Title: Towards genome-engineering in complex cyanobacterial communities: RNA-guided transposition in *Anabaena*

Authors:

Sergio Arévalo[1,2,3,4], Daniel Pérez Rico[1], María Dolores Abarca[1], Laura W. Dijkhuizen[1], Peter Lindblad[2], Enrique Flores[3], Sandra Nierzwicki-Bauer[4] and Henriette Schluepmann[1]

Affiliations:

[1] Biology Department, Utrecht University, Utrecht, The Netherlands.

[2] Microbial chemistry, Department of Chemistry-Ångström Laboratory, Uppsala University, Uppsala, Sweden.

[3] Instituto de Bioquímica Vegetal y Fotosíntesis, CSIC and Universidad de Sevilla, Seville, Spain.

[4] Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA.

Author for correspondence:

H. Schluepmann, h.schlupmann@uu.nl

Running Head:

RNA-guided transposition for cyanobacteria

Abstract (220 words)

In genome engineering, integration of incoming DNA has been dependent on enzymes produced by dividing cells which has been a bottle neck towards increasing DNA-insertion frequencies and accuracy. RNA-guided transposition with CRISPR-associated transposase (CAST) was shown to be highly effective and specific in *Escherichia coli*. Here we developed Golden-Gate vectors to test this approach in filamentous cyanobacteria and show that CAST is effective in *Anabaena* sp. strain PCC 7120. The comparatively large plasmids containing the CAST and the engineered transposon were successfully transferred into *Anabaena* via conjugation using either suicide or replicative plasmids. Single guide RNA that target the leading, but not the reverse complement strand were effective with the protospacer associated motif (PAM) sequence included in the single guide RNA. In four out of six cases analyzed over two distinct target loci, the insertion site was exactly 63 bases after the PAM. CAST on a replicating plasmid was toxic which could be used to cure the plasmid. In all six cases analyzed, only the transposon defined by the sequence ranging from left and right elements was inserted at the target loci, therefore, RNA-guided transposition resulted from cut and paste. No endogenous transposons were remobilized by exposure to CAST enzymes. This work is foundational for genome editing by RNA-guided transposition in filamentous cyanobacteria, whether in culture or in complex communities.

KEYWORDS: *Anabaena*; CRISPR-associated transposon (CAST); genome engineering; RNA-guided transposition; minion sequencing; *de novo* genome assembly.

53 **Introduction**

54 Cyanobacteria are of critical importance for the biogeochemical cycling of carbon and
55 nitrogen and therefore substantially influence climate and primary production on earth
56 (Sánchez-Baracaldo et al., 2022). Recent insights highlight the importance of symbioses
57 in these cycles (Zehr and Capone, 2020). Moreover, cyanobacteria are capable of
58 forming complex communities; they form symbioses with eukaryotic organisms
59 including algae, plants, fungi, protozoa and invertebrate animals such as sponges and
60 ascidians. This is in part because of their versatile secondary metabolism (Mutalipassi et
61 al., 2021). Filamentous cyanobacteria from the order Nostocales typically differentiate
62 O_2-tight heterocysts that can fix dinitrogen in symbioses with plants from all land plant
63 lineages (Rikkinen, 2017). Large sequencing datasets are growing in number confirming
64 and expanding our understanding of those associations. However, understanding of
65 biogeochemical interdependences at the molecular level within these associations has
66 been hampered by the genetic intractability of some particular cyanobacteria, whether
67 in culture or in symbiotic communities.

68 Genetic alteration in filamentous cyanobacteria has been accomplished mostly in a few
69 species from the *Nostoc/Anabaena* genus complex. DNA cargo transfer into these
70 filamentous cells was achieved by natural competence, electroporation, *E. coli*-
71 mediated conjugation and *Agrobacterium*-mediated transfer (Gutiérrez et al., 2021).
72 Novel approaches for gene transfer have included the use of cell penetrating peptides.
73 Stabilization of the DNA-cargo was further achieved by methylation of the cargo in donor
74 cells (Elhai et al., 1997), but also by engineering the DNA sequence such that the cargo
75 may replicate (using broad spectrum origins of replication) and/or be used as a substrate
76 for homologous recombination. This allowed for the integration into a target locus on
77 either a plasmid or the chromosome. In all cases, integration of the cargo DNA was
78 catalyzed by enzymes from the target cell either by homology directed repair (HDR) or
79 the predominant non-homologous end-joining repair. The efficacy of the HDR, however,
80 is notoriously low and requires cells in the S to-G phase of the cell-cycle (Jiang, 2017).

81 Increased efficiencies have been obtained recently in gram-negative bacteria by
82 catalyzing insertion of the cargo DNA using CRISPR-associated transposases. Two
83 systems, the types 1-F and V-K, have been studied in *Escherichia coli* (Klompe et al.,
84 2019; Strecker et al., 2019). The type I-F from *Vibrio cholerae* uses a multi-protein
85 effector consisting of the type-characteristic Cas3 protein complexing with CASCADE
86 (complex for antiviral defense)-proteins (Klompe et al., 2019). The type V-K uses the
87 single effector protein Cas12-K which was discovered in filamentous cyanobacteria,
88 including *Scytonema hofmannii* (Strecker et al., 2019; Saito et al., 2021). The CAST
89 increased the frequency of insertion of the donor DNA up to 80 % without selection.
90 Importantly, it furthermore afforded its guided insertion because the effector protein
91 binds small RNA (including the crRNA encoded by the spacer) that guide the
92 transposition complex to the sequence specified by the crRNA. The mechanism of

93 transposition was not thoroughly investigated but both cut- and copy-paste have been
94 reported (Strecker et al., 2019; Vo et al., 2021).
95 The CASCADE requires more proteins to be expressed than the type V-K, did not
96 transpose DNA cargo over 1 kbp in size at high efficiency, and reproducibly inserted it
97 46-55 bases after the PAM, yet in varying orientations (Klompe et al., 2019). In contrast,
98 the type V-K from *Scytonema hofmanni* inserted cargo DNA up to 10 kbp long, 60-66 bp
99 after the PAM in the orientation 5' left-end (LE) to right end (RE) 3'at high frequency.
100 The ends of the Tn7-derived type V-K cargo include the 150 bp LE and the 90 bp RE with
101 3 and 4 transposase binding sites respectively (Peters et al., 2001). The type V-K (CAST)
102 system was plagued with significant off-target insertions (Strecker et al., 2019), and may
103 be affected strongly by transcription (Rodrigo González Linares et al., 2020). Transcripts
104 likely titrated away the guide RNA in a way similar to DNA oligonucleotides or, possibly,
105 the RNApolymerase complex displaces the CAST complex (Bialk et al., 2015; Zhang et al.,
106 2020).
107 Type V-K transposition was reconstructed *in vitro*; it required TnsB known to join the 3'
108 ends of Tn7 with target DNA, TnsC an ATP-dependent transposase activator known to
109 form heptameric rings on DNA and TniQ known to recruit TnsC to the target DNA
110 (Stellwagen and Craig, 1997; Peters and Craig, 2001; Park et al., 2021; Querques et al.,
111 2021; Hoffmann et al., 2022). It also needed cas12k which, like Cas9, is related to TnpB
112 from the IS605 transposons (Faure et al., 2019). Unlike TnsA, cas12k did not have the
113 active site known to break the 5' of the Tn7 ends; the Cas12k required two small RNAs:
114 the constant transactivating RNA that formed a duplex with a part of the otherwise
115 variant crRNA sequence. To ease specifying of the CAST target, the two small RNA
116 binding to Cas12k have been expressed as a single RNA (sgRNA) fusion with a hair-pin
117 loop and type two restriction enzyme cutting sites that allow seamless cloning of the
118 variable target sequence; the sgRNA design has been validated and optimized (Strecker
119 et al., 2019).
120 Here we present the development of synthetic biology vectors with the elements from
121 type V-K CAST designed to study genome editing by way of RNA-guided and catalyzed
122 insertion in filamentous cyanobacteria. We test RNA-guided transposition into
123 expressed gene fusions in mutants of *Anabaena* sp. strain PCC 7120 (*Anabaena*), varying
124 the target loci whilst keeping the sgRNA constant; we further varied the sgRNA
125 sequence, PAM and strand. We characterize the insertions obtained in transconjugants
126 by PCR, confocal microscopy and whole genome sequencing/*de novo* assembly to
127 determine the mechanism of insertion and unravel eventual off-target effects.
128
129 **Results**
130 **Cloning (Golden Gate) elements to share for studies on RNA-guided transposition in**
131 **cyanobacteria**
132 Preliminary studies had shown that CAST elements were toxic when *E. coli* containing
133 vectors encoding them were left on plates for over one week. To be able to rapidly test

134 toxicity and efficacy of combinations of CAST elements in cyanobacteria, we chose to
135 transfer the CAST elements into vectors for Golden Gate cloning. This was also the most
136 efficient approach to eventually share all the individual elements and intermediate
137 vectors in other synthetic biology studies. It further allowed the construction of large
138 vectors systematically that included both, the expression cassettes of CAST elements
139 and the DNA cargo between the LE and RE (Figure 1).
140 We generated Golden-Gate Level 0 vectors containing the operon *tnsB, tnsC* and *tniQ,*
141 the *cas12k*, and the LE and RE after removal of BsaI and BpiI sites (domestication) by
142 PCR (Figure 1, Level 0). Aiming at restricting the expression of the CAST elements to the
143 cyanobacterial host and further decreasing it with nitrogen in the culture medium of
144 cyanobacteria, we similarly domesticated the *glnA* promoter from *Anabaena* (Valladares
145 et al., 2004) and transferred it into a level 0 vector (Figure 1, *glnA*).
146 The *glnA* promoter was only used to generate the level 1 expression cassettes of the
147 CAST transposase proteins, in the first and second position in the final level T vector
148 (Figure 1). The level 1 vector for the sgRNA expression cassette was chosen because its
149 strong promoter would express in the reverse antisense direction and it contained the
150 sgRNA scaffold with LguI restriction sites allowing for the facile insertion of annealed
151 primers with the spacer sequence specifying the sequence targeted by the sgRNA
152 (Figure 1, level 1 position 3).
153 Finally, using the Cyano-Gate level T plasmids we assembled conjugative and replicative
154 vectors as well as non-conjugative replicative vectors, and later complemented these
155 with plasmids that were conjugative but not replicative (suicide) (Figure 1, level T). Due
156 to the limited number of positions in the level T assemblies, we fused the LE and RE to
157 the expression cassettes inside the DNA cargo that expressed YFP and erythromycin
158 resistance (Figure 1, Level 1 positions 4 and 5).
159 The combinations of vectors listed in Supplemental Table S1 proved stable in *E. coli*. We
160 thus attempted transfer of the conjugative replicative vectors listed in Supplemental
161 Table S2 into *Anabaena* by triparental conjugation and subsequent antibiotic selection.
162 We concentrated on selection with erythromycin (Em) because this antibiotic is
163 compatible with the plant hosts in the cyanobacterial symbioses of our specific interest,
164 the floating ferns from the genus *Azolla*.
165
166 **RNA-guided transposition efficiently targets expressed genes irrespective of locus**
167 **position in the chromosome**
168 Reference vectors (Supplemental Table S1, pAzUT.3 and .4) containing only the cargo
169 YFP and erythromycin or spectinomycin/streptomycin resistance expression cassettes
170 yielded clones in all the *Anabaena* strains tested: the wild-type PCC7120, the derived
171 strains containing the GFP fused to the ammonium transporter protein Amt1 (*amt1::gfp*,
172 CSVT15; (Merino-Puerto et al., 2010) or the septal-protein SepJ (*sepJ::gfp*, CSAM137;
173 (Flores et al., 2007).

