Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation

Jered M Wendte et al
Figure 1. Expression of AtCMT3 in E. salsugineum results in increased CHG methylation. (A) Schematic of the experiment. Two E. salsugineum lines derived from transformations with genomic A. thaliana CMT3 were propagated by single seed decent for six generations (T1–T6). The two lines are continued on next page.
referred to as AtCMT3-L1 and AtCMT3-L2, followed by the generation (T1–T6). For additional lines analyzed in this study see Figure 1—figure supplement 1 and Supplementary file 1. (B) Genome browser view of CHG methylation levels derived from whole genome bisulfite sequencing. The image illustrates the gains in CHG methylation that occur in regions that are methylated in wild type, as well as over gene bodies with no pre-existing DNA methylation (boxed in red). Scales on tracks designate the weighted percent methylation, with 1 = 100% on the top strand and –1 = 100% on the bottom strand. Only CHG methylation is shown. For methylation in all contexts see Figure 1—figure supplement 2.

DOI: https://doi.org/10.7554/eLife.47891.002
Figure 1—figure supplement 1. Diagram of all experimental *E. salsugineum* lines analyzed in this study with associated data collected. DOI: https://doi.org/10.7554/eLife.47891.003
Figure 1—figure supplement 2. CHG methylation increases in AtCMT3-expressing E. salsugineum lines. (A) Genome browser view of CHG methylation levels derived from whole genome bisulfite sequencing. The image illustrates the gains in CHG methylation that occur in regions that are methylated in wild type, as well as over gene bodies with no pre-existing DNA methylation (boxed in red). Scales on tracks designate the weighted percent methylation, with 1 = 100% on the top strand and −1 = 100% on the bottom strand. Cytosine methylation is divided into CG (red), CHG (blue), and CHH (yellow) sequence contexts. Corresponds to the region shown in Figure 1B.

DOI: https://doi.org/10.7554/eLife.47891.004
Figure 1—figure supplement 3. CG, CHG, and CHH DMRs co-localize. (A–L) Upset plots showing the number and overlap between hyper DMRs identified for each sequence context, CG, CHG, and CHH, within each generation of each lineage, AtCMT3-L1 and AtCMT3-L2: (A) AtCMT3-L1^{T1}; (B) AtCMT3-L1^{T2}; (C) AtCMT3-L1^{T3}; (D) AtCMT3-L1^{T4}; (E) AtCMT3-L1^{T5}; (F) AtCMT3-L1^{T6}; (G) AtCMT3-L2^{T1}; (H) AtCMT3-L2^{T2}; (I) AtCMT3-L2^{T3}; (J) AtCMT3-L2^{T4}; (K) AtCMT3-L2^{T5}; (L) AtCMT3-L2^{T6}.

DOI: https://doi.org/10.7554/eLife.47891.005
Figure 1—figure supplement 4. AtCMT3 expression results in gains in CHG methylation over similar regions between lineages and across generations. Matrix showing the number and overlap between hyper-CHG DMRs identified in each individual in AtCMT3-L1 and AtCMT3-L2. The top half of the matrix shows the number of overlapping DMRs and the bottom half shows the p-value calculated based on fisher’s exact test (***p significant <0.0004, Bonferroni correction). See also Figure 2A and Supplementary file 3.

DOI: https://doi.org/10.7554/eLife.47891.006
Figure 1—figure supplement 5. Genome wide percent CHG methylation is correlated with AtCMT3 transgene expression levels. AtCMT3 expression was determined by RNA-seq (first panel) or qRT-PCR (second panel). For qRT-PCR AtCMT3 expression was normalized to TUB4 expression (Thhalv10003210m). See also Supplementary file 4.
DOI: https://doi.org/10.7554/eLife.47891.007
Figure 2. AtCMT3 expression results in CHG methylation over repeats and a subset of gene bodies. (A) Heatmap of % CHG methylation over hyper-CHG differentially methylated regions (DMRs) defined by comparing all AtCMT3-L1 and L2 generations (T1–T6) to wild type. DMRs are ranked by % CHG methylation levels in wild type plants showing that the majority of gains in CHG methylation occurred over regions with pre-existing CHG methylation classified as repeats or intergenic regions. A subset of regions with no pre-existing CHG methylation, mainly classified as genes, also showed gains in CHG methylation, especially in AtCMT3-L2 individuals. See also Supplementary file 2 and Supplementary file 3. (B) Metaplot summarizing % CHG methylation over repetitive sequences for each Line. (C) Metaplot summarizing % CHG methylation over gene bodies for each Line. (D) The number of genes gaining a minimum of 5% CHG methylation (CHG-gain genes) in each lineage. See Supplementary file 5 for lists of CHG-gain genes.

