, Kaustubh Gokhale, Jürgen Liebig, and Michael A D Goodisman. The caste- and sex-specific DNA methylome of the termite Zootermopsis nevadensis. *Scientific reports*, 6:37110, nov 2016a. ISSN 2045-2322. doi: 10.1038/srep37110. URL http://www.ncbi.nlm.nih.gov/pubmed/27848993.

Karl M. Glastad, Michael A. D. Goodisman, Soojin V. Yi, and Brendan G. Hunt. Effects of DNA Methylation and Chromatin State on Rates of Molecular Evolution in Insects. *G3: Genes|Genomes|Genetics*, 6(2):357–363, feb 2016b. ISSN 2160-1836. doi: 10.1534/g3.115.023499. URL http://g3journal.org/lookup/doi/10.1534/g3.115.023499.

Aaron D. Goldberg, C. David Allis, and Emily Bernstein. Epigenetics: A Landscape Takes Shape. *Cell*, 128(4):635–638, feb 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.02.006. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867407001869.

M. G. Goll. Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science*, 311(5759):395–398, jan 2006. ISSN 0036-8075. doi: 10.1126/science.1120976. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1120976.

Mary Grace Goll and Timothy H. Bestor. EUKARYOTIC CYTOSINE METHYLTRANSFERASES. *Annual Review of Biochemistry*, 74(1):481–514, jun 2005. ISSN 0066-4154. doi: 10.1146/annurev.biochem.74.010904.153721. URL http://www.annualreviews.org/doi/10.1146/annurev.biochem.74.010904.153721.

Annika Grimmer. Analysis of DNA methylation dynamics during Marmorkrebs (Procambarus fallax forma virginalis) development. *Lyko Lab Master Thesis*, 6(2):100, 2015. ISSN 0178-7888. doi: 10.1007/BF03192151. URL http://www.ncbi.nlm.nih.gov/pubmed/21365933.

Julian Gutekunst, Ranja Andriantsoa, Cassandra Falckenhayn, Katharina Hanna, Wolfgang Stein, Jeanne Rasamy, and Frank Lyko. Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nature Ecology & Evolution*, 2(3):567–573, mar 2018. ISSN 2397-334X. doi: 10.1038/s41559-018-0467-9. URL http://www.nature.com/articles/s41559-018-0467-9.

A. Noble Hendrix and William F Loftus. Distribution and relative abundance of the crayfishes Procambarus alleni (Faxon) and P. fallax (Hagen) in Southern Florida. *Wetlands*, 20(1):194–199, 2000.

Hendrik Herman Johann Hobbs. The crayfishes of Georgia. *Smithson Contrib Zool.*, 318: 1–549, 1981.

R Holliday and J E Pugh. DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.)*, 187(4173):226–32, jan 1975. ISSN 0036-8075. URL http://www.ncbi.nlm.nih.gov/pubmed/1111098.

Carson Holt and Mark Yandell. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, 12: 491, dec 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-491. URL http:

//www.ncbi.nlm.nih.gov/pubmed/22192575http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3280279.

R D Hotchkiss. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry*, 175(1):315–32, aug 1948. ISSN 0021-9258. URL http://www.ncbi.nlm.nih.gov/pubmed/18873306.

Yun Huang, William A. Pastor, Yinghua Shen, Mamta Tahiliani, David R. Liu, and Anjana Rao. The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS ONE*, 5 (1):e8888, jan 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0008888. URL http://dx.plos.org/10.1371/journal.pone.0008888.

Iksoo Huh, Jia Zeng, Taesung Park, and Soojin V Yi. DNA methylation and transcriptional noise. *Epigenetics & Chromatin*, 6(1):9, 2013. ISSN 1756-8935. doi: 10.1186/1756-8935-6-9. URL http://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/1756-8935-6-9.

Rudolf Jaenisch and Adrian Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33(3s):245–254, mar 2003. ISSN 10614036. doi: 10.1038/ng1089. URL http://www.nature.com/doifinder/10.1038/ng1089.

Julia P G Jones, Jeanne R. Rasamy, Andrew Harvey, Alicia Toon, Birgit Oidtmann, Michele H. Randrianarison, Noromalala Raminosoa, and Olga R. Ravoahangimalala. The perfect invader: A parthenogenic crayfish poses a new threat to Madagascar's freshwater biodiversity. *Biological Invasions*, 11:1475–1482, 2009. ISSN 13873547. doi: 10.1007/s10530-008-9334-y.

Peter A. Jones. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7):484–492, may 2012. ISSN 1471-0056. doi: 10.1038/nrg3230. URL http://www.nature.com/doifinder/10.1038/nrg3230.