174 Vectors containing all of the CAST and the cargo DNA but inactive sgRNA, containing the
175 sgRNA in isolation, or containing all the CAST elements and active sgRNA targeting the
176 GFP sequence yielded transconjugants on the nitrogen-rich BG11 medium
177 (Supplemental Table S1). Expression of the CAST components, therefore, was well
178 enough controlled behind the P_{glnA} or the inherent toxicity of the CAST proteins was low
179 enough for the transconjugant plasmids to be maintained. Strain CSAM137 did not grow
180 on $BG11_0$ medium without nitrogen, suggesting that the SepJ::Gfp could not completely
181 substitute for SepJ function under these conditions, although the strain did not show
182 the characteristic filament fragmentation of *sepJ* inactivation mutants in BG11 (not
183 shown, CSAM; Flores et al., 2007). The strain survived the short periods of induction for
184 P_{glnA}-dependent CAST-expression on $BG11_0$ as described below. Nevertheless, using
185 strain CSVT15 the effect of the plasmids pAzUT3, 7 and 9 on the growth of the
186 transconjugants could be tested (Supplemental Figure S1). In BG11 medium when
187 plasmids were selected with erythromycin, all transconjugants grew equally well
188 (Supplemental Figure S1 8d). In contrast, in $BG11_0$ medium when P_{glnA} strongly drives
189 the expression of the Cas12K from pAzUT7, transconjugant growth was inhibited
190 (Supplemental Figure S1 28 d, pAzUT7). Expression of the all the CAST components in
191 transconjugants with pAzUT9 did not seem to affect growth as in the case when Cas12k
192 was expressed with the sgRNA scaffold only. Absence of sgRNA target is known to
193 promote high-frequency off-target transposition which could be toxic (Xiao et al., 2021);
194 a toxic effect of Cas12K in the absence of TnsB, TnsC and TniQ was not reported.
195 Homogenous clones cured of the donor plasmids were obtained at high frequency after
196 sonication and growth on BG11 medium under erythromycin selection when using
197 pAzUT14 in either the CSVT15 or CSAM137 strains. pAzUT14 expressed the sgRNA1
198 targeting the sense strand of the GFP 163 bp into the 563 bp protein (Supplemental
199 Figure S2). Genotyping of the clones obtained at the *amt1::gfp* locus of CSVT15 was
200 caried out by PCR amplification spanning either ends of the transposon (Figure 2A, PCR1,
201 PCR2), the region targeted by the sgRNA (PCR3) and the pAzUT14 backbone (PCR4): in
202 all three clones the expected size fragments were amplified at either ends and the total
203 size of the inserted transposons was consistent with a single 5'LE to RE3'transposon
204 insertion (Figure 2B, PCR1,2,3). The backbone of the donor vector could not be detected
205 confirming that the clones were cured of it (Figure 2B, PCR4). Genotyping the clones
206 obtained at the *sepJ::gfp* locus after curing pAzUT14 from transconjugants of CSAM137
207 gave a similar result (Figure 3A): amplicons spanning either ends (Figure 3B, PCR1, PCR2),
208 amplicon spanning the entire length of the insertion (Figure 3B, PCR3) and absence of
209 the backbone (Figure 3B, PCR4).
210 Using the vector pAzUT18 which encodes an sgRNA targeting the sense strand at the
211 wild-type locus *alr3727*, all transconjugant colonies tested by PCR, as described in
212 Supplemental Figure S3A, had traces of the cargo transposon insertion (Supplemental
213 Figure S3B) unlike the parental wild-type strain (Figure 3B, WT), which confirms efficacy
214 of the CAST on a wild-type gene.

215 Therefore, PCR-genotyping demonstrated that CAST was able to accurately guide the
216 cargo transposon insertion into the same location, specified by the same sgRNA,
217 independently of the *Anabaena* loci chosen.
218
219 **Integration of the cargo DNA inactivated the targeted gene functions**
220 Whilst confocal viewing was not ideal for screening because it was not able to distinguish
221 whether the fluorescence detected from the expression of GFP was encoded in the
222 plasmid or the DNA cargo transposon insertion in a chromosomal locus. In strains cured
223 of the plasmids however, it verified strong YFP expression: this was seen as yellow
224 fluorescence in the cytosol of all cells of the clones UU1 and UU2 (Figure 4 YFP, BF+YFP)
225 compared to the respective parents CSVT15 and CSAM137 which expressed the GFP
226 fusions. This confirmed that the single copy insertion of RNA-guided transposon cargo
227 into the chromosome loci may be used to tag *Anabaena,* which is known to be polyploid
228 (average number of chromosome 8.2, Hu et al., 2007).
229 The filaments from strain CSVT15 differed little from the wild-type *Anabaena* (Figure 4
230 *Anabaena*). In contrast, the filaments were much shorter and the cells less regularly
231 spaced in the UU1 clone (Figure 4, UU1) compared with its parental strain (CSVT15),
232 consistent with disruption of a function in the *amt1::gfp* locus under the conditions
233 tested here.
234 Similarly, clone UU2 with the cargo transposon inserted in the sepJ::gfp exhibits shorter
235 filaments, irregularly shaped cells and cells dividing in an envelope resembling that of
236 heterocysts a phenotype characteristic for loss of function SepJ protein (Flores et al.,
237 2007). We conclude therefore that the RNA-guided transposition in *Anabaena* may
238 serve to generate tagged inactivation at expressed loci specified by the sgRNA.
239
240 **The sgRNA targeting the sense strand of expressed GFP and containing the GTT PAM**
241 **was effective**
242 We next explored how the sgRNA sequences affect the specificity and efficacy of RNA-
243 guided transposition (Supplemental Figure S2). The sgRNA from pAzUT14 efficiently
244 targeted the sense-strand of the expressed *gfp* fusions and contained in its 5 prime the
245 PAM GTT 162 bases into the *gfp* sequence (Figure 2, 3). Including the PAM in the sgRNA
246 sequence for strong binding of the sgRNA to Cas12k was attempted because the sgRNA
247 locus on the donor plasmid is protected by the proximity of the LE of the transposon
248 from DNA cargo insertions and thus inactivation. Insertions of the cargo transposon
249 were observed when sgRNA was used from pAzUT10 that contained the PAM GGTT and
250 targeted the sense strand of *gfp*, 271 bp into the coding sequence of *gfp*, but these were
251 recovered at much lower frequency (Supplemental Figure S2, and data not shown). No
252 insertions were observed for pAzUT12 where the sgRNA contained the PAM GGTT and
253 targeted the antisense strand of *gfp* 563 bp into the *gfp* sequence (Supplemental Figure
254 S2, and data not shown).
255

256 **The cargo transposon inserted mostly 63 b after the PAM and led to 2-5 base**
257 **duplications at the insertion sites**
258 PCR-results suggested that, in all cases, the cargo transposon was inserted in the 5' LE
259 to RE 3'orientation, but PCR could not distinguish whether the insertions had been a
260 result of cut- or copy-paste mechanisms. For accurate sequence information on the
261 integration loci, DNA was extracted from the six independently obtained clones
262 exhibiting RNA-guided transposition (Figure 3 and 4), their respective parental strains,
263 and from the wild-type *Anabaena*. The DNA was then sequenced using Minion long-
264 reads with a minimum 50 x coverage which allowed assemblies to be resolved into a
265 single full-length chromosome with high-accuracy sequence information.
266 The sequences immediately adjacent to the cargo transposon were extracted and
267 aligned for comparison (Figure 5). In four of the six genomes analyzed, the LE inserted
268 exactly 63 bases behind the PAM. When comparing sequences at the LE with RE,
269 duplications of the bases at the vicinity of the insertion due to the resolution of
270 transposase complex were two- to four bases long, and in five out of six cases they were
271 exactly five bases (Figure 5, duplications highlighted in purple). RNA-guided cargo
272 transposon insertions were therefore very reproducible.
273
274 **RNA-guided transposition was unidirectional 5' LE to RE 3' and single copy insertion**
275 **without co-integration of donor plasmid**
276 To inspect the overall structure of the loci targeted by the cargo transposon in each
277 clone sequenced, 13 kb regions of the consensus assembly sequences spanning the
278 insertions were viewed in IGV along with aligned reads from the clone and its parent. A
279 typical result is shown for the clone UU1.4 in Figure 6. Alignment of reads obtained from
280 the parental strain CSVT15 revealed some single nucleotide polymorphisms between
281 the clone and the reference strain, yet the foremost difference was the 2,918 bp
282 insertion in the clone UU1.4 corresponding in size with the cargo transposon. Automatic
283 annotation (Figure 6, Annotation (prokka)) identified *amt1* as *amtB* based on its
284 homology to *amtB* from *E. coli*. Individual alignments using known sequences (Figure 6,
285 BLAT alignments of known sequences) identified the start of the GFP, the gap caused by
286 the insertion (line with arrows), and the remainder of *gfp*; it furthermore identified the
287 *yfp* (VECTOR_GFPLIKE) and erythromycin resistance. Additionally, downstream of the
288 cargo transposon insertion, it identified the vector sequences used to generate the
289 *amt1::gfp* fusion in the parental strain CSVT15: these vector sequences inserted through
290 a single cross-over homologous recombination event. Results obtained from the
291 analyses of the other six clones investigated revealed an identical mechanism of
292 insertion of the RNA-guided cargo transposon. From the restricted analyses carried out
293 here we conclude, therefore, that cargo transposon insertions guided by CAST were
294 unidirectional from 5'LE to RE 3' with a precise resolution of the 5'and 3'end with cut.
295

