1 Comparative analysis of the effect of genomic isolators flanking transgenes to
2 avoid positional effects in Arabidopsis

3

4 Ana Pérez-González and Elena Caro*

5

6

7

8 Centre for Plant Biotechnology and Genomics

9 Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y

10 Tecnología Agraria y Alimentaria (INIA)

11 Campus Montegancedo UPM

12 Pozuelo de Alarcón (Madrid), Spain

13

14

15 ana.perez@upm.es

16

17 *Corresponding author

18 elena.caro@upm.es

19 [0034 3364568](https://orcid.org/0034-3364-568)

20

21

22 Submission: August 3rd 2018

23 Number of figures: 5 (4 main figures + 1 supplementary figure)

24 Number of tables: 2

25 Word count: 4011

26

27

28

29

30

31 Title

32 Comparative analysis of the effect of genomic isolators flanking transgenes to
33 avoid positional effects in Arabidopsis

35 Running title

36 Genomic insulators and transgene expression in Arabidopsis

38 Highlight

39 We have studied the effect of different insulator sequences over transgene
40 expression levels and variability, and over transgene integration, using NGS.
41 Our results compare the benefits obtained by their use.

43 Abstract

44 For more than 20 years, plant biologists have tried to achieve complete control
45 of transgene expression, but until gene targeting techniques become routine,
46 flanking transgenes with genetic insulators can help avoid positional effects.
47 Insulators are DNA sequences with barrier activity that protect transgenes from
48 interferences with the host genome. We have, for the first time, compared the
49 effect of three insulator sequences previously described in the literature and of
50 a matrix attachment region from Arabidopsis never tested before. Our results
51 indicate that the use of all sequences increases transgene expression, but only
52 the last one reduces variability between lines and between individuals to a
53 minimum. We have analyzed the integration of insulator-flanked T-DNAs using
54 whole genome re-sequencing (to our knowledge, also the first time) and found
55 chiMAR lines with insertions located within heterochromatic regions of the
56 genome, characterized by DNA methylation that did not spread into the T-DNA,
57 suggesting that chiMAR can shelter transgene insertions from neighboring
58 repressive epigenetic states. Finally, we could also observe a loss of accuracy
59 of the RB insertion in the lines harboring insulators, evidenced by a high
60 frequency of truncation of T-DNAs and of insertion of vector backbone that,
61 however, did not affect transgene expression.

62

63 **Keywords**

64 Genetic insulators, MARs, enhancer-blocking, transgene expression, construct
65 design, Arabidopsis, silencing.

66

67 **Abbreviations**

68 MAR: matrix attachment region

69 pNOS: nopaline synthase Agrobacterium gene promoter

70 TBS: transformation boost sequence

71 pMAS: mannopine synthase Agrobacterium gene promoter

72 S/MAR: scaffold/matrix attachment region

73 WGR: whole genome re-sequencing

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97 **Introduction**

98 Due to the random nature of transgene insertion in the majority of higher
 99 eukaryotes, transgenic DNA may integrate into regions of the genome that are
 100 transcriptionally repressed (heterochromatin), which can result in many cases in
 101 transgene silencing. Additionally, transgenes may be incorporated near
 102 endogenous regulatory elements, such as transcriptional enhancers or
 103 repressors, which can cause their miss-expression (reviewed by Pérez-
 104 González & Caro 2016).

105 Chromatin insulator sequences, or boundary elements, are DNA sequences
 106 with the capacity to define a chromatin domain because of two key activities,
 107 the first is the ability to interfere with enhancer-promoter communication when
 108 placed between the two (enhancer blocking activity) and the second one is the
 109 ability to protect a flanked transgene from position-dependent silencing (barrier
 110 activity) (Matzat and Lei, 2014).

111 These barrier elements have been characterized extensively in animals. In
 112 plants, possibly the best studied elements with potential applications are
 113 scaffold or matrix attachment regions (S/MARs), which have been suggested to
 114 trigger the formation of chromatin loops, and thus delimit the boundaries of
 115 discrete chromosomal domains (Butaye et al., 2004). Much of the research
 116 carried out concerning the use of transgene-flanking MARs as genetic
 117 insulators has shown that the use of these elements results in an increase in
 118 the level of transgene expression and/or a reduction in plant-to-plant variability
 119 (Butaye et al., 2005). However, in some cases, reports of success using this
 120 technique have been followed by negative results (De Bolle *et al.*, 2003, De
 121 Bolle *et al.*, 2007).

122 One of the most studied MARs is the one localized upstream the chicken
 123 lysozyme gene (chiMAR) (Loc and Strätling, 1988). Its role as insulator was
 124 shown in studies with animal cell lines where its presence near a reporter gene
 125 produced an increase in transgene expression and a decrease in variability
 126 among different lines (Stief et al., 1989). The use of the chiMAR in plant
 127 constructs has been somehow controversial, leading to reports with different

128 conclusions. Mlynarova *et al.*, 1994 showed that the chicken sequence was
 129 able to bind to the tobacco nuclear matrix and that when it flanked a T-DNA
 130 containing a GUS reporter gene, the variability of its expression decreased in
 131 full-grown primary transformants of tobacco. The same group later found that a
 132 significant reduction in variation of gene expression was conferred upon the
 133 GUS gene driven by the double cauliflower mosaic virus 35s promoter, but not
 134 to the NPTII gene, driven by the nopaline synthase (pNOS) promoter
 135 (Mlynarova *et al.*, 1995). These results could, however, not be replicated in
 136 *Arabidopsis thaliana* first generation plants, where the chiMAR was found to
 137 have no influence on the level or variability of expression of transgenes driven
 138 by the 35S promoter (De Bolle *et al.*, 2003). In fact, later studies applying
 139 different transformation methods and plant species reported no boost effect on
 140 transgene expression of wild type plants (De Bolle *et al.*, 2007), but an increase
 141 in silencing mutant backgrounds (Butaye *et al.*, 2004).