DOI: https://doi.org/10.7554/eLife.47891.008
## Figure 2—figure supplement 1.

AtCMT3 expression results in gains in CHG methylation over similar genes between lineages and across generations. Matrix showing the number and overlap between genes gaining a minimum of 5% CHG methylation identified in each individual in AtCMT3-L1 and AtCMT3-L2. The top half of the matrix shows the number of overlapping genes and the bottom half shows the p-value calculated based on a hypergeometric test (***p significant <0.0004, Bonferroni correction). See also Figure 2D and Supplementary file 5.

DOI: https://doi.org/10.7554/eLife.47891.009

<table>
<thead>
<tr>
<th></th>
<th>AtCMT3-L1T1 (99)</th>
<th>AtCMT3-L1T2 (126)</th>
<th>AtCMT3-L1T3 (257)</th>
<th>AtCMT3-L1T4 (239)</th>
<th>AtCMT3-L1T5 (401)</th>
<th>AtCMT3-L1T6 (583)</th>
<th>AtCMT3-L2T1 (718)</th>
<th>AtCMT3-L2T2 (2996)</th>
<th>AtCMT3-L2T3 (3553)</th>
<th>AtCMT3-L2T4 (4055)</th>
<th>AtCMT3-L2T5 (1937)</th>
<th>AtCMT3-L2T6 (88)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCMT3-L1T1</td>
<td></td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L1T2</td>
<td>66</td>
<td>103</td>
<td>93</td>
<td>94</td>
<td>94</td>
<td>121</td>
<td>115</td>
<td>114</td>
<td>103</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L1T3</td>
<td>***</td>
<td>***</td>
<td>175</td>
<td>194</td>
<td>214</td>
<td>185</td>
<td>235</td>
<td>237</td>
<td>244</td>
<td>222</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L1T4</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>203</td>
<td>208</td>
<td>159</td>
<td>217</td>
<td>214</td>
<td>223</td>
<td>202</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L1T5</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>355</td>
<td>242</td>
<td>349</td>
<td>356</td>
<td>359</td>
<td>315</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L1T6</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T1</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>679</td>
<td>670</td>
<td>674</td>
<td>580</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T2</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T3</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T4</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T5</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCMT3-L2T6</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p value of overlap: ***p < 0.0004
Figure 2—figure supplement 2. The number of genes gaining a minimum of 5% CHG methylation (CHG-gain genes) is correlated with AtCMT3 transgene expression levels. AtCMT3 expression was determined by RNA-seq (first panel) or qRT-PCR (second panel). For qRT-PCR AtCMT3 expression was normalized to TUB4 expression (Thhalv10003210m). See also Figure 2D and Supplementary file 4.

DOI: https://doi.org/10.7554/eLife.47891.010
Figure 3. Gains in CHG methylation do not alter H3K9me2 levels or distribution. (A) Genome browser view of CHG methylation levels and H3K9me2 ChIP sequencing levels in the T3 and T5 generations of the AtCMT3-L1 and L2 lineages. Arrows indicate gains of CHG methylation over gene bodies in Figure 3 continued on next page.
Figure 3 continued

the AtCMT3-L2 generations that do not show H3K9me2 enrichment. Scales on methylation tracks designate the weighted percent methylation, with 1 = 100% on the top strand and −1 = 100% on the bottom strand. Scales on the H3K9me2 tracks indicate the number of mapped reads and are not adjusted for library size (See C–F for comparison of normalized reads). DNA methylation is only shown in the CHG context. For DNA methylation in all contexts see Figure 3—figure supplement 1. (B) Metaplot of % CHG methylation over H3K9me2 ChIP peaks identified in wild type plants. (C–F) Metaplot of H3K9me2 ChIP-sequencing enrichment over H3K9me2 ChIP peaks defined in wild type plants and over CHG-gain genes in AtCMT3-L1T3 (C), AtCMT3-L1T5 (D), AtCMT3-L2T3 (E), and AtCMT3-L2T5 (F). Reads were normalized to library size. See Supplementary file 5 for lists of CHG-gain genes in each lineage.