Damian Kao, Alvina G Lai, Evangelia Stamataki, Silvana Rosic, Nikolaos Konstantinides, Erin Jarvis, Alessia Di Donfrancesco, Natalia Pouchkina-Stantcheva, Marie Semon, Marco Grillo, Heather Bruce, Suyash Kumar, Igor Siwanowicz, Andy Le, Andrew Lemire, Cassandra Extavour, William Browne, Carsten Wolff, Michalis Averof, Nipam H Patel, Peter Sarkies, Anastasios Pavlopoulos, and Aziz Aboobaker. The genome of the crustacean Parhyale hawaiensis : a model for animal development, regeneration, immunity and lignocellulose digestion. *bioRxiv*, page 065789, 2016. ISSN 2050-084X. doi: 10.1101/065789. URL http://biorxiv.org/lookup/doi/10.1101/065789.

Taiji Kawakatsu, Shao-shan Carol Huang, Florian Jupe, Eriko Sasaki, Robert J. Schmitz, Mark A. Urich, Rosa Castanon, Joseph R. Nery, Cesar Barragan, Yupeng He, Huaming Chen, Manu Dubin, Cheng-Ruei Lee, Congmao Wang, Felix Bemm, Claude Becker, Ryan O'Neil, Ronan C. O'Malley, Danjuma X. Quarless, Nicholas J. Schork, Detlef Weigel, Magnus Nordborg, Joseph R. Ecker, Carlos Alonso-Blanco, Jorge Andrade, Claude Becker, Felix Bemm, Joy Bergelson, Karsten Borgwardt, Eunyoung Chae, Todd Dezwaan, Wei Ding, Joseph R. Ecker, Moisés Expósito-Alonso, Ashley Farlow, Joffrey Fitz, Xiangchao Gan, Dominik G. Grimm, Angela Hancock, Stefan R. Henz, Svante

Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, Pamela Korte, Christa Lanz, Chen-Ruei Lee, Dazhe Meng, Todd P. Michael, Richard Mott, Ni Wayan Muliyati, Thomas Nägele, Matthias Nagler, Viktoria Nizhynska, Magnus Nordborg, Polina Novikova, F. Xavier Picó, Alexander Platzer, Fernando A. Rabanal, Alex Rodriguez, Beth A. Rowan, Patrice A. Salomé, Karl Schmid, Robert J. Schmitz, Ümit Seren, Felice Gianluca Sperone, Mitchell Sudkamp, Hannes Svardal, Matt M. Tanzer, Donald Todd, Samuel L. Volchenboum, Congmao Wang, George Wang, Xi Wang, Wolfram Weckwerth, Detlef Weigel, and Xuefeng Zhou. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell*, 166(2):492–505, jul 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.06.044. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867416308522.

Thomas E. Keller, Priscilla Han, and Soojin V. Yi. Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate–Vertebrate Boundary. *Molecular Biology and Evolution*, 33(4):1019–1028, apr 2016. ISSN 0737-4038. doi: 10.1093/molbev/msv345. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msv345.

Carly D. Kenkel and Mikhail V. Matz. Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nature Ecology & Evolution*, 1(1):0014, nov 2016. ISSN 2397-334X. doi: 10.1038/s41559-016-0014. URL http://www.nature.com/articles/s41559-016-0014.

Daehwan Kim, Ben Langmead, and Steven L Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4):357–60, apr 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3317. URL http://www.ncbi.nlm.nih.gov/pubmed/25751142http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4655817.

Felix Krueger. Trim Galore!, 2012.

Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–9, mar 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923. URL http://www.ncbi.nlm.nih.gov/pubmed/22388286http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3322381.

Julie A. Law and Steven E. Jacobsen. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3):204–220, mar 2010. ISSN 1471-0056. doi: 10.1038/nrg2719. URL http://www.nature.com/articles/nrg2719.

Amanda J. Lea, Tauras P. Vilgalys, Paul A. P. Durst, and Jenny Tung. Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nature Ecology & Evolution*, 1(8):1074–1083, aug 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0229-0. URL http://www.nature.com/articles/s41559-017-0229-0.

Gerald A. LeBlanc. Crustacean endocrine toxicology: a review. *Ecotoxicology*, 16(1):61–81, feb 2007. ISSN 0963-9292. doi: 10.1007/s10646-006-0115-z. URL http://link.springer.com/10.1007/s10646-006-0115-z.

Carine Legrand, Francesca Tuorto, Mark Hartmann, Reinhard Liebers, Dominik Jacob, Mark Helm, and Frank Lyko. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Research*, 27(9):1589–1596, sep 2017. ISSN 1088-9051. doi: 10.1101/gr.210666.116. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.210666.116.