142 Allen *et al.*, 1996 showed that stably transformed cell lines in which a GUS
 143 reporter gene was flanked by the tobacco MAR isolated from a genomic clone
 144 containing a root specific gene (Rb7) (Hall *et al.*, 1991) produced more than 140
 145 times more GUS enzyme activity than control transformants without it. However,
 146 the use of Rb7 did not reduce variation between different transformants.

147 The effect of the Rb7 MAR increasing transgene expression was also reported
 148 by Mankin *et al.*, 2003, that analyzed in depth the specificity of the results
 149 depending on the promoter used. They reported that highly active promoters
 150 exhibited significant increases in GUS activity in constructs flanked by Rb7
 151 compared to controls, but its presence did not significantly increase GUS
 152 activity when driven by weak promoters. Importantly, most transgenes flanked
 153 by the insulator showed a large reduction in the number of low expressing GUS
 154 transformants, suggesting that MARs can reduce the frequency of gene
 155 silencing.

156 Following that line, Abranches *et al.*, 2005 tested the effects of Rb7 in
 157 conjunction with regulated transcription using a doxycycline-inducible luciferase
 158 transgene. The Rb7 lines showed higher reporter gene expression levels and

159 avoided silencing apparition in the absence of active transcription from
160 condensed chromatin spreading.

161 Another well characterized genetic insulator, defined initially by its ability to
162 block interactions between enhancers and promoters when positioned between
163 them, is the petunia transformation boost sequence (TBS) (Hily et al., 2009).
164 This sequence has been shown to function in Arabidopsis and tobacco, and a
165 detailed analysis of the motifs it contains showed that several specific regions
166 are required for maximum enhancer-blocking function (Singer et al., 2011).

167 It was only a few years ago that another work showed that the TBS could
168 similarly function in synthetic constructs sheltering transgenes promoters from
169 the host plant genome regulatory elements. The TBS sequence was found to
170 produce enhanced transgene expression, but did not prevent gene silencing in
171 transformants with multiple and rearranged gene copies (Dietz-Pfeilstetter et al.,
172 2016).

173 Almost 25 years after the description of some of these DNA sequences, their
174 use is still not common practice in plant engineering projects due to their big
175 size that makes troublesome cloning them through traditional methods, and
176 because the reports on their effect are scattered over different organisms and
177 transformation methods with no comparisons to allow for comparison between
178 them.

179

180 Targeting transgenes to a specific integration site in the plant genome might
181 rule out chromosomal position effects, but until there are routine efficient
182 techniques for plant directed gene targeting, another alternative method needs
183 to be developed.

184

185 With the advent of modular cloning techniques that allow rapid and straight
186 forward generation of multigene constructs, the incorporation of genetic
187 insulators to the flanks of T-DNAs is no longer a problem. Therefore, we
188 decided to perform a systematic and parallel study comparing the activity and
189 effectivity of incorporating different boundary elements flanking transgenes as a
190 strategy in T-DNA design to maximize and stabilize transgene expression. We

191 have, moreover, used whole genome re-sequencing for the molecular
192 characterization of the insertion of insulator-flanked T-DNAs, finding interesting
193 results that point to previously unknown functions of the barrier sequences.

194 **Material and Methods**

195

196 *Modular cloning*

197 Modular pieces AtS/MAR10 and Rb7 were amplified by PCR using Phusion
198 High-Fidelity DNA Polymerase (NEB) from *A. thaliana* and *N. tabacum* genomic
199 DNA using primers 359/348 and 269/270, respectively; chiMARs and TBS were
200 amplified using KAPA2G Fast HotStart DNA Polymerase (Sigma) from chicken
201 liver tissue and *P. hybrida* genomic DNA using primers 1724/1725/1726/1727
202 and 275/276/277/278, respectively.

203

1724	GCGCCGTCTCGCTCGGGAGGCTCAGAAAACGGCAGTTGG
1725	GCGCCGTCTCGACCGCTCTAGGAAATTTAAGG
1726	GCGCCGTCTCGCGGTGCTCAGTAAGGCGGGT
1727	GCGCCGTCTCGCTCGAGCGCACACCAGAGCCTACACCTG
275	GCGCCGTCTCGCTCGGGAGTTCCTAACACCTGGAGAACC
276	GCGCCGTCTCGGCGACCAAAGTGTGCAGGCT
277	GCGCCGTCTCGTCGCCCCCTTGGCTGTGAAAA
278	GCGCCGTCTCGCTCGAGCGAAGTTGTAATGAGTTGCTGGC
359	GCGCCGTCTCGCTCGGGAGTGGCTATTGTTGTTATCATCA
348	GCGCCGTCTCGCTCAAGCGGGGTTTAGCCATTAACATCGT
269	GCGCCGTCTCGCTCGGGAGTCGATTAAAAATCCCAATTATATTT GG
270	GCGCCGTCTCGCTCGAGCGACTATTTTCAGAAGAAGTTCCCAA

204

205 Modular pieces were cloned into pFranki (chiMARs and TBS) or into
206 GoldenBraid pUPD2 (Rb7 and AtS/MAR10) vectors, as described in (Sarrion-
207 Perdigones et al., 2011). pFranki is a home-made vector adapted to clone
208 pieces originally designed for GB2.0 so they can be compatible with GB3.0 and
209 MoClo cloning systems. pFranki vector is composed by the cloning cassette of
210 the GoldenBraid pUPD vector and the backbone of the pUPD2 vector. To

