Aus dem:
Institute of Computational Biology
Institut des Helmholtz Zentrums München
Vorstand: Fabian Theis

Computational approaches for the quantification of transgenerational and somatic epimutations

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Johanna Denkena

aus
Hannover

Jahr
2024

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuerin:	Prof. Dr. Maria Colomé-Tatché
Zweitgutachter:	Prof. Dr. Philipp Korber
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	16. Oktober 2024

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dekanat Medizinische Fakultät
Promotionsbüro

Eidesstattliche Versicherung

Denkena, Johanna

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

Computational approaches for the quantification of transgenerational and somatic epimutations

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 18.10.2024
Ort, Datum

Johanna Denkena
Unterschrift Doktorandin/Doktorand

Contents

Index of Abbreviations

Abbreviations	Definition
4mC	4-methylcytosine
5mC	5-methylcytosine
6mA	N6-methyladenin
A. thaliana	*Arabidopsis thaliana*
α	DNA methylation gain rate
β	DNA methylation loss rate
A	Adenin
bp	basepairs
CMT2	CHROMOMETHYLASE 2
CMT3	CHROMOMETHYLASE 3
CpG	Cytosine-phosphate-Guanin
CG	Cytosine, which is followed by a Guanin in 5' $\rightarrow$ 3' direction
CHG	Cytosine, which followed by a C, T or A and then a G
CHH	Cytosine, which followed by two C, T or A nucleotides
D	divergence
ΔT	divergence time
DEG	Differentially Expressed Gene
DMR	Differentially Methylated Region
DMS	Differentially Methylated Site
DNA	deoxyribonucleic acid
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
G	Guanin
GC content	percentage of bases that are G or C
gbM	genebody Methylation
I	Intermediate Methylation Status
IGR	Intergenic Region
M	Methylated

Abbreviations	Definition
mC	cytosine methylation
mCG	cytosine methylation in CG context
mCHG	cytosine methylation in CHG context
mCHH	cytosine methylation in CHH context
mRNA	messenger RNA
MA	Mutation Accumulation
MET1	DNA METHYLTRANSFERASE 1
P. trichocarpa	*Populus trichocarpa*
HMM	Hidden Markov model
RdDM	RNA-directed DNA methylation
SNP	Single Nucleotide Polymorphism
TE	transposable element
T	Thymin
U	Unmethylated
UTR	untranslated region

List of Publications

Articles in Refereed Journals

Papers included in the Dissertation

1. Johanna Denkena, Frank Johannes & Maria Colomé-Tatché. **Region-level epimutation rates in *Arabidopsis thaliana***. *Heredity* 127, 190–202 (2021)

2. Brigitte T. Hofmeister, Johanna Denkena, Maria Colomé-Tatché, Yadollah Shahryary, Rashmi Hazarika, Jane Grimwood, Sujan Mamidi, Jerry Jenkins, Paul P. Grabowski, Avinash Sreedasyam, Shengqiang Shu, Kerrie Barry, Kathleen Lail, Catherine Adam, Anna Lipzen, Rotem Sorek, Dave Kudrna, Jayson Talag, Rod Wing, David W. Hall, Daniel Jacobsen, Gerald A. Tuskan, Jeremy Schmutz, Frank Johannes & Robert J. Schmitz. **A genome assembly and the somatic and epigenetic mutation rate in a long-lived perennial *Populus trichocarpa***. *Genome Biology* 21, 259 (2020)

Additional Journal Articles

3. Yadollah Shahryary, Aikaterini Symeonidi, Rashmi Hazarika, Johanna Denkena, Talha Mubeen, Brigitte Hofmeister, Thomas van Gurp, Maria Colomé-Tatché,Koen Verhoeven, Gerald Tuskan, Robert Schmitz & Frank Johannes. **AlphaBeta: Computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants**. *Genome Biology* 21, 260 (2020)

Manuscripts in Preparation

4. Johanna Denkena, Alexander Graf, Johannes Hellmuth, Heinrich Leohnardt, Oliver Weigert & Maria Colomé-Tatché. **epiDELO: Deconvolution of intra-sample DNA methylation signatures using long-read sequencing data**.

Talks at Conferences

1. Johanna Denkena, Alexander Graf, Johannes Hellmuth, Oliver Weigert & Maria Colomé-Tatché. **Epiclonal reconstruction of DNA methylation patterns using single molecule long reads** *System Biology of Human Diseases*, Berlin (2023)

2. Johanna Denkena, Brigitte T. Hofmeister, Yadollah Shahryary, Rashmi Hazarika, Jane Grimwood, Sujan Mamidi, Jerry Jenkins, Paul P. Grabowski, Avinash Sreedasyam, Shengqiang Shu, Kerrie Barry, Kathleen Lail, Catherine Adam, Anna Lipzen, Rotem Sorek, Dave Kudrna, Jayson Talag, Rod Wing, David W. Hall, Daniel Jacobsen, Gerald A. Tuskan, Jeremy Schmutz, Robert J. Schmitz, Frank Johannes & Maria Colomé-Tatché. **Estimates of Region-level Epimutation Rates in *A. thaliana* and *P. trichocarpa***. *Population Genetics Group 55*, Norwich (2022)

Posters at Conferences

1. Johanna Denkena, Sebastian Mackowiak, Marey Messingschlager, Laura Matzner, Monika Zaufall, Rihab Omer Hamid, Zhuoxin Peng, Britta Steffens, Marvin Häberle, Naveed Ishaque, Sven Wellmann, Marc Pfister, Saskia Trump, Jon Genueit, Michael Kabesch, Roland Eils & Irina Lehmann. **Prediction of Emerging Food Allergies at Birth using Methylation Biomarkers**. *Epigenetics of Common Diseases*, Hinxton (2022)

2. Johanna Denkena, Frank Johannes & Maria Colomé-Tatché. **Epimutation Rates in *Arabidopsis thaliana***. *Abcam Epigenetics meeting on Inheritance and Memory*, Munich (2019)

3. Johanna Denkena, Frank Johannes & Maria Colomé-Tatché. **Epimutation Rates in *Arabidopsis thaliana***. *Plant epi-/genetics symposium*, Angers (2018)

1 Introduction

1.1 DNA methylation: an epigenetic mark

The term "epigenetics" refers to all heritable biochemical alterations of the genome and its chromatin structure that might affect gene expression, but do not change the genetic sequence itself [DAB09; Li21]. Epigenetic marks do not only modulate gene expression and aid in commitment of a cell to a particular function, but also play an important role in maintaining genome integrity [DAB09]. The most common epigenetic mechanisms include histone modifications and DNA methylation [DAB09; Li21]. Histones can exhibit posttranslational modifications at their N- and C-terminal tails, such as acetylation, methylation, phosphorylation and ubiquitination among others [YQC21]. Posttranslational acetylation and methylation take part in the regulation of gene transcription by modulating how accessible regulatory regions like promoters are to transcription factors [Li21]. Furthermore, histone phosporylation participates in the regulation of chromatin structure, cell cycle regulation, DNA repair and active transcription, while ubiquitination has been shown to be important to the cellular DNA damage response for the maintenance of genome stability [MP21; YQC21].

DNA methylation describes a covalent bond of a methyl-group to a nucleotide, most commonly to a cytosine as 5-methylcytosine (5mC) or 4-methylcytosine(4mC), but also to adenine as N6-methyladenine (6mA)[LT21]. Of these types of DNA methylation 4mC and 6mA are mainly found in prokaryots, while 5mC ocurrs ubiquitously in eukaryots and has been extensively studied [Li21]. Generally, as in this thesis, the term ‘cytosine methylation‘ or ‘DNA methylation‘ references the 5mC type of methylation. DNA methylation serves a variety of functions in the genome. The most well studied function is gene silencing, whereby methylation of cytosines overlapping the promoter region and/or the transcription start site of a gene is associated with transcriptional repression [LHE18; LT21]. Although the exact mechanism remains unclear, studies have suggested that the presence of 5mC causes lowered binding affinity of certain transcription factors to the respective promoters [LHE18]. A further mechanism of gene silencing may be that hypermethylation inside regulatory sequences causes a reorganization of the chromatin structure, which leads to lower accessibility for the transciptional

machinery [LT21] Moreover, DNA methylation has been shown to impact gene transcriptional activity through nucleosome repositioning [LT21].