DOI: https://doi.org/10.7554/eLife.47891.011
A. 

Figure 3—figure supplement 1. Gains in CHG methylation do not alter H3K9me2 levels or distribution. (A) Genome browser view of CHG methylation levels and H3K9me2 ChIP sequencing levels in the T3 and T5 generations of the AtCMT3-L1 and AtCMT3-L2 lineages. Arrows indicate gene bodies that gain CHG methylation in the AtCMT3-L2 generations but do not show H3K9me2 enrichment. Scales on methylation tracks designate the weighted percent methylation, with 1 = 100% on the top strand and /C0 1 = 100% on the bottom strand. Scales on the H3K9me2 tracks indicate the number of mapped reads and are not normalized. Cytosine methylation is divided into CG (red), CHG (blue), and CHH (yellow) sequence contexts. Corresponds to the region shown in Figure 3A.

DOI: https://doi.org/10.7554/eLife.47891.012
Figure 3—figure supplement 2. Gains in CHG methylation do not alter H3K9me1 levels or distribution. (A) Genome browser view of CHG methylation levels and H3K9me1 ChIP sequencing levels in the T2c generation of the AtCMT3-L3 lineage. In wild type and AtCMT3-L3 T2c, H3K9me1 can be found enriched in regions also enriched for H3K9me2. The arrow indicates a gene body that gains CHG methylation in AtCMT3-L3 but does not show H3K9me1 enrichment. Scales on methylation tracks designate the weighted percent methylation, with 1 = 100% on the top strand and −1 = 100% on the bottom strand. Scales on the H3K9me1 tracks indicate the number of mapped reads and are not normalized. Cytosine methylation is divided into CG (red), CHG (blue), and CHH (yellow) sequence contexts. (B) Metaplot of H3K9me1 ChIP-sequencing enrichment over H3K9me2 ChIP peaks defined in wild type plants and over CHG-gain genes in AtCMT3-L3 T2c. (C) Metaplot of H3K9me1 ChIP-sequencing enrichment comparing enrichment over CHG-gain genes in AtCMT3-L3 T2c to genes that do not gain CHG methylation (UM genes).

DOI: https://doi.org/10.7554/eLife.47891.013
Figure 4. Genes that gain CHG methylation have A. thaliana gbM gene characteristics. (A) Relationship between levels of CHG methylation gain and log₂ fold change (FC) in expression for CHG-gain genes in AtCMT3-L2^T4 relative to wild type. Genes with zero FPKM values were removed from the analysis. See Figure 4—figure supplement 1 for additional individuals analyzed. (B–E) Comparison of the (B) distribution of gene lengths, (C) number of exons, (D) expression levels, and (E) frequency of CHG sites between CHG-gain genes and unmethylated genes (UM). P-values were calculated using a Wilcoxon rank-sum test. Boxes indicate the first and third quartiles, with the center line indicating the median and notches the 95% confidence interval of the median. Whiskers show 1.0 times the interquartile range and outliers beyond this range were excluded for visualization purposes, but included in all calculations.

DOI: https://doi.org/10.7554/eLife.47891.014
Figure 4—figure supplement 1. Changes in expression are not related to CHG methylation levels. Relationship between fold change in expression relative to wild type and CHG methylation levels for genes identified as gaining a minimum of 5% CHG methylation in each respective line. Lines analyzed include those with both RNA sequencing and whole genome bisulfite sequencing data: (A) AtCMT3-L1<sup>T3</sup>; (B) AtCMT3-L1<sup>T4</sup>; (C) AtCMT3-L1<sup>T5</sup>; (D) AtCMT3-L2<sup>T3</sup>; (E) AtCMT3-L2<sup>T5</sup>; (F) AtCMT3-L2<sup>T6</sup>; (G) AtCMT3-L3<sup>T2</sup>; (H) AtCMT3-L3<sup>T2b</sup>. Genes with zero FPKM values were removed from the analysis.