296 **No sign of off-target insertions or remobilization of endogenous mobile elements**
297 We next examined whether exposure to the CAST machinery led to the remobilization
298 of mobile genetic elements (MGE) already present within the genome of the parental
299 strains.
300 The automatic detection of indels and deletions in genomes of the transconjugants
301 compared to their respective parental strain gave the results listed in Supplemental
302 Table S3. Manual inspection of the regions corresponding to these events using evidence
303 from the aligned long-reads, however, could not verify any changes but instead assigned
304 the indels to computation artefacts. We therefore conclude that exposure to the CAST
305 machinery did not cause the remobilization of endogenous mobile genetic elements in
306 the six cases that were analyzed in this study.
307
308 **Discussion**
309 **A Golden Gate toolbox extension for genome engineering by RNA-guided**
310 **transposition in filamentous cyanobacteria**
311 The discovery of type II restriction enzymes with three- to four-base overhangs opened
312 the way to seamless cloning of many fragments simultaneously. Yet the most important
313 advance towards synthetic biology was their use in a cloning system with a controlled
314 "vocabulary" in which overhang sequences specify the position and orientation in an
315 expression cassette, or cloning assembly of multiple cassettes or non-coding elements:
316 this allowed sharing of the individual elements in plasmids between laboratories and
317 new combinations to be generated at great speed. Here we used the vocabulary from
318 the "Golden-Gate" system expanding on an existing set of vectors named CyanoGate
319 (Vasudevan et al., 2019): we added a conjugative suicide vector, all the CAST elements
320 and the P_{glnA} promoter for nitrogen-regulated expression in cyanobacteria under
321 nitrogen starvation.
322 We combined CAST elements and the cargo DNA in a single conjugative vector (Figure
323 1), but unlike in Rubin et al. (2021), we inserted the sgRNA cassette in the reverse
324 complement orientation such that RNA polymerase on the strongly expressed sgRNA
325 gene would not affect the transposase binding to the LE. In addition, we designed 34 nt-
326 long sgRNA spacers, well over the minimum required for specific targeting of RNA-
327 guided transposition *in vitro* (Xiao et al., 2021). Also, the PAM is not cleaved by the CAST
328 as in the case of Cas9 (Heler et al., 2015) and insertions near the PAM of the sgRNA
329 would be suppressed by the LE and RE borders of the donor plasmid (Strecker et al.,
330 2019). We therefore included the PAM in the sgRNA spacer to test whether the binding
331 of the sgRNA PAM sequence to the N-terminal grove of Cas12k would lead to tighter
332 binding and therefore repression of random transposition (Xiao et al., 2021). As
333 presented in the results, this configuration resulted in the specific RNA-guided
334 integration at either of the loci targeted by the sgRNA in *Anabaena* (Figure 2 and 3).
335

336 **Accurate targeting of the RNA-guided transposon cargo in *Anabaena***
337 In all six transconjugant *Anabaena* clones analyzed in depth, no off-target effects were
338 detected (Supplemental Table S3). These effects could have been off-target insertions
339 of cargo DNA, or deletions and remobilization of endogenous mobile elements inside
340 the *Anabaena* genome. Absence of off-target effects contrasted with previous reports
341 on poor target specificity in a variety of gram-negative bacteria (Strecker et al., 2019;
342 Rubin et al., 2021). Results from Xiao et al. (2021) showed that when Cas12 lacks the
343 sgRNA, RNA-transposition occurs at a high rate and in random locations, therefore, the
344 authors suggested that sgRNA binding suppresses the random transposition. In our
345 system, sgRNA was thus expressed at a sufficiently high level such that when bound to
346 the Cas12k it suppressed random transposon insertions.
347 It is recognized that, higher throughput analysis will be required to investigate off-target
348 effects of the engineered CAST system used here for *Anabaena*. We observed poor
349 growth of clones on $BG11_0$ medium, when the P_{glnA} directs optimum expression of the
350 Cas12k and an sgRNA scaffold, suggesting that the constitutive expression of these CAST
351 components may be toxic to the *Anabaena* cells (Supplemental Figure S1). We observed
352 toxicity of the CAST system in *E. coli* (results not shown). The relative growth handicap
353 imparted by the plasmids encoding the CAST in $BG11_0$ medium was exploited to cure
354 the replicative plasmid containing the CAST away from inside the clones exhibiting
355 transposon integration (Figure 2, 3).
356 We used the unmodified LE sequence which is suspected to direct homing via
357 delocalized CRISPR RNA to the tRNA-leu in *Scytonema hofmanii (sh)*: it contains a 17 bp
358 motif matching the *sh*tRNA-Leu gene. However, in our construct the CRISPR direct
359 repeat found upstream of the 17-bp motif in the *sh*CAST locus is missing (Saito et al.,
360 2021). A more thorough understanding of the LE sequence is critically needed in many
361 respects, but most importantly for the design of intron boundaries in the LE and cargo
362 DNA so as to engineer protein fusions in coding-loci targeted by the RNA guided
363 transposition. TnsB binding sites are the characteristic features of the LE and RE and
364 TnsB was found to bind the backbone of the DNA only and to have a not very strict DNA-
365 sequence preference upon binding (Kaczmarska et al., 2022).
366
367 **RNA-guided transposition was unidirectional from LE to RE, reflected a copy and paste**
368 **mechanism, and was likely influenced by RNA**
369 Unidirectional insertion of the cargo (Figure 5) is consistent with the unidirectional
370 formation of the polymeric TnsC complex with the target DNA and Cas12k serving to
371 recruit the TnsB transposase at the target site (Park et al., 2021; Querques et al., 2021).
372 Co-integration of the plasmid sequences supplied along with the transposon cargo was
373 reported to occur at frequencies ranging from some 85% (Vo et al., 2020; Vo et al., 2021)
374 to 19.4%, and 0,6% in the specific case of the Cas12k-homing endonuclease fusion (Tou
375 et al., 2022). An explanation for co-integration was also proposed: the CAST lacks the
376 TnsA required to excise the transposon and the Cas12k lacks the endonuclease activity

377 to substitute for the TnsA. In *Anabaena*, we recovered only single copies of the cargo
378 transposon and there were no plasmid sequence co-integrates (Figure 6). This may be
379 due to the presence of accessory proteins in the cyanobacteria in which the Cas12k has
380 evolved to be devoid of the nickase activity typically found in TnsA (Xiao et al 2021 and
381 ref therein). Our results are of insufficient throughput to conclude definitely.
382 The sgRNA targeting the antisense strand of the *gfp*-gene fusions at either loci was not
383 effective. This may have been due to the sense transcript RNA duplexing efficiently with
384 the sgRNA and consequently less efficient RNA/DNA hetero-duplexing at the target
385 location. For example, the gRNA depletion with DNA oligonucleotides was reported to
386 be effective in the case of Cas9 bound gRNA (Bialk et al., 2015). The lesser efficacy of the
387 sgRNA with a spacer containing GTT PAM may stem from a suboptimal fit at the T287
388 from the WED domain and the R421 from the PI domain (Xiao et al., 2021).
389
390 **RNA-guided transposition to understand genetic features in complex communities**
391 **such as symbioses**
392 RNA-guided transposition may be used to target a locus in a specific organism in a
393 mixture effectively: this has been recently demonstrated in bacterial communities of
394 various complexities (Rubin et al., 2021). The advantage of this approach does not only
395 reside in the catalysis, and therefore the efficiency with which the cargo DNA is inserted
396 into the DNA, but also in the insertion of tags at precise locations resulting in engineered
397 loci that may be selected or followed visually in complex mixtures. We used YFP to trace
398 whether the colonies were homogenous genetically after sonication and prolonged
399 selection (Figure 4). The method lends itself to study fitness of the cells with engineered
400 loci in mixed communities so as to study the role of the original versus engineered loci
401 in microbe interactions.
402 We used a replicative plasmid in this initial study to be assured of sufficient expression
403 of the CAST elements and sufficient cargo substrate for transposition. A replicative
404 plasmid is not desired for the precise engineering of loci in microbes within complex
405 communities because this may lead to a larger proportion of off-target insertions of the
406 cargo. Because curing the replicative plasmid used for the delivery of the CAST and cargo
407 is cumbersome, we tested a conjugative suicide vector (pAzUT.17) for delivery of the
408 cargo DNA in a shortened protocol: after consecutive conjugation and sonication (to
409 reduce the length of the filaments), cells were immediately transferred onto $BG11_0$
410 medium for two days, and then on BG11-medium and selection. RNA-guided cargo DNA
411 transposition events from conjugation with suicide plasmids were detected in clones
412 that had no trace of the suicide plasmid (Supplemental Figure S4).
413 We conclude that filamentous cyanobacteria, such as *Anabaena*, are amenable to
414 genome engineering using RNA-guided transposition with the Cas12k-based CAST
415 system. We next will need to test the approach in other species and using alternative
416 methods of DNA-transfer. We urgently will test the method on symbiotic species, such

417 as *Nostoc punctiforme* and *Nostoc azollae*, especially when present in complex microbial
418 consortia.
419
420
421 **Materials and Methods**
422 **Bacterial strains and growth conditions**

423 *Anabaena* sp. (also known as *Nostoc* sp.) strain PCC 7120 and mutants CSVT15 (Merino-
424 Puerto et al., 2010) and CSAM137 (Flores et al., 2007) were cultured photoautrophically
425 in BG11 or $BG11_0$ (without $NaNO_3$) medium (Rippka et al., 1974) at 30ºC, under constant
426 white light (35 $\mu E \cdot m^{-2} \cdot s^{-1}$) and shaking. For solid cultures, 1% (w/v) agar (Bacto-Agar,
427 Difco) was added. When required, media were supplemented with 5 μg/ml (solid
428 medium) or 2.5 μg/ml (liquid medium) streptomycin sulfate (Sm), spectinomycin (Sp) or
429 erythromycin (Em).