Bo Li and Colin N Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12:323, aug 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-323. URL http://www.ncbi.nlm.nih.gov/pubmed/21816040http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3163565.

E Li, T H Bestor, and R Jaenisch. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–26, jun 1992. ISSN 0092-8674. URL http://www.ncbi.nlm.nih.gov/pubmed/1606615.

En Li and Yi Zhang. DNA methylation in mammals. *Cold Spring Harbor perspectives in biology*, 6(5):a019133, may 2014. ISSN 1943-0264. doi: 10.1101/cshperspect.a019133. URL http://www.ncbi.nlm.nih.gov/pubmed/24789823http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3996472.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–9, aug 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp352. URL http://www.ncbi.nlm.nih.gov/pubmed/19505943http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2723002http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723002{\&}tool=pmcentrez{\&}rendertype=abstract.

Romain Libbrecht, Peter Robert Oxley, Laurent Keller, and Daniel Jan Christoph Kronauer. Robust DNA Methylation in the Clonal Raider Ant Brain. *Current Biology*, 26(3):391–395, feb 2016. ISSN 09609822. doi: 10.1016/j.cub.2015.12.040. URL https://linkinghub.elsevier.com/retrieve/pii/S0960982215015717.

Boris Lipták, Agata Mrugała, Ladislav Pekárik, Anton Mutkovič, Daniel Gruľa, Adam Petrusek, and Antonín Kouba. Expansion of the marbled crayfish in Slovakia: beginning of an invasion in the Danube catchment? *Journal of Limnology*, jan 2016. ISSN 1723-8633. doi: 10.4081/jlimnol.2016.1313. URL http://www.jlimnol.it/index.php/jlimnol/article/view/jlimnol.2016.1313.

R. Lister and J. R. Ecker. Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Research*, 19(6):959–966, jun 2009. ISSN 1088-9051. doi: 10.1101/gr.083451.108. URL http://genome.cshlp.org/cgi/doi/10.1101/gr.083451.108.

Matthew C Lorincz, David R Dickerson, Mike Schmitt, and Mark Groudine. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural & Molecular Biology*, 11(11):1068–1075, nov 2004. ISSN 1545-9993. doi: 10.1038/nsmb840. URL http://www.nature.com/articles/nsmb840.

Frank Lyko. The marbled crayfish (Decapoda: Cambaridae) represents an independent new species. *Zootaxa*, 4363(4):544–552, dec 2017a. ISSN 1175-5334. doi: 29245391. URL http://www.ncbi.nlm.nih.gov/pubmed/29245391.

Frank Lyko. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2):81–92, oct 2017b. ISSN 1471-0056. doi: 10.1038/nrg.2017.80. URL http://www.nature.com/doifinder/10.1038/nrg.2017.80.

Frank Lyko, Sylvain Foret, Robert Kucharski, Stephan Wolf, Cassandra Falckenhayn, and Ryszard Maleszka. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biology*, 8(11), 2010. ISSN 15449173. doi: 10.1371/journal.pbio.1000506.

Andor Lőkkös, Tamás Müller, Krisztián Kovács, Levente Várkonyi, András Specziár, and Peer Martin. The alien, parthenogenetic marbled crayfish (Decapoda: Cambaridae) is entering Kis-Balaton (Hungary), one of Europe's most important wetland biotopes. *Knowledge and Management of Aquatic Ecosystems*, (417):16, may 2016. ISSN 1961-9502. doi: 10.1051/kmae/2016003. URL http://www.kmae-journal.org/10.1051/kmae/2016003.

Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, 17(1):10–12, 2011.

Peer Martin, Klaus Kohlmann, and Gerhard Scholtz. The parthenogenetic Marmorkrebs (marbled crayfish) produces genetically uniform offspring. *Die Naturwissenschaften*, 94(10):843–6, 2007. ISSN 0028-1042. doi: 10.1007/s00114-007-0260-0. URL http://www.ncbi.nlm.nih.gov/pubmed/17541537.

Peer Martin, Nathan J. Dorn, Tadashi Kawai, Craig van der Heiden, and Gerhard Scholtz. The enigmatic Marmorkrebs (marbled crayfish) is the parthenogenetic form of Procambarus fallax (Hagen, 1870). *Contributions to Zoology*, 79(3):107–118, 2010. ISSN 13834517.

Peer Martin, Sven Thonagel, and Gerhard Scholtz. The parthenogenetic Marmorkrebs (Malacostraca: Decapoda: Cambaridae) is a triploid organism. *Journal of Zoological Systematics and Evolutionary Research*, (September):n/a–n/a, 2015. ISSN 09475745. doi: 10.1111/jzs.12114. URL http://doi.wiley.com/10.1111/jzs.12114.

