Final Report: Harnessing cytosine DNA hypomethylation to explore the potential for crop improvement in wheat

Team: Dr Natasha Yelina[1§], Dr Mark Winfield[3], Dr Ian Henderson[1], Dr Alison Bentley[2] and Prof Keith Edwards[3]

[1]Department of Plant Sciences, University of Cambridge; [2]NIAB, Cambridge; [3]University of Bristol

[§]correspondence: ne240@cam.ac.uk

1. Background: The **overall aim** of this project was to generate a genome-wide hypomethylated wheat to study the effects of cytosine DNA methylation on meiotic recombination and hybrid vigour, or heterosis. Manipulating both, meiotic recombination and hybrid vigour, have potential to improve the yields of cultivated wheat. A DNA hypomethylated wheat line would help our team to understand the role of epigenetic information behind these two processes. It would also serve as a valuable tool for the crop research community to study the role of epigenetics in gene regulation and genome stability in wheat, a complex eukaryote genome with a high DNA methylation content. In *Arabidopsis* and some other plant species including monocot cereals, a loss of DDM1, a SNF2-family nucleosome remodeler, results in a genome-wide loss of cytosine DNA methylation[1–3]. Our **strategy** was to edit putative wheat *TaDDM1* homologues via CRISPR/Cas9. Since wheat transformation is a long-term and expensive experiment and not all *in silico* predicted guide RNAs are equally efficient in CRISPR/Cas9 gene editing, we used the **OpenPlant funding** to test a number of candidate guide RNAs and select the most efficient ones for editing *TaDDM1* via stable transformation.

2. Results:

2.1. *In silico* guide RNA design: Since the submission of our OpenPlant Proposal, wheat genome assembly has been updated. In the latest genome version, we identified wheat *TaDDM1* homologues on chromosomes 2A, 2B, 2D, 4A (2 homologues), 4B, 4D, 7A and 7D (Table 1).

Table 1. *TaDDM1* homologues

Gene in *T. aestivum*	Genome Location
TraesCS2A02G368200	Chromosome 2A: 611,934,295-611,940,619 forward strand
TraesCS2D02G365000	Chromosome 2D: 470,202,238-470,211,751 forward strand
TraesCS2B02G385300	Chromosome 2B: 548,943,161-548,951,290 forward strand
TraesCS4A02G372200	Chromosome 4A: 647,087,353-647,093,572 reverse strand
TraesCS4A02G261000	Chromosome 4A: 573,358,241-573,365,523 reverse strand
TraesCS4B02G053600	Chromosome 4B: 42,317,502-42,324,556 forward strand
TraesCS4D02G053600	Chromosome 4D: 29,798,552-29,804,972 forward strand
TraesCS7A02G074600	Chromosome 7A: 40,360,598-40,366,793 reverse strand.
TraesCS7D02G069700	Chromosome 7D: 40,449,197-40,455,468 reverse strand

We reasoned that designing guide RNAs targeting three *TaDDM1* homeologues at once rather than each *TaDDM1* gene individually would greatly reduce the number of guide RNAs for CRISPR/Cas9-mediated *TaDDM1* editing. We, therefore, aligned 9 putative *TaDDM1* genomic sequences and generated a phylogenetic tree using Clustal Omega on-line alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Figure 1). We could identify regions of high sequence homology between *TaDDM1* homeologues that were ideal targets for gene editing (Figure 2). Base on the nucleotide sequence alignment and the phylogenetic tree data, we hypothesize that TraesCS4A02G372200 has been mis-assigned to chromosome 4A and represents the 'missing' *TaDDM1* homeologue on chromosome 7B (Figure 1). We used WheatCRISPR on-line tool (https://crispr.bioinfo.nrc.ca/WheatCrispr/) to *in silico* design a total of 17 guide RNAs targeting putative *TaDDM1* homologs on chromosomes 2, 4 and 7. We designed guide RNAs targeting *TaDDM1* exons, upstream or within the putative *TaDDM1* essential SNF2 helicase domain (Table 2 and Figures 2 and 3).

