Promilleafgiftsfonden for landbrug

DOI: 10.1002/tpg2.20049

#### ORIGINAL RESEARCH

The Plant Genome 🛛 🕮 🛛 🕬

### The cytosine methylation landscape of spring barley revealed by a new reduced representation bisulfite sequencing pipeline, WellMeth

Marta Malinowska <sup>1,2</sup> 💿   Istvan Nagy <sup>1,2</sup> 💿   Cornelis A. M. Wagemaker <sup>3</sup> 💿	
Anja K. Ruud <sup>1,2</sup> 💿   Simon F. Svane <sup>4</sup> 💿   Kristian Thorup-Kristensen <sup>4</sup> 💿	
Christian S. Jensen <sup>7</sup>   Birger Eriksen <sup>5</sup>   Lene Krusell <sup>5</sup>   Ahmed Jahoor <sup>6</sup>	
Jens Jensen <sup>6</sup> Lars Bonde Eriksen <sup>8,9</sup> Torben Asp <sup>1,2</sup>	

 $^{\rm 1}$  Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

 $^2$  Center for Quantitative Genetics & Genomics, Aarhus University, Slagelse, Denmark

<sup>3</sup> Department of Experimental Plant Ecology, Radboud University, Nijmegen, The Netherlands

- <sup>4</sup> Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark
- <sup>5</sup> Sejet Plant Breeding, Horsens, Denmark

<sup>6</sup> Nordic Seed A/S, Galten, Denmark

- <sup>7</sup> Research Division, DLF Seeds A/S, Store Heddinge, Denmark
- <sup>8</sup> Landbrug & Fødevarer, SEGES, Aarhus, Denmark
- <sup>9</sup> LIMAGRAIN A/S, Horsens, Denmark

#### Correspondence

Torben Asp, Center for Quantitative Genetics and Genomics; Aarhus University, Forsøgsvej 1, DK-4200 Slagelse. Email: torben.asp@qgg.au.dk

**Funding information** This work was supported by Innovation Fund Denmark (grant number 5184-00035B) and Promilleafgiftsfonden for Landbrug.

#### Abstract

Patterns and level of cytosine methylation vary widely among plant species and are associated with genome size as well as the proportion of transposons and other repetitive elements in the genome. We explored epigenetic patterns and diversity in a representative proportion of the spring barley (*Hordeum vulgare* L.) genome across several commercial and historical cultivars. This study adapted a genotyping-by-sequencing (GBS) approach for the detection of methylated cytosines in genomic DNA. To analyze the data, we developed WellMeth, a complete pipeline for analysis of reduced representation bisulfite sequencing. WellMeth enabled quantification of context-specific DNA methylation at the single-base resolution as well as identification of differentially methylated sites (DMCs) and regions (DMRs). On average, DNA methylation levels were significantly higher than what is commonly observed in many plants species, reaching over 10-fold higher levels than those in *Arabidopsis thaliana* 

**Abbreviations:** DMC, differentially methylated site; DMR, differentially methylated region; FPKM, fragments per kilobase million; GBS, genotyping-by-sequencing; GO, gene ontology; LTR, long-terminal repeat; PC, principal component; PCA, principal component analysis; PCR, polymerase chain reaction; RRBS, reduced representation bisulfite sequencing; TE, transposable element; TES, transcription end sites; TSS, transcription start sites; UMI, unique molecular identifier; UTR, untranslated region; WGBS, whole-genome bisulfite sequencing.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. The Plant Genome published by Wiley Periodicals, Inc. on behalf of Crop Science Society of America

(L.) Heynh. in the CHH methylation. Preferential methylation was observed within and at the edges of long-terminal repeats (LTR) retrotransposons *Gypsy* and *Copia*. From a pairwise comparison of cultivars, numerous DMRs could be identified of which more than 5,000 were conserved within the analyzed set of barley cultivars. The subset of regions overlapping with genes showed enrichment in gene ontology (GO) categories associated with chromatin and cellular structure and organization. A significant correlation between genetic and epigenetic distances suggests that a considerable portion of methylated regions is under strict genetic control in barley. The data presented herein represents the first step in efforts toward a better understanding of genome-level structural and functional aspects of methylation in barley.

### **1** | INTRODUCTION

Barley is one of the four major cereals grown over about 47 million ha worldwide and with a world production over 147 Tg in 2017 (FAOSTAT, 2018). It is used for human and animal consumption and for producing malt for the distilling and brewing industry. Barley is cultivated across a wide range of environments, which can be attributed to its adaptability to different climates and high tolerance to various abiotic stresses (Newton et al., 2011). Anthropogenic climate change is a threat to food security, and through direct (water and nutrient availability and abiotic stresses) and indirect (diseases and pests) impacts is expected to cause further barley yield stagnation (Ray et al., 2019; Schauberger et al., 2018). In the face of changing climate and a growing human population, quality and yield improvement under adverse environmental conditions are important goals in barley breeding programs in northern Europe. Since genetic diversity is limited, it is of high interest to explore the prospect of including epigenetics to assist barley breeding strategies.

DNA methylation is a dynamic epigenetic modification playing a role in the regulation of the plant genome during environmental adaptation and development and is associated with cellular processes like the regulation of transposon silencing or expression (Law & Jacobsen, 2010; Miura et al., 2001; Pikaard & Scheid, 2014; Zhang, Lang, & Zhu, 2018). Patterns in DNA methylation can manifest as a result of genetic control, environmental stimuli, or stochastic epigenetic alteration. Most of the changes to patterns of DNA methylation are reversible or of no consequence but some could be passed down through meiosis and persist between generations (Heard & Martienssen, 2014). The reversible nature of DNA methylation is thought to be a source of phenotypic plasticity, which, under very dynamic environments, may benefit the survival and reproduction of populations (Springer, 2013). Heritable epigenetic variants that affect phenotypes could be under natural selection and contribute to adaptation independently of genetic variation (Richards et al., 2017).

In plants, DNA methylation occurs in both symmetric CG and CHG, and asymmetric CHH (where H is A, T or C) contexts (Meyer, Niedenhof, & Ten Lohuis, 1994). The methylation state reflects a balance between distinct molecular pathways that establish de novo, maintain, or remove methylation and, in the cases of CHG and CHH, are unique to plants (Law & Jacobsen, 2010; Stroud et al., 2014). De novo establishment of DNA methylation is mediated by the RNA-directed DNA methylation pathway (Law & Jacobsen, 2010). After DNA replication, maintenance of DNA methylation, depending on the cytosine sequence context, is regulated by multiple methyltransferases (Law & Jacobsen, 2010). The role and effect of methylation depend on the pattern, context, and complexity of the genome as well as chromatin configuration (Law & Jacobsen, 2010). Studies have shown that DNA methylation plays an important role in (a) the regulation of plant developmental processes and (b) the response to biotic and abiotic environmental stimuli (Cubas et al., 1999; Le et al., 2014; Manning et al., 2006).

Despite a growing interest in the impact of epigenetic variation on crop performance under different environmental conditions (Gallusci et al., 2017; Richards et al., 2017; Springer, 2013), there have been very few studies focusing on DNA methylation landscape diversity in barley. Studies of epigenetic variation published in the last 10 yr have used methylation-sensitive amplification polymorphism (Chwialkowska, Nowakowska, Mroziewicz, Szarejko, & Kwasniewski, 2016) for the detection of methylome modulation under drought, methylation-sensitive restriction enzyme, multiplexed polymerase chain reaction (PCR) for responses to salt stress, and bisulfite sequencing for responses to salt (Konate et al., 2018) or heavy metals (Kashino-fujii et al., 2018).

