DNA methylation analysis in plant giga-genomes: comparing two bisulfite sequencing techniques in *Abies alba* trees affected by dieback

Isabel García-García [1*], Belén Méndez-Cea [2], Jose Luis Horreo [1], Juan Carlos Linares [2], Francisco Javier Gallego [1]

[1] Dpto de Genética, Fisiología y Microbiología, UD Genética, Facultad de CC Biológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

[2] Dpto de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, 41013 Sevilla, Spain

* Corresponding Author: Isabel García-García, E-mail: isabel06@ucm.es

Abstract

Phenotypic plasticity is a fundamental mechanism that enables plants to adapt to shifting environmental conditions, such as those induced by climate change. Epigenetic modifications, notably DNA methylation, may play a pivotal role in such process. However, this field remains largely unstudied in non-model organisms with large, complex genomes. Here, we focus on silver fir (*Abies alba*), more precisely on a natural population subjected to climate stress, comparing the results obtained from two different bisulfite sequencing techniques in the study of the epigenetic patterns of its giga-genome. DNA was extracted from two non-declining and two declining *A. alba* individuals and subjected to whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS). General methylation levels in each cytosine context (CpG, CHG and CHH, where H stands for A, C, or T) were determined and differentially methylated regions (DMRs) were identified. WGBS sequenced a higher amount of cytosines, while RRBS had a higher number of CpG sites passing filtering conditions. Both techniques showed high methylation percentages in CpG and CHG contexts, but disagreed on CHH sites, likely due to RRBS having a bias towards highly methylated CHH contexts. Differentially methylated sites were only identified using WGBS data but could not be functionally annotated. In conclusion, while successful in methylation assessment, neither WGBS nor RRBS produced the desired outcome regarding the identification of DMRs. Thus, the fragmented nature of the available reference genomes for species with large genomes seems to hinder the attainment of meaningful results from WGBS and RRBS.

Keywords: *Whole Genome Bisulfite Sequencing (WGBS); Reduced Representation Bisulfite Sequencing (RRBS); Non-model organism; Epigenetic; Forest decline; Abies alba*

Introduction

Phenotypic plasticity is the ability of a single genotype to produce different phenotypes when exposed to different environmental conditions (Pigliucci et al. 2006). Thus, it can serve as a safeguard against rapid climate variations and aid in quick adaptation (Nicotra et al. 2010), which could be a crucial mechanism for species survival to ongoing climate change. In fact, phenotypic plasticity is recognized as a key process through which plants can cope with climate change (Stots et al. 2021). They are able to undergo dynamic physiological and molecular changes to adapt to a changing environment, in which modifications at the epigenetic level are involved (Dar et al. 2022), namely DNA methylation, histone modifications and the action of non-coding RNAs.

DNA methylation, usually understood as an addition of a methyl group onto a cytosine to form 5-methylcytosine (Bartels et al. 2018), is a widely studied epigenetic mark in plants (reviewed in Zhang et al. 2018). DNA methylation is known to play a crucial role in various biological processes, including silencing of transposable elements (TEs), maintaining genome stability, regulation of gene transcription, and stress tolerance (Kumar and Mohapatra, 2021). Specifically, plants methylate cytosines in three different contexts: CpG, CHG and CHH (where H = A, T or G) (Takuno et al. 2016). These different contexts are usually studied through bisulfite sequencing, which is considered the gold standard technique to assess DNA methylation (Kurdyukov and Bullock, 2016). Whole genome bisulfite sequencing (WGBS) stands out as the most exhaustive method among these techniques, as it enables the assessment of DNA methylation throughout the genome (Suzuki et al. 2018). However, such extensive sequencing might not be necessary given that only certain genomic regions can undergo differential methylation. Thus, a variation of this approach, called reduced representation bisulfite sequencing (RRBS), was also developed. Hence, RRBS uses an enzymatic digestion to reduce the

DOI:10.2478/sg-2024-0020
edited by the Thünen Institute of Forest Genetics

targeted portion of the genome, enriched in CpG content (Beck et al. 2022), offering a more cost-effective approach and enhancing sequencing coverage (Kurdyukov and Bullock, 2016).