Alexander Meissner, Andreas Gnirke, George W. Bell, Bernard Ramsahoye, Eric S. Lander, and Rudolf Jaenisch. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18):5868–5877, 2005. ISSN 03051048. doi: 10.1093/nar/gki901.

Huaiyu Mi, Anushya Muruganujan, and Paul D. Thomas. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1):D377–D386, nov 2012. ISSN 0305-1048. doi: 10.1093/nar/gks1118. URL http://academic.oup.com/nar/article/41/D1/D377/1060482/PANTHER-in-2013-modeling-the-evolution-of-gene.

Huaiyu Mi, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D. Thomas. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, 45(D1):D183–D189, jan 2017. ISSN 0305-1048. doi: 10.1093/nar/gkw1138. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1138.

Francesco Neri, Stefania Rapelli, Anna Krepelova, Danny Incarnato, Caterina Parlato, Giulia Basile, Mara Maldotti, Francesca Anselmi, and Salvatore Oliviero. Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, 543(7643):72–77, mar 2017. ISSN 0028-0836. doi: 10.1038/nature21373. URL http://www.nature.com/articles/nature21373.

John R. S. Newman, Sina Ghaemmaghami, Jan Ihmels, David K. Breslow, Matthew Noble, Joseph L. DeRisi, and Jonathan S. Weissman. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–846, jun 2006. ISSN 0028-0836. doi: 10.1038/nature04785. URL http://www.nature.com/articles/nature04785.

Masaki Okano, Daphne W Bell, Daniel A Haber, and En Li. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99(3):247–257, oct 1999. ISSN 00928674. doi: 10.1016/S0092-8674(00)81656-6. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867400816566.

Nuala A. O'Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O'Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, and Kim D. Pruitt. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745, jan 2016. ISSN 0305-1048. doi: 10.1093/nar/gkv1189. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1189.

Vicki Pandey, Robert C. Nutter, and Ellen Prediger. Applied Biosystems SOLiD™ System: Ligation-Based Sequencing. In *Next Generation Genome Sequencing*, pages 29–42. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527625130.ch3. URL http://doi.wiley.com/10.1002/9783527625130.ch3.

Lucian Pârvulescu, Andrei Togor, Sandra-Florina Lele, Sebastian Scheu, Daniel Șinca, and Jörn Panteleit. First established population of marbled crayfish Procambarus fallax

(Hagen, 1870) f. virginalis (Decapoda, Cambaridae) in Romania. *BioInvasions Records*, 6(4):357–362, 2017. ISSN 22421300. doi: 10.3391/bir.2017.6.4.09. URL http://www.reabic.net/journals/bir/2017/Issue4.aspx.

Solenn Patalano, Anna Vlasova, Chris Wyatt, Philip Ewels, Francisco Camara, Pedro G Ferreira, Claire L Asher, Tomasz P Jurkowski, Anne Segonds-Pichon, Martin Bachman, Irene González-Navarrete, André E Minoche, Felix Krueger, Ernesto Lowy, Marina Marcet-Houben, Jose Luis Rodriguez-Ales, Fabio S Nascimento, Shankar Balasubramanian, Toni Gabaldon, James E Tarver, Simon Andrews, Heinz Himmelbauer, William O H Hughes, Roderic Guigó, Wolf Reik, and Seirian Sumner. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45):13970–5, nov 2015. ISSN 1091-6490. doi: 10.1073/pnas.1515937112. URL http://www.ncbi.nlm.nih.gov/pubmed/26483466http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4653166.

Jiří Patoka, Miloš Buřič, Vojtěch Kolář, Martin Bláha, Miloslav Petrtýl, Pavel Franta, Robert Tropek, Lukáš Kalous, Adam Petrusek, and Antonín Kouba. Predictions of marbled crayfish establishment in conurbations fulfilled: Evidences from the Czech Republic. *Biologia*, 71(12), jan 2016. ISSN 0006-3088. doi: 10.1515/biolog-2016-0164. URL https://www.degruyter.com/view/j/biolog.2016.71.issue-12/biolog-2016-0164/biolog-2016-0164.xml.

Veronica J Peschansky and Claes Wahlestedt. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics*, 9(1):3–12, jan 2014. ISSN 1559-2294. doi: 10.4161/epi.27473. URL http://www.tandfonline.com/doi/abs/10.4161/epi.27473.