Multiple methods have been developed to investigate DNA methylation patterns, levels, or variation on a genome-wide scale Kurdyukov & Bullock, (2016). Since the first demonstration of using bisulfite conversion of 5-

the first demonstration of using bisulfite conversion of 5methylcytosine in sequencing Frommer et al., (1992), the application of high-throughput, next-generation bisulfite sequencing technologies has accelerated the development of powerful epigenetic tools. With the advantage of targeting essentially every cytosine base in a genome, the whole-genome bisulfite sequencing (WGBS) (Cokus et al., 2008; Lister et al., 2010) is considered the gold standard for a comprehensive analysis of DNA methylation variation. However, aside from being restricted to species where there is a high-quality genome available, it is also the most expensive method to study genome-wide methylation. A cost-effective alternative to WGBS is reduced representation bisulfite sequencing (RRBS) that uses a fraction of the genome and generates a scalable DNA methylation profile at single-nucleotide resolution (Gu et al., 2011).

By using double-enzyme digestion in the RRBS method, it is possible to increase coverage of the detected regions more accurately and thus reflect their average methylation levels (Wang et al., 2013). In recent years, updated reduced representation approaches such as bsRADseq (Trucchi, Mazzarella, Gilfillan, Lorenzo, & O, 2016) and epiGBS (Van Gurp et al., 2016) have been developed for species lacking a reference genome. While these methods may be attractive for particular population and metagenomic studies, the tedious steps of constructing synthetic references (reconstructing the original sequences from stranded bisulfite-treated reads), and subsequent difficulties of resolving their genomic positions, are rather objectionable for species with complex genomes and available genomic resources.

Here we present WellMeth, a complete RRBS analysis pipeline for species with high-quality reference genomes. In the study, we examine and compare the methylation landscape of 12 commercial and historical spring barley cultivars and lay the groundwork to better understand variation in the barley genome beyond differences in DNA sequence.

#### 2 | MATERIALS AND METHODS

### 2.1 | Plant material

Twelve spring barley commercial cultivars (Evelina, Invictus, Kenia, Prisma, Tocada, Evergreen, Laurikka, Flair, Laureate, RGTPlanet, Kaarle and Alvari) were grown in the RadiMax facility located at Copenhagen University's experimental farm, west of Copenhagen, Denmark (Svane, Jensen, & Kristensen, 2019). Flag leaves

- Presents WellMeth, a complete, reduced representation bisulfite sequencing analysis pipeline.
- Assess context-specific DNA methylation patterns in 12 cultivars.
- Spring barley has a very high proportion of CHH methylation throughout the genome.
- Preferential methylation of transposable elements was observed.
- Variation in methylation profiles partially mirrored the genetic diversity of the cultivars.

from seven different plants per cultivar, were collected before heading. Extraction and further analysis of DNA and RNA was conducted for one pool of seven individuals per cultivar (n = 12).

### 2.2 | Library preparation

Genomic DNA was extracted by using the standard cetyltrimethyl-ammonium-bromide (CTAB) method, quantified and checked for quality using the PicoGreen dsDNA Assay Kit. EpiGBS libraries were prepared as described by Van Gurp et al. (2016). Briefly, genomic DNA (400 ng) was digested using two methylation insensitive restriction endonucleases PacI (New England BioLabs, Inc.) and Nsil (New England BioLabs, Inc.). After digestion, barcoded adaptors were ligated to the fragments using T4 DNA ligase (New England BioLabs, Inc.). An equal volume of 12 ligates were pooled and purified using PCR cleanup (Marchery Nagel) followed by size selection using AMPureXP (Beckman Coulter, Inc.) according to the manufacturer's protocol and favoring DNA fragments >200 bp,. Prior to bisulfite treatment, we carried out nick translation with DNA polymerase I (New England BioLabs, Inc.) and 5-methylcytosine dNTP mix (Zymo Research). A nick, in each fragment-adapter connection, was created because the barcoded adapters were not phosphorylated to prevent the formation of adapter dimers. The nick-translated libraries were treated with bisulfite to convert unmethylated (but not methylated) cytosines to uracil using EZ DNA Methylation Lightning Kit (Zymo Research). Samples were PCR amplified in four individual 10-µL reactions with KAPA HiFi HotStart Uracil+ ReadyMix (Kapa Biosystems) and universal Illumina PE PCR primers. The PCR products were combined and purified together using a PCR cleanup kit and solid-phase reversible immobilization beads. Library quality and quantity (>2 nM) was checked using a PicoGreen dsDNA Assay Kit and qPCR, while the consistency of the size profiles (150–600 bp with a peak around 450) was determined using a Bioanalyzer (Agilent; 5067-4626). Pooled libraries were sequenced at Novogene (UK) Ltd. on the Illumina HiSeq4000 platform in 2- by 150-bp paired-end mode distributed across multiple sequencing lanes.

### 2.3 | Reference sequence

The IBSCv2 barley genome assembly (Mascher et al., 2017) was used, which contains seven pseudochromosomes (chr1H to chr7H) plus an additional genomic sequence (chrUn) with concatenated unassigned scaffolds. As all of the seven barley pseudochromosomes exceed the 512 Mb size limit of bai indexing by SAMtools (Li et al., 2009), a custom Unix script was applied to split long chromosomes at the center, while avoiding splitting within annotated features. Split pseudochromosomes were used as references throughout all steps of the present experiment. Feature coordinates obtained on split pseudochromosomes were recalculated to original coordinates using another custom script.

#### 2.4 | Gene model

RNA sequencing reads representing 150 barley leaf samples (including the 12 barley lines of the present study) were mapped onto the split-chromosome version of the IBSCv2 barley genome using HISAT2 (Kim, Langmead, & Salzberg, 2015). Transcript structures were reconstructed using StringTie (Pertea et al., 2015) and high-quality exon junction data were obtained using the Portcullis tool (Mapleson, Swarbreck, Venturini, & Kaithakottil, 2018). These data, along with the published annotation data of the IBSCv2 assembly, were used to produce a polished gene model with Mikado pipeline (Venturini, Caim, Kaithakottil, Mapleson, & Swarbreck, 2018).

### 2.5 | Identification of repetitive DNA sequences

Transposable elements and repeats idenwere tified using RepeatMasker open-4.0.6, (version http://www.repeatmasker.org) using the Liliopsida species model and RepBase Update 20160829. A custom pipeline was used to annotate the RepeatMasker output including the following main steps: (i) assign RM class and target to each RepeatMasker hit, (ii) merge overlapping features of identical class and target categories, and (iii) produce separate BED format files for each RM class

### 2.6 | Restriction fragment representation assessment

To assess the proportion of genomic regions covered by bisulfite-treated short reads originating from PacI-NsiI restriction fragments, sites of a minimum read depth of 5 were collected from bam-format alignments after removing PCR-duplicates using the coverage command in the BamTools toolkit (Barnett, Garrison, Quinlan, Strömberg, & Marth, 2011). Contiguous regions were created for intersecting and size estimation using the merge function from BEDTools (Quinlan & Hall, 2010).

### 2.7 | Read alignment and measuring methylation level

Paired-end reads were demultiplexed and barcode sequences were trimmed using the first script of the WellMeth pipeline WM-1\_preprocessing.sh. After demultiplexing and quality filtering, bisulfite-treated short reads were aligned to the split-chromosome version of the barley reference genome version Hv\_IBSC\_PGSB\_v2 using WM-2 mapping.sh. The second step also included removal of the PCR duplicates. Reads that were left after quality control and PCA duplicated methylation variants were called by the execution of the third script WM-3\_call\_methylation.sh (Supplemental Figure S1). Methylation information ratios (number of methylated regions vs. all reads) for each cytosine were obtained in three sequence contexts (CG, CHG, and CHH). Following mapping and calling of the raw sequences of the methylation variants, data were filtered and only sites with minimum 5× read coverage were considered for further analysis.