Another function of DNA methylation is the prevention of transposon movement within the genome through maintenance of high DNA methylation levels in transposable elements. This mechanisms is crucial for the preservation of genome integrity [LJ10; LT21]. In contrast, hypermethylation of cytosines in gene bodies has been found to correlate with transcriptional activity of genes, although the mechanism linking hypermethylation to hyperexpression and whether this is causal has not been established [Jon12; LT21]. Hypotheses for the functionality of gene-body methylation range from it being a neutral by-product of active transcription, to silencing of repetitive sequences inside genes [Bew+16; Li+17].

During embryonic development, pluripotent cells show very low levels of methylation following a combination of passive demethylation, through reduced maintenance and *de novo* methylation pathways and active demethylation, through DNA glycolases in plants or hydroxylation in mammals [LJ10; WZ17]. As these cells differentiate, methylation is gained globally but also lost in specific emerging cell types. Genes associated with pluripotency and gametogenesis are silenced through methylation gain, while genes that define cellular identity are activated through methylation loss. This leads to the formation of lineage-specific DNA methylation patterns that shape and maintain gene expression for specialized cell types [Sue+16].

1.2 DNA methylation in plants

In plants, cytosine methylation is observed in the nucleotide contexts CG, CHG and CHH in 5' → 3' direction, whereby 'H' represents A, T or C nucleotides. Both CG and CHG are symmetric between both strands, i.e. a CG or CHG will always be present on the opposite strand as a mirror image, while the CHH context is always asymmetric [Che+21; LJ10]. In the model plant *Arabdopsis thaliana* DNA methylation has been observed at approximately 24% of CGs, 6.7% of CHGs and 1.7% of CHHs genome-wide [Cok+08]. While DNA methylation in plants occurs primarily in transposable elements and repeat-rich regions (at CG, CHG and CHH sites), it is also observed in a subset of gene bodies of ubiquitously expressed genes, mainly in CG context [Bew+16; LJ10; Vid+16].

Establishment and maintenance of methylation in plants is mediated by pathways that are conserved across plants species [JS19]. *De novo* methylation is established by DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE 2) via the RNA-directed DNA methylation (RdDM) pathway [FDG18]. For each of the 3 cytosine

contexts methylation is maintained by a different pathway. While methylation in CG context is maintained through MET1 (DNA METHYLTRANSFERASE 1), CHG methylation is kept up by CMT3 (CHROMOMETHYLASE 3). CHH methylation in the other hand is consistently established *de novo* via DRM2 and the RdDM pathway, while only a subset of CHHs are maintained through CMT2 (CHROMOMETHYLASE 2) [GB16; LJ10]. During gametogenesis global methylation levels are temporarily lost and transposable elements reactivated, likely to reveal the expressed transposons and priming *de novo* methylation pathways to thoroughly silence them [LJ10].

1.3 Epimutations and epigenetic evolution

Although cytosine methylation is tightly maintained by pathways conserved across species, cytosines do at times lose or gain methylation during cell division. These stochastic changes are generally termed "epimutations" and originate in somatic tissue as well as in germline cells [Gai+19; JS19; Sha21]. Epimutations are defined as being independent of genetic mutations and may be inherited to subsequent generations [JS19; Sha21]. While it is rare, epimutations have also been shown to cause phenotypic effects. Two examples of this are *Linaria vulgaris*, where the symmetry of the flower was observed to change as a result of increased spontaneous methylation at the promoter of the *Lcyc* gene [CVC99] and *Solanum lycopersicum*, where an epimutation in the promoter of *SQUAMOSA promoter binding protein–like* gene caused an alteration of color and cell-to-cell adhesion [Man+06]. Both epimutations were found to be inherited by subsequent generations [CVC99; Man+06]. However, in *Linaria* a few branches were observed that produced flowers resembling a mixture from mutant and wild type, suggesting that epimutations are also prone to somatic reversion [CVC99]. More recently epimutations were discovered in the "mantling" phenotype of the African oil palm *Elaeis guineensis*, which is characterized by aborted fruit development and very low oil yield. In this case, CHG hypomethylation of the *Karma* retrotransposon near a splice site of the *EgDEF1* gene was found to cause the expression of a novel truncated *EgDEF1* transcript. At times this mantled phenotype has been shown to spontaneously revert to normal [Ong+15].

To study extent to which epimutations occur spontaneously in the absence of environmental pressure, researchers use Mutation Accumulation (MA) experiments. MA lines are made up of single founder inbred or clonal lines that are propagated over many generation in stable conditions and measurements are taken at multiple generations [HK09; JS19]. While MA lines were originally developed to study genetic mutation rates, the samples taken from the same experimental design can

also be used for genome-wide methylome profiling. This is typically achieved through whole-genome bisulfite sequencing, where treatment of genomic DNA with sodium bisulfite converts unmethylated cytosines to uracil, which is further converted to thymine by PCR amplification. Meanwhile methylated cytosines remain unchanged, thus allowing for the computation of methylation levels as cytosine-to-thymine ratios [JS19; LT21]. From these epigenetic measurements and knowing at which generations epimutations are first observed researchers can then extrapolate at which rate epimutations arise [Gra+15].

For the self-fertilizing plant *A. thaliana*, several MA lines grown under stable conditions have been generated [SBD00]. In 2015, Van der Graaf et al. [Gra+15] first estimated genome-wide rates of methylation gain and loss per CG site per haploid genome and generation at $2.56 \cdot 10^{-4}$ and $6.3 \cdot 10^{-4}$, respectively. In 2020, Shahryary et al. [Sha+20] formalized the pre-processing and estimation procedure, which was published in the R package AlphaBeta. The estimation method implemented in AlphaBeta fits a model of time-dependent accumulation of epimutations to pedigree-based methylation data to account for measurement noise [Gra+15; Sha+20]. More concretely, AlphaBeta calculates the average difference in DNA methylation (divergence D) as well as the number of independent time units that passed since the most recent common ancestor (divergence time ΔT) for every pair of samples in the MA pedigree. To model the divergence D as a function of divergence time ΔT, it is assumed that in one time unit $[t,t+1]$ unmethylated cytosines gain methylation stochastically with a probability of α, while cytosines lose methylation with a probability of β. Genome-wide the methylation levels are assumed to be at an equilibrium, since α and β only act on their substrates [Gra+15]. The transitions of the diploid epigenotypes during $[t,t+1]$ are modelled though a transition matrix. Shahryary et al. [Sha+20] applied AlphaBeta to the same MA lines as Graaf et al. [Gra+15] as well as other *A. thaliana* MA lines. The resulting epimutation rates were found to be in the same order of magnitude (10^{-4}) as those estimated by Van der Graaf et al. [Gra+15] and at the same time the rate of methylation loss always considerably exceeded the rate of methylation gain, mirroring the genome-wide methylation levels. This indicates that spontaneous epimutations have an influence on global methylation patterns [Gra+15; JS19]. In contrast to the genetic mutation rate in *A. thaliana* ($7 \cdot 10^{-9}$ base substitutions per site per generation), CGs accumulate methylation changes much more quickly over time, suggesting that epigenetic variation becomes uncoupled from genetic variation over short evolutionary timescales [Gra+15; Oss+10]. For cytosines in CHG and CHH context, methylation changes could also be detected, but the accumulation over generations is so low that significant epimutation rates could only be estimated in 2020 by Shahryary et al. [Sha+20] using a particularly extensive

MA line spanning a large number of generations, showing that in these contexts epimutations accumulate 1-2 magnitudes slower than in CG context.

However, the extent of epimutations in *A. thaliana* does not only vary for different cytosine contexts but also in relation to cytosines' genomic environment. For example, methylation changes observed in *A. thaliana* MA lines by Hofmeister et al. [Hof+17] were mostly restricted to only 13 regions, while more than 99.997% of the methylome was inherited stably, suggesting the existence of regions with high methylome instability or "hotspots". In fact, CG epimutations are most frequently observed in genes and chromosome arms, while they are mostly depleted in TEs and (peri-)centromeres. Additionally, local chromatin composition as well as sparse methylation have been shown to be highly predictive of regions with high epimutability [Gra+15; Haz+22].

Further MA experiments also investigate the impact of different genotypes and environmental conditions such as drought on the emergence of epimutations in *A. thaliana* and other plants [Gan+17; JS19; Zhe+17] Experiments like these are expected to expand out understanding of epimutation dynamics in the future, but so far it can only be hypothesized that epimutations play a minor role in short-term adaption and long-term evolution [Sar20].

1.4 Region-level changes of DNA methylation

When investigating possible functional effects of epigenetic variation, CpG methylation is often analyzed in the context of differentially methylated regions (DMRs). A DMR is defined as multiple adjacent CpG sites with differential methylation states between different samples or conditions [Rak+11]. In plants, DMRs have been observed in association with environmental stressors such as salt, drought and heat and well as developmental stages [Hos+17; Jia+14; Kom+18; ZLZ18].