J Pósfai, A S Bhagwat, G Pósfai, and R J Roberts. Predictive motifs derived from cytosine methyltransferases. *Nucleic acids research*, 17(7):2421–35, apr 1989. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pubmed/2717398http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC317633.

R Development Core Team. R: A Language and Environment for Statistical Computing. *http://www.r-project.org*, 2013.

G. Raddatz, P. M. Guzzardo, N. Olova, M. R. Fantappie, M. Rampp, M. Schaefer, W. Reik, G. J. Hannon, and F. Lyko. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences*, 110(21):8627–8631, may 2013. ISSN 0027-8424. doi: 10.1073/pnas.1306723110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1306723110.

E. J. Radford, M. Ito, H. Shi, J. A. Corish, K. Yamazawa, E. Isganaitis, S. Seisenberger, T. A. Hore, W. Reik, S. Erkek, A. H. F. M. Peters, M.-E. Patti, and A. C. Ferguson-Smith. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science*, 345(6198):1255903–1255903, aug 2014. ISSN 0036-8075. doi: 10.1126/science.1255903. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1255903.

V. K. Rakyan, T. A. Down, N. P. Thorne, P. Flicek, E. Kulesha, S. Graf, E. M. Tomazou, L. Backdahl, N. Johnson, M. Herberth, K. L. Howe, D. K. Jackson, M. M. Miretti, H. Fiegler, J. C. Marioni, E. Birney, T. J.P. Hubbard, N. P. Carter, S. Tavare, and S. Beck. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Research*, 18(9):1518–1529, jul 2008. ISSN 1088-9051. doi: 10.1101/gr.077479.108. URL http://www.genome.org/cgi/doi/10.1101/gr.077479.108.

S. Ramchandani, S. K. Bhattacharya, N. Cervoni, and M. Szyf. DNA methylation is a reversible biological signal. *Proceedings of the National Academy of Sciences*, 96(11): 6107–6112, may 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.11.6107. URL http://www.pnas.org/cgi/doi/10.1073/pnas.96.11.6107.

Oliver J. Rando and Kevin J. Verstrepen. Timescales of Genetic and Epigenetic Inheritance. *Cell*, 128(4):655–668, 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.01.023.

A.D. Riggs. X inactivation, differentiation, and DNA methylation. *Cytogenetic and Genome Research*, 14(1):9–25, 1975. ISSN 1424-859X. doi: 10.1159/000130315. URL https://www.karger.com/Article/FullText/130315.

Arthur Riggs, Robert Martienssen, and Vincenzo Russo. Epigenetic Mechanisms of Gene Regulation. *Woodbury: Cold Spring Harbor Laboratory Press*, 1996.

Nathan R. Rose and Robert J. Klose. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(12):1362–1372, dec 2014. ISSN 18749399. doi: 10.1016/j.bbagrm.2014.02.007. URL https://linkinghub.elsevier.com/retrieve/pii/S1874939914000285.

Silvana Rošić, Rachel Amouroux, Cristina E. Requena, Ana Gomes, Max Emperle, Toni Beltran, Jayant K. Rane, Sarah Linnett, Murray E. Selkirk, Philipp H. Schiffer, Allison J. Bancroft, Richard K. Grencis, Albert Jeltsch, Petra Hajkova, and Peter Sarkies. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature Genetics*, 50(3):452–459, mar 2018. ISSN 1061-4036. doi: 10.1038/s41588-018-0061-8. URL http://www.nature.com/articles/s41588-018-0061-8.

S. Sarda, J. Zeng, B. G. Hunt, and S. V. Yi. The Evolution of Invertebrate Gene Body Methylation. *Molecular Biology and Evolution*, 29(8):1907–1916, aug 2012. ISSN 0737-4038. doi: 10.1093/molbev/mss062. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss062.

Gerhard Scholtz, Anke Braband, Laura Tolley, André Reimann, Beate Mittmann, Chris Lukhaup, Frank Steuerwald, and Günter Vogt. Parthenogenesis in an outsider crayfish. *Nature*, 421(6925):806–806, feb 2003. ISSN 0028-0836. doi: 10.1038/421806a. URL http://www.nature.com/articles/421806a.

Robert Seitz, Kathia Vilpoux, Ulrich Hopp, Steffen Harzsch, and Gerhard Maier. Ontogeny of the Marmorkrebs (Marbled Cray sh): a Parthenogenetic Cray sh With Unknown Origin and Phylogenetic Position. *Reproduction*, 405(October 2004):393–405, 2005. doi: 10.1002/jez.a.143.394.

M W Simmen, S Leitgeb, J Charlton, S J Jones, B R Harris, V H Clark, and A Bird. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science (New York, N.Y.)*, 283(5405):1164–7, feb 1999. ISSN 0036-8075. URL http://www.ncbi.nlm.nih.gov/pubmed/10024242.