Employing these techniques to study methylation patterns in non-model species can be specially challenging due to the lack of high-quality references and annotations. This task can become even more complex when dealing with large genomes, such as those of conifers (García-García et al. 2022), with sizes ranging from ~6.5 to ~37 Gb, being ~15 Gb the most common 1C-value (Ahuja and Neale, 2005). Indeed, only three species of the Pinaceae family (Picea abies, Pinus tabuliformis and Pseudotsuga menziesii), which are among the few conifers with reference genome, have had their methylomes studied so far (Ausin et al. 2016; Niu et al. 2022; Vu et al. 2023). P. abies and P. menziesii had their reference genomes pre-processed before the bioinformatic analysis in order to reduce the computational costs derived from their highly fragmented state (~9–10 million scaffolds each), being subjected to a merging step or a filtering step to keep contigs with high continuity. On the other hand, P. tabuliformis is the only conifer with a chromosome-level reference genome, facilitating any analysis based on it. Here, we analyze cytosine methylation in silver fir, *Abies alba*, which is the only species of the genus *Abies* with an available reference genome. This assembly remains in its initial version and is highly fragmented, consisting of approximately 37 million scaffolds (available at https://treegenesdb.org/FTP/Genomes/Abies_alba/v1.1/, last accessed in February 2025). In the present study, we focus on a natural silver fir population from the Pyrenees, which shows signs of drought-induced forest decline and dieback (Linares & Camarero, 2012; González de Andrés et al. 2022). As no genetic differentiation was found between declining and non-declining individuals in a previous study (García-García et al. 2023), we hypothesized that epigenetic processes could be contributing to the phenotypic diversity observed across vigor classes. We show results obtained from both whole WGBS and RRBS and compare the ability of these techniques to produce data with enough quality to (1) assess general methylation levels of each methylation context (CpG, CHG and CHH) and (2) allow the identification of DMRs between declining and non-declining trees in silver fir forests.

Materials and Methods

Plant material

Leaf samples were collected from two declining *A. alba* trees (aged 73 and 81 years) and two non-declining trees (aged 91 and 106 years), sampled in the Western Spanish Pyrenees (Paco Mayor, 42.71ºN, 0.64ºW), where the species has been showing long-term symptoms of dieback (González de Andrés et al. 2022; García-García et al. 2024). Tree vigor was assessed based on visual estimation of crown transparency (Dobbertin, 2005; González de Andrés et al. 2022). In order to minimize potential confounding variables and ensure the reliability of epigenomic analyses, all samples were collected during the same morning under consistent weather conditions (cloud-covered sky, snowing), thereby maintaining environmental uniformity across the experimental set. Two samples per vigor group were used for this pilot study to explore preliminary outcomes before choosing the best sequencing methodology for a large-scale investigation.

Bisulfite sequencing and data analysis

Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen®, Germany) and subjected to WGBS and RRBS (enzymatic digestion with MspI and TaqI). The obtained raw reads underwent quality and adapter trimming using Trim Galore v0.5.0 (The Babraham Institute), with default options, before being mapped to the complete *A. alba* reference genome v1.1, including its ~37 million scaffolds (Mosca et al. 2019), using Bismark v0.21.0 (Krueger and Andrews, 2011) with default options for single-end (SE) reads. The SE mode was chosen due to two reasons: (1) the RRBS data was generated through single-read sequencing as RRBS libraries are made of short fragments (50 bp) which do not require paired-end (PE) sequencing to achieve high accuracy levels, and (2) Bismark is a very strict aligner which only reports concordant PE reads with end-to-end mapping, which is acceptable for model organism and high-quality DNA but can be problematic for non-model organisms.

Bismark was then used to perform the methylation calling step, producing files containing information about the methylation state of the sequenced cytosines in each sample, which allowed to calculate their general methylation levels in each context (CpG, CHG and CHH). Yates' continuity corrected chi-squared test was used to identify differences between vigor classes using GraphPad Prism v8.2.1 for Windows (GraphPad Software, Boston, Massachusetts USA, http://www.graphpad.com).