# 2.8 | Identification of differentially methylated sites and regions

Differentially methylated sites and regions were identified by all-to-all pairwise comparison of methylation patterns of the 12 barley cultivars using the WellMeth script WM-4\_diff\_methylation.sh that implements a hidden Markov model-based differential methylation detection framework adopted from the BisulFighter package (Saito, Tsuji, & Mituyama, 2014).

To compare the level of methylation within DMRs at the population level, consensus regions were called using the merge function from BEDTools (Quinlan & Hall, 2010). Only regions with a minimum of five cytosines and a frequency (number of methylated cytosines per length of the DMR) higher than 0.2 were kept for further analysis. The level of DNA methylation was averaged independently for each region and cultivar.

### 2.9 | Relatedness analysis

Total RNA was isolated from the same pool of seven flagleaf samples described above using the Sigma-Aldrich Total Plant RNA kit. The Beijing Genomics Institute performed RNA sequencing library construction and sequencing. Paired-end sequencing was performed on the Illumina HiSeq4000 platform (2- by 100-bp, ~20 M reads per sample). Clean reads were mapped on the barley reference genome using HISAT2 (Kim et al., 2015). Alignments were subsequently processed by StringTie (Pertea et al., 2015) and normalized read count data (fragments per kilobase million [FPKM]) were collected for each sample for 80,855 transcripts. The FPKM values were then log2+1 transformed in order to stabilize the variance. The SNP variants were called from the bam alignments using a haplotypebased variant detector Freebayes (Garrison & Marth, 2012). The SNPs were filtered using the R package SNPrelate (Zheng et al., 2012), where SNPs in high linkage disequilibrium  $(r^2 > 0.9)$  and >30% of missing values were filtered out, leaving a set of 72,642 polymorphic SNPs that were used to perform relatedness analysis.

### 2.10 | Statistical analysis

A Mantel test was conducted with the VEGAN package (Dixon, 2003) in R to explore the correlation between epigenetic and genetic distance and between epigenetic distance and gene expression variation. For the Mantel test, Euclidean distance matrices were calculated for genetic markers, transformed FPKM values, and methylation data sets. The significance of the correlations from the Mantel tests was determined through the Pearson product-moment correlation coefficient after 9,999 permutations. Principal component analysis (PCA) was performed in R using the PCA function from the mixOmics package (Rohart, Gautier, Singh, & Cao, 2017).

### 2.11 | Gene ontology term enrichment

To investigate the biological functions of genes with DMR within the gene body or promoter region, GO term enrichment analysis was carried out with the Python library GOATOOLS (Klopfenstein et al., 2018). Gene ontologyterm categories that were statistically overrepresented in the differentially methylated regions compared with the total set of 13,433 genes covered by the sequencing were corrected for multiple comparisons with a false discovery rate of 0.05 using the Benjamini–Hochberg method (Benjamini & Hochberg, 1995).

### 3 | RESULTS

#### 3.1 | Restriction enzymes

For double-digestion, methylation insensitive, frequent (NsiI) and rare cutter (PacI) were selected based on expected coverage with 2 million reads. The summed length of genomic regions in the bwa-meth alignments covered short reads with a minimum read depth of 5,and varied between 25 and 28 Mbp.

# 3.2 | General characterization of the spring barley methylome

We performed global methylome analysis of the 12 spring barley cultivars using RRBS. The design captured 1.4 million reads (0.7% of the barley genome) and 3.2 million unique methylated sites among which 20% (0.67 million) were cultivar specific. Interrogation of the conservation of methylated sites between cultivars showed that 0.38 million methylated sites were consistently methylated in 90% of the cultivars, and only these sites were considered for any statistics and summary data of the analyzed population (Supplemental Table S1). Out of all 19 million cytosines covered by the analysis, 27% were fully methylated. In the analyzed barley cultivars, the CHH context was identified as the predominant type of DNA methylation contributing to >70% of total methylated cytosine, while 14% were in the CG and CHG contexts. Globally, in all cultivars, we found DNA methylation levels of 88, 74, and 49% in the CG, CHG, and CHH sequence context, respectively (Supplemental Table S1).

# 3.3 | Genome-wide distribution of DNA methylation

Mapped fragments were relatively equally distributed along the chromosomes, and the proportion of methylated cytosines captured by the analysis reflected the genome architecture (Supplemental Figure S2). Methylation within the gene bodies occurred at the highest densities in the distal chromosome regions that are characterized by a



**FIGURE 1** The percentage of methylated cytosines in the three main genomic regions: transposable elements (TEs), intergenic, genic (intron, exon), and the overall covered region. Each locus that could be identified in at least six cultivars was classified into three methylation levels: unmethylated (<10%), intermediate (>10% but <90%), and methylated (>90%)

high gene copy number. Methylation within retroelements dominated at a greater distance from genes toward the centromeric regions, which follows the occurrence of the LTR retrotransposons (Mascher et al., 2017). The higher density of cytosine methylation within transposons was found in distal chromosome regions, mirroring the proximity to genes (Supplemental Figure S2).

Analysis of the genomic features of the methylated sites showed that 72% of methylated cytosines were found in repetitive DNA of which 65% were retroelements, particularly in LTR Gypsy (Supplemental Figure S3; Supplemental Table S2). About 3.5% of the methylated cytosines were found in gene bodies (including introns, exons, and untranslated regions [UTRs]) that directly reflect the proportion of the genic space in the barley genome (Wicker et al., 2017). Over 13% of methylated sites were located in the flanking regions specified 5 kb up- and 2 kb downstream of the genes and so-called intergenic regions (between 5-100 kb upstream of genes). Approximately 11% of methylated cytosines could not be classified to any genomic feature, which is unsurprising considering 16% of the barley genome remains unannotated (Wicker et al., 2017).

Genic regions (introns and exons) showed a substantially high proportion of unmethylated cytosines (33– 38%; Figure 1) and a low fraction of highly methylated sites. In comparison, transposable elements showed almost reversed proportions with 18% of cytosines (coverage > 4) having zero or very low methylation (down to 18%) and over 34% of cytosines being highly methylated (Figure 1).

Methylation profiles were examined in all three sequence contexts within the gene body (i.e. introns, exons) and its flanking regions (up to 100% of the gene length) as well as within and around transposons and retrotransposons (up to 100% of the transposable elements [TEs] length) focusing on the three superfamilies: DNA Cacta, LTR Gypsy, and LTR Copia. The average methylation level (Table 1) and distribution (Figure 2) showed overall reduced DNA methylation in gene bodies and extensive methylation of transposable elements. For each feature, we identified three different methylation patterns. Genes were characterized by moderate CG and very low CHG and CHH methylation within the gene itself and increased level in the surrounding area (Figure 2a). The methylation level in all three contexts decreased sharply with proximity to genes, reaching a minimum at transcription start sites (TSS) and transcription end sites (TES). Many species show a similar trend of low (or no) gene body methylation in the CHG and CHH context (Niederhuth et al., 2016) but in the majority, the occurrence of CHH is significantly lower than CHG. Our results showed that generally, genes have a very similar level and pattern in both CHG and CHH context (Table 1, Figure 2). A very high level of methylation and its increase in comparison to the flanking regions in the CG context characterized both classes of transposable elements (Table 1, Figure 2). Differences in the methylation patterns between retrotransposons and DNA transposons were visible in the CHG and CHH context. The LTR retrotransposons were characterized by moderate peaks of CHG methylation and robust peaks of CHH methylation

TABLE 1	The average methylation	level according to local	annotation and	sequence context.	Upstream or d	lownstream f	lanking regions
are defined as	5 kbp before transcription	start sites or 2 kbp after	r transcription e	nd sites			

	Mean methylation	Mean methylation level				
Features	CG	CHG	СНН			
Flank (downstream)	0.70	0.51	0.42			
Flank (upstream)	0.77	0.61	0.47			
Gene body	0.60	0.35	0.36			
Intergenic	0.90	0.75	0.47			
Retroelement	0.94	0.80	0.46			
DNA transposon	0.93	0.75	0.52			
Unclassified	0.87	0.70	0.47			



**FIGURE 2** Methylation level distribution of CG, CHG, and CHH context along genes, retroelements, and DNA transposons. Upstream and downstream regions are proportionally the same length as the features

at the edges of TE (Figure 2b), while the pattern in DNA *Copia* showed only a moderate to low increase within the transposon body and no peaks in the flanking region (Figure 2c).