V J Simpson, T E Johnson, and R F Hammen. Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic acids research*, 14(16):6711–9, aug 1986. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pubmed/3748820http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC311675.

A F Smit and A D Riggs. Tiggers and DNA transposon fossils in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1443–8, feb 1996. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/pubmed/8643651http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC39958.

Zachary D Smith and Alexander Meissner. DNA methylation: roles in mammalian development. *Nature reviews. Genetics*, 14(3):204–20, 2013. ISSN 1471-0064. doi: 10.1038/nrg3354. URL http://www.ncbi.nlm.nih.gov/pubmed/23400093.

Zachary D Smith, Michelle M Chan, Tarjei S Mikkelsen, Hongcang Gu, Andreas Gnirke, Aviv Regev, and Alexander Meissner. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 484(7394):339–44, 2012. ISSN 1476-4687. doi: 10.1038/nature10960. URL http://www.nature.com/nature/journal/v484/n7394/pdf/nature10960.pdf.

Miho M. Suzuki and Adrian Bird. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9(6):465–476, jun 2008. ISSN 1471-0056. doi: 10.1038/nrg2341. URL http://www.nature.com/articles/nrg2341.

Miho M Suzuki, Akiko Yoshinari, Madoka Obara, Shohei Takuno, Shuji Shigenobu, Yasunori Sasakura, Alastair Rw Kerr, Shaun Webb, Adrian Bird, and Atsuo Nakayama. Identical sets of methylated and nonmethylated genes in Ciona intestinalis sperm and muscle cells. *Epigenetics & chromatin*, 6(1): 38, 2013. ISSN 1756-8935. doi: 10.1186/1756-8935-6-38. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3827831{\&}tool=pmcentrez{\&}rendertype=abstract.

S. Takuno and B. S. Gaut. Body-Methylated Genes in Arabidopsis thaliana Are Functionally Important and Evolve Slowly. *Molecular Biology and Evolution*, 29(1):219–227, jan 2012. ISSN 0737-4038. doi: 10.1093/molbev/msr188. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr188.

Hisashi Tamaru, Xing Zhang, Debra McMillen, Prim B. Singh, Jun-ichi Nakayama, Shiv I. Grewal, C. David Allis, Xiaodong Cheng, and Eric U. Selker. Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. *Nature Genetics*, 34(1):75–79, may 2003. ISSN 1061-4036. doi: 10.1038/ng1143. URL http://www.nature.com/articles/ng1143.

K. H. Taylor, R. S. Kramer, J. W. Davis, J. Guo, D. J. Duff, D. Xu, C. W. Caldwell, and H. Shi. Ultradeep Bisulfite Sequencing Analysis of DNA Methylation Patterns in Multiple Gene Promoters by 454 Sequencing. *Cancer Research*, 67(18):8511–8518, sep 2007. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-07-1016. URL http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-07-1016.

Peter Tessarz and Tony Kouzarides. Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*, 15(11):703–708, nov 2014. ISSN 1471-0072. doi: 10.1038/nrm3890. URL http://www.nature.com/articles/nrm3890.

Christa Geeke Toenhake, Sabine Anne-Kristin Fraschka, Mahalingam Shanmugiah Vijayabaskar, David Robert Westhead, Simon Jan van Heeringen, and Richárd Bártfai. Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying Plasmodium falciparum Blood-Stage Development. *Cell Host & Microbe*, 23 (4):557–569.e9, apr 2018. ISSN 19313128. doi: 10.1016/j.chom.2018.03.007. URL https://linkinghub.elsevier.com/retrieve/pii/S1931312818301367.

G van Rossum. Python tutorial. *Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam*, 1995.

L Vardimon, A Kressmann, H Cedar, M Maechler, and W Doerfler. Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proceedings of the National Academy of Sciences of the United States of America*, 79(4):1073–7, feb 1982. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/pubmed/6951163http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC345902.

Koen J. F. Verhoeven, Bridgett M. VonHoldt, and Victoria L. Sork. Epigenetics in ecology and evolution: what we know and what we need to know. *Molecular Ecology*, 25(8):1631–1638, apr 2016. ISSN 09621083. doi: 10.1111/mec.13617. URL http://doi.wiley.com/10.1111/mec.13617.

G. Vogt. The marbled crayfish: A new model organism for research on development, epigenetics and evolutionary biology. *Journal of Zoology*, 276(1):1–13, 2008. ISSN 09528369. doi: 10.1111/j.1469-7998.2008.00473.x.