430 *Escherichia coli* DH5α (Invitrogen) was used for cloning techniques, while strains HB101
431 and ED8634, which contain plasmid pRL623 and pRL443 (Elhai et al., 1997) respectively,
432 were used for conjugation as described in (Elhai and Wolk, 1988). All *E. coli* strains were
433 grown in Luria-Bertani (LB) medium supplemented with the appropriate antibiotics,
434 incubated at 37ºC and shaking for liquid cultures.

435 **DNA tools and vectors**

436 *Scytonema hofmannii* Tn7-like transposase components including shCas12k, the operon
437 encoding TnsB, TnsC and TniQ and the optimized sgRNA scaffold were obtained from
438 pHelper_ShCAST; the left- and right-end (LE and RE) sequences of the Tn7-like
439 transposon were obtained from pDonor_ShCAST (Strecker et al., 2019b). The pJ23119
440 promoter and T7Te terminator were PCR amplified from pHelper_ShCAST together with
441 the optimized single guide (sg) RNA-scaffold sequence that contained the Igu1 sites for
442 insertion of the target-specific spacer. The P_{glnA} was amplified from the pRL3845 plasmid
443 (Valladares et al., 2004). All other promoters and terminators were obtained from the
444 CyanoGate system (Vasudevan et al., 2019). The spacer part of the sgRNA was
445 assembled by hybridization of two partially complementary synthetic oligonucleotides
446 (Integrate DNA Technologies; Figure S2). The coding sequences for the erythromycin and
447 Sm/Sp resistance genes were obtained from the CyanoGate system (Vasudevan et al.,
448 2019).

449 Vectors to assemble the plasmids to test CAST in cyanobacteria were obtained from the
450 MoClo Plant Tool kit following the pipeline described in (Engler et al., 2014), where level
451 0 plasmids contain individual components (promoters, coding regions, terminators, etc.)
452 and expression cassettes are assembled in level 1 plasmids. The final level T plasmids,
453 containing all the expression cassettes, were assembled using the replicative and
454 conjugative vector pCAT.000 or the replicative but not conjugative vector pCAT.334
455 form the CyanoGate system (Vasudevan et al., 2019). In addition, cyanobacterial
456 replication encoded by OriT in pCAT.000 was replaced with the ColE1 from pEERM3

457 (Englund et al., 2015) to obtain a T-level backbone allowing conjugation but not
458 replication (committing suicide) in the cyanobacterial host.

459 **Plasmid construct and bacterial colony screening**

460 All plasmids generated in this work were named pAzUX.Y (plasmid Azolla Utrecht),
461 where X indicates the level and Y is the specific ID (Supplemental Table S1). They will be
462 submitted to Addgene (reference number upon acceptance of the manuscript) to ease
463 sharing.

464 To obtain plasmids pAzU0.1, pAzU0.2 pAzU0.3, sequences from the ShCAST components
465 were PCR amplified to remove internal *Bsa*I and *Bpi*I and introduce flanking *Bsa*I
466 restriction sites. The coding sequences for the erythromycin and Sm/Sp resistance genes
467 were also amplified to introduce the flanking *Bsa*I sites. In all cases, Phusion High-Fidelity
468 DNA Polymerase (ThermoFisher) was used following the manufacturer instructions. All
469 other individual components were obtained using CyanoGate or MoClo level 0 plasmids
470 (Figure 1).

471 Level 1 plasmids were assembled by digestion with *Bsa*I and ligation, following the
472 MoClo cloning protocol (Weber et al., 2011). To obtain pAzU1.3.3, the annealed
473 oligonucleotide spacers were introduced in plasmid pAzU1.3 containing the sgRNA
474 scaffold by *Lgu*I digestion and ligation (Figure 1). To obtain level T plasmids, level 1
475 inserts were introduced in pCAT.000 or pCAT.334, together with an end-linker (L), by
476 *Bpi*I digestion and ligation. All restriction enzymes and T4-DNA ligase were from
477 ThermoFisher.

478 Plasmids were introduced in *E. coli* DH5α and HB101 by heat shock. Putative positive
479 colonies were selected using the appropriate antibiotics. Positive colonies were
480 confirmed by PCR with Dream Taq polymerase using specific primers. Amplified
481 fragments were purified and sequenced (Macrogen Europe).

482 **Cyanobacterial transformation and sonication of exconjugants**

483 *E. coli* strains ED8634 (containing pRL443 encoding the conjugation machinery) and
484 HB101 (containing pRL623 and the cargo plasmid) were used for tri-parental conjugation
485 as described in (Elhai and Wolk, 1988). The mixture of *E. coli* and cyanobacteria was
486 spread and cultured on filters deposited on solid BG11 + 5% LB plates for 24 hours and
487 then transferred to BG11 medium for after another 24 hours. Filters were moved to
488 BG11 (or $BG11_0$ for the rapid conjugation protocol) plates supplemented with Sm, Sp
489 and Em for 48 hours, then transferred back to BG11 medium supplemented with the
490 corresponding antibiotics. Potential positive colonies growing after two weeks were re-
491 streaked and confirmed by PCR amplification followed by sequencing, as described
492 above.
493 To generate clonal strains of exconjugants, exconjugants that had integrated the eYFP-
494 encoding sequence into the genomic DNA were grown in 25 ml of BG11 medium to an
495 optical density (OD) of 1 at 750 nm. Then, under sterile conditions, 1 ml of culture was

496 removed and placed in a sterile plastic tube for sonication. During the sonication step,
497 the filament-length was monitored until most of the filaments were broken down to 2-
498 3 cells. Finally, serial dilutions were made and 50 μl of each dilution was plated on solid
499 BG11 medium supplemented with the appropriate antibiotic, and allowed to grow
500 under the conditions described above. The genotypes of the clonal colonies thus
501 obtained were analyzed by PCR (Supplemental Table S4).

502 To test relative growth rates, cultures grown in liquid BG11 medium with the
503 corresponding antibiotic for 1 week were washed with $BG11_0$ medium, inoculated with
504 the chlorophyll equivalents indicated and incubated in the light at 30°C for 8, 12 and 28
505 days.

506 **Confocal microscopy**

507 Samples were grown on a plate of BG11 medium supplemented with the corresponding
508 antibiotic under conditions previously described. Biomass was taken with a toothpick
509 and suspended in 100 μl of sterile distilled water. For fluorescence detection, drops of
510 10 μl were placed on a new BG11 plate, cut out of the agar and covered with a cover
511 slip. Images were photographed using a Leica SP5 microscope (40X oil immersion
512 objective). The eYFP was excited using a 514 nm laser and the sf-GFP using a 488 nm
513 laser, both lasers were used at 20% power and the irradiation came from an argon ion
514 laser. The fluorescence of eYFP was visualized with a window of 515-545 nm, for sf-GFP
515 a window of 500-525 nm was used. Autofluorescence from the natural pigments of
516 cyanobacterial cells was collected using a window of 640-740 nm. ImageJ software was
517 then used to remove background as well as overlapping images (Schindelin et al., 2015).

518 **DNA extraction and sequencing**

519 Cyanobacterial genomic DNA was isolated using the GeneJET Genomic DNA Purification
520 Kit (ThermoFisher) following the manufacturer's instructions. DNA quality and
521 concentration were analyzed by UV absorption, q-bit and on gel.

522 Minion sequencing libraries were generated using 0.3-3 μg DNA with the SQK-LSK109 kit
523 following instructions by Nanopore Technologies (version
524 NBE_9065_v109_revAK_14Aug2019); in case of multiplexing the EXP-NBD104 extension
525 was combined with NEB Blunt/TA Ligase Master Mix (M0367, New England BioLabs
526 (NEB)). Briefly, the DNA was first repaired (NEBNextFFPE Repair mix (M6630) and
527 NEBNext Ultra II End repair/dA-tailing Module (E7546)), then bound on AMPure XP
528 beads (Beckman & Coulter) and cleaned twice with 75% v/v ethanol, before elution in
529 water at 50°C for 10 min. Subsequently, when barcoded, the barcode adapters were
530 ligated and the DNA cleaned once more using the AMPure XP beads and 75% v/v
531 ethanol. Finally, the DNA was ligated to the sequencing primers in the presence of the
532 tether, and washed with Short Fragment Buffer (SQK-LSK109 kit, Nanopore
533 Technologies) bound once more on the AMPure XP beads, before elution in elution
534 buffer at 50°C for 10 min. Priming and loading of the recycled MinION flowcell (R9.4.1
535 Nanopore Technologies, starting with 600 active pores) was as per the manufactuers

536 instruction using reagents from EXP-FLP002 (Nanopore Technologies). The flowcell was
537 washed between the loading of different libraries using the EXP-WSH004 (Nanopore
538 Technologies) reagents that included DNase.