211 generate transcriptional units, MoClo Level1 destination vectors were used
 212 (pICH47732-L1P1, pICH47742-L1P2, pICH47751-L1P3, pICH47761-L1P4).
 213 Insulators modular pieces were cloned into L1P1 and L1P4 in all cases.
 214 Luciferase transcriptional unit was cloned into L1P2 vector using the following
 215 modular pieces: pICH85281 (pMAS), pICSL80001 (luciferase CDS), pICH41421
 216 (tNOS) (Engler et al., 2014). Bialaphos resistance cassette (pICSL70005) was
 217 cloned into L1P3. Level2 destination vector pAGM4673 (Weber et al., 2011)
 218 was used for multigene assembly, and a rule of 2:1 molar ratio of
 219 inserts:acceptor was applied for adding Level1 plasmids to the reaction. Level1
 220 and Level2 digestion/ligation reactions were performed in a thermocycler as
 221 follows: 20 seconds at 37°C, [3 minutes at 37°C, 4 minutes at 16°C] for 26
 222 cycles, 5 minutes at 50°C, 5 minutes at 80°C, hold 16°C (adapted from Weber
 223 et al. 2011). *E. coli* DH5α quimiocompetent cells were transformed with the
 224 ligation products from either level and grown in LB medium containing X-Gal
 225 (20µg/mL) (Duchefa) and IPTG (1mM) (Anatrace), supplemented with ampicillin
 226 (100µg/mL) (Formedium) for GB pUPD and MoClo Level1, chloramphenicol
 227 (50µg/mL) (Formedium) for GB pUPD2, or kanamycin (50µg/mL) (IBIAN
 228 Technologies) for MoClo Level2. Sequencing (Macrogen) was done previously
 229 to plant transformation for correct sequence confirmation.

230

231 *Plant transformation*

232 Level 2 transformation plasmids were introduced into *Agrobacterium*
 233 *tumefaciens* LBA4404 quimiocompetent cells and plated in LB medium
 234 supplemented with Rifampicin (25µg/mL) (Sigma-Aldrich), Streptomycin
 235 (100µg/mL) (sigma-Aldrich) and Kanamycin (50µg/ml). A single transformant
 236 colony was grown in 200mL LB medium supplemented with the same antibiotics
 237 at 28°C under constant shaking to perform Col0 plant transformation (Clough
 238 and Bent, 1998).

239

240 *Plant growth conditions and selection*

241 T1 seeds were put into soil and grown in an environment controlled room
 242 (FitoClima HP, Aralab) under 16/8 hours light/dark conditions, at 22°C and 65%

243 RH. After 10-20 days, seedlings were sprayed with Basta herbicide (200mg/L).
244 Resistant plants were grown in the same conditions for T2 seeds recovering.

245

246 Seedlings were grown in plates in MS medium (Murashige and Skoog, 1962)
247 with 1% sucrose, supplemented with 6µg/mL of DL-Phosphinothricin (Basta)
248 herbicide (DL-Phosphinothricin, Sigma-Aldrich) for selection when needed, in a
249 growth chamber under 16/8 hours light/dark conditions at 22°C.

250

251 *Luciferase reporter assay*

252 For luciferase imaging, 16 seedlings per line were sowed in plates to analyze
253 LUC activity. D-Luciferin Firefly, potassium salt (Biosynth) was dissolved in
254 sterile H₂O with 0.01% Triton X-100 to a final concentration of 0.2µM and
255 sprayed over. After 6 minutes in the dark, luciferase activity was measured in a
256 NightOWL II LB 983 (Berthold Technologies), with 3 minutes of exposition.

257

258 *Whole Genome Re-sequencing*

259 Isolation of Arabidopsis genomic DNA was performed using a DNeasy Plant
260 Mini Kit (Qiagen). Samples were sent to Novogene Co., Ltd. for library
261 construction and sequencing. There, genomic DNA of each sample was
262 randomly sheared into short fragments of about 350bp. These fragments were
263 subjected to library construction using the Illumina TruSeq Library Construction
264 Kit, strictly following manufacturer's instructions. As followed by end-repairing,
265 dA-tailing and further ligation with Illumina adapters, the requirement fragments
266 (between 300bp and 500bp) were selected by PCR and amplified. After gel
267 electrophoresis and subsequent purification, the required fragments were
268 obtained for library construction.

269

270 Quality control of the constructed libraries were performed afterwards. Qubit 2.0
271 fluorometer (Life Technologies) was used to determine the concentration of the
272 DNA libraries. After that, a dilution to 1 ng/µl was done and the Agilent 2100
273 bioanalyzer was used to assess the insert size. Finally, a quantitative real-time
274 PCR (qPCR) was performed to detect the effective concentration of each

275 library. Pair-end sequencing was performed on the Illumina platform, with the
276 read length of 150bp at each end.

277

278 *Bisulfite conversion and sequencing*

279 Genomic DNA of 12 days-old plants of line chiMARs 6.13 was extracted using a
280 DNeasy Plant Mini Kit (Qiagen). Bisulfite treatment was done using the EZ DNA
281 Methylation Gold kit (Zymo Research) following the manufacturer's instructions.
282 Amplification from converted DNA was performed with NXT Taq PCR kit
283 (EURx) using primers 642 and 635. PCR fragments were checked on an 1%
284 agarose gel for size verification. 4µl of PCR product was cloned into pGEM-T
285 Easy (Promega) and transformed into chemically competent *E. coli* DH5α cells.
286 Nine clones were selected for the analysis. Plasmid DNA of each clone was
287 sent for sequencing (GATC), and results were checked using Geneious version
288 10.2.2 software (Kearse et al., 2012). Comparison of the converted clones to
289 the original unconverted sequences was done using CyMate software (Hetzl et
290 al., 2007), to count the converted/unconverted cytosines at each site.
291 Percentatge of methylation was calculated as (number of methylated C residues
292 in each context (CG, CHG or CHH)/total number of C residues in that
293 context)*100.

