**DNA methylation**

Methylation of cytosine bases is involved in defining regions of chromatin, silencing transposons, maintaining genome integrity, and can affect gene expression. DNA methylation itself is shaped by the factors such as gene expression and the underlying sequence [1]. To understand how DNA methylation evolves during polyploidy and contributes to subgenome dominance, DNA methylation was assessed using whole-genome bisulfite sequencing (WGBS)[2, 3]. In plants, three sequence contexts are generally recognized for DNA methylation, depending on second and third bases downstream of the methylated cytosine: CG (or CpG), CHG, and CHH (where H=A, T, or C). Methylation of these contexts are established and maintained by different molecular pathways and their associations with gene expression can differ based on the pattern of DNA methylation, hence they are typically analyzed separately [1].

All samples, including the parental TO1000 and IMB218 were mapped to a combined genome of both *B. oleracea* and *B. rapa.* As a further control and to account for any biases from mapping the two parental samples were also combined to mimic the allopolyploid *in silico* (referred to as "mock" sample) (see Methods). Total methylation levels summarize methylation for all sites within a genome and are determined by a number of factors, such as genome size and TE content [4-6]. These were calculated for the combined genome and each sub-genome (Figure XX). CG methylation showed little change in the resynthesized lines, regardless of generation, and typically were comparable to levels in the mock allopolyploid or to the highest methylated parent. This was true for the individual subgenomes as well. In general, TO1000 was more highly methylated than IMB218 and this is reflected in the subgenomes of the allopolyploids as well.

More variance in total methylation was observed for the non-CG contexts. In the first generation, CHG methylation was lower than would be expected for the entire genome, based on the mock genome sample and were more comparable to levels in the less-methylated IMB218 parent. However, when compared at the subgenome level, both subgenomes showed reduced CHG methylation in the first generation compared to either parent. Variation in CHG methylation increased in generations five and ten, with some samples increasing in methylation level, while others remaining similar to the first generation. CHH methylation levels were even more striking. For the combined genome, a slight in CHH methylation was observed in generation one and this trend became more pronounced in generations 5 and 10. The increased CHH methylation in generation one occurred mostly on the *B. oleracea* subgenome, but in later generations it increased in both subgenomes. Sample 600 was the sole exception in this trend; CHH methylation was slightly higher in the first generation, increased significantly in the fifth generation, but dropped again in the tenth.

When looking at all genes and Long-Terminal Repeat (LTR) TEs, CG methylation typically lie between the two parents. First generation allopolyploids often show the least variance and closely resemble the mock allopolyploid. Later generations show increased variance, and higher levels, but these rarely exceed that of the TO1000. At a subgenome level, however, a number of samples show increased CG methylation in genes and flanking regions for *B. oleracea* genes, but little difference is observed for *B. rapa*. Increases in *B. oleracea* genic CG methylation occurred primarily in non-syntenic genes, but for flanking regions occurred for both syntenic and non-syntenic genes. For CHG, methylation levels first generation samples again showed lower levels than both in and in flanking regions of both genes and LTRs, increasing in some subsequent generations to parental and mock levels.

CHH methylation again showed the most striking differences. First generation allopolyploids showed little difference from the parents, when all genes were examined, however, when examined by subgenome, *B. oleracea* genes show increased CHH methylation in flanking regions, while *B. rapa* genes did not. In subsequent generations, CHH methylation increased in both LTRs and gene flanking regions. For *B. oleracea* genes, CHH methylation of flanking regions increased beyond the initial increase in the first generation and in addition, increased CHH methylation was now found in flanking regions of *B. rapa* genes. These changes in flanking CHH methylation occurred for both syntenic and non-syntenic genes.