539 **Genome assemblies and annotation**

540 Data acquisition of the nanopore sequencing used the Minknow program (Oxford
541 Nanopore Technologies) until 50 times genome coverage was achieved for the barcode
542 with the lowest reads. Base calling was carried out separately. Assemblies were
543 computed *de novo* using Flye (Kolmogorov et al., 2019) with default settings, then
544 visualized using Bandage (Wick et al., 2021); polishing using medaka proved not to
545 improve the assemblies; annotation of the assemblies was carried out with bakta
546 (Schwengers et al., 2021) and prokka (Seemann, 2014) for comparison. *Anabaena*
547 genomes, and alignment files of the minion reads aligned to them with minimap2 (Li,
548 2018), were visualized using Integral Genome Viewer (IGV; Thorvaldsdóttir et al., 2013).
549 The BLAT function inside the IGV was used to locate YFP, GFP, Amt1, SepJ as well as the
550 left and right end sequences of the Tn7 provided in the original plasmid pAzUT14. Large
551 structural variations between the transconjugant after RNA-guided transposition and
552 the reference genomes were programmatically detected using Sniffles, after mapping of
553 the reads using NGMLR (Sedlazeck et al., 2018). The log of the analyses is detailed at
554 https://github.com/lauralwd/anabaena_nanopore_workflow/blob/main/script.sh.
555 When comparing all of the strains with the transposon targeted into the *amt1::gfp* locus,
556 sniffles identified 14 insertions, 3 deletions and 6 recombinations (Supplemental Table
557 S2). The insertions were then inspected by extracting the FASTA files of the insertions
558 and deletions, then locating these regions in the assemblies of the transconjugants using
559 BLAT in IGV for further evaluation.
560

561 **Material and Data availability**

562 CASTGATE vectors listed in Supplemental Table S1 have been deposited with Addgene
563 (reference numbers provided upon acceptance). Minion sequencing data were
564 deposited at the European Nucleotide Archive (ENA) and are made available at the
565 accession number (provided upon acceptance of the manuscript).
566

567 **Accession numbers**

568 The DNA sequencing data from the article may be found at the ENA accession number
569 (provided upon acceptance of the manuscript).
570

571 **Acknowledgements**

572 We thank Gracia Benítez for maintaining the cyanobacterial stocks and cultures. We
573 further thank the Gordon and Betty Moore Foundation's Symbiosis in Aquatic Systems
574 Initiative (Prime Contract No. 9355) for funding and the Netherlands Science
575 Organization (NWO-ALWGS.2016.020) for funding L.W. Dijkhuizen.
576

577 **Author contributions**

578 SN-B, PL and HS conceived and oversaw the project. EF provided reference strains and
579 material. SA, DPR, MDA carried out the cloning, SA generated the transconjugants and
580 verified RNA-guided transposition by PCR and confocal microscopy. EF analyzed the
581 growth fitness of the exconjugants. HS and LWD sequenced, assembled and analyzed
582 the genomes. All authors contributed to generating and writing the manuscript and
583 agreed to its final form.

584

585

586 **References**

587 **Bialk P, Rivera-Torres N, Strouse B, Kmiec EB** (2015) Regulation of Gene Editing
588 Activity Directed by Single-Stranded Oligonucleotides and CRISPR/Cas9 Systems.
589 PLoS One **10**: e0129308
590 **Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, Wolk CP** (1997) Reduction of conjugal
591 transfer efficiency by three restriction activities of Anabaena sp. strain PCC 7120. J
592 Bacteriol **179**: 1998–2005
593 **Elhai J, Wolk CP** (1988) [83] Conjugal transfer of DNA to cyanobacteria. Methods
594 Enzymol **167**: 747–754
595 **Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JDG, Patron NJ,**
596 **Marillonnet S** (2014) A Golden Gate modular cloning toolbox for plants. ACS
597 Synth Biol **3**: 839–843
598 **Englund E, Andersen-Ranberg J, Miao R, Hamberger B, Lindberg P** (2015) Metabolic
599 Engineering of Synechocystis sp. PCC 6803 for Production of the Plant Diterpenoid
600 Manoyl Oxide. ACS Synth Biol **4**: 1270–1278
601 **Faure G, Shmakov SA, Yan WX, Cheng DR, Scott DA, Peters JE, Makarova KS, Koonin E**
602 **v.** (2019) CRISPR–Cas in mobile genetic elements: counter-defence and beyond.
603 Nature Reviews Microbiology 2019 17:8 **17**: 513–525
604 **Flores E, Pernil R, Muro-Pastor AM, Mariscal V, Maldener I, Lechno-Yossef S, Fan Q,**
605 **Wolk CP, Herrero A** (2007) Septum-localized protein required for filament
606 integrity and diazotrophy in the heterocyst-forming cyanobacterium Anabaena sp.
607 strain PCC 7120. J Bacteriol **189**: 3884–3890
608 **Gutiérrez S, Lauersen KJ, Fabris M, Abbriano R** (2021) Gene Delivery Technologies
609 with Applications in Microalgal Genetic Engineering. Biology 2021, Vol 10, Page
610 265 **10**: 265
611 **Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA** (2015)
612 Cas9 specifies functional viral targets during CRISPR–Cas adaptation. Nature 2015
613 519:7542 **519**: 199–202
614 **Hoffmann FT, Kim M, Beh LY, Wang J, Vo PLH, Gelsinger DR, George JT, Acree C,**
615 **Mohabir JT, Fernández IS, et al** (2022) Selective TnsC recruitment enhances the
616 fidelity of RNA-guided transposition. Nature 2022 1–10
617 **Hu B, Yang G, Zhao W, Zhang Y, Zhao J** (2007) MreB is important for cell shape but not
618 for chromosome segregation of the filamentous cyanobacterium Anabaena sp.
619 PCC 7120. Mol Microbiol **63**: 1640–1652
620 **Jiang F** (2017) CRISPR-Cas9 Structures and Mechanisms. Article in Annual Review of
621 Biophysics. doi: 10.1146/annurev-biophys-062215-010822
622 **Kaczmarska Z, Czarnocki-Cieciura M, Górecka-Minakowska KM, Wingo RJ, Jackiewicz**
623 **J, Zajko W, Poznański JT, Rawski M, Grant T, Peters JE, et al** (2022) Structural
624 basis of transposon end recognition explains central features of Tn7 transposition
625 systems. Mol Cell **82**: 2618-2632.e7
626 **Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH** (2019) Transposon-encoded
627 CRISPR–Cas systems direct RNA-guided DNA integration. Nature **571**: 219–225
628 **Kolmogorov M, Yuan J, Lin Y, Pevzner PA** (2019) Assembly of long, error-prone reads
629 using repeat graphs. Nature Biotechnology 2019 37:5 **37**: 540–546
630 **Li H** (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:
631 3094–3100

632 **Merino-Puerto V, Mariscal V, Mullineaux CW, Herrero A, Flores E** (2010) Fra proteins
633 influencing filament integrity, diazotrophy and localization of septal protein SepJ
634 in the heterocyst-forming cyanobacterium Anabaena sp. Mol Microbiol **75**: 1159–
635 1170
636 **Mutalipassi M, Riccio G, Mazzella V, Galasso C, Somma E, Chiarore A, De Pascale D,**
637 **Zupo V** (2021) Symbioses of cyanobacteria in marine environments: Ecological
638 insights and biotechnological perspectives. mdpi.com. doi: 10.3390/md19040227
639 **Park JU, Tsai AWL, Mehrotra E, Petassi MT, Hsieh SC, Ke A, Peters JE, Kellogg EH**
640 (2021) Structural basis for target site selection in RNA-guided DNA transposition
641 systems. Science (1979) **373**: 768–774
642 **Peters J, development NC-G&, 2001 undefined** (2001) Tn7 recognizes transposition
643 target structures associated with DNA replication using the DNA-binding protein
644 TnsE. genesdev.cshlp.org. doi: 10.1101/gad.870201
645 **Peters JE, Craig NL** (2001) Tn7: smarter than we thought. Nature Reviews Molecular
646 Cell Biology 2001 2:11 **2**: 806–814
647 **Querques I, Schmitz M, Oberli S, Chanez C, Jinek M** (2021) Target site selection and
648 remodelling by type V CRISPR-transposon systems. Nature 2021 599:7885 **599**:
649 497–502
650 **Rikkinen J** (2017) Cyanobacteria in terrestrial symbiotic systems. Modern Topics in the
651 Phototrophic Prokaryotes: Environmental and Applied Aspects 243–294
652 **Rippka R, Waterbury J, Cohen-Bazire G** (1974) A cyanobacterium which lacks
653 thylakoids. Archives of Microbiology 1974 100:1 **100**: 419–436
654 **Rodrigo González Linares B, Rita Costa Cristóbal Almendros Romero Stan J Brouns**
655 **Thesis committee AJ, J Brouns Chirlmin Joo Christophe Danelon SJ** (2020) A
656 CRISPR-associated transposase presents null cargo integration efficiency when
657 targeting a transcriptionally highly active region.
658 **Rubin BE, Diamond S, Cress BF, Crits-Christoph A, Lou YC, Borges AL, Shivram H, He C,**
659 **Xu M, Zhou Z, et al** (2021) Species- and site-specific genome editing in complex
660 bacterial communities. Nature Microbiology 2021 7:1 **7**: 34–47
661 **Saito M, Ladha A, Strecker J, Faure G, Neumann E, Altae-Tran H, Macrae RK, Zhang F**
662 (2021) Dual modes of CRISPR-associated transposon homing. Cell **184**: 2441-
663 2453.e18
664 **Sánchez-Baracaldo P, Bianchini G, Wilson JD, Knoll AH** (2022) Cyanobacteria and
665 biogeochemical cycles through Earth history. Trends in Microbiology **30**: 143–157
666 **Schindelin J, Rueden CT, Hiner MC, Eliceiri KW** (2015) The ImageJ ecosystem: An open
667 platform for biomedical image analysis. Mol Reprod Dev **82**: 518–529
668 **Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A** (2021)
669 Bakta: rapid and standardized annotation of bacterial genomes via alignment-free
670 sequence identification. Microb Genom. doi: 10.1099/MGEN.0.000685
671 **Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz**
672 **MC** (2018) Accurate detection of complex structural variations using single-
673 molecule sequencing. Nat Methods **15**: 461–468
674 **Seemann T** (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**:
675 2068–2069
676 **Stellwagen AE, Craig NL** (1997) Avoiding self: Two Tn7-encoded proteins mediate
677 target immunity in Tn7 transposition. EMBO Journal **16**: 6823–6834