Generally, DMRs are determined *de novo* for every comparison between samples. Thereby DMRs are established in such a way that is maximizes the combined methylation differences between (groups of) samples. This results in regions with genomic coordinates that are different for every comparison of groups of samples. At time of writing, at least 27 DMR calling algorithms have been published for WGBS alone [Aka+12; BB15; Cat+18; Che+21; Con+18; DS14; GP19; Har17; HLI12; Hüt+22; Jüh+16; Kis+15; Li+13; Par+14; Pet+21; Sch+15; Son+13; Sri+19; STM14; Su+13; Sun+14; Tra+18; Wan+15; Wen+16; Wu+15]. They utilize a variety of approaches. Some of the more straightforwards methods use a sliding window to calculate differential methylation between samples per window. These are extended to the adjoining similarly differentially methylated windows (SwDMR

from Wang et al. [Wan+15]). More complex methods employ Hidden Markov models (HMM-DM by Yu and Sun [YS16] or HMM-Fisher by Sun and Yu [SY16]) or unsupervised machine learning (MethylScore by Hüther et al. [Hüt+22]). Many DMR callers are usually specific to species (plant vs. mammals), quality of data and data type [Cat+18; Kre+20].

Using DMR calling methods for the analysis of region-level DNA methylation changes in large data sets, such as MA lines, would require to first perform a great number of pairwise DMR callings. Then consolidating the DMRs between comparisons to establish population-wide DMRs would likely diminish the formerly maximized methylation differences between samples. Additionally, variable region boundaries will limit the comparability between different MA lines. Population studies often utilize a set of pre-established regions and average methylation levels in the defined regions are compared between samples/groups. This is frequently accomplished by a straightforward partitioning of the genome into non-overlapping bins of the same size [Eic+13; Gan+17]. However, methylation levels at cytosines have been found to exhibit strong spatial correlation in plants as well as mammals [Cok+08; Lis+08; Lis+09]. This indicates that when constructing regions *a priori*, one could leverage this aspect and identify methylation units that are more meaningful, by clustering neigbhouring cytoines in areas with high cytosine density. In mammals, this is very routinely done by tracking CpG islands, which are mostly defined as a region of minimally 200bp length, with at least 50% GC content and an observed-to-expected CpG ratio of at least 0.6 or variations thereof. Especially promoters are particularly enriched in CpG islands, while other genomic regions remain CpG-depleted, which creates clear cut boundaries for region calling [Ash01; Wu+10]. However, in contrast to the mammalian genome, cytosines and particularly CpGs in plants occur more frequently and CpG-rich regions are distributed fairly evenly over the genome [Ash01]. This poses the challenge of creating regions that represent meaningful clusters of cytosines that can be expected to show high agreement in methylation level due to spatial correlation, which is discussed in Denkena et al. [DJC21].

1.5 Objectives for the thesis

The scope of this thesis was to investigate how the methylome evolves in plant populations, more specifically how quickly the DNA methylation of an organism changes and how these changes can be quantified and used for biological insight. Using plant models, I investigated two different methylation propagation systems: the emergence of epimutations meiotically over multiple generations of plant propagation as well as mitotically within one generation over multiple years. To address these two questions, I estimated transgenerational epimutation rates using short-lived *Arabidopsis thaliana* populations as well as the somatic epimutation rates for the long-lived perennial *Populus trichocarpa*. In particular, I emphasized investigating differences and/or similarities in how both species accumulate epimutations in different genomic features such as genes, TEs, chromosome arms and centromeres among others. I specifically aimed to estimate the epimutation rates for regions or clusters of cytosines, to complement the epimutation rates per single cytosine. For this, I investigated different ways to establish genomic regions or cytosine clusters that would represent potentially functional epigenetic changes. In my thesis I contributed to the understanding we have about the sources and time scale of epigenetic heterogeneity.

2 Contributions to Publications

2.1 Region-level epimutation rates in *Arabidopsis thaliana*

The concept of pre-establishing regions from genetic information was conceived by Maria Colomé-Tatché, Frank Johannes and me. I came up with and implemented the method for establishing regions and combining methylation counts per region including which quality measures would be applied. All subsequent analyses, including annotation-specific methylation accumulation accumulation, were planned and implemented by me and epimutation rate estimation was run by me using the method published in Graaf et al. [Gra+15]. All figures were made by me with input from Maria Colomé-Tatché. Finally, I wrote the first draft of the publication, which was then finalized in a joint effort from Maria Colomé-Tatché, Frank Johannes and me.

2.2 A genome assembly and the somatic and epigenetic mutation rate in a wold long-lived perennial *Populus trichocarpa*

How to view the tree and its branches in the context of epimutation accumulation analogous to MA line pedigrees was conceptualized by Maria Colomé-Tatché, Frank Johannes and me. In this very collaborative project, I performed the analysis of cytosine-level methylation and estimation of epimutation rates genome-wide and per genomic annotation. After Brigitte Hofmeister identified and provided the DMRs, I conceptualized and performed the subsequent analysis for the region-level methylation levels and epimutation rates. Also, I investigated whether transcription levels accumulated over time. I wrote the methods and results sections for these parts of the project and I also generated several figures (Fig. 3, 5a). Additionally, I generated the Fig. 2 about proportions of transition and transversion mutations as well as the proportions of genomic features exhibiting mutations.

2.3 AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants (Appendix)

I performed testing of the *AlphabBeta* package and estimated the epimutation rates for *P. trichocarpa* from the pre-processed methylome data. Also the model comparison for poplar was done by me.

3 Zusammenfassung

In Pflanzen wird die DNA-Methylierung durch spezialisierte Mechanismen aufrechterhalten. Trotzdem wird der Methylierungsstatus von Cytosinen nicht immer zuverlässig über meiotische und mitotische Zellteilungen hinweg beibehalten. Solche stochastischen Methylierungsgewinne und -verluste, die als "spontane Epimutationen" bezeichnet werden, können sowohl transgenerational als auch somatisch vererbt werden. Dies führt im Laufe der Zeit zu einer Akkumulation im Pflanzengenom. Die Raten, mit denen einzelne Cytosine (im CG-Kontext) über Generationen hinweg Methylierung gewinnen und verlieren, wurden zuvor in verschiedenen Mutationsakkumulationslinien in *A. thaliana* untersucht, wobei sich zeigte, dass die Epimutationsraten um 5 Größenordnungen höher sind als die genetischen Mutationsraten (Gewinnrate = $1,4 \cdot 10^{-4}$ pro Generation pro haploidem Methylom, Verlustrate = $5,7 \cdot 10^{-4}$, genetische Mutationsrate = $7 \cdot 10^{-9}$ [Sha+20]). Da jedoch funktionell relevante Effekte wie phänotypische Variation und Transkriptionsaktivität im Allgemeinen mit Methylierungsänderungen in größeren Genomregionen und nicht in einzelnen Cytosinen in Verbindung gebracht werden, haben wir versucht, solche Epimuationsraten auf Regionsebene für *A. thaliana* zu berechnen (siehe Kapitel 5 [DJC21]). Um funktionell aussagekräftige Epimutationsraten für Cytosin-Cluster zu untersuchen, die auch über verschiedene MA-Linien hinweg vergleichbar sind, haben wir das *A. thaliana*-Methylom in Methylierungseinheiten auf Grundlage des *A. thaliana*-Referenzgenoms und nicht auf Grundlage von Methylierungsdaten segmentiert. Ausgehend von der Beobachtung, dass die Korrelation der Methylierungsniveaus zwischen zwei Cytosinen mit dem Abstand (gemessen in Basenpaaren) zwischen diesen Cytosinen skaliert, wurden Regionen durch Zusammenfassen von nahe beieinander liegenden Cytosinen konstruiert. Ausgehend vom geringsten Abstand wurden die Cytosine iterativ mit ihren benachbarten Cytosinen (oder bereits verketteten Clustern von Cytosinen) verkettet, bis die Größe dieser Cluster bzw. der Abstand zwischen ihnen einen Schwellenwert von 185 Basenpaaren erreichte. So konnten wir zeigen, dass die Epimutationsraten in der Region in der gleichen Größenordnung lagen wie die Epimutationsraten auf Cytosin-Ebene, wenn auch etwas niedriger (Gewinnrate = $1,2 \cdot 10^{-4}$, Verlustrate = $4,6 \cdot 10^{-4}$). Diese Epimutationsraten waren nur geringfügig von der Größe und Dichte der Cytosinregionen abhängig, aber stark von den genomischen Eigenschaften. Konkret akkumulierten sich Epimutationen schnell in Genkörpern,

während Transposons vergleichsweise wenig Methylierungs-unterschiede zeigten. Darüber hinaus wiesen Chromosomenarme eine höhere Epimutationsakkumulation auf als zentromerische Regionen, und dieser Trend blieb auch dann bestehen, wenn Gene und Transposons je nach chromosomaler Herkunft getrennt auf Epimutationen untersucht wurden. Zusätzlich zu den Regionen im CG-Kontext fanden wir Hinweise darauf, dass auch Regionen mit Cytosinen im Nicht-CG-Kontext auf genomweiter Ebene Epimutationen akkumulieren. Diese Anhäufung erfolgt jedoch wesentlich langsamer als im CG-Kontext.