### 3.4 | Methylation variation between cultivars

Methylation patterns in analyzed cultivars were not significantly different when the proportion of methylated sites in each context was considered (Supplemental Table S3). Average methylation of each cultivar was also comparable (Supplemental Table S1); however, differences could be observed in methylation of different genomic features (Supplemental Table S3). Cultivar-specific differences in methylation patterns were particularly evident in the gene body and LTR retrotransposons (Supplemental Figure S4; Figure 3). In the genic and its flanking regions, three cultivars (Alvari, Kaarle, and Evelina) were separated from the rest and showed increased methylation especially in CG context (Supplemental Figure S4). There was an interesting dynamic between cultivars in CHH methylation in the immediate flanking regions of LTR retrotransposons; its



**FIGURE 3** Methylation level distribution in CHH context along genes, retroelements, and DNA transposons. Upstream and downstream regions are proportionally the same length as the features

methylation pattern separated cultivars into three groups (Figure 3). In the first group, Alvari, Evelina, and Kaarle showed the most striking peaks of CHH methylation at the edges of the LTR. Cultivars in the second group, Invictus and Laureate, had the lowest methylation levels in both retrotransposon body and surrounding area. The rest of the cultivars showed intermediate levels between the two groups.

## 3.5 | Genetic diversity of the twelve cultivars

Genetic relatedness was analyzed using a subset of 72,642 polymorphic SNPs identified in transcribed regions. The first three principal components (PCs) explained 42% of the variance. Based on these components, the cultivars clustered into three groups (Figure 4a). The most distant cultivars were Kaarle and Alvari. Kenia and Evelina were gathered together and separated from the Finnish lines on both PCs. The additional group consisted of highmalting cultivars for Scandinavian countries and northern Europe. The second PCA (Figure 4b) was performed using the transformed FPKM values for the transcript expression data. The first three PCs explained 34.5% of the variance (Figure 4b), which is less than for PCAs based on SNP data. The cultivars separated into three clusters, with Kenia and Evelina in one group, the Finnish cultivars Alvari and Kaarle in another, and the remaining (conventional and more recently released) cultivars in a third group.

### 3.6 Differential methylation

Differential methylation analysis between each pair of cultivars was conducted to compare methylation patterns of the commercial cultivars. We identified nearly 21,000 unique DMRs and 0.7 million DMCs between all the pairs of cultivars analyzed. On average, over 3,172 significant DMRs and 0.3 million DMCs could be detected between each pair. On the extremes, over 4,300 DMRs were identified between Laurikka and Invictus, while only 2,035 DMRs were identified between Evelina and RGTPlanet (Figure 5a). There were over 0.46 million DMCs identified between Prisma and Laurikka, while only 0.1 million DMCs were identified between Evelina and Evergreen or Alvari. Many DMRs did not necessarily match a high number of DMCs for each pair (Figure 5b). The limited overlap between DMCs and DMRs can explain this discrepancy. Only 45% of DMCs were found within regions that showed significant differences, the remaining DMCs (independent to DMRs) were scattered across the covered proportion of the genome. The majority of DMRs (78%) and DMCs (53%) were in CHG context. This is in contrast to the overall proportion of methylated sites, where CHH methylation was the most abundant with CG and CHG comprising <30% of all identified methylated sites. Only ~1% of DMRs and 4%



**FIGURE 4** (a) Principal component analysis (PCA) showing the first and second principal components based on 72 642 SNPs identified from transcriptomes; (b) PCA showing the cultivars clustering based on normalized (log2+1 transformed) transcript expression data for the 12 cultivars and a total of 80855 transcripts



**FIGURE 5** Methylation differences between different cultivars. Numbers of (a) differentially methylated regions and (b) differentially methylated sites (in millions) identified between each pair of cultivars

of DMCs identified between all pairs of cultivars were in CHH context.

### 3.7 | Epigenetic diversity of the twelve cultivars

We evaluated the consistency of methylation within regions to investigate whether the number of DMRs between cultivars was due to real differences or similarities or variation in coverage. Differentially methylated sites methylated in at least six cultivars ( $\geq$ 50%) were classified as consistent, those occurring in less than six cultivars (<50%) as rare (Figure 6a), and regions methylated in all cultivars were classified as highly represented. The majority (82.1%) of 20,828 DMRs were consistently methylated in at least 50% of the cultivars, out of which 5578 were methylated in all of them (Figure 6a). Nearly 18% of the highly represented DMRs overlapped the intergenic regions and over 4% occurred within the gene bodies. Both proportions were significantly higher when compared with the distribution of the methylated loci in the genome (Figure 6b; Supplemental Figure S2). The PCA was performed using the average methylation in DMRs common for all cultivars (Figure 6c) to assess if methylation profiles could distinguish the cultivars and mirror their genetic diversity. The three first components described over 57% of the variance. In the middle of PC2 was a cluster of nine cultivars including Kenia that was coupled with Evelina in the genetic variance analysis.

Similar to the PCA of the transcriptome variants, Kaarle, Alvari and, Evelina were divergent from the main group. On the other hand, Kenia differed more from the main cluster based on the genetic variants than on the differential methylation pattern. The similarities and differences between genetic variants and methylation of highly represented regions were echoed in the transcript expression data. A Mantel test with 9,999 permutations showed that DNA methylation differences between the lines are significantly associated with DNA sequence differences (R = .62, P = .02). The test also revealed a positive correlation between epigenetic distance and gene expression variation (R = .54, P = .02). Expression differentiation of the analyzed cultivars was also significantly correlated with its genetic distance (R = .73,P < .001).

### 3.8 | Gene ontology of methylated genes in barley cultivars

The genes that were covered by the sequenced reads in the promoters (1 kbp upstream of the TSS) and gene body regions were taken into consideration and termed as the methylated genes. The analysis covered over 13,000 genes. A total of 2,118 methylated genes (of the 13,433 genes covered in total by the RRBS) overlapped with DMRs within the gene body and promoter region, of which 467 were identified in all 12 cultivars. Gene ontology term enrichment analysis of functional domains in these 467 genes identified 17 significantly enriched GO terms, including FIGURE 6 Methylation differences between different cultivars. (a) The number of differentially methylated regions (DMRs) identified in pairwise comparisons between each cultivar in the analyzed population. The DMRs were considered rare when <50% and rare when >50% of cultivars were methylated in the region. (b) Genome-wide distribution of highly represented DMRs (regions methylated in all cultivars). (c) Summary of a principal component analysis showing the first and second principal components based on the average methylation in 5,578 DMRs that were consistently methylated in all cultivars

100%

75%

50%

25%

Gene

TEs

(a)

Percent of highly represented DMRs



All

26.8%



GO terms associated with chromatin structure and organization and metal ion transport (Supplemental Figure S5).