678 **Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin E v., Zhang F**
679 (2019) RNA-guided DNA insertion with CRISPR-associated transposases. Science
680 (1979). doi: 10.1126/science.aax9181
681 **Thorvaldsdóttir H, Robinson JT, Mesirov JP** (2013) Integrative Genomics Viewer (IGV):
682 high-performance genomics data visualization and exploration. Brief Bioinform
683 **14**: 178–192
684 **Tou CJ, Orr B, Kleinstiver BP** (2022) Cut-and-Paste DNA Insertion with Engineered Type
685 V-K CRISPR-associated Transposases. bioRxiv 2022.01.07.475005
686 **Valladares A, Muro-Pastor AM, Herrero A, Flores E** (2004) The NtcA-dependent P1
687 promoter is utilized for glnA expression in N2-fixing heterocysts of Anabaena sp.
688 strain PCC 7120. J Bacteriol **186**: 7337–7343
689 **Vasudevan R, Gale GAR, Schiavon AA, Puzorjov A, Malin J, Gillespie MD, Vavitsas K,**
690 **Zulkower V, Wang B, Howe CJ, et al** (2019) CyanoGate: A Modular Cloning Suite
691 for Engineering Cyanobacteria Based on the Plant MoClo Syntax. Plant Physiol
692 **180**: 39–55
693 **Vo PLH, Acree C, Smith ML, Sternberg SH** (2021) Unbiased profiling of CRISPR RNA-
694 guided transposition products by long-read sequencing. Mob DNA. doi:
695 10.1186/S13100-021-00242-2
696 **Vo PLH, Ronda C, Klompe SE, Chen EE, Acree C, Wang HH, Sternberg SH** (2020) CRISPR
697 RNA-guided integrases for high-efficiency, multiplexed bacterial genome
698 engineering. Nature Biotechnology 2020 39:4 **39**: 480–489
699 **Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S** (2011) A Modular Cloning
700 System for Standardized Assembly of Multigene Constructs. PLoS One **6**: e16765
701 **Wick R, Schultz M, Zobel J, Bioinformatics KH-, 2015 undefined** Bandage: interactive
702 visualization of de novo genome assemblies. academic.oup.com
703 **Xiao R, Wang S, Han R, Li Z, Gabel C, Mukherjee IA, Chang L** (2021) Structural basis of
704 target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. Mol
705 Cell **81**: 4457-4466.e5
706 **Zehr JP, Capone DG** (2020) Changing perspectives in marine nitrogen fixation. Science.
707 doi: 10.1126/SCIENCE.AAY9514
708 **Zhang S, Guo F, Yan W, Dai Z, Dong W, Zhou J, Zhang W, Xin F, Jiang M** (2020) Recent
709 Advances of CRISPR/Cas9-Based Genetic Engineering and Transcriptional
710 Regulation in Industrial Biology. Frontiers in Bioengineering and Biotechnology **7**:
711 459
712
713
714 **Figure legends**

715 **Figure 1** Cloning strategy to obtain the modules and vectors of the CASTGATE. The Level
716 0 plasmids with names beginning with pC0 and the donor for the Level T, pCAT.000, a
717 replicative and conjugative vector were from the CyanoGate kit (Vasudevan R. at al.,
718 2019). The level 0 pICH41308 and the Level 1 pICH41800 providing the linker (L) at
719 position 6 in the T-level assemblies were from the MoClo kit (Weber et al., 2011). Firstly,
720 sequences were domesticated and transferred in the Level 0 vectors: pAzU0.1 for P_{glnA},
721 pAzU0.2 for the operon encoding TnsB, TnsC and TniQ (*tnsBCQ*), and pAzU0.3 for Cas12k
722 which originated from pHelper_ShCAST (Strecker et al., 2019). Alternatively, sequences

723 were PCR amplified for direct insertion into Level 1 vectors, as in the case of the cargo
724 transposon LE and RE which originated from pDonor_ShCAST (Strecker et al., 2019), and
725 the expression cassette of the sgRNA-scaffold from pHelper_ShCAST that allows to ligate
726 target-specific sequences in LguI restriction sites. Secondly, expression cassettes were
727 assembled in Level 1 for the specific positions (1.1 to 1.5) of the Level T assemblies:
728 pAzU1.1, low nitrogen-inducible expression of the TnsBCQ; pAzU1.2, low nitrogen-
729 inducible expression of Cas12k; pAzU1.3, the sgRNA-scaffold, which, when the GFP-
730 specific target sequence (AAAGTT-GFPgRNA-GCT) was ligated in the LguI site, yielded
731 pAz1.3.3; pAzU1.4, with at the 5' the LE, then the cassette for constitutive expression of
732 eYFP; pAzU1.5, with at the 5' the cassette for constitutive erythromycin resistance
733 followed at the 3' with the RE. Thirdly, the level 1 plasmids were combined to generate
734 the Level T pAzUT.14, encoding the CAST machinery followed by the cargo DNA flanked
735 by the LE and RE. CAST component Level 1 plasmids were replaced with linkers during
736 the final Level T plasmid assemblies to test their individual toxicity and efficacy.

737

738 **Figure 2** Scheme of the RNA-guided transposon insertion in the *gfp* of the *amt1::gfp*
739 locus of the parental strain CSVT15, and its detection by PCR. A, Scheme of the *amt1::gfp*
740 locus from strain CSVT15 that, after conjugation and selection on erythromycin, also
741 contained pAzUT.14. pAzUT.14 encoded the RE and LE flanked transposon cargo, which
742 would be mobilized by the separately encoded CAST enzymes and sgRNA targeting the
743 *gfp*. When mobilized with complete resolution of the transposase complex, the cargo
744 transposon inserts inside the *gfp* of the *amt1:gfp* fusion. B, PCR detection of cargo
745 insertion and full segregation in the different exconjugant clones. PCR 1 detects the eYFP
746 integration; PCR 2 detects erythromycin-resistance gene integration; PCR 3 detects *gfp*
747 and thus segregation of transposed genotype; PCR 4 detects whether the exconjugant
748 had been cured of pAzUT.14. The letters and numbers correspond with: (M) 1-kb DNA
749 ladder; (W) *Anabaena* sp. PCC 7120; (P) CSVT15; (B) pAzUT.14; (1) AzUU1 clone 1; (2)
750 AzUU1 clone 4; and (3) AzUU1 clone 8. *Anabaena* sp. PCC 7120 was used as negative
751 control in all PCR tests. In the PCR 3, the CSVT15 strain was used as positive control.
752

753 **Figure 3** Scheme of the RNA-guided transposon insertion in the *gfp* of the *sepJ::gfp* locus
754 of the parental strain CSAM137, and its detection a by PCR. A, Scheme of the *sepJ::gfp*
755 locus from strain CSAM137 that, after conjugation and selection on erythromycin, also
756 contained pAzUT.14. pAzUT.14 encoded the RE and LE flanked transposon cargo, which
757 would be mobilized by the CAST enzymes and sgRNA targeting the *gfp,* also encoded in
758 pAzUT.14. When mobilized with resolution of the transposase complex, the cargo
759 transposon inserts inside the *gfp* of the *sepJ:gfp* fusion. B, PCR detection of cargo
760 insertion and full segregation in the different exconjugants. PCR 1 detects the eYFP
761 integration; PCR 2 detects erythromycin-resistance gene integration; PCR 3 detects *gfp*
762 and thus segregation of transposed genotype; PCR 4 detects whether the exconjugant

763 had been cured of pAzUT.14. The letters and numbers correspond with: (M) 1-kb DNA
764 ladder; (W) *Anabaena* sp. PCC 7120; (P) CSAM137; (B) pAzUT.14; (1) AzUU2 clone 3; (2)
765 AzUU2 clone 8; and (3) AzUU2 clone 7. *Anabaena* sp. PCC 7120 was used as negative
766 control in all PCR tests. In the PCR 3, the CSAM137 strain was used as positive control.
767
768 **Figure 4** Fluorescence detection of the GFP-protein fusions and of the YFP in the strains
769 fully segregated for the insertion and cured of the pAzUT.14, compared to their
770 respective parental strain. Fluoresce microscopy under brightfield (BF) settings, or
771 fluorescence settings to detect red autofluorescence (RFP), yellow fluorescent protein
772 (YFP), and the image obtained by superimposition of BF and YFP. *Anabaena*, wild-type
773 strain. CSVT15, the parental strain with the *amt1::gfp* locus; UU1, a CSVT15 derived
774 exconjugant that was homozygous for the insertion in the *gfp* and cured of the pAzUT.14
775 as shown in Figure 2. CSAM137, the parental strain with the *sepJ::gfp* locus; UU2, a
776 CSAM137 derived exconjugant that was homozygous for the insertion in the *gfp* and
777 cured of the pAzUT.14, as shown in Figure 3. The bar corresponds to 10 mm.
778
779 **Figure 5** Sequences at the insertion sites of the cargo transposon after RNA-guided
780 transposition into the *gfp* of the *amt1::gfp* or *sepJ::gfp* loci. A, Sequences that encode
781 the GFP (green), or the LE and RE of the cargo transposon (yellow), are highlighted. In
782 addition, sequences of the *gfp* that were a part of the sgRNA encoded by pAzUT.14 are
783 highlighted that encode the PAM (blue) and the remainder sequence specific part of the
784 sgRNA (red). Resolution of the transposase complex leads to the insertion of nucleotides,
785 causing the small insertions highlighted in purple. The position (Pos) of the insertions
786 was counted starting from the first base after the PAM. (B) The insertion was exactly at
787 position 63 for 4 of the six independently recovered clones derived from either parent.
788 T in red denotes the LE start base. Similarly, resolution of the transposase led in 4 of the
789 6 six clones to a five base repeat, CCAGA in purple.
790
791 **Figure 6** The *amt1:gfp* locus in the strain UU1.4 compared to the parental strain CSVT15.
792 Strains were sequenced using MinIon with minimally 50 times coverage and their
793 genomes assembled *de novo*. The assembly was automatically annotated with prokka,
794 and known sequences were aligned using the BLAT-aligner to the *amt1::gfp* locus. In
795 addition, MinIon reads obtained from sequencing strain UU1.4 and the parental strain
796 CSV15 were aligned to the assembled genome of UU1.4. The assembly, automated and
797 manual annotation, and the alignments were then visualized in Integral Genome Viewer
798 (IGV) in the 13 kbp region spanning the 2918 bp cargo transposon.
799
800