294

642	AATTTCCCGGACGTAGCGTA
635	ATCCAAGCTTTCAAGCCACAC

295

296

297 **Results**

298 Since the advent of plant genetic transformation, plant biologists have tried to
299 maximize transgene expression level and minimize variability by flanking
300 transgenes with genetic insulators. There are numerous studies that describe
301 the use of a certain insulator sequence in a host organism and analyze different
302 aspects of its barrier and enhancer-blocking ability, but they are performed in
303 such diverse conditions that do not allow for comparison and their results are
304 sometimes contradictory. Our work consists on the use four different insulator
305 sequences flanking a LUC transgene with the aim of conducting a definitive

306 parallel and systematic analysis of their effect on transgene integration,
307 expression level and variance in Arabidopsis.

308 Taking advantage of the capacities of modular cloning systems, we generated
309 five identical constructs harboring the firefly luciferase transgene driven by the
310 constitutive mannopine synthase *Agrobacterium* gene promoter (pMAS) and
311 followed by the Basta resistance selection marker cassette. One of these
312 constructs was used as a control, and the other four were flanked by different
313 sequences reported in the literature to have some type of insulator activity
314 (Figure 1A). The insulator sequences used in this work were the MAR located
315 next to the tobacco root specific gene Rb7 (Rb7) (Hall et al., 1991), the chicken
316 lysozyme A MAR region (chiMAR) (Loc and Strätling, 1988), the petunia
317 transformation booster sequence (TBS) (Hily et al., 2009) and one of the
318 scaffold/matrix attachment region sequences isolated from Arabidopsis
319 chromosome 4 (AtS/MAR10) (Pascuzzi et al., 2014).

320 A time course study of the LUC expression conferred by the pMAS showed that
321 its activity was maximum in young seedlings, and decreased rapidly as plants
322 matured and formed the rosette (Figure 1B). Given these results, for the
323 following experiments, LUC activity was always measured in 12 day old
324 seedlings. Eight 3:1 segregating Arabidopsis Col0 T2 lines were randomly
325 selected and a 100% Basta resistant T3 line coming from each of them was
326 used for LUC activity imaging to assess their levels of transgene expression
327 (Figure 1C). Our results confirmed previous reports, indicating that all
328 constructs flanked by insulator elements led to plants with increased transgene
329 expression (Figure 1D).

330 Another property of insulator sequences is their ability to decrease variability
331 between transgenic lines transformed with the same construct. When the
332 transgene was flanked by Rb7, chiMAR or TBS, the increase in LUC expression
333 described above was accompanied also by a statistically significant increase in
334 the coefficient of variation between lines, which measures the extent of variation
335 in relation to the mean within a population (Figures 2A and B). Line 40.01 from
336 AtS/MAR10 behaved very differently from the rest in terms of expression
337 (Figure 1B). We confirmed it was an outlier (expression value above

338 Q3 + 1.5×InterQuartileRange) and thus, did not consider it for this analysis.
 339 When the outlier line data was removed, the presence of AtS/MAR10 flanking
 340 the transgene led to the opposite effect than the rest of insulators, a statistically
 341 significant reduction in the coefficient of variation between lines, or what is the
 342 same, a reduction in inter-line variation (Figures 2A and B).

343 To measure the level of variation between genetically identical individuals within
 344 a population, we measured the expression of 16 seedlings from each line, and
 345 analyzed the effect of insulators on inter-individual (intra-line) variation (Figure
 346 2C). For Rb7, chiMAR and TBS, the increase in expression induced was not
 347 homogeneous between individuals and, as a result, there was a greater
 348 variance in these lines compared to the control. For AtS/MAR10, there was a
 349 small variance, similar to that of the control with no insulator (CV around 25%)
 350 (Figure 2D).

351 Next, we compared LUC expression in segregating lines from the T2 generation
 352 with homozygous lines from the T3 generation, in an effort to establish if, in our
 353 system in study, LUC expression was dependent on gene dosage. Our
 354 experiments confirm an increase in expression in all T3 lines compared to T2,
 355 consistent with the establishment of homozygous populations. No differences
 356 could be observed due to the presence of insulators (Figure 2E).

357 In an effort to further characterize the insulators lines in more detail than
 358 previous works, we proceeded to perform whole genome re-sequencing (WGR)
 359 in some of the lines obtained by transformation with each construct (Figure 3A).
 360 The results allowed us to select 21 lines with a single T-DNA insertion locus.
 361 Even though all the lines showed a 3:1 Basta resistance segregation in the T2,
 362 we found three T3 lines in which there were multiple insertions in different
 363 chromosomes, suggesting that some of them were not expressing the
 364 transgenes properly. An interesting finding was that AtS/MAR10 40.01, the
 365 outlier line that showed abnormally high LUC expression, had two insertions
 366 very close to each other in chromosome 1, what could explain their behavior as
 367 a single locus in our segregation analysis and the increased transgene
 368 expression. The WGR data also allowed us to map the T-DNA insertion site of
 369 each line and to identify the deletions in the host genome associated with the

370 insertion (Figure 3B and Table 1). Surprisingly, integration was not
 371 homogeneous among all chromosomes (we found none of the mapped
 372 insertions to be located in chromosome 2), and for Rb7 lines there was a clear
 373 preference for insertion within chromosome 3 (60%, 3 out of 5 lines) and with
 374 the T-DNA in the 3'→5' direction (100%, 5 out of 5 lines), while for the rest of
 375 the lines chromosome 3 integrations and reverse T-DNA insertions only
 376 represented a 31% in each case (5 out of 16 for each) (Table 1).