**Methods**

**DNA methylation analysis**

The genomes of both *B. oleracea* TO1000 (EnsemblPlants 43[7])[8] and *B. rapa* R500 were combined into a single fasta file to create an *in silico* allopolyploid genome. Whole genome bisulfite sequencing (WGBS) was mapped to the combined genome using methylpy v1.2[9] (see Supplementary Table X); using cutadapt v [10] for adaptor trimming, Bowtie2 v [11] for alignment, and Picard tools v for marking duplicates. The chloroplast genome is unmethylated in plants and can be used as in internal control for calculating the non-conversion rate of bisulfite treatment (percentage of unmethylated sites that fail to be converted to uracil)[2]. In mapping the parental TO1000 and R500 genomes, a small fraction of reads (XXX% TO1000 and XXX% IMB218) mapped to the wrong genome. While the percentage of cross-mapping was small in both cases, it is likely that this issue also affects mapping of the resynthesized allopolyploids. To account for this, we randomly down-sampled TO1000 to an equal number of read pairs as IMB218 and combined these with the IMB218 reads into a single fastq. This combined dataset was then mapped to the combined genome to mimic an *in silico* allopolyploid. By including this *in silico* allopolyploid alongside the individually mapped parents, any we should be able to observe the influence of mismapping. DNA methylation levels in this combined dataset were either approximately half-way between the two parents for the whole genome or at a sub-genome level, approximately equal to that of the respective parent. This indicates that mismapping has little effect on genome-wide analyses, although it may still be a factor at specific regions.

Genome-wide levels of DNA methylation and DNA methylation metaplots were analyzed as previously described[4] using python v3.7.3. Pybedtools v [12] and Bedtools v2.25.0[13]. Briefly, for genome-wide DNA methylation levels, the weighted methylation level[14], which accounts for sequencing coverage, was calculated for each sequence context (CG, CHG, and CHH). This was done for the combined genome, and each individual subgenome. For gene metaplots, cytosines from 2 kbps upstream, 2 kbs downstream and within the gene body were extracted. Cytosines within gene bodies were restricted to those found in coding sequences, as the presence of TEs in introns and problems of proper UTR annotation can obscure start/stop sites and introduce misleadingly high levels DNA methylation[4]. Each of these three regions were then divided into 20 windows and the weighted methylation level for each window calculated and average for all genes. For LTR metaplots, the same analysis was performed, except the all cytosines within the LTR body were included. Metaplots were created for all genes in the combined genome, all genes within each subgenome, syntenic genes in each subgenome, and non-syntenic genes in each subgenome. All plots and statistics were done in R v3.6.0[15], plots made using ggplot2[16]. All code and original analyzed data and plots are available on Github (https://github.com/niederhuth/Replaying-the-evolutionary-tape-to-investigate-subgenome-dominance).

1. Niederhuth, C.E. and R.J. Schmitz, *Putting DNA methylation in context: from genomes to gene expression in plants.* Biochim Biophys Acta Gene Regul Mech, 2017. **1860**(1): p. 149-156.

2. Lister, R., et al., *Highly integrated single-base resolution maps of the epigenome in Arabidopsis.* Cell, 2008. **133**(3): p. 523-36.

3. Cokus, S.J., et al., *Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.* Nature, 2008. **452**(7184): p. 215-9.

4. Niederhuth, C.E., et al., *Widespread natural variation of DNA methylation within angiosperms.* Genome Biol, 2016. **17**(1): p. 194.

5. Takuno, S., J.H. Ran, and B.S. Gaut, *Evolutionary patterns of genic DNA methylation vary across land plants.* Nat Plants, 2016. **2**: p. 15222.

6. Alonso, C., et al., *Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms.* Front Genet, 2015. **6**(4): p. 4.

7. Kersey, P.J., et al., *Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species.* Nucleic Acids Res, 2018. **46**(D1): p. D802-D808.

8. Parkin, I.A., et al., *Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea.* Genome Biol, 2014. **15**(6): p. R77.

9. Schultz, M.D., et al., *Human body epigenome maps reveal noncanonical DNA methylation variation.* Nature, 2015. **523**(7559): p. 212-6.

10. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, 2011. **17**(1).

11. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

12. Dale, R.K., B.S. Pedersen, and A.R. Quinlan, *Pybedtools: a flexible Python library for manipulating genomic datasets and annotations.* Bioinformatics, 2011. **27**(24): p. 3423-3424.

13. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-842.

14. Schultz, M.D., R.J. Schmitz, and J.R. Ecker, *'Leveling' the playing field for analyses of single-base resolution DNA methylomes.* Trends Genet, 2012. **28**(12): p. 583-5.

15. Team, R.C. *R: A Language and Environment for Statistical Computing*. Available from: [http://www.r-project.org/](http://www.R-project.org/).

16. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York.