Only CpG coverage files were used for subsequent analyses, as CpG sites are usually located near/in genes, having effects on their expression, and they need less computational resources to be studied than CHG and CHH sites. Finally, edgeR v3.34.0 (Robinson et al. 2010) was used to identify DMRs, following the steps indicated in the edgeR user's guide: filtering (only sites which were not always methylated or unmethylated and had at least 5 counts per sample were kept), normalization, dispersion estimation, negative binomial generalized lineal model fitting, and performance of a likelihood ratio test with 5 % false discovery rate. Finally, either the available *A. alba* annotation or NCBI's blastn (Altschul et al. 1990) was used to identify nearby genomic features. The datasets for the WGBS and RRBS pilot phases generated and analyzed during the current study are available from the corresponding author on request.

Results

The amount of sequenced cytosines (Cs) was approximately 5.5 times higher for the WGBS approach (Table 1), which is an expected outcome as the sequenced fraction of a genome should be larger using a whole-genome technique than using

a reduced-representation technique. However, the number of CpG sites that passed the filtering conditions was higher (~4.5 times) in RRBS sequences (Table 1), as this technique increases the sequencing depth of the targeted regions, making it easier to obtain at least 5 counts per sample.

Regarding general methylation levels, both techniques showed high methylation percentages in CpG and CHG contexts (although significant differences were found between the results obtained from WGBS and RRBS data, as shown in Table S1), while they strongly disagree on CHH sites (Fig 1). Significant differences between declining and non-declining trees were found in all contexts except for CpG sites when studied through RRBS (Tables S2, S3). Despite the higher amount of high quality CpG sites identified in the RRBS data, differentially methylated sites were only found in the WGBS data (Table 1). 14 CpG sites were found to have significant higher methylation levels in the declining group. 13 of these CpG sites are found in the same region (scaffold 5,292,684 : 43–148), which constitute a DMR. Unfortunately, this region is part of a 180 bp-long scaffold, which does not contain any annotations and shows no similarity with any of the sequences in the NCBI database.

Table 1
Comparison of the WGBS and the RRBS results, including methylation calling, filtering, and the identification of differentially methylated sites.

	WGBS	**RRBS**
Methylation calling	~850 million Cs / sample	~150 million Cs / sample
Filtering	10,250 CpG sites	47,674 CpG sites
Differential methylation	14 CpG sites	0 CpG sites

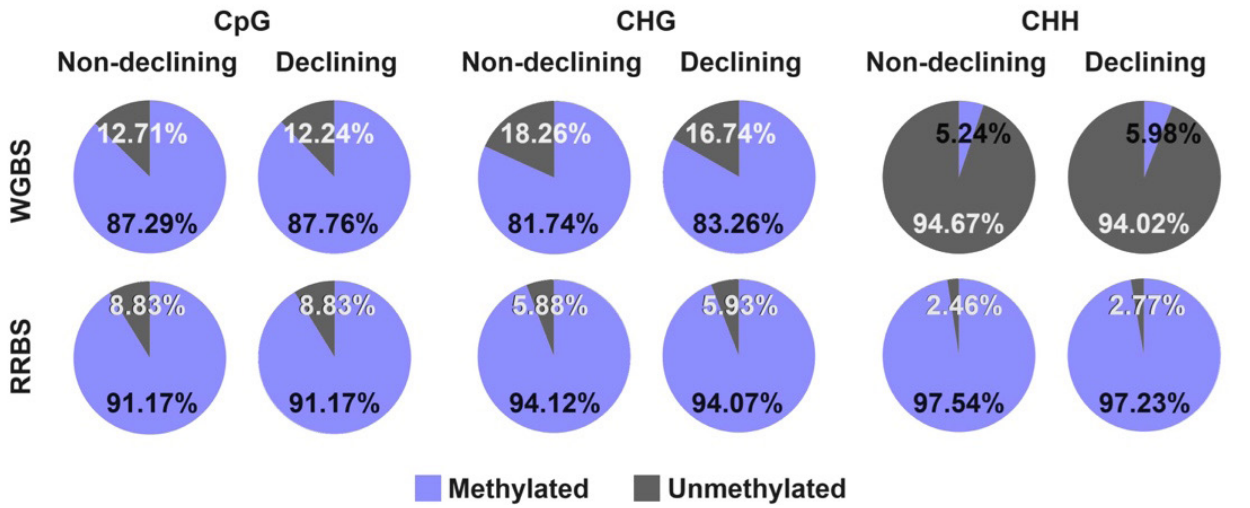

Figure 1
General methylation levels found in each vigor class (non-declining and declining trees) in each context (CpG, CHG, CHH) obtained from WGBS and RRBS data