377 The existence of a selection bias towards T-DNA integrations in euchromatin
 378 where the transgenes used for selection of transformants are expressed has
 379 been reported previously in the literature (Francis and Spiker, 2005). This was
 380 the case for most of the insertions we mapped (insertion sites in euchromatin,
 381 chromatin states 1 to 7 described in Sequeira-Mendes *et al.*, 2014, Figure 3C),
 382 and when we plotted LUC activity versus state of the chromatin at the T-DNA
 383 insertion site, we could observe that lines grouped high or low depending on the
 384 construct they belonged to, and not left or right depending on the chromatin
 385 state where the T-DNA integration was located (Figure 3C). However, 2 lines
 386 carrying the chiMAR insulator presented T-DNA insertions in regions of the host
 387 genome featuring “chromatin state 8”, described as an A/T rich heterochromatic
 388 region characterized by methylated DNA and chromatin modifications such as
 389 H3K9me2 and H3K27me1 (Sequeira-Mendes *et al.*, 2014).

390 We performed an analysis of the DNA methylation levels in the junction
 391 between the host genome and the T-DNA insertion for chiMAR line 6.13 and
 392 our results show that the DNA at the insertion site is indeed heavily methylated
 393 while the DNA of the T-DNA remains devoid of this chromatin modification even
 394 in the T3 generation, consistent with a boundary role of the insulator (Figure 4).

395 The data from WGR also allowed us to characterize the genomic sequence
 396 generated as a result of the T-DNA integration, and we could observe that for 8
 397 out of 17 of the lines that contained insulator sequences, we had evidence of a
 398 lack of precision in the insertion of the RB, while that was not the case for any of
 399 the 4 control lines (Table 2). 3 out of 5 of the AtS/MAR10 lines contained vector
 400 backbone DNA (from outside the T-DNA region) integrated into the plant
 401 genome, while 3 out of 5 of the Rb7 lines, one AtS/MAR10 and one TBS line

402 showed different degrees of truncation of the inserted T-DNA in the right border
403 region. There was no evidence of truncation in the LB for any of the lines
404 analyzed.

405

406 **Discussion**

407 *Effect of insulators on transgene expression level and variation between lines*

408 Most previous works have reported positive evidence of the effects of insulators
409 on transgene expression, although some works can be found in the literature
410 that report no such effect. The experiments were, however, very diverse in
411 terms of species (some experiments had been done in tobacco and others in
412 Arabidopsis) and in terms of method of transformation (some performed in
413 primary transformants after regeneration and some in floral-dipped
414 Arabidopsis).

415 It was an important motivation for this study to compare the effects of the
416 different isolators in the same conditions: organism, developmental stage and
417 transformation method. Our results do in fact support most results from
418 literature, since we detect an increase in expression for lines where LUC is
419 flanked by any of the four insulators, and previous negative results could reflect
420 a dependency of the function of insulators on the experiment conditions.

421 Noteworthy, the use of AtS/MAR10, that had never been tested before for
422 insulator activity, resulted in a moderate but very consistent increase in LUC
423 expression.

424 In our hands, neither chiMAR, Rb7 nor TBS had an effect on reducing inter-line
425 or inter-individual variation, in fact they increased them significantly. However,
426 previous studies on the effect of chiMAR had highlighted its effect on the
427 reduction of expression variability among transgenic lines (Mlynarova et al.,
428 1995, 1994). This inconsistency could derive from a few factors in which our
429 study differs basically from these other works. First, in our system we have used
430 the pMAS promoter (versus the p35S used by Mlynarova et al. 1994 and

431 Mlynarova et al. 1995) which never reaches such high levels of expression as
 432 the p35S, but that results in normally distributed expression levels in
 433 populations of transformants (De Bolle et al., 2003). It might be possible that the
 434 chiMAR works reducing the variance of strong promoters but its effect is not so
 435 apparent in promoters with an intrinsically low level of variation such as pMAS,
 436 like Mankin *et al.*, 2003 described for Rb7. Second, in our study we have
 437 analyzed expression in homozygous T3 lines, that are already established lines
 438 with low variance in comparison with the T1 transformants analyzed by
 439 Mlynarova et al. 1994 and Mlynarova et al. 1995. It is interesting to note that the
 440 levels of variability between lines in the LUC control are in the same range as
 441 the variability between genetically identical individuals (around 30%), supporting
 442 the consistency and small intrinsic variance of our experimental set up in which
 443 we analyze T3.

444 In fact, it is striking that AtS/MAR10 is able to diminish inter-line variance,
 445 proving efficient in modifying both of the parameters measured, increasing
 446 transgene expression and reducing variability between lines, what makes it the
 447 best performing of the insulators analyzed.

448

449 *Effect of insulators on T-DNA insertion*

450 Two interesting observations have been made regarding the effect of insulators
 451 on the insertion of T-DNAs. On the one hand, it is reported that T-DNA
 452 integrations recovered by selection are mostly located in “open chromatin” or
 453 euchromatin, while, without selection, integration is biased towards regions with
 454 marks of heterochromatin (Francis and Spiker, 2005). This is explained by the
 455 silencing of the selection genes when integration takes place within
 456 heterochromatin, a phenomenon that prevents transformant recovery. Our
 457 results show the ability of chiMAR to shelter T-DNAs from heterochromatin
 458 spreading and to allow for transgene expression regardless of the position
 459 effect.

460 On the other hand, the observation of an increased frequency of truncated T-
 461 DNAs in the lines containing insulators had been reported before by Li *et al.*,

2008. Our results can be interpreted in the light of a role of insulators in the protection of transgenes at the right border end of the T-DNA from deletions. This would also explain the low correlation of expression between reporter genes located within the same T-DNA observed in many previous studies, and shown to improve by the use of insulators flanking them (Mlynarova et al., 1995). The preferential insertion of vector backbone in constructs harboring AtS/MAR10 cannot be explained by this rationale, though, and further experiments will be necessary to understand it.