801 **Supplemental Data**
802
803 The following supplemental data is available in the online version of this article.
804
805 **Supplemental Figure S1.** Relative toxicity of the CAST plasmids in *Anabaena*.
806
807 **Supplemental Figure S2.** The three different sgRNAs targeting *gfp*.
808
809 **Supplemental Figure S3.** Efficient targeting at locus *alr3727* in wild-type *Anabaena*.
810
811 **Supplemental Figure S4.** Rapid conjugation protocol for RNA-guided transposition using
812 the suicide plasmid pAzUT.17.
813
814 **Supplemental Table S1.** CASTGATE vectors generated in this study.
815
816 **Supplemental Table S2.** CASTGATE vectors transferred to and tested in *Anabaena* wild-
817 type, and the CSVT15 and CSAM137 strains in this study.
818
819 **Supplemental Table S3.** Insertions detected by Snifles comparing genome assemblies
820 from the parental strains with those from clones obtained after RNA-guided
821 transposition.
822
823 **Supplemental Table S4.** Primers used for PCR assays and key cloning steps in this study.
824
825
826
827
828
829
830
831
832
833
834
835 **Figures**

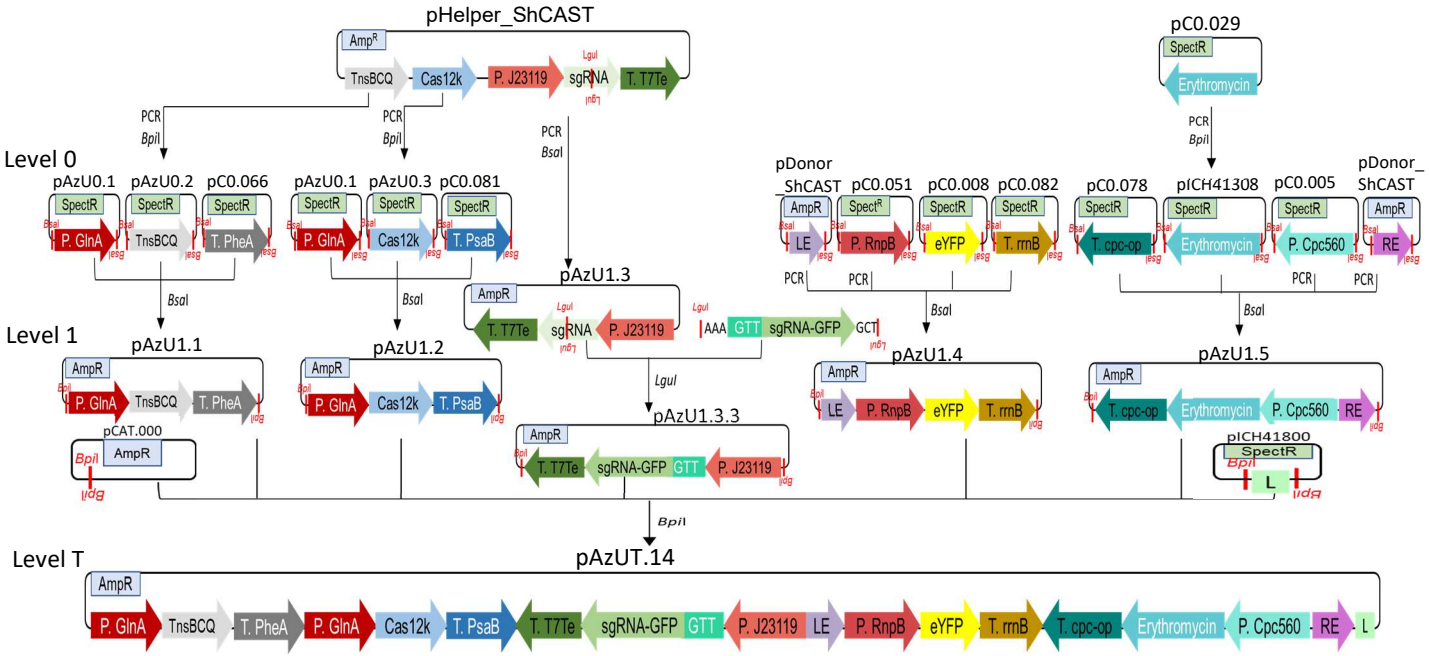

Figure 1 Cloning strategy to obtain the modules and vectors of the CASTGATE. The Level 0 plasmids with names beginning with pC0 and the donor for the Level T, pCAT.000, a replicative and conjugative vector were from the CyanoGate kit (Vasudevan R. at al., 2019). The level 0 pICH41308 and the Level 1 pICH41800 providing the linker (L) at position 6 in the T-level assemblies were from the MoClo kit (Weber et al., 2011). Firstly, sequences were domesticated and transferred in the Level 0 vectors: pAzU0.1 for P_{glnA}, pAzU0.2 for the operon encoding TnsB, TnsC and TniQ (*tnsBCQ*), and pAzU0.3 for Cas12k which originated from pHelper_ShCAST (Strecker et al., 2019). Alternatively, sequences were PCR amplified for direct insertion into Level 1 vectors, as in the case of the cargo transposon LE and RE which originated from pDonor_ShCAST (Strecker et al., 2019), and the expression cassette of the sgRNA-scaffold from pHelper_ShCAST that allows to ligate target-specific sequences in LguI restriction sites. Secondly, expression cassettes were assembled in Level 1 for the specific positions (1.1 to 1.5) of the Level T assemblies: pAzU1.1, low nitrogen-inducible expression of the *tnsBCQ*; pAzU1.2, low nitrogen-inducible expression of Cas12k; pAzU1.3, the sgRNA-scaffold, which, when the GFP-specific target sequence (AAAGTT-GFPgRNA-GCT*) was ligated in the LguI site, yielded pAz1.3.3; pAzU1.4, with at the 5' the LE, then the cassette for constitutive expression of eYFP; pAzU1.5, with at the 5' the cassette for constitutive erythromycin resistance followed at the 3' with the RE. Thirdly, the level 1 plasmids were combined to generate the Level T pAzUT.14, encoding the CAST machinery followed by the cargo DNA flanked by the LE and RE. CAST component Level 1 plasmids were replaced with linkers during the final Level T plasmid assemblies to test their individual toxicity and efficacy. * three different target sequences within the gfp sequence were tested as described in Supplemental Figure S2.

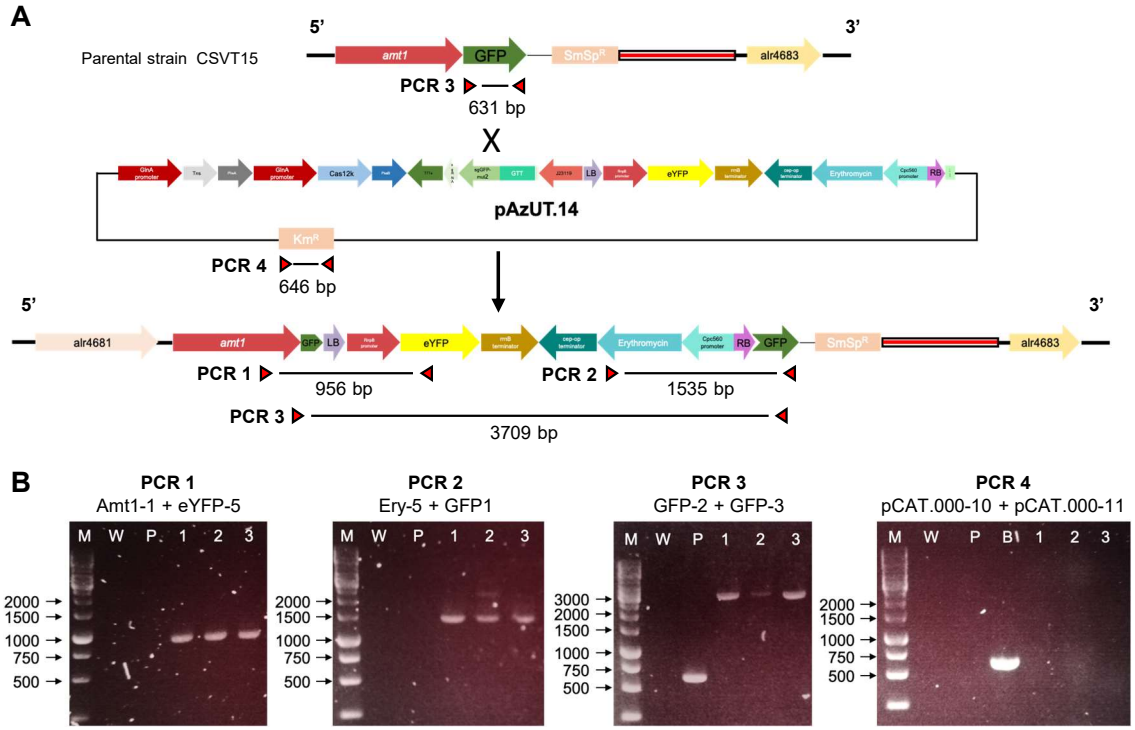

Figure 2 Scheme of the RNA-guided transposon insertion in the *gfp* of the *amt1::gfp* locus of the parental strain CSVT15, and its detection by PCR. A, Scheme of the *amt1::gfp* locus from strain CSVT15 that, after conjugation and selection on erythromycin, also contained pAzUT.14. pAzUT.14 encoded the RE and LE flanked transposon cargo, which would be mobilized by the separately encoded CAST enzymes and sgRNA targeting the *gfp*. When mobilized with complete resolution of the transposase complex, the cargo transposon inserts inside the *gfp* of the *amt1:gfp* fusion. (B) PCR detection of cargo insertion and full segregation in the different exconjugant clones. PCR 1 detects the eYFP integration; PCR 2 detects erythromycin-resistance gene integration; PCR 3 detects *gfp* and thus segregation of transposed genotype; PCR 4 detects whether the exconjugant had been cured of pAzUT.14. The letters and numbers correspond with: (M) 1-kb DNA ladder; (W) *Anabaena* sp. PCC 7120; (P) CSVT15; (B) pAzUT.14; (1) AzUU1 clone 1; (2) AzUU1 clone 4; and (3) AzUU1 clone 8. *Anabaena* sp. PCC 7120 was used as negative control in all PCR tests. In the PCR 3, the CSVT15 strain was used as positive control.