Discussion

While both techniques showed high methylation percentages in CpG and CHG contexts, WGBS and RRBS disagreed on CHH sites. Methylated CHH islands are likely to appear close to genes with nearby TEs (Martin et al. 2021), flanking them and playing a role in mitigating their adverse effects in gene expression (Kenchanmane Raju et al. 2019). In addition, conifer species have many ultra-long genes with TE insertions, which create lengthy introns which are likely to be heavily methylated (Li et al. 2023). Therefore, as RRBS targets CpG-rich regions (Beck et al. 2022), which are generally associated with genes (Ashikawa, 2021), the results obtained using this technique could be biased towards highly methylated CHH contexts. Thus, the percentage estimated from the WGBS data could be a more realistic representation of the global levels of methylation in CHH areas in *A. alba* as they were sequenced at random throughout the whole genome. Hence, our WGBS results seem to be in line with previous studies that report lower methylation levels in CHH sites than in CpG and CHG contexts (Bartels et al. 2018), while our RRBS results agree with the hypothesis of higher methylation levels within the CHH regions situated in the proximity of genes.

Significant methylation differences between declining and non-declining trees were observed across all contexts (except for CpG sites in RRBS analyses). However, significant results are easy to find in contingency tests when working with high values (millions of cytosines in each test). For example, the 0.05 % difference in methylation in CHG contexts using the RRBS data produces a highly significant result (p-value < 0.0001), which might seem odd, but considering the number of cytosines included in this percentage (~20,000), this outcome does not appear to be illogical. Thus, statistical significance should be combined with the identification of DMRs and their functional annotation to obtain meaningful biological significance about the role of these differently methylated sites in the context of drought-induced decline of silver fir forests, which could be involved in processes such as adaptation to changing environments (e.g. Sáez-Laguna et al. 2014; Alakärppä et al. 2018; García-García et al. 2024).

Finally, although a DMR could be identified using the WGBS data, it could not be functionally annotated, hindering the biological interpretation of this finding. The *A. alba* genome sequence is composed of 37 million scaffolds, sorted by descending size, of which the last 35 million have a length of less than 500 bp. In other words, instead of having a continuous sequence for each of the 12 chromosomes (2n = 24) present in its genome (Puizina et al. 2008), it is fragmented into small pieces. This lack of continuity adds difficulty to the task of finding genes and other genetic structures, making it almost unattainable and highlighting the need of improvement of both the reference genome and its annotation, which is ~3 times more fragmented than those available for other conifers that have had their methylomes studied. It also complicates the use of candidate regions proposed in other species, as they may not be conserved in *A. alba* or hold the same biological significance. Additionally, identifying methylation contexts that could regulate the expression of previously described genes involved in processes such as local adaptation in this species is challenging, as this fragmentation results in short scaffolds, limiting access to upstream and downstream regions. For these reasons, while thousands of DMRs are found experiments with model species, such as Arabidopsis thaliana (Huang et al. 2016) and Lycopersicon esculentum (Tian et al. 2021), this task is much more complicated in giga-genomes without high quality references, and using samples of the same tissue, like the ones used in this study. For example, several hundreds of DMRs were found in CpG contexts comparing needles with somatic embryogenic cultures in P. abies (Ausin et al. 2016), while only 83 DMRs were found comparing needles from different trees in P. menziesii (Vu et al. 2023), likely due to the existence of more diverse methylation patterns in different types of sample.