Ergänzend zu den transgenerationalen Epimutationsraten wollten wir auch somatische epigenetische und genetische Mutationen untersuchen, die sich im Laufe des Lebenszyklus einer Pflanze ansammeln (siehe Kapitel 6 [Hof+20]). Insbesondere bei langlebigen mehrjährigen Bäumen ist die Untersuchung des Umfangs somatischer (Epi-)Mutationen wichtig für das Verständnis wie sich Bäume lokal anpassen. Hier wurde ein neues Referenzgenom für den Wildtyp von Populus trichocarpa aus einem Baum mit zwei Hauptstämmen generiert. Das Alter der verschiedenen Verzweigungspunkte des Baums wurde durch Entnahme von Bohrkernen bestimmt, und an den Enden der Hauptäste wurden Blattproben entnommen. Aus den Blattproben wurden genomische, epigenomische und transkriptomische Sequenzierungsdaten gewonnen und zur Quantifizierung von somatischen Mutationen und Epimutationen zwischen den Zweigen sowie zur Abschätzung der Geschwindigkeit ihrer Akkumulation pro Jahr und pro Generation verwendet. Wir beobachteten, dass CG-Epimutationen mit Raten in der Größenordnung von 10^{-6} pro Jahr akkumulierten (Gewinnrate = $1,8 \cdot 10^{-6}$, Verlustrate = $5,8 \cdot 10^{-6}$), was bei einer durchschnittlichen Generationsdauer von 15-150 Jahren bei Pappeln auf CG-Epimutationsraten zwischen 10^{-5} und 10^{-4} pro Generation schließen lässt. Dies war - wie bereits anhand von *A.thaliana* gesehen - um ein Vielfaches höher als die genetische Mutationsrate mit $1,33 \cdot 10^{-10}$ pro Jahr und $1,99 \cdot 10^{-9}$ pro Generation von Saatgut zu Saatgut. Darüber hinaus ermittelten wir zwischen den Zweigen vorhandene differenziell methylierte Regionen (DMRs) und unterteilten den verbleibenden genomischen Raum in Regionen, die der Größenverteilung der DMRs entsprachen. Die Methylierungsgrade wurden für alle Regionen aggregiert und zur Schätzung der Epimutationsraten auf Regionsebene verwendet. Wie bei *A. thaliana* waren auch bei *P. trichocarpa* die regionsweisen CG-Epimutationsraten mit $2,1 \cdot 10^{-6}$ (Gewinnrate) und $6,1 \cdot 10^{-6}$ (Verlustrate) den einzelnen CG-Epimutationsraten sehr ähnlich, lagen aber etwas höher. Diese Beobachtungen könnten nicht nur darauf hinweisen, dass Epimutationen und die Geschwindigkeit, mit der diese sich in beiden Pflanzenarten ansammeln, auf evolutionärer Ebene von genetischen Mutationen entkoppelt sind, sondern auch darauf, dass diese Epimutationen eine funktionelle Auswirkung auf den Organismus haben könnten, wenn sich ganze Regionen ändern. Im Fall der

somatischen Epimutationen in *P. trichocarpa* konnten wir DMRs identifizieren, die mit der unterschiedlichen Expression eines nahegelegenen Gens korrelierten, auch wenn insgesamt kein signifikanter Zusammenhang zwischen Transkription und Methylierung gefunden werden konnte.

Eine weitere Schlussfolgerung, die aus der Ähnlichkeit der Epimutationsraten bei *A. thaliana* und *P. trichocarpa* gezogen werden kann, ist, dass die Rate und das Spektrum der Epimutationen nicht durch transgenerationale Zellteilungsprozesse, sondern eher während der somatischen Zellteilung bestimmt werden. Dies wird auch durch die für verschiedene genomische Annotationen berechneten Epimutationsraten untermauert, die bei *A. thaliana* und *P. trichocarpa* ähnlich angeordnet waren, wobei die Gene am schnellsten Epimutationen anhäufen, während Transposons am zuverlässigsten erhalten werden, was möglicherweise die unterschiedlichen Mechanismen widerspiegelt, durch die die Methylierung an diesen verschiedenen Genomstrukturen während der Mitose erhalten werden. Darüber hinaus haben wir gezeigt, wie wir das Gesamtalter des von uns analysierten Baums vorhersagen können, indem wir die sich stetig akkumulierenden genomweiten Methylierungsänderungen als epigenetische Uhr verwenden. Durch die Anpassung des Modells der Epimutationsakkumulation mit verschiedenen Gesamtbaumaltern fanden wir heraus, dass das Alter des Baums ~ 330 Jahre betrug, was mit dem angenommenen Altersfenster von 250 bis 350 Jahren übereinstimmte. Dies deutet darauf hin, dass genomweite zufällige CG-Epimutationen, die mit einer konstanten Rate auftreten, dazu verwendet werden können, Alterung zu verfolgen.

4 Abstract

In plants, DNA methylation is maintained through specialized DNA methylation maintenance pathways. In spite of this, the methylation status of cytosines is not always faithfully maintained across meiotic and mitotic cell divisions. Epimutations, defined as stochastic gains and losses of methylation, may be inherited transgenerationally as well as somatically. The rates at which single cytosines (CG context) gain and lose methylation over generations have been previously studied in different mutation accumulation (MA) lines in *A. thaliana*, showing that epimutation rates are 5 magnitudes higher than genetic mutation rates (gain rate = $1.4 \cdot 10^{-4}$, loss rate = $5.7 \cdot 10^{-4}$ per generation per haploid methylome, genetic mutation rate = $7 \cdot 10^{-9}$, see Appendix A [Sha+20]). But since functionally relevant effects linked to phenotypic variation and transcriptional activity have generally been associated with methylation changes in larger genomic regions rather than single cytosines, we have sought to estimate such regions-level epimutation rates for *A. thaliana* (see chapter 5 [DJC21]). In order to study functionally meaningful epimutation rates for clusters of cytosines that would also be comparable across different MA lines, we segmented the *A. thaliana* methylome into methylation units on the basis of the *A. thaliana* reference genome, rather than methylation data. Based on the observation that the correlation of methylation levels between two cytosines scales with the distance (measured in basepairs) between these cytosines, regions were constructed by aggregating close cytosines. Starting from the lowest distance, cytosines were iteratively concatenated with their neighboring cytosines (or already concatenated clusters of cytosines) until the size of these clusters reached a threshold of 185 bp. Using these regions, we were able to show that region-wise epimutation rates were of the same magnitude as cytosine-level epimutation rates, albeit slightly lower (gain rate = $1.2 \cdot 10^{-4}$, loss rate = $4.6 \cdot 10^{-4}$). These epimutation rates were only marginally dependent of size and density of cytosine regions, but depended heavily on genomic features. More concretely, epimutations accumulated rapidly in genes bodies, while transposable elements showed comparatively faithful maintenance. Moreover, chromosome arms showed higher epimutation accumulation than centromeric regions and this trend held true even when genes and transposable elements were investigated separately for epimutations specific to chromosomal location. In addition to the regions in CG context, we found evidence that cytosine

regions with non-CG context also accumulate epimutations at the genome-wide scale. This accumulation, however, occurs at much slower rates than in CG context.