#### DISCUSSION 4

In this study, we assessed the DNA methylation landscape at single-base resolution in 12 commercial and historical spring barley cultivars developed for the European climate. Our results revealed features regarding the distribution of methylated cytosines, differential methylation patterns, and their relationship with genetic diversity. Furthermore, we offer a reproducible and complete pipeline to analyze epiGBS data for species with a high-quality genome

The RRBS approach is a cost-effective alternative to WGBS for large experimental designs and species with a large genome such as barley. We selected a combination genes.

of methylation insensitive, rare, and frequent cutting restriction enzymes, thereby avoiding bias toward any particular genic regions. Identified fragments were distributed equally and randomly along the chromosomes, which allowed for the representative identification of potential DMRs of interest found mostly outside of

### **4.1** | DNA methylation patterns in spring barley

We found that spring barley cultivars analyzed had much higher levels of methylation than other members of the Poaceae family, like rice (Oryza sativa L.) and maize (Zea mays L.), or in Arabidopsis (Cokus et al., 2008; Niederhuth et al., 2016; Figure 1 and 2; Table 1). Additionally, CHG and CHH methylation, which tends to be relatively low in most species, was unusually high in barley, with the level of methylation in CHH context being over 10-fold higher than maize or green foxtail [Setaria viridis (L.) P. Beauv.] (Niederhuth et al., 2016). We also observed that the majority of cytosines methylated throughout the barley genome were in CHH context, similar to sugar beet (Beta vulgaris L. subsp. vulgaris) (Zakrzewski et al., 2014). These two trends are likely a consequence of the barley genome size, frequency of each context (Wicker et al., 2017), TEs composition, and mechanisms by which CHG and CHH methylation are established and maintained. Variation in methylation patterns between species has been observed before (Niederhuth et al., 2016; Takuno, Ran, & Gaut, 2016). Non-CG methylation shows relatively little conservation at the level of individual methylation pathways, which seem to be lineage specific (Li et al., 2014; Takuno et al., 2016; Zemach et al., 2013). Despite the general similarities between plant species, uniquely high non-CG methylation in barley suggests significant distinction in the predominance of methylation pathways and a species-specific distribution of CHH in the genome.

# 4.2 | DNA methylation patterns in genic regions

In contrast to heavy methylated transposable elements and intergenic regions, DNA methylation in gene bodies was decreased in all three contexts. Similarly to poplar (*Populus trichocarpa* Torr. & A. Gray), rice, or *Arabidopsis* (Feng et al., 2010; Zemach, McDaniel, Silva, & Zilberman, 2010), the CG methylation profile at the start and end of the gene body sequence colocalized with a prominent dip in average methylation. We found that the level of non-CG methylation was low throughout the gene body, similar to observations in rice and soybean [*Glycine max* (L.) Merr.] (Garg, Narayana Chevala, Shankar, & Jain, 2015; Song et al., 2013). Interestingly, although methylation in UTRs was at a very low level in all three contexts (Supplemental Table S4), it was still significantly higher than methylation in UTRs typically observed in many plants species (Feng et al., 2010; Garg et al., 2015; Song et al., 2013).

Even though it was suggested that gene body methylation regulates splicing and gene expression, its function in plants remains mostly unclear (Bewick & Schmitz, 2017; Candaele et al., 2014; Zhou et al., 2016). In contrast to previous reports, which hypothesize that preferential methylation of the gene body occurs at exons (Gelfman, Cohen, Yearim, & Ast, 2013; Takuno & Gaut, 2013), we found that in spring barley, levels of DNA methylation are higher in genomic regions encoding introns than the exon sequences. However, these correlations hold true only for CG methylation and not for CHG or CHH methylation, highlighting an invariably low level of methylation across introns, exons, and UTRs. The pattern and proportion of intron-exon methylation suggest that in barley there is selection against intragenic TE methylation, possibly to avoid its negative effect on gene expression. Alternatively, TEs present in introns could be targeted by CG and not non-CG methylation. It may be that the unique CG methylation pattern in introns and exons is due to gene body methylation in spring barley prioritizing CG context and, similarly to rice, this may play an essential role in pre-messengerRNA splicing (Wang et al., 2016).

# **4.3** | Methylation of the transposable elements

Epigenetic silencing driven by DNA methylation and histone modification is central for the repression of transposition and expression of TEs (Law & Jacobsen, 2010). Transposable elements' repressive activity of DNA methylation is particularly important in plants with large genomes, for example, maize or barley, in which the sheer abundance of transposable elements determine chromosome structure and genome organization (Anderson et al., 2019; SanMiguel et al., 1996; Wicker et al., 2017). In the analyzed cultivars, pericentromeric heterochromatin, as well as repeats and TEs containing euchromatin regions, were heavily methylated in all cytosine contexts (Supplemental Figure S2; Table 1; Mascher et al., 2017). What is more, the CHH methylation in both LTRs and DNA transposons reached up to 10-fold the level of the average methylation in many plant species (Cokus et al., 2008; Song et al., 2013; Niederhuth et al., 2016; Noshay et al., 2019). We observed distinct non-CG methylation profiles within and at the

edges of retrotransposons and transposon families. The difference is likely a consequence of the abundance of the identified TE families, their chromosomal position, and chromatin configuration (Haag & Pikaard, 2011; Law & Jacobsen, 2010; Zemach et al., 2013). While <6% of the methylated positions that we identified were within DNA transposons (Supplemental Table S2) and located predominantly in the gene-rich regions (Supplemental Figure S2), TE methylation occurred mostly within the retrotransposons LTR Gypsy and LTR Copia. These two families of retroelements comprise almost 50% of the barley genome (Mascher et al., 2017; Wicker et al., 2017). It has been shown that RNA-directed non-CG methylation pathways target the edges of LTR retrotransposons, often located close to expressed genes (Li et al., 2015, Zemach et al., 2013). However, in spring barley, nearly 60% of highly methylated sites were identified in LTR Gypsy, which is enriched in pericentromeric regions (Wicker et al., 2017). This suggests that while very high TE body methylation confirms preferential suppression of the transposable elements, the dynamics of non-CG methylation at the edges of TEs must have a distinct role. A dramatic increase in CHG and CHH methylation at the boundaries of TEs has been discussed before (Li et al., 2015; Noshay et al., 2019; Wang, Liang, & Tang, 2018; Zemach et al., 2013). However, barley LTRs (borders are targeted by CHH methylation) are mostly located in heterochromatin and do not show lowered CG and CHG methylation (quite the opposite). We hypothesize that the peaks near LTR act as markers for long retrotransposons and partition between heteroand euchromatin, rather than prevent the spread of these families.

# 4.4 | Epigenetic and genetic diversity in an analyzed spring barley population

The genome shock hypothesis proposes that under stress, TEs can threaten the stability of the genome through relocation and insertion of new copies but this TE reactivation may also facilitate genetic adaptation (McClintock, 1984). Therefore, changes in TE methylation are particularly interesting in species with large genomes that harbor a high proportion of TEs such as barley. In the cultivars examined in this study, the number of DMRs that overlapped with transposons (Figure 6) indicates a possible degree of reversibility of TE methylation in this species. A balance between hypo- and hyper-TE methylation through the promotion of genome plasticity (Fedoroff, 2012; Raffaele & Kamoun, 2012) may be a driver of epigenetic and phenotypic diversity and potentially the ability to adapt to different environments. Alternatively, the differences between cultivars observed in the methylation profiles in transposons and retrotransposons, as well as DMRs and DMCs identified within TEs, might be simply a consequence of transposon-based genetic diversity within the analyzed population (Singh, Nandha, & Singh, 2017).