Conclusions

Significant differences in DNA methylation levels were observed between declining and non-declining trees, but their biological significance remains unclear due to the lack of functional annotation. Thus, reference genomes might not be enough to obtain meaningful results from WGBS or RRBS data when they are highly fragmented and not fully annotated, which is, unfortunately, the case of the majority of the plant species with large genomes. Overall, chromosome-level references seem to be fundamental to carry out successful single-base resolution DNA methylation studies in plant species with giga-genomes. Alternatively, there are other methods that enable the detection of general methylation percentages at a lower cost, such as those based on colorimetric quantification via enzyme-linked immunosorbent assay (ELISA), which quantify global DNA methylation, regardless of cytosine context, by measuring levels of 5-methylcytosine (e.g. García-García et al. 2024), or reversed-phase ultra-performance liquid chromatography (UPLC) coupled with sensitive mass spectrometry (MS) (Boulias and Greer, 2021). Additionally, if previously identified genomic regions involved in relevant biological processes are available for the species under study, targeted bisulfite sequencing of these specific loci could yield insightful results.

Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation (MICINN) [grant numbers TED2021-129770B-C22 and PID2021-123675OB-C44].

References

Ahuja MR, Neale DB (2005) Evolution of Genome Size in Conifers. Silvae Genetica 54(3):126–137. https://doi.org/10.1515/sg-2005-0020

Alakärppä E, Salo HM, Valledor L, et al (2018) Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. Journal of Experimental Botany 69(21):5293–5305. https://doi.org/10.1093/jxb/ery292

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ashikawa I (2001) Gene-associated CpG islands in plants as revealed by analyses of genomic sequences. The Plant Journal 26(6):617–25. https://doi.org/10.1046/j.1365-313x.2001.01062.x

Ausin I, Feng S, Yu C, et al (2016) DNA methylome of the 20-gigabase Norway spruce genome. Proceedings of the National Academy of Science of the USA 113(50):E8106–E8113. https://doi.org/10.1073/pnas.1618019113

Bartels A, Han Q, Nair P, et al (2018) Dynamic DNA Methylation in Plant Growth and Development. International Journal of Molecular Sciences 19(7):2144. https://doi.org/10.3390/ijms19072144

Beck D, Ben Maamar M, Skinner MK (2022) Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons. Epigenetics 17(5):518–530. https://doi.org/10.1080/15592294.2021.1924970

Boulias K, Greer EL (2021) Detection of DNA Methylation in Genomic DNA by UHPLC-MS/MS. Methods in Molecular Biology 2198:79–90. https://doi.org/10.1007/978-1-0716-0876-0_7

Dar FA, Mushtaq NU, Saleem S, et al (2022) Role of Epigenetics in Modulating Phenotypic Plasticity against Abiotic Stresses in Plants. International Journal of Genomics 2022:1092894. https://doi.org/10.1155/2022/1092894

Dobbertin M (2005) Tree growth as indicator of tree vitality and of tree reaction to environmental stress: A review. European Journal of Forest Research 124:319–333. https://doi.org/10.1007/s10342-005-0085-3

García-García I, Méndez-Cea B, Gallego FJ, et al (2024) Genomic insights into climate change-induced forest dieback in Abies alba hotspots of decline. European Journal of Forest Research. https://doi.org/10.1007/s10342-024-01737-2

García-García I, Méndez-Cea B, González de Andrés E, et al (2023) Climate and Soil Microsite Conditions Determine Local Adaptation in Declining Silver Fir Forests. Plants 12(14):2607. https://doi.org/10.3390/plants12142607

García-García I, Méndez-Cea B, Martín-Gálvez D, et al (2022) Challenges and Perspectives in the Epigenetics of Climate Change-Induced Forests Decline. Frontiers in Plant Science 12:797958. https://doi.org/10.3389/fpls.2021.797958

González de Andrés E, Gazol A, Querejeta JI, et al (2022) The role of nutritional impairment in carbon-water balance of silver fir drought-induced dieback. Global Change Biology 28(14):4439–4458. https://doi.org/10.1111/gcb.16170

Huang XY, Chao DY, Koprivova A, et al (2016) Nuclear Localised MORE SULPHUR ACCUMULATION1 Epigenetically Regulates Sulphur Homeostasis in Arabidopsis thaliana. PLoS Genetics 12(9):e1006298. https://doi.org/10.1371/journal.pgen.1006298