To complement the transgenerational epimutations rates, we also aimed to explore somatic epigenetic and genetic mutations, which accumulate during a plant's life cycle (see chapter 6 [Hof+20]). Especially in long-lived perennial trees, investigating the scope of somatic (epi-)mutations is important for our understanding of trees capacity for local adaption. In this work a novel reference genome for wild type *Populus trichocarpa* was generated from a tree with two main stems. The ages of different tree branching events were determined through coring, and leafs were sampled from the ends of the main branches. From the leaf samples, genomic, epigenomic and transcriptomic sequencing data was obtained and used to quantify somatic mutations and epimutations between the branches and to estimate how quickly they accumulated per year as well as per generation. We observed that CG epimutations accumulated at rates in the magnitude of 10^{-6} per year (gain rate = $1.8 \cdot 10^{-6}$, loss rate = $5.8 \cdot 10^{-6}$), which (assuming an average generation time of 15-150 years in poplar) amounts to per-generation CG epimutation rates between 10^{-5} and 10^{-4}. This was - as already seen over generations in *A. thaliana* - multiple orders of magnitudes higher than the genetic mutation rate at $1.33 \cdot 10^{-10}$ per year and $1.99 \cdot 10^{-9}$ per seed-to-seed generation. Furthermore, we established Differentially Methylated Regions (DMRs) present between branches and partitioned the remaining genomic space into regions that resembled the size distribution of the DMRs. Methylation levels were aggregated for all regions and used to estimate region-level epimutation rates. The *P. trichocarpa* region-level CG epimutation rates, at $2.1 \cdot 10^{-6}$ (gain rate) and $6.1 \cdot 10^{-6}$ (loss rate), were also very similar to the single CG epimutation rates, but were slightly higher, as had been observed in *A. thaliana*. These observations did not only indicate that epimutations and the speed at which they accumulate in both plant species are decoupled from genetic mutations on an evolutionary scale, but also that these epimutations might have a functional impact on the organism by emerging in the context of regions. In the case of the somatic epimutations in *P. trichocarpa*, we were able to identify DMRs that correlated with differential expression of a nearby gene, even though globally a significant association between transcription and methylation could not be found. Another conclusion that can be drawn from the similarity of epimutation rates in *A. thaliana* and *P. trichocarpa* is that the rate and spectrum of epimutations is not determined by transgenerational cell division processes but rather during somatic cell division. This is further underpinned by the epimutation rates calculated for different genomic features, which were also similarly ordered between *A. thaliana* and *P. trichocarpa* with the genes accumulating epimutations the fastest while transposable elements are maintained the most faithfully, possibly reflecting the

distinct pathways maintaining these different features during mitotis. Additionally, we showed how we could predict the total age of the tree that we analysed by using the steadily accumulating genome-wide methylation changes as an epigenetic clock. Through fitting the model of epimutation accumulation with different total tree ages we found that the age of the tree was ~ 330 years, which was in accordance with the hypothesized age window of 250 to 350 years. This suggests that genome-wide random CG epimutations that occur at a steady rate can be used to track aging over time.

5 Paper I

Johanna Denkena, Frank Johannes & Maria Colomé-Tatché.
Region-level epimutation rates in *Arabidopsis thaliana*.
Heredity 127, 190–202 (2021)
https://doi.org/10.1038/s41437-021-00441-w

My contributions to this paper are detailed in chapter 2.1

6 Paper II

Brigitte T. Hofmeister, **Johanna Denkena**, Maria Colomé-Tatché, Yadollah Shahryary, Rashmi Hazarika, Jane Grimwood, Sujan Mamidi, Jerry Jenkins, Paul P. Grabowski, Avinash Sreedasyam, Shengqiang Shu, Kerrie Barry, Kathleen Lail, Catherine Adam, Anna Lipzen, Rotem Sorek, Dave Kudrna, Jayson Talag, Rod Wing, David W. Hall, Daniel Jacobsen, Gerald A. Tuskan, Jeremy Schmutz, Frank Johannes & Robert J. Schmitz.

A genome assembly and the somatic and epigenetic mutation rate in a wold long-lived perennial *Populus trichocarpa*.

Genome Biology 21, 259 (2020)

https://doi.org/10.1186/s13059-020-02162-5

My contributions to this paper are detailed in chapter 2.2

Bibliography

[Aka+12] Altuna Akalin, Matthias Kormaksson, Sheng Li, Francine E. Garrett-Bakelman, Maria E. Figueroa, Ari Melnick, and Christopher E. Mason. **MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles**. *Genome Biology* 13:10 (2012), R87. ISSN: 1474760X. DOI: 10.1186/gb-2012-13-10-R87. URL: http://genomebiology.com/2012/13/10/R87 (see page 5).

[Ash01] I. Ashikawa. **Gene-associated CpG islands in plants as revealed by analyses of genomic sequences**. *Plant Journal* 26:6 (2001), 617–625. ISSN: 09607412. DOI: 10.1046/j.1365-313X.2001.01062.x (see page 6).

[BB15] Lee M. Butcher and Stephan Beck. **Probe Lasso: A novel method to rope in differentially methylated regions with 450K DNA methylation data**. *Methods* 72:C (2015), 21–28. ISSN: 10959130. DOI: 10.1016/j.ymeth.2014.10.036. URL: http://dx.doi.org/10.1016/j.ymeth.2014.10.036 (see page 5).

[Bew+16] Adam J. Bewick, Lexiang Ji, Chad E. Niederhuth, Eva Maria Willing, Brigitt T. Hofmeister, Xiuling Shi, Li Wang, Zefu Lu, Nicholas A. Rohr, Benjamin Hartwig, Christiane Kiefer, Roger B. Deal, Jeremy Schmutz, Jane Grimwood, Hume Stroud, Steven E. Jacobsen, Korbinian Schneeberger, Xiaoyu Zhang, and Robert J. Schmitza. **On the origin and evolutionary consequences of gene body DNA methylation**. *Proceedings of the National Academy of Sciences of the United States of America* 113:32 (2016), 9111–9116. ISSN: 10916490. DOI: 10.1073/pnas.1604666113 (see page 2).

[Cat+18] Marco Catoni, Jonathan M.F. Tsang, Alessandro P. Greco, and Nicolae Radu Zabet. **DMRcaller: A versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts**. *Nucleic Acids Research* 46:19 (2018). ISSN: 13624962. DOI: 10.1093/nar/gky602 (see pages 5, 6).

[Che+21] Yingyu Chen, Chin Kiu Kwok, Hangjin Jiang, and Xiaodan Fan. **Detect differentially methylated regions using non-homogeneous hidden Markov model for bisulfite sequencing data**. *Methods* 189:September 2020 (2021), 34–43. ISSN: 10959130. DOI: 10.1016/j.ymeth.2020.09.009. URL: https://doi.org/10.1016/j.ymeth.2020.09.009 (see pages 2, 5).

[Cok+08] 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 (2008), 215–219. ISSN: 14764687. DOI: 10.1038/nature06745 (see pages 2, 6).

[Con+18] David E. Condon, Phu V. Tran, Yu Chin Lien, Jonathan Schug, Michael K. Georgieff, Rebecca A. Simmons, and Kyoung Jae Won. **Defiant: (DMRs: Easy, fast, identification and ANnoTation) identifies differentially Methylated regions from iron-deficient rat hippocampus**. *BMC Bioinformatics* 19:31 (2018), 1–12. ISSN: 14712105. DOI: 10.1186/s12859-018-2037-1 (see page 5).

[CVC99] Pilar Cubas, Coral Vincent, and Enrico Coen. **An epigenetic mutation responsible for natural variation in floral symmetry**. *Nature* 401 (1999), 157–161 (see page 3).

[DAB09] Cathérine Dupont, D. Randall Armant, and Carol A. Brenner. **Epigenetics: Definition, mechanisms and clinical perspective**. *Seminars in Reproductive Medicine* 27:5 (2009), 351–357. ISSN: 15268004. DOI: 10.1055/s-0029-1237423 (see page 1).

[DJC21] Johanna Denkena, Frank Johannes, and Maria Colomé-Tatché. **Region-level epimutation rates in Arabidopsis thaliana**. *Heredity* 127 (2021), 190–202. ISSN: 13652540. DOI: 10.1038/s41437-021-00441-w. URL: http://dx.doi.org/10.1038/s41437-021-00441-w (see pages 6, 11, 15).

[DS14] Egor Dolzhenko and Andrew D. Smith. **Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments**. *BMC Bioinformatics* 15:1 (2014), 1–8. ISSN: 14712105. DOI: 10.1186/1471-2105-15-215 (see page 5).

[Eic+13] Steven R. Eichten, Roman Briskine, Jawon Song, Qing Li, Ruth Swanson-Wagner, Peter J. Hermanson, Amanda J. Waters, Evan Starr, Patrick T. West, Peter Tiffin, Chad L. Myers, Matthew W. Vaughn, and Nathan M. Springer. **Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations**. *Plant Cell* 25:8 (2013), 2783–2797. ISSN: 10404651. DOI: 10.1105/tpc.113.114793 (see page 6).

[FDG18] Fang Fang Fu, R. Kelly Dawe, and Jonathan I. Gent. **Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants**. *Plant Cell* 30:7 (2018), 1617–1627. ISSN: 1532298X. DOI: 10.1105/tpc.18.00053 (see page 2).