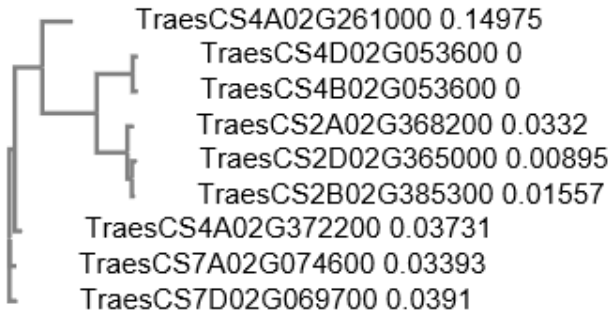

Figure 1. Phylogenetic tree of *TaDDDM1* genomic sequences

```
4B      GCCAACCTTGTTCCATTGTTGACTGGGGGAAAGTTGAAATCTTACCAGATAAAGGGTGTT    1662
4D      GCCAATCTTGTTCCATTGTTGACTGGGGGAAAGTTGAAATCTTACCAGATAAAGGGTGTT    1448
4A      GCCAACCTTGTTCCATTGTTGACTGGGGGAAAGTTGAAATCTTACCAGATAAAGGGTGTT    1791
        ***** **************************************************
```

Figure 2. Example of a guide RNA targeting three *TaDDM1* homeologues on chromosome 4. A fragment of genome sequence alignment is shown; grey - PAM, green - guide RNA sequence, yellow - single nucleotide polymorphisms (SNPs) within the guide RNA target sequence. Perfect complementarity of a guide RNA to 10 nucleotides adjacent to PAM is essential for Cas9-mediated gene targeting[4].

Table 2. Guide RNA sequences designed in this study

Guide RNA sequence	Target Chromosome	Target Domain
CACCAGTCATTAATGGTACA	2	upstream of SNF2 helicase domain
GACACCTCTTCAGAATAACC	2	within SNF2 helicase domain
GGTTATTCTGAAGAGGTGTC	2	within SNF2 helicase domain
ATGTGGAGCAGATGCTTCCA	2	within SNF2 helicase domain
TTGCTTGAGAGCACAGACAT	2	within SNF2 helicase domain
CGGGAAAGCGGGCGAGCTCA	4	upstream of SNF2 helicase domain
CCTTCAGCGCCTCCGCAATG	4	upstream of SNF2 helicase domain
TCTTGTTCCATTGTTGACTG	4	upstream of SNF2 helicase domain
ATACGAAGTGCAAATTACTG	4	within SNF2 helicase domain
TGATATATTCTCATCCCATG	4	within SNF2 helicase domain
AAGCAGAATGGAAAAGATCA	4	within SNF2 helicase domain
TATAAATATTGTTCAGAAGA	7	upstream of SNF2 helicase domain
TTCACGCGATCTTGTAAGCA	7	upstream of SNF2 helicase domain
CAGTTTGACAGAGTAGAAAG	7	within SNF2 helicase domain
CTGATTGGTCTGATTTATCA	7	within SNF2 helicase domain
GTAGTGACCACAGCTCTGCT	7	within SNF2 helicase domain
CATGAAGCTTTGAGACAACA	7	within SNF2 helicase domain

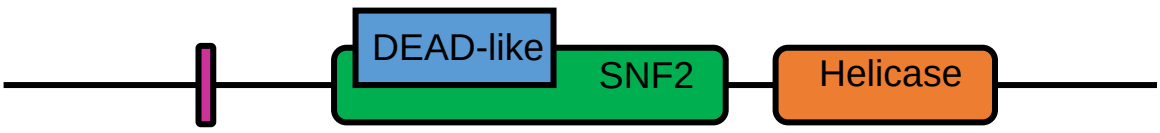

Figure 3. A schematic representation of TaDDM1 functional domains.

2.2. Testing guide RNA genome editing efficiencies: We reasoned that the most informative and time- and cost-effective way would be to test gene editing efficiencies of *in silico* predicted guide RNAs in wheat protoplasts[5]. A pBUN421 vector which contains a maize codon-optimized Cas9 and has been successfully used for CRISPR/Cas9 genome editing in wheat protoplasts and via stable transformation[5]. Previous studies have also shown that multiple guide RNAs can be expressed from a polycistronic tRNA-guide RNA gene[6]. We followed the above strategy and split 17 *in silico* designed

guide RNAs targeting *TaDDM1* into 4 groups, each containing 2-6 guide RNAs. We co-transfected wheat protoplasts of *Cadenza* variety with one of the resulting pBUN421-TaDDM1#1 - pBUN421-TaDDM1#4 vectors and a GFP-expressing plasmid, an indicator of protoplast transformation efficiency. We reached a ~70% transfection efficiency with the GFP plasmid (Figure 4).