As a general conclusion, we can state that there are many different insulators described in the literature with very different properties. Their functions might reflect differences in their action mechanisms and their use in transgenic constructs should depend on the needs of a specific experiment.

In our experimental setup, the best performing insulators were Rb7 in terms of increase of transgene expression, and AtS/MAR10 in terms of reducing variance.

Plant biologists should invest more efforts in the development of technologies that can render transgenes with high and stable expression with rapidity and ease. The future of synthetic biology and biotechnology projects depends on our ability to stabilize transgene expression and alleviate interference with the host genome regulation. In this work we show that the use of genetic insulators can help achieve these objectives with their simple addition at the flanks of the constructs used for transformation.

Acknowledgements

Technical help from Cristina Vaca and Diana Coroian is greatly appreciated.

Author contributions

AP and EC designed the experiments and analyzed the results. AP carried out the experiments. EC wrote the manuscript.

495

496

Conflict of interest

498 The authors have no conflict of interest to report.

499

500

References

502 Abranches, R., Shultz, R.W., Thompson, W.F., Allen, G.C., 2005. Matrix

503 attachment regions and regulated transcription increase and stabilize

504 transgene expression. *Plant Biotechnol. J.* 3, 535–543.

505 <https://doi.org/10.1111/j.1467-7652.2005.00144.x>

506 Allen, G.C., Hall, G., Michalowski, S., Newman, W., Spiker, S., Weissinger,

507 A.K., Thompson, W.F., 1996. High-Level Transgene Expression in Plant

508 Cells: Effects of a Strong Scaffold Attachment Region from Tobacco.

509 *Plant Cell* 8, 899–913. <https://doi.org/10.1105/tpc.8.5.899>

510 Butaye, K.M.J., Cammue, B.P.A., Delauré, S.L., De Bolle, M.F.C., 2005.

511 Approaches to minimize variation of transgene expression in plants. *Mol.*

512 *Breed.* <https://doi.org/10.1007/s11032-005-4929-9>

513 Butaye, K.M.J., Goderis, I.J.W.M., Wouters, P.F.J., Pues, J.M.-T.G., Delauré,

514 S.L., Broekaert, W.F., Depicker, A., Cammue, B.P. a, De Bolle, M.F.C.,

515 2004. Stable high-level transgene expression in *Arabidopsis thaliana* using

516 gene silencing mutants and matrix attachment regions. *Plant J.* 39, 440–9.

517 <https://doi.org/10.1111/j.1365-313X.2004.02144.x>

518 Clough, S.J., Bent, A.F., 1998. Floral dip: a simplified method for

519 *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.*

520 16, 735–43.

521 De Bolle, M.F.C., Butaye, K.M.J., Coucke, W.J.W., Goderis, I.J.W.M., Wouters,

522 P.F.J., van Boxel, N., Broekaert, W.F., Cammue, B.P.A., 2003. Analysis of

523 the influence of promoter elements and a matrix attachment region on the

524 inter-individual variation of transgene expression in populations of

525 *Arabidopsis thaliana*. *Plant Sci.* 165, 169–179.

526 [https://doi.org/10.1016/S0168-9452\(03\)00156-0](https://doi.org/10.1016/S0168-9452(03)00156-0)

527 De Bolle, M.F.C., Butaye, K.M.J., Goderis, I.J.W.M., Wouters, P.F.J., Jacobs,

528 A., Delauré, S.L., Depicker, A., Cammue, B.P.A., 2007. The influence of

529 matrix attachment regions on transgene expression in *Arabidopsis thaliana*
530 wild type and gene silencing mutants. *Plant Mol. Biol.* 63, 533–543.
531 <https://doi.org/10.1007/s11103-006-9107-x>

532 Dietz-Pfeilstetter, A., Arndt, N., Manske, U., 2016. Effects of a petunia
533 scaffold/matrix attachment region on copy number dependency and
534 stability of transgene expression in *Nicotiana tabacum*. *Transgenic Res.* 25,
535 149–162. <https://doi.org/10.1007/s11248-015-9924-2>

536 Francis, K.E., Spiker, S., 2005. Identification of *Arabidopsis thaliana*
537 transformants without selection reveals a high occurrence of silenced T-
538 DNA integrations. *Plant J.* 41, 464–477. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-313X.2004.02312.x)
539 [313X.2004.02312.x](https://doi.org/10.1111/j.1365-313X.2004.02312.x)

540 Hall, G., Allen, G.C., Loer, D.S., Thompson, W.F., Spiker, S., 1991. Nuclear
541 scaffolds and scaffold-attachment regions in higher plants. *Proc. Natl.*
542 *Acad. Sci. U. S. A.* 88, 9320–4.

543 Hetzl, J., Foerster, A.M., Raidl, G., Scheid, O.M., 2007. CyMATE: A new tool for
544 methylation analysis of plant genomic DNA after bisulphite sequencing.
545 *Plant J.* <https://doi.org/10.1111/j.1365-313X.2007.03152.x>

546 Hily, J.M., Singer, S.D., Yang, Y., Liu, Z., 2009. A transformation booster
547 sequence (TBS) from *Petunia hybrida* functions as an enhancer-blocking
548 insulator in *Arabidopsis thaliana*. *Plant Cell Rep.* 28, 1095–1104.
549 <https://doi.org/10.1007/s00299-009-0700-8>

550 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S.,
551 Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B.,
552 Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and
553 extendable desktop software platform for the organization and analysis of
554 sequence data. *Bioinformatics.*
555 <https://doi.org/10.1093/bioinformatics/bts199>

556 Li, J., Brunner, A.M., Meilan, R., Strauss, S.H., 2008. Matrix attachment region
557 elements have small and variable effects on transgene expression and
558 stability in field-grown *Populus*. *Plant Biotechnol. J.* 6, 887–896.
559 <https://doi.org/10.1111/j.1467-7652.2008.00369.x>

560 Loc, P. V., Strätling, W.H., 1988. The matrix attachment regions of the chicken
561 lysozyme gene co-map with the boundaries of the chromatin domain.
562 *EMBO J.* 7, 655–64.