DOI: https://doi.org/10.7554/eLife.47891.015
Figure 5. CHG methylation over gene bodies is associated with gains in non-CHG methylation. (A) Heatmap of % methylation levels across CHG DMRs overlapping the CHG-gain genes in AtCMT3-L2\(^{14}\) divided into selected trinucleotide contexts. See Figure 5—figure supplement 1A for further parsing of the data into all 16 possible trinucleotide contexts. (B) Assessment of the relationship of genic cytosine methylation to AtCMT3 expression across AtCMT3-L2 generations. Line plots show the number of methylated cytosines in each context relative to AtCMT3-L2\(^{14}\) across CHG DMRs overlapping CHG-gain genes identified in AtCMT3-L2\(^{14}\). Bar plots show the expression of the AtCMT3 transgene. For further parsing of the CHG and CHH contexts into CWG vs. CCG and CWA vs. other CHH, see Figure 5—figure supplement 1B–C. (C) Assessment of the relationship of repeat and intergenic cytosine methylation to AtCMT3 expression across AtCMT3-L2 generations. Line plots show the number of methylated cytosines relative to AtCMT3-L2\(^{14}\) across hyper CHG DMRs identified in AtCMT3-L2\(^{14}\) that overlap repeats or intergenic regions. Bar plots show the expression of the AtCMT3 transgene.

DOI: https://doi.org/10.7554/eLife.47891.016
Figure 5—figure supplement 1. CHG methylation over gene bodies is associated with gains in non-CHG methylation. (A) Heatmap of % methylation levels across CHG hyper-DMRs overlapping the CHG-gain genes in AtCMT3-L2$^{T4}$ divided into all 16 possible trinucleotide contexts. (B–C) Assessment...
of the relationship of genic cytosine methylation to AtCMT3 expression across AtCMT3-L2 generations. Line plots show the number of methylated cytosines in (B) CWG vs. CCG contexts and (C) CWA vs. other CHH contexts relative to AtCMT3-L2\textsuperscript{T4} across CHG DMRs overlapping CHG-gain genes defined in AtCMT3-L2\textsuperscript{T4}. Bar plots show the expression of the AtCMT3 transgene. See also Figure 5A–B. (D) Assessment of the relationship of genic cytosine methylation to AtCMT3 expression across AtCMT3-L1 generations. Line plots show the number of methylated cytosines in each context relative to AtCMT3-L1\textsuperscript{T4} across CHG DMRs overlapping CHG-gain genes identified in AtCMT3-L1\textsuperscript{T4}. Bar plots show the expression of the AtCMT3 transgene. (E–F) Assessment of the number of methylated cytosines relative to AtCMT3-L1\textsuperscript{T5} across hyper CHG DMRs identified in AtCMT3-L1\textsuperscript{T5} that overlap CHG-gain genes identified in AtCMT3-L1\textsuperscript{T5} (E) or repeats or intergenic regions (F). AtCMT3-L1\textsuperscript{T5} was crossed to wild type (non-transgenic) and three F2 progeny were assessed: one progeny that retained the transgene (L1T5XWT F2 (+CMT3)) and two where the transgene segregated out (L1T5XWT F2 (-CMT3) #1 and #2). The gel image below the line plot is the result of PCR conducted on genomic DNA from each line with primers to detect the AtCMT3 transgene.