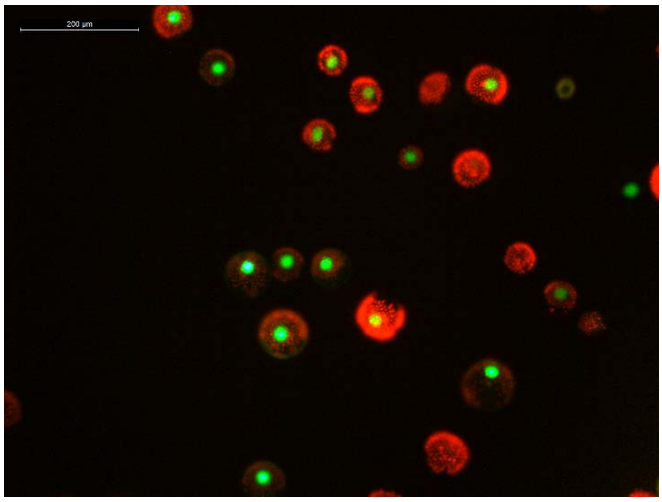

Figure 4. Visualizing GFP expression as a proxy of protoplast transformation efficiency. GFP is visible as green circles, chlorophyll in protoplasts shows red autofluorescence.

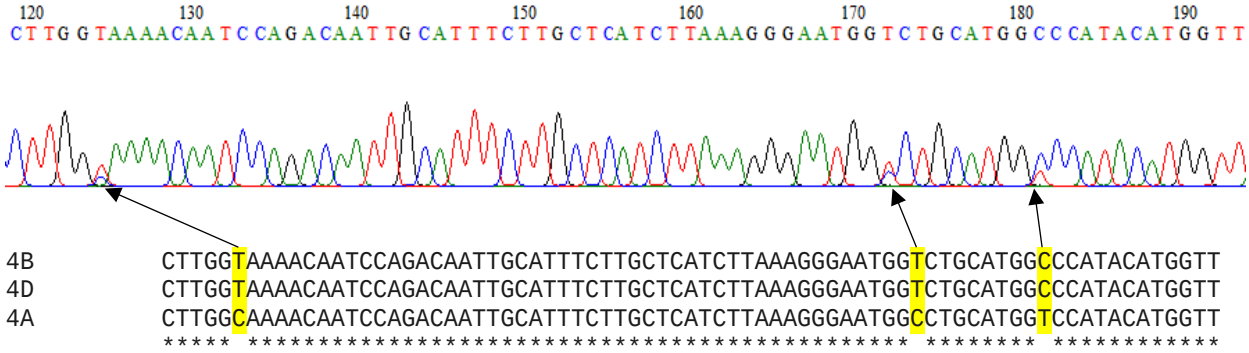

Figure 5. Sanger sequencing (top) and a fragment of genomic sequence alignment (bottom) of 4A *TaDDM1* homeologues. Arrows point to the mixed sequencing traces corresponding to single nucleotide polymorphisms (highlighted in yellow) between the homeologuous genes.

We extracted genomic DNA from protoplasts following transfection and PCR amplified target regions with oligonucleotides specific for three homoeloguous genes. Prior to amplification of protoplast DNA, we tested oligonucleotide specificity using *Cadenza* wild type DNA and confirmed amplification of the expected genome regions by Sanger sequencing (Figures 5 and 6).

We then generated next generation sequencing libraries from the resulting amplicons to be sequenced on a MiSeq instrument. We expect to obtain ~100,000 next generation sequencing reads per amplicon which is consistent with the coverage depth required to detect CRISPR/Cas9-generated mutations in wheat protoplasts[5].

We performed next generation sequencing on a MiSeq instrument and analysed gene editing efficiency of *TaDDM1* gene in wheat protoplasts using on-line software (http://crispr-ga.net http://www.rgenome.net/be-analyzer/#!result and http://www.rgenome.net/cas-analyzer/#!). As shown in Table 3, guide RNAs varied in their gene editing efficiency, but we could identify at least one efficiently targeting *TaDDM1* homeologues on chromosomes 2, 4 and 7. Guide RNAs selected for stable wheat transformation to generate a *taddm1* line are highlighted in grey. Figure 7 shows examples of CRISPR/Cas9-mediated editing of *TaDDM1* in wheat protoplasts.