- 563 Mankin, S.L., Allen, G.C., Phelan, T., Spiker, S., Thompson, W.F., 2003.
564 Elevation of Transgene Expression Level by Flanking Matrix Attachment
565 Regions (MAR) is Promoter Dependent: A Study of the Interactions of Six
566 Promoters with the RB7 3' MAR. *Transgenic Res.* 12, 3–12.
567 <https://doi.org/10.1023/A:1022194120518>
- 568 Matzat, L.H., Lei, E.P., 2014. Surviving an identity crisis: a revised view of
569 chromatin insulators in the genomics era. *Biochim. Biophys. Acta* 1839,
570 203–14. <https://doi.org/10.1016/j.bbagr.2013.10.007>
- 571 Mlynarova, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., Nap, J.P., 1995.
572 The MAR-Mediated Reduction in Position Effect Can Be Uncoupled from
573 Copy Number-Dependent Expression in Transgenic Plants. *PLANT CELL*
574 *ONLINE* 7, 599–609. <https://doi.org/10.1105/tpc.7.5.599>
- 575 Mlynarova, L., Loonen, A., Heldens, J., Jansen, R.C., Keizer, P., Stiekema,
576 W.J., Nap, J.P., 1994. Reduced Position Effect in Mature Transgenic Plants
577 Conferred by the Chicken Lysozyme Matrix-Associated Region. *PLANT*
578 *CELL ONLINE* 6, 417–426. <https://doi.org/10.1105/tpc.6.3.417>
- 579 Murashige, T., Skoog, F., 1962. A Revised Medium for Rapid Growth and Bio
580 Assays with Tobacco Tissue Cultures. *Physiol. Plant.* 15, 473–497.
581 <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
- 582 Pascuzzi, P.E., Flores-Vergara, M.A., Lee, T.-J., Sosinski, B., Vaughn, M.W.,
583 Hanley-Bowdoin, L., Thompson, W.F., Allen, G.C., 2014. In vivo mapping
584 of arabidopsis scaffold/matrix attachment regions reveals link to
585 nucleosome-disfavoring poly(dA:dT) tracts. *Plant Cell* 26, 102–20.
586 <https://doi.org/10.1105/tpc.113.121194>
- 587 Pérez-González, A., Caro, E., 2016. Hindrances to the Efficient and Stable
588 Expression of Transgenes in Plant Synthetic Biology Approaches, in:
589 Systems Biology Application in Synthetic Biology. Springer India, New
590 Delhi, pp. 79–89. https://doi.org/10.1007/978-81-322-2809-7_7
- 591 Sequeira-Mendes, J., Araguez, I., Peiro, R., Mendez-Giraldez, R., Zhang, X.,
592 Jacobsen, S.E., Bastolla, U., Gutierrez, C., 2014. The Functional
593 Topography of the Arabidopsis Genome Is Organized in a Reduced
594 Number of Linear Motifs of Chromatin States. *Plant Cell* 26, 2351–2366.
595 <https://doi.org/10.1105/tpc.114.124578>
- 596 Singer, S.D., Hily, J.M., Cox, K.D., 2011. Analysis of the enhancer-blocking

597 function of the TBS element from *Petunia hybrida* in transgenic *Arabidopsis*
 598 *thaliana* and *Nicotiana tabacum*. *Plant Cell Rep.* 30, 2013–2025.
 599 <https://doi.org/10.1007/s00299-011-1109-8>
 600 Stief, A., Winter, D.M., Strätling, W.H., Sippel, A.E., 1989. A nuclear DNA
 601 attachment element mediates elevated and position-independent gene
 602 activity. *Nature* 341, 343–345. <https://doi.org/10.1038/341343a0>
 603 Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., 2011. A
 604 Modular Cloning System for Standardized Assembly of Multigene
 605 Constructs. *PLoS One* 6, e16765.
 606 <https://doi.org/10.1371/journal.pone.0016765>

607

608

609 **Figure legends:**

610

611 **Figure 1. Analysis of insulator effect over LUC activity**

612 A) Schematic representation of the constructs used for studying the effect
 613 genomic insulators flanking transgenes. The above scheme represents the
 614 construction used as a control (LUC) while the scheme below represents the
 615 four constructions flanked by the four different insulators. *pMAS*: mannopine
 616 synthase gene promoter; LUC: firefly luciferase; Tnos: nopaline synthase
 617 terminator; *pNos*: nopaline synthase promoter; Tocs: octopine synthase
 618 terminator. “Insulator” represents Rb7, chiMAR, TBS or AtS/MAR10.

619 B) Time course of LUC activity when expressed under the *pMAS* promoter.
 620 Lines were assayed for LUC imaging at 12, 22 and 28 days-old. Results for
 621 control line LUC 14.9 (indicated with an arrow) are shown, but similar data was
 622 obtained for the rest of the lines. d.o.s: day-old seedlings; cps: counts per
 623 second.

624 C) Box plots showing LUC activity. ** represents Student’s test significant
 625 differences ($p < 0.005$); *** represents Student’s test highly significant differences
 626 ($p < 0.001$); cps: counts per second.

627 D) LUC activity imaging of the T3 homozygous lines, eight lines per
 628 construction.

629

Figure 2. Analysis of insulator effect over inter-line, inter-individual and inter-generation variation of LUC activity.

A) Scattergrams showing LUC activity in the selected eight lines obtained after transformation with each construct. The CV of each population was calculated as $(\text{standard deviation}/\text{mean}) \times 100$.