Because of its genome size, the analysis of barley methylome at the single cytosines level can be expensive, cumbersome, and time-consuming. Nevertheless, we have shown that epigenetic analysis using the data obtained from RRBS allows for the characterization of DNA methylation profiles and to capture differences between individuals. The 12 barley epigenomes compared in this study represent cultivars adapted to mostly Nordic climate and represent only a small subset of barley diversity. However, a high degree of TE polymorphism has been reported before between just four maize inbred lines (Anderson et al., 2019). Therefore, with over 80% of the genome consisting of transposable elements (Mascher et al., 2017) and numerous quantitative trait loci already found in TEs of a population of 233 spring barley genotypes (Abdel-Ghani et al., 2019), it can be expected that sufficient TE polymorphism is present in the analyzed cultivars. Our findings emphasize the potential importance of the methylation of retrotransposons in the promotion of genome plasticity and epigenetic diversity.

Although <1% of the genome was captured in this study, we managed to identify over 20,000 unique DMRs across all cultivars. Very often, an identified DMR was unique for just one pair of cultivars. However, analysis of the methylation patterns within all identified regions indicated that significant fractions of sites were conserved between the cultivars, showing the degree of epigenetic variation within a population grown under the same conditions. A subset of these conserved regions showed fluctuating levels of methylation and overlapped with nearly 500 genes involved in various biological processes and molecular functions. We found no evidence that the differences in DNA methylation within those genes were associated with the changes in their expression. However, those genes exhibited enrichment in the categories associated with chromatin and cellular structure and organization and in the GO analysis (Supplemental Figure S5), suggesting the interdependence of gene expression and epigenetic marks. Nonetheless, in the present study, it is not possible to decipher whether differences in methylation are a cause or a result of gene expression.

The interplay between genetic and epigenetic variation is yet to be fully understood. Various studies have demonstrated that the relationship is not straightforward and that multiple factors play a role in DNA methylation variation including genetic variants (Dubin et al., 2015; Richards, 2006; Richards, Bossdorf, & Verhoeven, 2010). In this study, PCA (Figure 4, 5 & 6) and Mantel tests revealed a significant relationship between epigenetic and genetic variation. To some extent, methylation profiles mirrored the genetic diversity of the cultivars, which was particularly visible in the separation of Finnish from Nordic lines and central European cultivars. Genetic, transcriptomic, and epigenetic distance similarities show that in the absence of the pressure of differential growth conditions, environmental cues cannot mask the effect of genetic relatedness and variability on methylation patterns. The number of DMRs conserved between the cultivars together with a relatively strong relationship between genetic and epigenetic distance suggests that a proportion of methylated regions are genetically controlled and should, therefore, be stably inherited across generations. This hypothesis can be supported by the fact that even though CHH was the most abundant methylation context, the majority of DMR and DMCs were in CG and CHG context. It has been demonstrated before that CG and CHG methylation is more stable in plant genome than CHH context in both responses to stress and daily methylation fluctuations (An et al., 2017; Liang et al., 2019). Therefore, we can believe that methylation differences between cultivars found mainly in CG and CHG context are, in fact, a consequence of genetic diversity within the analyzed population.

Once disentangled from genetic variation, the barley methylome may be a remarkable resource to identify epigenetic associations with phenotype, environment, and gene expression, acting as a powerful future tool to aid plant breeding. More extensive studies with more cultivars, various environments and several generations will be needed to find heritable methylation markers that could act as predictors of plant performance.

### 4.5 | WellMeth, a complete, reduced representation bisulfite sequencing analysis pipeline

In this paper, we present WellMeth, a new RRBS analysis pipeline. The design of adapters and primers and the library construction principles of WellMeth are adopted from the epiGBS pipeline (Van Gurp et al., 2016), applying 4- to 6-nt long asymmetrical barcodes on the forward and reverse adapters and a fixed number of random nucleotides (unique molecular identifiers [UMIs]) adjacent to the barcodes. The double-stranded adapters used in WellMeth contain 5-methylcytosine, except for an optional unmethylated C nucleotide that can be used for strand identification and assessment of bisulfite treatment efficiency. A combination of 12 forward and eight reverse adapters

accomplish a 96-plex design, thus reducing adapter-related costs. The asymmetric barcode and UMI arrangement ensures equal sequence representation at the 5' end of the sequenced fragments and minimizes phasing errors. While the epiGBS technology was originally designed as a reference-free application, WellMeth is mainly recommended for large and complex genomes, for which highquality, annotated reference sequences (long scaffolds or pseudochromosomes) are available. Paired-end sequencing technology and the combination of two methylation insensitive restriction enzymes is recommended. This allows for optimization of fragment size distribution and equal representation of different chromosome regions but does not create additional presence or absence differences resulting from methylation in restriction sites. As the forward and reverse adapters of each read pair carry a fixed combination of two different barcodes and two different cohesive ends, demultiplexing of sequence pools will yield double-digested fragments only that will map in a prefixed orientation on genomic reference sequences. The UMI information can be efficiently used for postalignment identification of PCR duplicates.

WellMeth is based on the direct mapping of bisulfitetreated short reads to genomic reference sequences. Its downstream analysis steps are entirely different to those of the epiGBS pipeline: WellMeth applies Je suite (Girardot, Scholtalbers, Sauer, Su, & Furlong, 2016) for fastq deconvolution, as well as for identification of PCR duplicates; bwa-meth (https://github.com/brentp/bwa-meth) for mapping of bisulfite-treated reads; and MethylDackel (https://github.com/dpryan79/MethylDackel) for calculating per-base methylation metrics. The detection of DMCs and DMRs is carried out through pair-wise comparisons using a hidden Markov model-based framework provided by the ComMet tool of the BisulFighter package (Saito et al., 2014). The whole analysis procedure of WellMeth is organized into four overlapping modules: preprocessing, mapping, methylation variant detection, and differential methylation). Custom scripts, allowing step-by-step or fully automated execution, interconnect the modules. WellMeth is fast; bisulfite-treated short-read data representing up to 300 samples (or a full dataset of an Illumina sequencing run) can be mapped and analyzed within 3 d on a high-performance Linux cluster. While being designed for reduced representation bisulfite sequencing analysis, upon modifications in the scripts of the preprocessing module (and keeping consistency with the downstream analysis modules) WellMeth can also be applied for whole-genome bisulfite analysis. We have made the Well-Meth manual, executables, and test data available (https:// sourceforge.net/projects/wellmeth/).

### AUTHOR CONTRIBUTIONS

The study was conceived, designed, and coordinated by TA, KTK, CSJ, and LBE. Plant material was provided by SFS, KTK, BE, LK, AJ, and JDJ. CW and MM performed lab work. IN provided bioinformatics support and developed WellMeth pipeline. MM performed the analysis with contributions from AKR. The manuscript was written by MM and IN with contributions from AKR, CW, and SFS. All authors read and approved the final manuscript.

#### ACKNOWLEDGMENTS

We thank Pernille Bjarup Hansen and Stephan Hentrup at Aarhus University for collecting and processing the leaves samples.

#### DATA AVAILABILITY STATEMENT

WellMeth detailed manual and executables, along with test data, are available on https://sourceforge.net/projects/ wellmeth/. The raw data is openly available in European Nucleotide Archive at https://www.ebi.ac.uk/ena/ data/view/PRJEB38305, reference number PRJEB38305

#### CONFLICT OF INTEREST DISCLOSURE

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study, in the collection, analyses, or interpretation of data, and the writing of the manuscript.