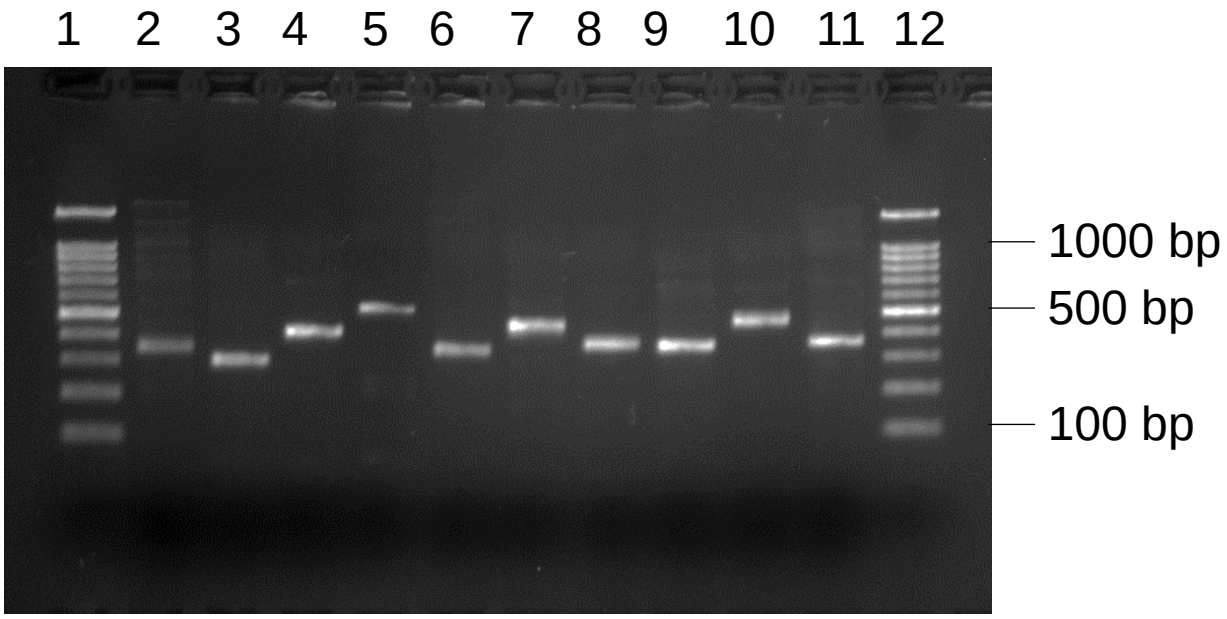

Figure 6. Amplification of guide RNA target sites from wheat protoplasts. Lanes 2-11: wheat amplicons harbouring guide RNA target sites. Lanes 1 and 12 - DNA size markers. Next generation sequencing libraries were constructed from amplicons in lanes 2-11.

Table 3. Efficiency of *TaDDM1* gene editing in wheat protoplasts

Guide RNA sequence	Target Chromosome	Total Reads Analysed	Mutated Reads	Proportion of Mutated Reads (%)
CACCAGTCATTAATGGTACA	2	70,695	0	0
GACACCTCTTCAGAATAACC	2	65,895	11	0.02
ATGTGGAGCAGATGCTTCCA	2	44,866	84	0.19
TTGCTTGAGAGCACAGACAT	2	27,506	20	0.07
CGGGAAAGCGGGCGAGCTCA	4	53,694	343	0.64
CCTTCAGCGCCTCCGCAATG	4	53,694	343	0.64
TCTTGTTCCATTGTTGACTG	4	44,474	2	0
ATACGAAGTGCAAATTACTG	4	59,166	0	0
TGATATATTCTCATCCCATG	4	59,166	0	0
AAGCAGAATGGAAAAGATCA	4	37,380	60	0.16
TATAAATATTGTTCAGAAGA	7	76,900	0	0
TTCACGCGATCTTGTAAGCA	7	76,900	0	0
CAGTTTGACAGAGTAGAAAG	7	76,900	0	0
CTGATTGGTCTGATTTATCA	7	76,900	0	0
GTAGTGACCACAGCTCTGCT	7	70,040	0	0
CATGAAGCTTTGAGACAACA	7	43,887	9	0.02

3. References

1. Vongs et al. *Science* 260, 1926–1928 (1993). **2**. Higo et al. *Mol Genet Genomics* 287, 785–792 (2012). **3**. Corem et al. *Plant Cell* 30, 1628-1644 (2018). **4**. Zheng, T. *et al. Scientific Reports* 7:40638 (2017). **5**. Wang et al. *CRISPR Journal* 1, 65-74 (2018). **6**. Xie et al. *PNAS* 112, 3570-3575 (2015).