G. Vogt, M. Huber, M. Thiemann, G. van den Boogaart, O. J. Schmitz, and C. D. Schubart. Production of different phenotypes from the same genotype in the same environment by developmental variation. *Journal of Experimental Biology*, 211(4):510–523, feb 2008. ISSN 0022-0949. doi: 10.1242/jeb.008755. URL http://jeb.biologists.org/cgi/doi/10.1242/jeb.008755.

Günter Vogt, Cassandra Falckenhayn, Anne Schrimpf, Katharina Schmid, Katharina Hanna, Jörn Panteleit, Mark Helm, Ralf Schulz, and Frank Lyko. The marbled crayfish as a paradigm for saltational speciation by autopolyploidy and parthenogenesis in animals. *Biology open*, 4(11):1583–94, oct 2015. doi: 10.1242/bio.014241. URL http://www.ncbi.nlm.nih.gov/pubmed/26519519http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4728364.

C. H. Waddington. The Epigenotype. *International Journal of Epidemiology*, 41(1): 10–13, feb 1942. ISSN 0300-5771. doi: 10.1093/ije/dyr184. URL https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyr184.

Colum P. Walsh, J. Richard Chaillet, and Timothy H. Bestor. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, 20(2):116–117, oct 1998. ISSN 1061-4036. doi: 10.1038/2413. URL http://www.nature.com/articles/ng1098{_}116.

R Y Wang, C W Gehrke, and M Ehrlich. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic acids research*, 8 (20):4777–90, oct 1980. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pubmed/7443525http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC324387.

Xianhui Wang, Xiaodong Fang, Pengcheng Yang, Xuanting Jiang, Feng Jiang, Dejian Zhao, Bolei Li, Feng Cui, Jianing Wei, Chuan Ma, Yundan Wang, Jing He, Yuan Luo, Zhifeng Wang, Xiaojiao Guo, Wei Guo, Xuesong Wang, Yi Zhang, Meiling Yang, Shuguang Hao, Bing Chen, Zongyuan Ma, Dan Yu, Zhiqiang Xiong, Yabing Zhu, Dingding Fan, Lijuan Han, Bo Wang, Yuanxin Chen, Junwen Wang, Lan Yang, Wei Zhao, Yue Feng, Guanxing Chen, Jinmin Lian, Qiye Li, Zhiyong Huang, Xiaoming Yao, Na Lv, Guojie Zhang, Yingrui Li, Jian Wang, Jun Wang, Baoli Zhu, and Le Kang. The locust genome provides insight into swarm formation and long-distance flight. *Nature communications*, 5:2957, 2014a. ISSN 2041-1723. doi: 10.1038/ncomms3957. URL http://www.ncbi.nlm.nih.gov/pubmed/24423660http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3896762.

Xiaotong Wang, Qiye Li, Jinmin Lian, Li Li, Lijun Jin, Huimin Cai, Fei Xu, Haigang Qi, Linlin Zhang, Fucun Wu, Jie Meng, Huayong Que, Xiaodong Fang, Ximing Guo, and Guofan Zhang. 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(1):1119, 2014b. ISSN 1471-2164. doi: 10.1186/1471-2164-15-1119. URL http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-1119.

Xu Wang, David Wheeler, Amanda Avery, Alfredo Rago, Jeong-Hyeon Choi, John K Colbourne, Andrew G Clark, and John H Werren. Function and evolution of DNA methylation in Nasonia vitripennis. *PLoS genetics*, 9(10):e1003872, 2013. ISSN 1553-7404. doi: 10.1371/journal.pgen.1003872. URL http://www.ncbi.nlm.nih.gov/pubmed/24130511http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3794928.

Didier Wion and Josep Casadesús. N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nature Reviews Microbiology*, 4(3):183–192, mar 2006. ISSN 1740-1526. doi: 10.1038/nrmicro1350. URL http://www.nature.com/articles/nrmicro1350.

Ping Wu, Wencai Jie, Qi Shang, Enoch Annan, Xiaoxu Jiang, Chenxiang Hou, Tao Chen, and Xijie Guo. DNA methylation in silkworm genome may provide insights into epigenetic regulation of response to Bombyx mori cypovirus infection. *Scientific Reports*, 7(1):16013, dec 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-16357-7. URL http://www.nature.com/articles/s41598-017-16357-7.

Yuanxin Xi and Wei Li. BSMAP: whole genome bisulfite sequence MAPping program. *BMC bioinformatics*, 10:232, jul 2009. ISSN 1471-2105. doi: 10.1186/1471-2105-10-232. URL http://www.ncbi.nlm.nih.gov/pubmed/19635165http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2724425.