DOI: https://doi.org/10.7554/eLife.47891.017
Figure 5—figure supplement 2. AtCMT3-induced genic CG methylation is maintained at higher levels than background following loss of AtCMT3 expression. (A) Comparison of the % methylation of CG sites that were not found to be methylated in the non-transgenic wild type (Shandong ecotype) parent of AtCMT3 transgenic lines. Shown is the % CG methylation calculated in each lineage across hyper CHG DMRs identified in AtCMT3-L2\textsuperscript{T4} that overlap CHG-gain genes identified in AtCMT3-L2\textsuperscript{T4} (black bars; same regions assessed in Figure 5B) compared to the average % CG methylation of an equal amount of sequence space extracted from five randomly chosen sets of unmethylated genes that did not gain CHG methylation in AtCMT3-L2\textsuperscript{T4}. The number of genes chosen in each set was equal to the number of AtCMT3-L2\textsuperscript{T4} CHG gain genes. Note that AtCMT3-L2\textsuperscript{T5} and T6 exhibited silencing of the AtCMT3 transgene, yet still maintained CG methylation levels higher than that detected on unmethylated genes. Methylation over the same regions was also assessed in additional, non-transgenic *E. salsugineum* accession (Yukon) to demonstrate that the levels of CG methylation over CHG gain genes in transgenic lines were unlikely to have occurred independently of AtCMT3. (B) Comparison of the % methylation of CG sites that were not found to be methylated in the non-transgenic wild type (Shandong ecotype) parent of AtCMT3 transgenic lines. Shown is the % CG methylation calculated in each lineage across hyper CHG DMRs identified in AtCMT3-L1\textsuperscript{T5} that overlap CHG-gain genes identified in AtCMT3-L1\textsuperscript{T5} (black bars; same regions assessed in Figure 5—figure supplement 1E) compared to the average % CG methylation of an equal amount of sequence space extracted from five randomly chosen sets of unmethylated genes that did not gain CHG methylation in AtCMT3-L1\textsuperscript{T5}. The number of genes chosen in each set was equal to the number of AtCMT3-L1\textsuperscript{T5} CHG gain genes. AtCMT3-L1\textsuperscript{T5} was crossed to wild type (non-transgenic) and three F2 progeny were assessed: one progeny that retained the transgene (L1T5XWT F2 (+CMT3)) and two where the transgene segregated out (L1T5XWT F2 (-CMT3) #1 and #2). Note that the F2 progeny where the transgene segregated out still maintained CG methylation levels higher than that detected on unmethylated genes. As in (A), methylation over the same regions was also assessed in additional, non-transgenic *E. salsugineum* accession (Yukon) to demonstrate that the levels of CG methylation over CHG gain genes in transgenic lines were unlikely to have occurred independently of AtCMT3. Error bars are ± one standard deviation of the mean.

DOI: https://doi.org/10.7554/eLife.47891.018
Figure 6. Hypothetical model for CMT3 establishment of gbM. The activity of CMT3 and histone methyltransferases (HMTs) maintains CWG methylation and H3K9me2, respectively, and is most readily detected at silenced loci. At silenced loci, methylation by CMT3 and HMTs is reinforced by other methyltransferase pathways, including MET1 (CG methylation), CMT2 (CWA methylation), and DRM2 (methylation in all contexts), which maintain constitutive heterochromatin. In contrast to silenced loci, CMT3 and HMTs can only transiently establish de novo CWG methylation and H3K9me at transcribed loci characterized by gbM. This process may be initiated by incorporation of H3K9me1 nucleosomes, which are bound exclusively by CMT3 and not CMT2 and normally removed by the H3K9 de-methylase IBM1 in a transcription coupled mechanism. However, on rare occasions CMT3 may bind H3K9me1 located in genes and de novo methylate CWG cytosines. De novo methylation of CMT3 is not a favored activity of the enzyme, so this happens only very rarely or when CMT3 is expressed at high levels. This temporally stabilizes H3K9me due to the self-reinforcing feedback loop between histone and DNA methyltransferases. Transient stabilization of H3K9me promotes H3K9me2 that can then recruit additional methyltransferases including MET1 and CMT2 to methylate CG and CWA cytosines, respectively. Heterochromatin formation is inhibited, however, through eventual removal of H3K9me by IBM1. Loss of H3K9me and/or loss of available CMT3 results in the failure of maintenance of DNA methylation in all contexts except CG following DNA replication. CG methylation is maintained due to the preferential targeting of the CG maintenance methyltransferase, MET1, to hemi-methylated cytosines following replication.

DOI: https://doi.org/10.7554/eLife.47891.019