[Gai+19] Federico Gaiti, Ronan Chaligne, Hongcang Gu, Ryan M. Brand, Steven Kothen-Hill, Rafael C. Schulman, Kirill Grigorev, Davide Risso, Kyu-Tae Kim, Alessandro Pastore, Kevin Y. Huang, Alicia Alonso, Caroline Sheridan, Nathaniel D. Omans, Evan Biederstedt, Kendell Clement, Lili Wang, Joshua A. Felsenfeld, Erica B. Bhavsar, Martin J. Aryee, John N. Allan, Richard Furman, Andreas Gnirke, Catherine J. Wu, Alexander Meissner, and Dan A. Landau. **Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia**. *Nature* (2019). ISSN: 0028-0836. DOI: 10.1038/s41586-019-1198-z. URL: http://www.nature.com/articles/s41586-019-1198-z (see page 3).

[Gan+17] Diep R. Ganguly, Peter A. Crisp, Steven R. Eichten, and Barry J. Pogson. **The arabidopsis DNA methylome is stable under transgenerational drought stress**. *Plant Physiology* 175:4 (2017), 1893–1912. ISSN: 15322548. DOI: 10.1104/pp.17.00744 (see pages 5, 6).

[GB16] Quentin Gouil and David C. Baulcombe. **DNA Methylation Signatures of the Plant Chromomethyltransferases**. *PLoS Genetics* 12:12 (2016), 1–17. ISSN: 15537404. DOI: 10.1371/journal.pgen.1006526 (see page 3).

[GP19] Boying Gong and Elizabeth Purdom. **MethCP: Differentially Methylated Region Detection with Change Point Models**. In: *Research in Computational Molecular Biology*. Ed. by Lenore J. Cowen. Cham: Springer International Publishing, 2019, 68–84. ISBN: 978-3-030-17083-7 (see page 5).

[Gra+15] Adriaan van der Graaf, René Wardenaar, Drexel A. Neumann, Aaron Taudt, Ruth G. Shaw, Ritsert C. Jansen, Robert J. Schmitz, Maria Colomé-Tatché, and Frank Johannes. **Rate, spectrum, and evolutionary dynamics of spontaneous epimutations**. *Proceedings of the National Academy of Sciences* 112:21 (2015), 6676–6681. ISSN: 0027-8424. DOI: 10.1073/pnas.1424254112. arXiv: arXiv:1410.5723. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1424254112 (see pages 4, 5, 9).

[Har17] Thomas J. Hardcastle. **Methods for discovering genomic loci exhibiting complex patterns of differential methylation**. *BMC Bioinformatics* 18:1 (2017), 1–11. ISSN: 14712105. DOI: 10.1186/s12859-017-1836-0 (see page 5).

[Haz+22] Rashmi R. Hazarika, Michele Serra, Zhilin Zhang, Yinwen Zhang, Robert J. Schmitz, and Frank Johannes. **Molecular properties of epimutation hotspots**. *Nature Plants* 8:2 (2022), 146–156. ISSN: 20550278. DOI: 10.1038/s41477-021-01086-7 (see page 5).

[HK09] Daniel L. Halligan and Peter D. Keightley. **Spontaneous mutation accumulation studies in evolutionary genetics**. *Annual Review of Ecology, Evolution, and Systematics* 40 (2009), 151–172. ISSN: 1543592X. DOI: 10.1146/annurev.ecolsys.39.110707.173437 (see page 3).

[HLI12] Kasper D. Hansen, Benjamin Langmead, and Rafael A. Irizarry. **BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions**. *Genome Biology* 13:10 (2012). ISSN: 1474760X. DOI: 10.1186/gb-2012-13-10-R83 (see page 5).

[Hof+17] Brigitte T. Hofmeister, Kevin Lee, Nicholas A. Rohr, David W. Hall, and Robert J. Schmitz. **Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation**. *Genome Biology* 18:1 (2017), 1–16. ISSN: 1474760X. DOI: 10.1186/s13059-017-1288-x (see page 5).

[Hof+20] Brigitte T. Hofmeister, Johanna Denkena, Maria Colomé-Tatché, Yadollah Shahryary, Rashmi Hazarika, Jane Grimwood, Sujan Mamidi, Jerry Jenkins, Paul P. Grabowski, Avinash Sreedasyam, Shengqiang Shu, Kerrie Barry, Kathleen Lail, Catherine Adam, Anna Lipzen, Rotem Sorek, Dave Kudrna, Jayson Talag, Rod Wing, David W. Hall, Daniel Jacobsen, Gerald A. Tuskan, Jeremy Schmutz, Frank Johannes, and Robert J. Schmitz. **A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial Populus trichocarpa**. *Genome Biology* 21:259 (2020), 1–27. ISSN: 1474760X. DOI: 10.1186/s13059-020-02162-5 (see pages 12, 16).

[Hos+17] Md Shakhawat Hossain, Taiji Kawakatsu, Kyung Do Kim, Ning Zhang, Cuong T. Nguyen, Saad M. Khan, Josef M. Batek, Trupti Joshi, Jeremy Schmutz, Jane Grimwood, Robert J. Schmitz, Dong Xu, Scott A. Jackson, Joseph R. Ecker, and Gary Stacey. **Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs**. *New Phytologist* 214:2 (2017), 808–819. ISSN: 14698137. DOI: 10.1111/nph.14421 (see page 5).

[Hüt+22] Patrick Hüther, Jörg Hagmann, Adam Nunn, Ioanna Kakoulidou, Rahul Pisupati, David Langenberger, Detlef Weigel, Frank Johannes, Sebastian J. Schultheiss, and Claude Becker. **MethylScore, a pipeline for accurate and context-aware identification of differentially methylated regions from population-scale plant whole-genome bisulfite sequencing data**. *Quantitative Plant Biology* 3:19 (2022), 1–13. DOI: 10.1017/qpb.2022.14 (see pages 5, 6).

[Jia+14] Caifu Jiang, Aziz Mithani, Eric J. Belfield, Richard Mott, Laurence D. Hurst, and Nicholas P. Harberd. **Environmentally responsive genome-wide accumulation of de novo Arabidopsis thaliana mutations and epimutations**. *Genome Research* 24:11 (2014), 1821–1829. ISSN: 15495469. DOI: 10.1101/gr.177659.114. arXiv: 15334406 (see page 5).

[Jon12] Peter A. Jones. **Functions of DNA methylation: Islands, start sites, gene bodies and beyond**. *Nature Reviews Genetics* 13:7 (2012), 484–492. ISSN: 14710056. DOI: 10.1038/nrg3230 (see page 2).

[JS19] Frank Johannes and Robert J. Schmitz. **Spontaneous epimutations in plants**. *New Phytologist* 221:3 (2019), 1253–1259. ISSN: 14698137. DOI: 10.1111/nph.15434 (see pages 2–5).

[Jüh+16] Frank Jühling, Helene Kretzmer, Stephan H. Bernhart, Christian Otto, Peter F. Stadler, and Steve Hoffmann. **Metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data**. *Genome Research* 26:2 (2016), 256–262. ISSN: 15495469. DOI: 10.1101/gr.196394.115 (see page 5).

[Kis+15] Kamal Kishore, Stefano de Pretis, Ryan Lister, Marco J. Morelli, Valerio Bianchi, Bruno Amati, Joseph R. Ecker, and Mattia Pelizzola. **methylPipe and compEpiTools: A suite of R packages for the integrative analysis of epigenomics data**. *BMC Bioinformatics* 16:1 (2015), 1–11. ISSN: 14712105. DOI: 10.1186/s12859-015-0742-6. URL: http://dx.doi.org/10.1186/s12859-015-0742-6 (see page 5).

[Kom+18] Dossa Komivi, Ali Mmadi Marie, Zhou Rong, Zhou Qi, Yang Mei, Cisse Ndiaga, Diouf Diaga, Wang Linhai, and Zhang Xiurong. **The contrasting response to drought and waterlogging is underpinned by divergent DNA methylation programs associated with transcript accumulation in sesame**. *Plant Science*: July (2018). ISSN: 18732259. DOI: 10.1016/j.plantsci.2018.09.012 (see page 5).

[Kre+20] Clemens Kreutz, Nilay S. Can, Ralf Schulze Bruening, Rabea Meyberg, Zsuzsanna Mérai, Noe Fernandez-Pozo, and Stefan A. Rensing. **A blind and independent benchmark study for detecting differentially methylated regions in plants**. *Bioinformatics* 36:11 (2020), 3314–3321. ISSN: 14602059. DOI: 10.1093/bioinformatics/btaa191 (see page 6).