#### ORCID

Marta Malinowska D https://orcid.org/0000-0002-2152-4183

Istvan Nagy https://orcid.org/0000-0001-6549-9525 Cornelis A. M. Wagemaker https://orcid.org/0000-0002-8968-4546

Anja K. Ruud https://orcid.org/0000-0003-2642-7566 Simon F. Svane https://orcid.org/0000-0002-8333-6524 Kristian Thorup-Kristensen https://orcid.org/0000-0001-5476-985X

#### REFERENCES

- Abdel-Ghani, A. H., Sharma, R., Wabila, C., Dhanagond, S., Owais, S. J., Duwayri, M. A., ... Neumann, K. (2019). Genome-wide association mapping in a diverse spring barley collection reveals the presence of QTL hotspots and candidate genes for root and shoot architecture traits at seedling stage. *BMC Plant Biology*, *19*, 216. https://doi.org/10.1186/s12870-019-1828-5
- An, Y. C., Goettel, W., Han, Q., Bartels, A., Liu, Z., & Xiao, W. (2017). Dynamic changes of genome-wide DNA methylation during soybean seed development. *Scientific Reports*, 7, 12263. https://doi. org/10.1038/s41598-017-12510-4
- Anderson, S. N., Stitzer, M. C., Brohammer, A. B., Zhou, P., Noshay, J. M., O'connor, C. H., ... Springer, N. M. (2019). Transposable ele-

ments contribute to dynamic genome content in maize. *Plant Journal*, 100, 1052–1065. https://doi.org/10.1111/tpj.14489

- Barnett, D. W., Garrison, E. K., Quinlan, A., Strömberg, M. P., & Marth, G. T. (2011). Bamtools: A C++ API and toolkit for analyzing and managing bam files, *Bioinformatics*, 27, 1691–1692. https:// doi.org/10.1093/bioinformatics/btr174
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
- Bewick, A. J., & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, 36, 103–110. https://doi. org/10.1016/j.pbi.2016.12.007
- Candaele, J., Demuynck, K., Mosoti, D., Beemster, G. T., Inzé, D., & Nelissen, H. (2014). Differential methylation during maize leaf growth targets developmentally regulated genes. *Plant Physiology*, 164, 1350–1364. https://doi.org/10.1104/pp.113.233312
- Chwialkowska, K., Nowakowska, U., Mroziewicz, A., Szarejko, I., & Kwasniewski, M. (2016). Water-deficiency conditions differently modulate the methylome of roots and leaves in barley (*Hordeum* vulgare L.), Journal of Experimental Botany, 67, 1109–1121, https:// doi.org/10.1093/jxb/erv552
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., ... Jacobsen, S. E. (2008). Shotgun bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 452, 215–219. https://doi.org/10.1038/nature06745
- Cubas, P., Vincent, C., Coen, E., Centre, J. I., Lane, C., & Nr, N. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401, 157–161. https://doi.org/10.1038/43657
- Dixon, P. (2003). Computer program review VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x
- Dubin, M. J., Zhang, P., Meng, D., Remigereau, M. S., Osborne, E. J., Casale, F. P., ... Nordborg, M. (2015). DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife*, 2015, e05255. https://doi.org/10.7554/elife.05255
- Food and Agriculture Organization of the United Nations. (2018). *FAOSTAT Database*. Rome, Italy: FAO. Retrieved from http:// www.fao.org/faostat/en/#data/QC
- Fedoroff, N. V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. *Science*, 338, 758–767. https:// doi.org/10.1126/science.338.6108.758
- Feng, S., Cokus, S. J., Zhang, X., Chen, P., Bostick, M., & Goll, M. G. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy* of Sciences of the United States of America, 107, 8689–8694 https:// doi.org/10.1073/pnas.1002720107
- Frommer, M., Mcdonald, L. E., Millar, D. S., Collist, C. M., Watt, F., Griggt, G. W., ... Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy* of Sciences of the United States of America, 89, 1827–1831. https:// doi.org/10.1073/pnas.89.5.1827
- Gallusci, P., Dai, Z., Génard, M., Gauffretau, A., Leblanc-Fournier, N., Richard-molard, C., & Brunel-muguet, S. (2017). Epigenetics for plant improvement: Current knowledge and modeling avenues. *Trends in Plant Science*, 22, 610–623. https://doi.org/10. 1016/j.tplants.2017.04.009

- Garg, R., Narayana Chevala, V., Shankar, R., & Jain, M. (2015). Divergent DNA methylation patterns associated with gene expression in rice varieties with contrasting drought and salinity stress response. *Scientific Reports*, *5*, 14922. https://doi.org/10.1038/srep14922
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arxiv*, *1207*, 3907. Retrieved from https://arxiv.org/abs/1207.3907
- Gelfman, S., Cohen, N., Yearim, A., & Ast, G. (2013). DNAmethylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Research*, 23, 789–799. https://doi.org/10.1101/gr.143503.112
- Girardot, C., Scholtalbers, J., Sauer, S., Su, S., & Furlong, E. E. M. (2016). Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics*, *17*, 419. https://doi.org/10.1186/s12859-016-1284-2
- Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, 6, 468–481. https://doi.org/10.1038/nprot.2010.190
- Haag, J. R., & Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: Purveyors of non-coding RNA for plant gene silencing. *Nature Reviews Molecular Cell Biology*, 12, 483–492. https://doi.org/ 10.1038/nrm3152
- Heard, E., & Martienssen, R. A. (2014). Review transgenerational epigenetic inheritance: Myths and mechanisms. *Cell*, 157, 95–109. https://doi.org/10.1016/j.cell.2014.02.045
- Kashino-fujii, M., Yokosho, K., Yamaji, N., Yamane, M., Saisho, D., & Sato, K. (2018). Retrotransposon insertion and DNA methylation regulate aluminum tolerance in European barley accessions. *Plant Physiology*, *178*, 716–727. https://doi.org/10.1104/pp.18.00651
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). Hisat: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*, 357– 360. https://doi.org/10.1038/Nmeth.3317.
- Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramirez, F., Vesztrocy, A. W., Naldi, A., ... Tang, H. (2018). Goatools: A Python library for gene ontology analyses. *Scientific Reports*, *8*, 10872. https://doi.org/ 10.1038/S41598-018-28948-Z
- Konate, M., Wilkinson, M., Mayne, B., Pederson, S., Scott, E., Berger, B., & Rodriguez Lopez, C. (2018). Salt stress induces non-CG methylation in coding regions of barley seedlings (*Hordeum Vulgare*). Epigenomes, 2, 12. https://doi.org/10.3390/ Epigenomes2020012
- Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology*, 5, 3. https://doi.org/10.3390/ Biology5010003
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11, 204–220. https://doi.org/10. 1038/Nrg2719
- Le, T., Schumann, U., Smith, N. A., Tiwari, S., Chi, P., Au, K., ... Wang, M. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis. Genome Biology*, 15, 458. https://doi.org/10.1186/S13059-014-0458-
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and Samtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/ Bioinformatics/Btp352