Hui Xiang, Jingde Zhu, Quan Chen, Fangyin Dai, Xin Li, Muwang Li, Hongyu Zhang, Guojie Zhang, Dong Li, Yang Dong, Li Zhao, Ying Lin, Daojun Cheng, Jian Yu, Jinfeng Sun, Xiaoyu Zhou, Kelong Ma, Yinghua He, Yangxing Zhao, Shicheng Guo, Mingzhi Ye, Guangwu Guo, Yingrui Li, Ruiqiang Li, Xiuqing Zhang, Lijia Ma, Karsten Kristiansen, Qiuhong Guo, Jianhao Jiang, Stephan Beck, Qingyou Xia, Wen Wang, and Jun Wang. Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*, 28(5):516–520, may 2010. ISSN 1087-0156. doi: 10.1038/nbt.1626. URL http://www.nature.com/articles/nbt.1626.

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, and Jussi Taipale. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356 (6337):eaaj2239, may 2017. ISSN 0036-8075. doi: 10.1126/science.aaj2239. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aaj2239.

A. Zemach, I. E. McDaniel, P. Silva, and D. Zilberman. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, 328(5980):916–919, may 2010. ISSN 0036-8075. doi: 10.1126/science.1186366. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1186366.

Assaf Zemach and Daniel Zilberman. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current Biology*, 20(17):R780–R785, 2010. ISSN 09609822. doi: 10.1016/j.cub.2010.07.007. URL http://dx.doi.org/10.1016/j.cub.2010.07.007.

Daniel Zilberman. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology*, 18(1):87, dec 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1230-2. URL http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1230-2.

Daniel Zilberman, Mary Gehring, Robert K Tran, Tracy Ballinger, and Steven Henikoff. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature genetics*, 39(1):61–9, 2007.

ISSN 1061-4036. doi: 10.1038/ng1929. URL http://www.ncbi.nlm.nih.gov/pubmed/17128275.

Michael J Ziller, Hongcang Gu, Fabian Müller, Julie Donaghey, Linus T-Y Tsai, Oliver Kohlbacher, Philip L De Jager, Evan D Rosen, David A Bennett, Bradley E Bernstein, Andreas Gnirke, and Alexander Meissner. Charting a dynamic DNA methylation landscape of the human genome. *Nature*, 500(7463):477–81, aug 2013. ISSN 1476-4687. doi: 10.1038/nature12433. URL http://www.ncbi.nlm.nih.gov/pubmed/23925113http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3821869.

Acknowledgements

My sincere thank you goes to Prof. Frank Lyko, first of all, for the opportunity to pursue this PhD project, for a door that was always open and the support provided during the course of these three years.

I sincerely thank Prof. Henrik Kaessmann, Dr. Lars Feuerbach, Prof. Stefan Wiemann and Dr. Darjus Tschaharganeh: for being my second examiner, for constructive and helpful comments during my TAC meetings, and for being part of my examination commission.

The DKFZ GPCF for their constant support, and the DKFZ career service, Karin Greulich-Bode especially, for personal support and encouragement.

A heartful thanks goes to former and current roommates and lab members of A130: the marbled crayfish team, most of all: Cassandra for her induction and cordiality, Julian G. for Gudekuns bioinformatics solutions and teaching us all proper humor, Laura and Ranja for becoming close friends and companions, Kathi for her kindness and support, Vitor for his profound support in scientific and human regards, his appreciation of good food and for showing me how exciting collaborative work can be, Matthias and Lena for the Neuenschuh-Gang times, Olena for inspiring discussions and for teaching us the Mayakovska opening, Jana and Lorena for being the best assistants, Sina for the many delicious cakes, Guenther for many constructive discussions and advice, and everybody else in A130.

To my family, who was always supportive, my mother, father and sister Frizzy, Julian F. for his kind heart, his support during this PhD, and the time spent together.

People who have accompanied me before already, or since the start of this PhD, or have proof-read this thesis: Anna-Maria S., Claudia P. and Lexi S. for the best neverending support during the course of this PhD, Katja R., Bob B., Jens P., Zach C., Ernest W., Franz-Reiner B., Cosima W., Alina B., the best housemates Rebecka and Florian W., the Leo Club in Heidelberg for a different focus, Justyna W. for some last-minute advice, Franni B., Ingolf S.. The Avantgarde for being the best travel companions, i.e., Thomas vB., Eva W., Raveesh M., Daniel S., Verena S. The crew in Bonn (Lisa, Amelie, Flo, Arne, Erik) and in Munich (Oli M., Martha S., Christl D.) for making me feel like home whenever I'm back.