[LHE18] Chongyuan Luo, Petra Hajkova, and Joseph R. Ecker. **Dynamic DNA methylation: In the right place at the right time**. *Science* 361:6409 (2018), 1336–1340. ISSN: 10959203. DOI: 10.1126/science.aat6806 (see page 1).

[Li+13] Sheng Li, Francine E. Garrett-Bakelman, Altuna Akalin, Paul Zumbo, Ross Levine, Bik L. To, Ian D. Lewis, Anna L. Brown, Richard J. D'Andrea, Ari Melnick, and Christopher E. Mason. **An optimized algorithm for detecting and annotating regional differential methylation**. *BMC Bioinformatics* 14:SUPPL.5 (2013), S10. ISSN: 14712105. DOI: 10.1186/1471-2105-14-S5-S10. URL: http://www.biomedcentral.com/1471-2105/14/S5/S10 (see page 5).

[Li+17] Yan Li, Qingyu Xu, Na Lv, Lili Wang, Hongmei Zhao, Xiuli Wang, Jing Guo, Chongjian Chen, Yonghui Li, and Li Yu. **Clinical implications of genome-wide DNA methylation studies in acute myeloid leukemia**. *Journal of Hematology and Oncology* 10:1 (2017), 1–10. ISSN: 17568722. DOI: 10.1186/s13045-017-0409-z (see page 2).

[Li21] Yuanyuan Li. **Modern Epigenetics Methods in Biological Research**. *Methods* 187 (2021), 104–113. DOI: 10.1016/j.ymeth.2020.06.022 (see page 1).

[Lis+08] Ryan Lister, Ronan C. O'Malley, Julian Tonti-Filippini, Brian D. Gregory, Charles C. Berry, A. Harvey Millar, and Joseph R. Ecker. **Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis**. *Cell* 133:3 (2008), 523–536. ISSN: 15378276. DOI: 10.1016/j.cell.2008.03.029.Highly. arXiv: NIHMS150003 (see page 6).

[Lis+09] Ryan Lister, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, Leonard Lee, Zhen Ye, Que Minh Ngo, Lee Edsall, Jessica Antosiewicz-Bourget, Ron Stewart, Victor Ruotti, A. Harvey Millar, James A. Thomson, Bing Ren, and Joseph R. Ecker. **Human DNA methylomes at base resolution show widespread epigenomic differences**. *Nature* 462:7271 (Nov. 2009), 315. ISSN: 00280836. DOI: 10.1038/NATURE08514. URL: /pmc/articles/PMC2857523/%20/pmc/articles/PMC2857523/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857523/ (see page 6).

[LJ10] Julie A Law and Steven E Jacobsen. **Establishing, maintaining and modifying DNA methylation patterns in plants and animals**. *Nature Reviews Genetics* 11:3 (2010), 204–220. ISSN: 1471-0056. DOI: 10.1038/nrg2719. URL: http://dx.doi.org/10.1038/nrg2719 (see pages 2, 3).

[LT21] Shizhao Li and Trygve O Tollefsbol. **DNA methylation methods: global DNA methylation and methylomic analyses**. *Methods* 187 (2021), 28–43. DOI: 10.1016/j.ymeth.2020.10.002.DNA (see pages 1, 2, 4).

[Man+06] Kenneth Manning, Mahmut Toer, Mervin Poole, Yiguo Hong, Andrew J Thompson, Graham J King, James J Giovannoni, and Graham B Seymour. **A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening**. *Nature Genetics* 38:8 (2006), 948–952 (see page 3).

[MP21] Francesca Mattiroli and Lorenza Penengo. **Histone Ubiquitination: An Integrative Signaling Platform in Genome Stability**. *Trends in Genetics* 37:6 (2021), 566–581. ISSN: 13624555. DOI: 10.1016/j.tig.2020.12.005. URL: https://doi.org/10.1016/j.tig.2020.12.005 (see page 1).

[Ong+15] Meilina Ong-Abdullah, Jared M. Ordway, Nan Jiang, Siew Eng Ooi, Sau Yee Kok, Norashikin Sarpan, Nuraziyan Azimi, Ahmad Tarmizi Hashim, Zamzuri Ishak, Samsul Kamal Rosli, Fadila Ahmad Malike, Nor Azwani Abu Bakar, Marhalil Marjuni, Norziha Abdullah, Zulkifli Yaakub, Mohd Din Amiruddin, Rajanaidu Nookiah, Rajinder Singh, Eng Ti Leslie Low, Kuang Lim Chan, Norazah Azizi, Steven W. Smith, Blaire Bacher, Muhammad A. Budiman, Andrew Van Brunt, Corey Wischmeyer, Melissa Beil, Michael Hogan, Nathan

Lakey, Chin Ching Lim, Xaviar Arulandoo, Choo Kien Wong, Chin Nee Choo, Wei Chee Wong, Yen Yen Kwan, Sharifah Shahrul Rabiah Syed Alwee, Ravigadevi Sambanthamurthi, and Robert A. Martienssen. **Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm**. *Nature* 525:7570 (2015), 533–537. ISSN: 14764687. DOI: 10.1038/nature15365 (see page 3).

[Oss+10] Stephan Ossowski, Korbinian Schneeberger, José Ignacio Lucas-Lledó, Norman Warthmann, Richard M. Clark, Ruth G. Shaw, Weigel Michael, and Detlef Lynch. **The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana**. *Science* 327:5961 (2010), 1–9. DOI: 10.1126/science.1180677.The (see page 4).

[Par+14] Yongseok Park, Maria E. Figueroa, Laura S. Rozek, and Maureen A. Sartor. **MethylSig: A whole genome DNA methylation analysis pipeline**. *Bioinformatics* 30:17 (2014), 2414–2422. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu339 (see page 5).

[Pet+21] Timothy J. Peters, Michael J. Buckley, Yunshun Chen, Gordon K. Smyth, Christopher C. Goodnow, and Susan J. Clark. **Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate**. *Nucleic Acids Research* 49:19 (2021). ISSN: 13624962. DOI: 10.1093/nar/gkab637 (see page 5).

[Rak+11] Vardhman K. Rakyan, Thomas A. Down, David J. Balding, and Stephan Beck. **Epigenome-wide association studies for common human diseases**. *Nature Reviews Genetics 2011 12:8* 12:8 (July 2011), 529–541. ISSN: 1471-0064. DOI: 10.1038/nrg3000. URL: https://www.nature.com/articles/nrg3000 (see page 5).

[Sar20] Peter Sarkies. **Molecular mechanisms of epigenetic inheritance: Possible evolutionary implications**. *Seminars in Cell and Developmental Biology* 97:June 2019 (2020), 106–115. ISSN: 10963634. DOI: 10.1016/j.semcdb.2019.06.005. URL: https://doi.org/10.1016/j.semcdb.2019.06.005 (see page 5).

[SBD00] Ruth G Shaw, Diane L Byers, and Elizabeth Darmo. **Spontaneous Mutational Effects on Reproductive Traits of**. *Genetics* 155 (2000), 369–378 (see page 4).

[Sch+15] Matthew D Schultz, Yupeng He, John W Whitaker, Manoj Hariharan, Eran A Mukamel, Danny Leung, Nisha Rajagopal, Joseph R Nery, Mark A Urich, Huaming Chen, Shin Lin, Yiing Lin, Inkyung Jung, and Anthony D Schmitt. **Human body epigenome maps reveal noncanonical DNAmethylation variation**. *Nature* 523:7559 (2015), 212–216. DOI: 10.1038/nature14465. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4499021/pdf/nihms-681054.pdf (see page 5).

[Sha+20] Yadollah Shahryary, Aikaterini Symeonidi, Rashmi Hazarika, Johanna Denkena, Talha Mubeen, Brigitte Hofmeister, Thomas van Gurp, Maria Colomé-Tatché, Koen Verhoeven, Gerald Tuskan, Robert Schmitz, and Frank Johannes. **AlphaBeta: Computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants**. *Genome biology* 21:1 (2020), 1–22. DOI: 10.1101/862243 (see pages 4, 11, 15).

[Sha21] Jasmine M. Shah. **Epimutations and mutations, nurturing phenotypic diversity**. *Genetica*: 0123456789 (2021). ISSN: 15736857. DOI: 10.1007/s10709-021-00124-8. URL: https://doi.org/10.1007/s10709-021-00124-8 (see page 3).