- Li, Q., Eichten, S. R., Hermanson, P. J., Zaunbrecher, V. M., Song, J., Wendt, J., ... Springer, N. M. (2014). Genetic perturbation of the maize methylome. *The Plant Cell*, *26*, 4602–4616. https://doi.org/ 10.1105/Tpc.114.133140
- Li, Q., Gent, J. I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C. D., ... Springer, N. M. (2015). RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 14728–14733. https://doi.org/10.1073/Pnas.1514680112
- Liang, L., Chang, Y., Lu, J., Wu, X., Liu, Q., Zhang, W., ... Zhang, B. (2019). Global methylomic and transcriptomic analyses reveal the broad participation of DNA methylation in daily gene expression regulation of *Populus trichocarpa*. *Frontiers in Plant Science*, 10, 243. https://doi.org/10.3389/Fpls.2019.00243
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Nery, J. R., ... Ecker, J. R. (2010). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462, 315– 322. https://doi.org/10.1038/Nature08514
- Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., ... Seymour, G. B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening, *Nature Genetics*, 38, 948–952. https://doi.org/10.1038/ng1841
- Mapleson, D., Swarbreck, D., Venturini, L., & Kaithakottil, G. (2018). Efficient and accurate detection of splice junctions from RNAseq with Portcullis, *Gigascience*, 7, giy131. https://doi.org/10.1093/ gigascience/giy131
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., ... Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 544, 427– 433. https://doi.org/10.1038/nature22043
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, *226*, 792–801. https://doi.org/10.1126/science. 15739260
- Meyer, P., Niedenhof, I., & Ten Lohuis, M. (1994). Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*. *The EMBO Journal*, *13*, 2084–2088. https://doi. org/10.1002/j.1460-2075.1994.tb06483.x
- Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., & Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis. Nature*, 411, 212– 214. https://doi.org/10.1038/35075612
- Newton, A. C., Flavell, A. J., George, T. S., Leat, P., Mullholland, B., Ramsay, L., & Russell, J. (2011). Crops that feed the world 4. Barley: A resilient crop? Strengths and weaknesses in the context of food security. *Food Security*, *3*, 141. https://doi.org/10.1007/s12571-011-0126-3
- Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. Do., Li, Q., ... Schmitz, R. J. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*, 17, 194. https:// doi.org/10.1186/s13059-016-1059-0
- Noshay, J. M., Anderson, S. N., Zhou, P., Ji, L., Ricci, W., Lu, Z., ... Springer, N. M. (2019). Monitoring the interplay between transposable element families and DNA methylation in maize. *PLoS Genetics*, *15*, e1008291. https://doi.org/10.1371/journal.Pgen.1008291
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruc-

tion of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33, 290–295. https://doi.org/10.1038/nbt.3122.stringtie

- Pikaard, C. S., & Scheid, O. M. (2014). Epigenetic regulation in plants. *Cold Spring Harbor Perspectives in Biology*, *6*, a019315. https://doi. org/10.1101/cshperspect.a019315
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033
- Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology*, 10, 417–430. https://doi.org/10.1038/nrmicro2790
- Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, V., & Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS ONE*, *14*, e0217148. https://doi.org/ 10.1371/journal.pone.0217148
- Richards, C. L., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colomé-Tatché, M., ... Verhoeven, K. J. F. (2017). Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology Letters*, 20, 1576–1590. https://doi.org/10.1111/ ele.12858
- Richards, C. L., Bossdorf, O., & Verhoeven, K. J. F. (2010). Understanding natural epigenetic variation. *New Phytologist*, 187, 562– 564. https://doi.org/10.1111/j.1469-8137.2010.03369.x
- Richards, E. J. (2006). Inherited epigenetic variation—Revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395–401. https://doi.org/ 10.1038/nrg1834
- Rohart, F., Gautier, B., Singh, A., & Cao, K. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, 13, e1005752. https://doi.org/10.1371/ journal.pcbi.1005752
- Saito, Y., Tsuji, J., & Mituyama, T. (2014). Bisulfighter: Accurate detection of methylated cytosines and differentially methylated regions. *Nucleic Acids Research*, 42, e45. https://doi.org/10.1093/ nar/gkt1373
- SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., ... Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274, 765–768. https://doi.org/10.1126/science.274. 5288.765
- Schauberger, B., Ben-ari, T., Makowski, D., Kato, T., Kato, H., & Ciais, P. (2018). Yield trends, variability and stagnation analysis of major crops in France over more than a century. *Scientific Reports*, *8*, 16865. https://doi.org/10.1038/s41598-018-35351-1
- Singh, S., Nandha, P. S., & Singh, J. (2017). Transposon-based genetic diversity assessment in wild and cultivated barley. *The Crop Journal*, 5, 296–304, https://doi.org/10.1016/j.cj.2017.01.003
- Song, Q., Lu, X., Li, Q., Chen, H., Hu, X., Ma, B., & Zhang, W. (2013). Genome-wide analysis of DNA methylation in soybean. *Molecular Plant*, 6, 1961–1974. https://doi.org/10.1093/mp/sst123
- Springer, N. M. (2013). Epigenetics and crop improvement. *Trends in Genetics*, 29, 241–247. https://doi.org/10.1016/j.tig.2012.10.009
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., ... Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis. Nature Structural and Molecular Biology*, *21*, 64–72. https://doi.org/10.1038/nsmb.2735
- Svane, S. F., Jensen, C. S., & Kristensen, K. T. (2019). Construction of a large-scale semi-field facility to study genotypic differences in deep root growth and resources acquisition. *Plant Methods*, 15, 26. https://doi.org/10.1186/s13007-019-0409-9

- Takuno, S., & Gaut, B. S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proceedings of the National Academy of Sciences of the United States of America, 110, 1797–1802. https://doi.org/10.1073/pnas.1215380110
- Takuno, S., Ran, J. H., & Gaut, B. S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. *Nature Plants*, 2, 15222. https://doi.org/10.1038/nplants.2015.222.
- Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., & O, P. S. C. H. (2016). BSRADseq: Screening DNA methylation in natural populations of non-model species. *Molecular Ecology*, 25, 1697– 1713. https://doi.org/10.1111/mec.13550
- Van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016). Epigbs: Referencefree reduced representation bisulfite sequencing. *Nature Methods*, 13, 322–324. https://doi.org/10.1038/nmeth.3763
- Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, L., & Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, 7, giy093. https://doi.org/10.1093/gigascience/giy093.
- Wang, J., Xia, Y., Li, L., Gong, D., Yao, Y., Luo, H., & Wu, H. (2013). Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. *BMC Genomics*, 14, 11. https:// doi.org/10.1186/1471-2164-14-11
- Wang, X., Hu, L., Wang, X., Li, N., Xu, C., Gong, L., & Liu, B. (2016). DNA methylation affects gene alternative splicing in plants: An example from rice. *Molecular Plant*, 9, 305–307. https://doi.org/10. 1016/j.molp.2015.09.016
- Wang, Y., Liang, W., & Tang, T. (2018). Constant conflict between Gypsy LTR retrotransposons and CHH methylation within a stressadapted mangrove genome. New Phytologist, 220, 922–935. https:// doi.org/10.1111/nph.15209
- Wicker, T., Schulman, A. H., Tanskanen, J., Spannagl, M., Twardziok, S., Mascher, M., ... Gundlach, H. (2017). The repetitive landscape of the 5100 Mbp barley genome. *Mobile DNA*, *8*, 22. https://doi.org/ 10.1186/s13100-017-0102-3
- Zakrzewski, F., Schubert, V., Viehoever, P., Minoche, A. E., Dohm, J., Himmelbauer, H., ... Schmidt, T. (2014). The CHH motif in sugar beet satellite DNA: A modulator for cytosine methylation. *The Plant Journal*, 78, 937–950. https://doi.org/10.1111/tpj.12519
- Zemach, A., Kim, M. Y., Hsieh, P., Colemanderr, D., Eshed-Williams, L., Thao, K., ... Zilberman, D. (2013). The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1containing heterochromatin. *Cell*, 153, 193–205. https://doi.org/10. 1016/j.cell.2013.02.033
- Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328, 916–919. https://doi.org/10.1126/science.1186366.
- Zhang, H., Lang, Z., & Zhu, J. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, *19*, 489–506. https://doi.org/10.1038/s41580-018-0016-z
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data, *Bioinformatics*, 28, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606
- Zhou, S., Liu, X., Zhou, C., Zhou, Q., Zhao, Y., Li, G., & Zhou, D. X. (2016). Cooperation between the H3K27me3 chromatin mark and non-CG methylation in epigenetic regulation. *Plant Physiol*ogy, *172*, 1131–1141. https://doi.org/10.1104/pp.16.01238

### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article. How to cite this article: Malinowska M, Nagy I, Wagemaker CAM, et al. The cytosine methylation landscape of spring barley revealed by a new reduced representation bisulfite sequencing pipeline, WellMeth. *Plant Genome*. 2020;13:e20049. https://doi.org/10.1002/tpg2.20049