Figure 7. Examples of CRISPR/Cas9-mediated mutations in *TaDDM1*. Guide RNA recognition sequence highlighted in green, PAM highlighted in grey.

Guide RNA ATGTGGAGCAGATGCTTCCA (Chr2) - 1-, 2- and 3-bp deletions

```
Wt  TTTTAAGACGGATGAAGGTGGATGTGGAACAGATGCTTCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGGTGGATGTGGAACAGATGC-TCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGATGGATGTGGAGCAGATGC-TCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGATGGATGTGGAGCAGATGC-TTCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGGTGGATGTGGAACAGAT--GCCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGATGGATGTGGAGCAGAT--GCCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGGATGAAGGTGGATGTGGAACAG---ATGCCACGGAAGAAAGAGATAATCA
Mut TTTTAAGACGAATGAAGGTGGATGTGGAACAG---ATGCCACGGAAGAAAGAGATAATCA
```

Guide RNA TTGTTTGAGAGCACAGACAT (Chr2) - 4-, 7-bp deletions

```
Wt  GAAAAAACATTTGACAACTACTTGCTTGAGAGCACAGACATAGGTATGCATATGTAAGCTCGTG
Mut GAAAAGACATTTGACAACTACTTGTTTGAG-------ACATAGGTATGCATATGTAAGCTTGTG
Mut GAAAAGACATTTGACAACTACTTGTTTGAGAGCAC----ATAGGTATGCATATGTAAGCTTGTG
```

Guide RNA GACACCTCTTCAGAATAACC (Chr 2) - a 49-bp insertion

```
Wt  TAAACTCCTTTTGACTGGGACACCTCTTCAGAATA-------------------------------------------------ACCTGGCAGAGTTGTGGTCATT
Mut TAAACTTCTTTTGACTGGGACACCTCTTCAGAATAAGCACGTTGCGCAGATTCTCGACTCACGGATGAACACTAAGTACGATGAACCTGGCAGAATTGTGGTCGTT
```

Guide RNAs TGGGAAAGCGGGCGAGCTCA and CCTTCAGCGCCTCCGCAATG (Chr 4) 31-33-bp deletions

```
Wt  GATGGGAAAGCGGGCGAGCTCAAGGCCAACGGCGGCGAGGCGCACCCCATTGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GACGGGAAAGCGGGCGAGC--------------------------------CGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GACGGGAAAGCGGGCGAGC-------------------------------CGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GATGGGAAAGCGGGCGAGC-------------------------------CGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GATGGGAAAGCGGGCGAG--------------------------------TGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GACGGGAAAGCGGGCGAG--------------------------------TGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
Mut GATGGGAAAGCGGGCGAGC-------------------------------TGCGGAGGCGCTGAAGGCTGAAGAACAGCTACTGAACTC
```

Guide AAGCAGAATGGAAAAGATCA (Chr 4) - 2-, 3-, 4- and 5-bp deletions

```
Wt  TTGTGGCCACCCTGATCTTTTCCATTCTGCTTTTGACTCAAACAGTGCGTCATCTCATTGGACATGTC
Mut TTGTGGCCACCCTGA--TTTTCCATTCTGCTTTTGACTCAAACAGTGCGTCATCTCATTGGACATGTC
Mut TTGTGGCCACCCTGA---TTTCCATTCTGCTTTTGACTCAAACAGTGCGTCATCTCATTGGACATGTC
Mut TTGTGGCCACCCTGA----TTCCATTCTGCTTTTGACTCAAACAGTGCGTCATCTCATTGGACATGTC
Mut TTGTGGCCACCCTGA----TTCCATTCTGCTTTTGACTCAAACAGTGCGTCATCTCATTGGACATGTC
Mut TTGTGGCCACCCTGA-----TCCATTCTGCCTTTGACTCAAACAGTGCGT---CTTATTGGACATGTC
```

Guide RNA CATGAAGCTTTGAGACAACA (Chr 7) - a 28-bp insertion

```
Wt  AACTGATGAGAACAAAAGAGTCCTTGT----------------------------TGTCTCAAAGCTTCATGCCATTTTGCGTCCATTCCTTCTA
Mut AACTGATGAGAACAAAAGAGTCCTTGTGCCGATCTGGGCCAGGAGGTTATCCAGGTGTCTCAAAGCTTCATGCCATTTTGCGTCCATTCCTTCTA
```