B) Comparison of the inter-line coefficient of variation. ** represents Student's test significant differences ($p < 0.005$); *** represents Student's test highly significant differences ($p < 0.001$); CV: coefficient of variation; cps: counts per second.

C) Scattergrams showing LUC activity in 16 seedlings of the eight selected lines obtained after transformation with each construct. CV was determined for each line and calculated as $(\text{standard deviation}/\text{mean}) \times 100$. cps/cm²: counts per second/cm². The arrow in the AtS/MAR10 graph represents the outlier line.

D) Comparison of the inter-individual coefficients of variation. CV for each insulator was calculated as $(\text{standard deviation}/\text{mean}) \times 100$. A great variance was overserved for the insulated lines compared to the control except for AtS/MAR10, that showed a small variation similar to the control, in agreement with the Student's test. ***represents highly significant differences ($p < 0.001$).

E) Box plots showing LUC activity in T2 and T3 generations of the 8 selected lines obtained after transformation eighth each construct. cps: counts per second.

Figure 3. Analysis of insulator effect over T-DNA insertion

A) Scheme of the WGR pipeline

B) Representation of the T-DNA insertion sites mapped within the five Arabidopsis chromosomes.

C) Graph showing LUC activity versus chromatin state (Sequeira-Mendes et al., 2014) at T-DNA integration site. cps: counts per second.

Figure 4. Analysis of the DNA methylation levels in the junction between the host genome and the T-DNA for line chiMAR 6.13

Upper part: schematic representation of the junction site. Middle part: graphical output of the methylation analysis (CyMate software) in 12 day-old seedlings of chiMARs 6.13 line. Red circles represent CG sites, blue squares represent

664 CHG sites and green triangles represent CHH sites. Filled symbols indicate
 665 methylated cytosines while empty ones represent non methylated cytosines.
 666 Lower part: the graph shows the DNA methylation quantification of CG (red
 667 bars), CHG (blue bars) and CHH (green bars) cytosine contexts for the flanking
 668 sequence (left) and the T-DNA (right).

669

670 **Table 1. Details of the T-DNA insertions for single-copy lines based on**
 671 **WGR results.**

672

673 **Table 2. Characterization of the genomic sequences generated as a result**
 674 **of the T-DNA integrations.**

675

676 **Supplementary files 1: WGR data at the 21 mapped single insertion sites.**

A

B

C

D

A

B

C

A. thaliana
Chromosome 1

T-DNA

Line	Chromosome	Insertion site (coordinates TAIR10)	T-DNA direction	Deletion of host genome at insertion site	Chromatin State (Sequeira-Mendes et al., 2014)	Additional File
LUC 1009	1	17,979,345	3' → 5'	3bp	1	Sup.Fig.1.1
LUC 1314	3	2,976,299	5' → 3'	63bp	3	Sup.Fig.1.2
LUC 1409	5	561,006	5' → 3'	0bp	5	Sup.Fig.1.3
LUC 3414	3	22,164,245	5' → 3'	35bp	3	Sup.Fig.1.4
At S/MAR10 0101	4	317,714	5' → 3'	2bp	7	Sup.Fig.1.5
At S/MAR10 1607	1	9,473,679	5' → 3'	4256bp	6	Sup.Fig.1.6
At S/MAR10 2308	1	27,233,367	3' → 5'	12bp	1	Sup.Fig.1.7
At S/MAR10 4101	3	22,571,686	5' → 3'	15bp	6	Sup.Fig.1.8
At S/MAR10 4203	4	1,427,639	5' → 3'	26bp	3	Sup.Fig.1.9
Rb7 0502	5	25,137,027	3' → 5'	48bp	1	Sup.Fig.1.10
Rb7 0902	1	2,741,409	3' → 5'	14bp	7	Sup.Fig.1.11
Rb7 1001	3	3,646,853	3' → 5'	26bp	6	Sup.Fig.1.12
Rb7 2707	3	19,513,996	3' → 5'	91bp	2	Sup.Fig.1.13
Rb7 4202	3	6,147,763	3' → 5'	34bp	2	Sup.Fig.1.14
TBS 0403	3	1,999,288	5' → 3'	2bp	4	Sup.Fig.1.15
TBS 1109	1	10,441,945	5' → 3'	30bp	6	Sup.Fig.1.16
TBS 1503	5	16,979,834	3' → 5'	27bp	2	Sup.Fig.1.17
TBS 1806	1	30,225,399	3' → 5'	21bp	1	Sup.Fig.1.18
chiMARs 0112	3	1,192,380	5' → 3'	96bp	1	Sup.Fig.1.19
chiMARs 0613	1	21,638,034	5' → 3'	11bp	8	Sup.Fig.1.20
chiMARs 0711	4	5,558,851	3' → 5'	1bp	8	Sup.Fig.1.21

Line	Chromosome	Insertion site (coordinates TAIR10)	T-DNA direction	T-DNA 5' insertion site	T-DNA 3' insertion site
At S/MAR10 0101	4	317,714	5' → 3'	-35 (LB)	-959 (Tocs)
At S/MAR10 1607	1	9,473,679	5' → 3'	0 (LB)	+3746 (RK2 TrfA)
At S/MAR10 4101	3	22,571,686	5' → 3'	-21 (LB)	+3825 (RK2 TrfA)
At S/MAR10 4203	4	1,427,639	5' → 3'	-17 (LB)	+1788 (pUC ori)
Rb7 0902	1	2,741,409	3' → 5'	-19 (LB)	-370 (Rb7)
Rb7 2707	3	19,513,996	3' → 5'	0 (LB)	-515 (Rb7)
Rb7 4202	3	6,147,763	3' → 5'	-21 (LB)	-1052 (Rb7)
TBS 1503	5	16,979,834	3' → 5'	-5 (LB)	-2454 (Tocs)