Kenchanmane Raju SK, Ritter EJ, Niederhuth CE (2019) Establishment, maintenance, and biological roles of non-CG methylation in plants. Essays in Biochemistry 63(6):743–755. https://doi.org/10.1042/EBC20190032

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11):1571–2. https://doi.org/10.1093/bioinformatics/btr167

Kumar S, Mohapatra T (2021) Dynamics of DNA Methylation and Its Functions in Plant Growth and Development. Frontiers in Plant Science 12:596236. https://doi.org/10.3389/fpls.2021.596236

Kurdyukov S, Bullock M (2016) DNA Methylation Analysis: Choosing the Right Method. Biology 5(1):3. https://doi.org/10.3390/biology5010003

Li J, Han F, Yuan T, et al (2023) The methylation landscape of giga-genome and the epigenetic timer of age in Chinese pine. Nature Communications 14(1):1947. https://doi.org/10.1038/s41467-023-37684-6

Linares JC, Camarero JJ (2012) Silver Fir Defoliation Likelihood Is Related to Negative Growth Trends and High Warming Sensitivity at Their Southernmost Distribution Limit. ISRN Forestry 2012:437690. https://doi.org/10.5402/2012/437690

Martin GT, Seymour DK, Gaut BS (2021) CHH Methylation Islands: A Nonconserved Feature of Grass Genomes That Is Positively Associated with Transposable Elements but Negatively Associated with Gene-Body Methylation. Genome Biology and Evolution 13(8):evab144. https://doi.org/10.1093/gbe/evab144

Mosca E, Cruz F, Gómez-Garrido J, et al (2019) A Reference Genome Sequence for the European Silver Fir (Abies alba Mill.): A Community-Generated Genomic Resource. Genes Genome Genetics (G3) 9(7):2039–2049. https://doi.org/10.1534/g3.119.400083

Nicotra AB, Atkin OK, Bonser SP, et al (2010) Plant phenotypic plasticity in a changing climate. Trends in Plant Science 15(12):684–92. https://doi.org/10.1016/j.tplants.2010.09.008

Niu S, Li J, Bo W, et al (2022) The Chinese pine genome and methylome unveil key features of conifer evolution. Cell 185(1):204–217.e14. https://doi.org/10.1016/j.cell.2021.12.006

Pigliucci M, Murren CJ, Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. Journal of Experimental Biology 209:2362–7. https://doi.org/10.1242/jeb.02070

Puizina J, Sviben T, Krajacić-Sokol I, et al (2008) Cytogenetic and molecular characterization of the Abies alba genome and its relationship with other members of the Pinaceae. Plant Biology 10(2):256–67. https://doi.org/10.1111/j.1438-8677.2007.00018.x

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616

Sáez-Laguna E, Guevara MÁ, Díaz LM, et al (2014) Epigenetic variability in the genetically uniform forest tree species Pinus pinea L. PLoS One 9(8):e103145. https://doi.org/10.1371/journal.pone.0103145

Stotz GC, Salgado-Luarte C, Escobedo VM, et al (2021) Global trends in phenotypic plasticity of plants. Ecology Letters 24(10):2267–2281. https://doi.org/10.1111/ele.13827

Suzuki M, Liao W, Wos F, et al (2018) Whole-genome bisulfite sequencing with improved accuracy and cost. Genome Research 28(9):1364–1371. https://doi.org/10.1101/gr.232587.117

Takuno S, Ran JH, Gaut B (2016) Evolutionary patterns of genic DNA methylation vary across land plants. Nature Plants 2:15222. https://doi.org/10.1038/nplants.2015.222

Vu GTH, Cao HX, Hofmann M, et al (2023) Uncovering epigenetic and transcriptional regulation of growth in Douglas-fir: identification of differential methylation regions in mega-sized introns. Plant Biotechnology Journal 22(4):863–875. https://doi.org/10.1111/pbi.14229

Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nature Reviews Molecular Cell Biology 19(8):489–506. https://doi.org/10.1038/s41580-018-0016-z

Zuo J, Wang Y, Zhu B, et al (2018) Comparative Analysis of DNA Methylation Reveals Specific Regulations on Ethylene Pathway in Tomato Fruit. Genes 9(5):266. https://doi.org/10.3390/genes9050266