[Son+13] Qiang Song, Benjamin Decato, Elizabeth E. Hong, Meng Zhou, Fang Fang, Jianghan Qu, Tyler Garvin, Michael Kessler, Jun Zhou, and Andrew D. Smith. **A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics**. *PLoS ONE* 8:12 (2013). ISSN: 19326203. DOI: 10.1371/journal.pone.0081148 (see page 5).

[Sri+19] Akanksha Srivastava, Yuliya V. Karpievitch, Steven R. Eichten, Justin O. Borevitz, and Ryan Lister. **HOME: A histogram based machine learning approach for effective identification of differentially methylated regions**. *BMC Bioinformatics* 20:1 (2019), 1–15. ISSN: 14712105. DOI: 10.1186/s12859-019-2845-y (see page 5).

[STM14] Yutaka Saito, Junko Tsuji, and Toutai Mituyama. **Bisulfighter: Accurate detection of methylated cytosines and differentially methylated regions**. *Nucleic Acids Research* 42:6 (2014). ISSN: 13624962. DOI: 10.1093/nar/gkt1373 (see page 5).

[Su+13] Jianzhong Su, Haidan Yan, Yanjun Wei, Hongbo Liu, Hui Liu, Fang Wang, Jie Lv, Qiong Wu, and Yan Zhang. **CpG-MPs: Identification of CpG methylation patterns of genomic regions from high-throughput bisulfite sequencing data**. *Nucleic Acids Research* 41:1 (2013), 1–15. ISSN: 03051048. DOI: 10.1093/nar/gks829 (see page 5).

[Sue+16] Mónica Suelves, Elvira Carrió, Yaiza Núñez-Álvarez, and Miguel A. Peinado. **DNA methylation dynamics in cellular commitment and differentiation**. *Briefings in Functional Genomics* 15:6 (2016), 443–453. ISSN: 20412657. DOI: 10.1093/bfgp/elw017 (see page 2).

[Sun+14] Deqiang Sun, Yuanxin Xi, Benjamin Rodriguez, Hyun J. Park, Pan Tong, Mira Meong, Margaret A. Goodell, and Wei Li. **MOABS: Model based analysis of bisulfite sequencing data**. *Genome Biology* 15:2 (2014), 1–12. ISSN: 1474760X. DOI: 10.1186/gb-2014-15-2-r38 (see page 5).

[SY16] Shuying Sun and Xiaoqing Yu. **HMM-Fisher: identifying differential methylation using a hidden Markov model and Fisher's exact test**. *Statistical Applications in Genetics and Molecular Biology* 15:1 (2016), 55–67. DOI: doi:10.1515/sagmb-2015-0076. URL: https://doi.org/10.1515/sagmb-2015-0076 (see page 6).

[Tra+18] Hong Tran, Hongxiao Zhu, Xiaowei Wu, Gunjune Kim, Christopher R. Clarke, Hailey Larose, David C. Haak, Shawn D. Askew, Jacob N. Barney, James H. Westwood, and Liqing Zhang. **Identification of differentially methylated sites with weak methylation effects**. *Genes* 9:2 (2018). ISSN: 20734425. DOI: 10.3390/genes9020075 (see page 5).

[Vid+16] Amaryllis Vidalis, Daniel Zivkovic, René Wardenaar, David Roquis, Aurélien Tellier, and Frank Johannes. **Methylome evolution in plants**. *Genome Biology* 17 (2016), 1–14. DOI: 10.1186/s13059-016-1127-5 (see page 2).

[Wan+15] Zhen Wang, Xianfeng Li, Yi Jiang, Qianzhi Shao, Qi Liu, Bing Yu Chen, and Dongsheng Huang. **SwDMR: A sliding window approach to identify differentially methylated regions based on whole genome bisulfite sequencing**. *PLoS ONE* 10:7 (2015). ISSN: 19326203. DOI: 10.1371/journal.pone.0132866 (see pages 5, 6).

[Wen+16] Yalu Wen, Fushun Chen, Qingzheng Zhang, Yan Zhuang, and Zhiguang Li. **Detection of differentially methylated regions in whole genome bisulfite sequencing data using local Getis-Ord statistics**. *Bioinformatics* 32:22 (2016), 3396–3404. ISSN: 14602059. DOI: 10.1093/bioinformatics/btw497 (see page 5).

[Wu+10] Hao Wu, Brian Caffo, Harris A. Jaffee, Rafael A. Irizarry, and Andrew P. Feinberg. **Redefining CpG islands using hidden Markov models**. *Biostatistics* 11:3 (2010), 499–514. DOI: 10.1093/biostatistics/kxq005 (see page 6).

[Wu+15] Hao Wu, Tianlei Xu, Hao Feng, Li Chen, Ben Li, Bing Yao, Zhaohui Qin, Peng Jin, and Karen N. Conneely. **Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates**. *Nucleic Acids Research* 43:21 (2015), 1–9. ISSN: 13624962. DOI: 10.1093/nar/gkv715 (see page 5).

[WZ17] Xiaoji Wu and Yi Zhang. **TET-mediated active DNA demethylation: Mechanism, function and beyond**. *Nature Reviews Genetics* 18:9 (2017), 517–534. ISSN: 14710064. DOI: 10.1038/nrg.2017.33. URL: http://dx.doi.org/10.1038/nrg.2017.33 (see page 2).

[YQC21] Jianhong Yang, Qiang Qiu, and Lijuan Chen. **Histone Mutations and Cancer**. Vol. 1283. 2021, 85–96. ISBN: 9789811581038 (see page 1).

[YS16] Xiaoqing Yu and Shuying Sun. **HMM-DM: identifying differentially methylated regions using a hidden Markov model**. *Statistical Applications in Genetics and Molecular Biology* 15:1 (2016), 69–81. DOI: doi:10.1515/sagmb-2015-0077. URL: https://doi.org/10.1515/sagmb-2015-0077 (see page 6).

[Zhe+17] Xiaoguo Zheng, Liang Chen, Hui Xia, Haibin Wei, Qiaojun Lou, Mingshou Li, Tiemei Li, and Lijun Luo. **Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition**. *Scientific Reports* 7:November 2016 (2017), 1–13. ISSN: 20452322. DOI: 10.1038/srep39843. arXiv: 0901.2517. URL: http://dx.doi.org/10.1038/srep39843 (see page 5).

[ZLZ18] Huiming Zhang, Zhaobo Lang, and Jian-kang Zhu. **Dynamics and function of DNA methylation in plants**. *Nature Reviews Molecular Cell Biology* 19:August (2018), 489–509. ISSN: 1471-0080. DOI: 10.1038/s41580-018-0016-z. URL: http://dx.doi.org/10.1038/s41580-018-0016-z (see page 5).

Appendix A: Paper III

Yadollah Shahryary, Aikaterini Symeonidi, Rashmi R. Hazarika, **Johanna Denkena**, Talha Mubeen, Brigitte Hofmeister, Thomas van Gurp, Maria Colomé-Tatché, Koen J.F. Verhoeven, Gerald Tuskan, Robert J. Schmitz & Frank Johannes
AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants
Genome Biology 21, 260 (2020)
https://doi.org/10.1186/s13059-020-02161-6

My contributions to this paper are detailed in chapter 2.3

Acknowledgments

I extend my deepest gratitude to those whose support and encouragement have been instrumental in the completion of this doctoral thesis.

First and foremost, I express my sincere appreciation to my supervisor, Maria Colomé-Tatché, whose guidance, expertise, and patience have been invaluable throughout this journey. Your mentorship has not only shaped the academic direction of this research but has also inspired personal and professional growth.

I am grateful to Frank Johannes as well as the members of my thesis committee, Matthias Heinig and Sebastian Bultmann, for their insightful feedback and constructive criticism. Their diverse perspectives enriched the quality of this work and contributed significantly to its refinement.

Special thanks are due to my colleagues and peers at the ICB, specifically the Computational Epigenomics work group, who fostered a stimulating intellectual environment. The exchange of ideas and collaborative spirit greatly enriched my research experience. But of course I also feel incredibly lucky to have gotten to know some of you on a personal level. Especially Anna Danese and Akshaya Ramakrishnan I want to thank for sharing this challenging journey with me, for fun trips, great food and passionate discussions about science, politics or pop culture over (not so) late night drinks.

Heartfelt appreciation goes to my family for their unwavering emotional and financial support. Your belief in my abilities sustained me during the challenging moments, and I am profoundly grateful for your love and encouragement.

To my partner Christian who provided a listening ear, words of encouragement and occasional distractions, thank you for being a source of strength and joy.

This dissertation represents the culmination of years of dedication and hard work, and it would not have been possible without the support of these remarkable individuals. Thank you for being an integral part of this academic journey.