204 **Loss of gene body methylation in *Eutrema salsugineum* is associated**
205 **with reduced gene expression**

206 Aline Muyle[1] and Brandon S. Gaut[1]

207 [1] *Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, CA 92697-2525*

208

209 **Supplementary Materials**

210

211 **Inference of gene methylation state:** For each gene, the methylation state was inferred
212 using exon methylation data only. An average exon methylation level was computed for
213 each site (CG, CHG, CHH) across all species and samples. This methylation level varies
214 across species, but the average across the five species is equivalent to taking the
215 methylation level in *A. thaliana* (Supplementary Table S2). The average methylation
216 level was used for binomial tests that determine the status (gbM or UM) of individual *A.*
217 *thaliana* and *E. salsugineum* genes (Takuno and Gaut 2012). P-values were corrected for
218 multiple tests using Benjamini and Hochberg (1995) correction. A gene was inferred to
219 be CHH (or CHG) methylated if it had more than 20 CHH (or CHG) sites and if CHH
220 methylation was higher than expected by chance (one-sided $p$-value $< 0.05$). A gene was
221 inferred to be gbM if it had more than 20 CG sites, if CG methylation was higher than
222 expected by chance ($p < 0.05$), and if CHG and CHH methylation were not significantly
223 higher than expected by chance ($p > 0.05$). A gene was inferred to be UM if it had more
224 than 20 CG sites, if CG methylation was lower than expected by chance ($p < 0.05$), and if
225 CHG and CHH methylation were not significantly higher than expected by chance ($p >$
226 $0.05$). We also ignored 1038 genes that were classified as promoter CG methylated – i.e.,
227 they had more CG methylation than expected by chance in the 200bp upstream from the
228 transcription start site ($p<0.05$ using the overall promoter methylation level across all
229 species and samples as the expectation).

| Species | Methylation data | RNA-seq data | Reference genome |
|---|---|---|---|
| *Arabidopsis thaliana* | GSM1942117 (Col-0 leaf), GSM1942127 (Col-0 leaf) | GSM1942140, GSM1942141, GSM1942142 (three Col-0 leaf replicates ) and GSM1942134, GSM1942135, GSM1942136 (three Col-0 aerial replicates) | TAIR10 |
| *Arabidopsis lyrata* | GSM1942118 (shoot) | GSM1942125 (shoot) | v1.0 |
| *Brassica rapa* | GSM1942115 (leaf) | GSM2096996 (leaf) | v1.3 |
| *Capsella rubella* | GSM1942126 (shoot) | GSM1942119 (shoot) | v1.0 |
| *Eutrema salsugineum* | GSM1942114 (shandong leaf), GSM1942120 (yukon leaf) | GSM1942123 (shandong leaf), GSM1942124 (yukon leaf) | v1.0 |

230 **Supplementary Table S1:** studied datasets for methylation, expression and genome
231 reference with GEO accession numbers.

| Species | CG | CHG | CHH |
|---|---|---|---|
| *Arabidopsis lyrata* | 0.34 | 0.054 | 0.026 |
| *Arabidopsis thaliana* | 0.21 | 0.012 | 0.0039 |
| *Brassica rapa* | 0.23 | 0.16 | 0.052 |
| *Capsella rubella* | 0.28 | 0.027 | 0.011 |
| *Eutrema salsugineum* | 0.040 | 0.018 | 0.011 |
| **Average across all samples** | 0.19 | 0.020 | 0.0087 |

232 **Supplementary Table S2:** average level of exon methylation in the three contexts for the
233 five species of the dataset.