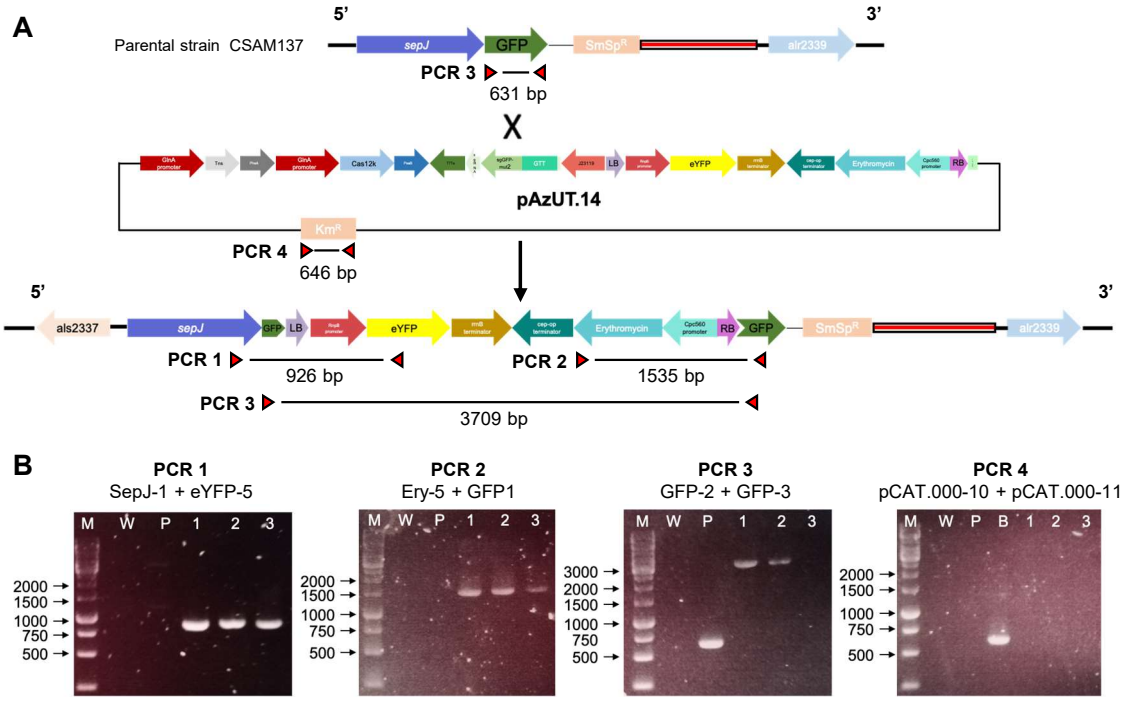

Figure 3 Scheme of the RNA-guided transposon insertion in the *gfp* of the *sepJ::gfp* locus of the parental strain CSAM137, and its detection a by PCR. A, Scheme of the *sepJ::gfp* locus from strain CSAM137 that, after conjugation and selection on erythromycin, also contained pAzUT.14. pAzUT.14 encoded the RE and LE flanked transposon cargo, which would be mobilized by the CAST enzymes and sgRNA targeting the *gfp,* also encoded in pAzUT.14. When mobilized with resolution of the transposase complex, the cargo transposon inserts inside the *gfp* of the *sepJ:gfp* fusion. B, PCR detection of cargo insertion and full segregation in the different exconjugants. PCR 1 detects the eYFP integration; PCR 2 detects erythromycin-resistance gene integration; PCR 3 detects *gfp* and thus segregation of transposed genotype; PCR 4 detects whether the exconjugant had been cured of pAzUT.14. The letters and numbers correspond with: (M) 1-kb DNA ladder; (W) *Anabaena* sp. PCC 7120; (P) CSAM137; (B) pAzUT.14; (1) AzUU2 clone 3; (2) AzUU2 clone 8; and (3) AzUU2 clone 7. *Anabaena* sp. PCC 7120 was used as negative control in all PCR tests. In the PCR 3, the CSAM137 strain was used as positive control.

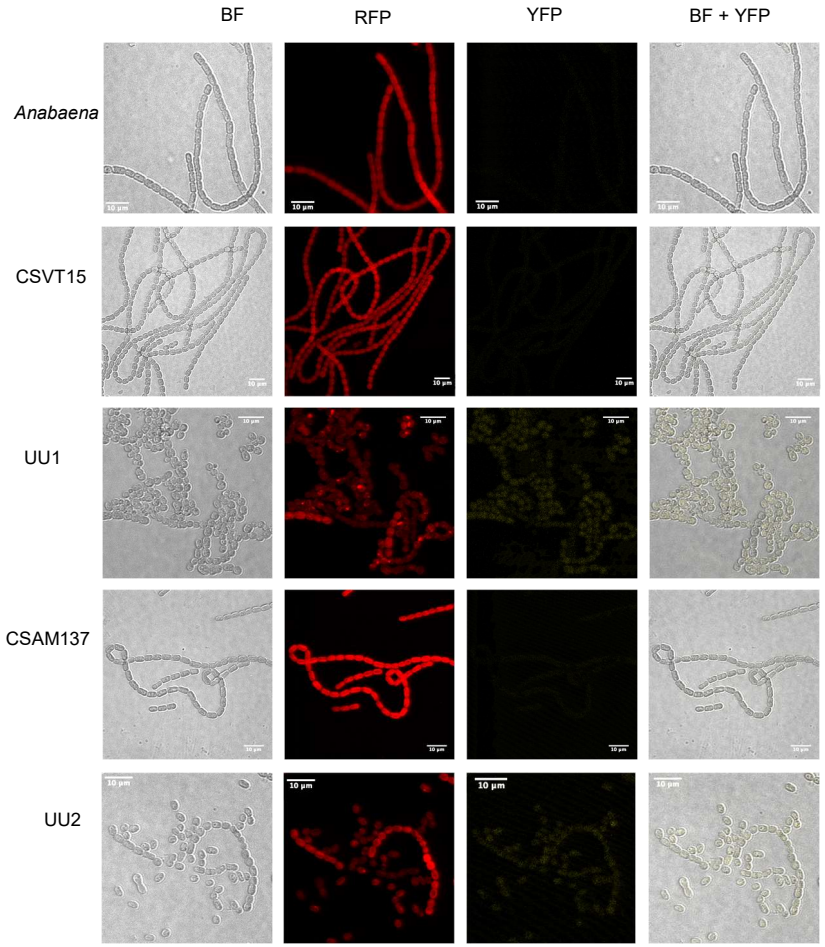

Figure 4 Fluorescence detection of the GFP-protein fusions and of the YFP in the strains fully segregated for the insertion and cured of the pAzUT.14, compared to their respective parental strain. Fluoresce microscopy under brightfield (BF) settings, or fluorescence settings to detect red autofluorescence (RFP), yellow fluorescent protein (YFP), and the image obtained by superimposition of BF and YFP. *Anabaena*, wild-type strain. CSVT15, the parental strain with the *amt1::gfp* locus; UU1, a CSVT15 derived exconjugant that was homozygous for the insertion in the *gfp* and cured of the pAzUT.14 as shown in Fig. 2. CSAM137, the parental strain with the *sepJ::gfp* locus; UU2, a CSAM137 derived exconjugant that was homozygous for the insertion in the *gfp* and cured of the pAzUT.14, as shown in Fig. 3. The bar corresponds to 10 µm.

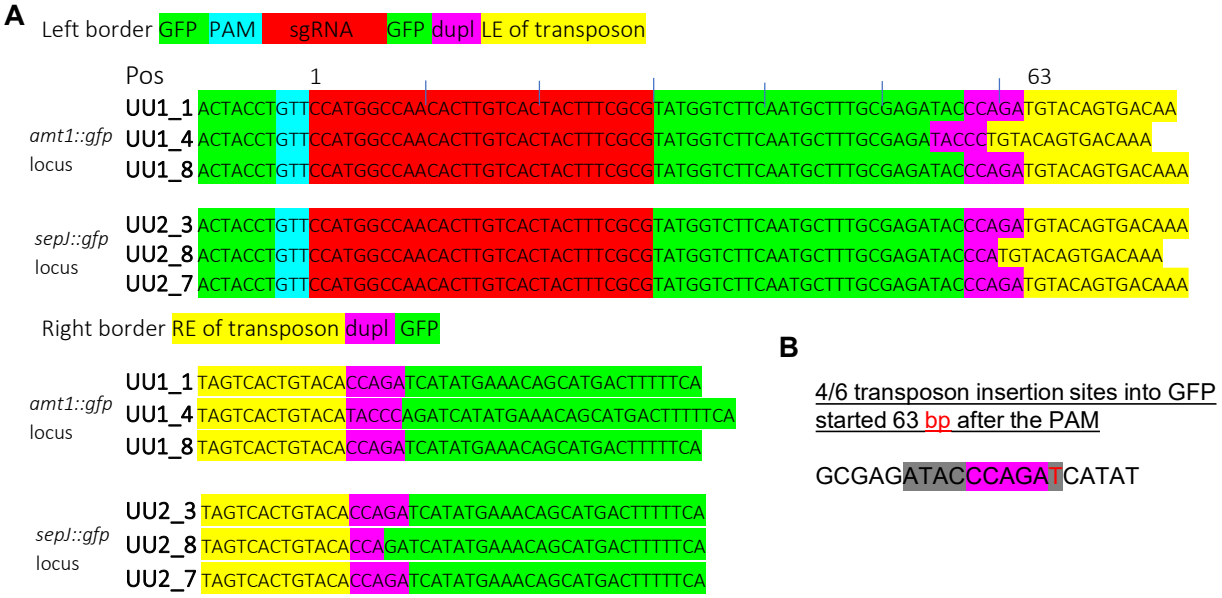

Figure 5 Sequences at the insertion sites of the cargo transposon after RNA-guided transposition into the *gfp* of the *amt1::gfp* or *sepJ::gfp* loci. A, Sequences that encode the GFP (green), or the LE and RE of the cargo transposon (yellow), are highlighted. In addition, sequences of the *gfp* that were a part of the sgRNA encoded by pAzUT.14 are highlighted that encode the PAM (blue) and the remainder sequence specific part of the sgRNA (red). Resolution of the transposase complex leads to the insertion of nucleotides, causing the small insertions highlighted in purple. The position (Pos) of the insertions was counted starting from the first base after the PAM. B, The insertion was exactly at position 63 for 4 of the six independently recovered clones derived from either parent. T in red denotes the LE start base. Similarly, resolution of the transposase led in 4 of the 6 six clones to a five-base repeat, CCAGA in purple.

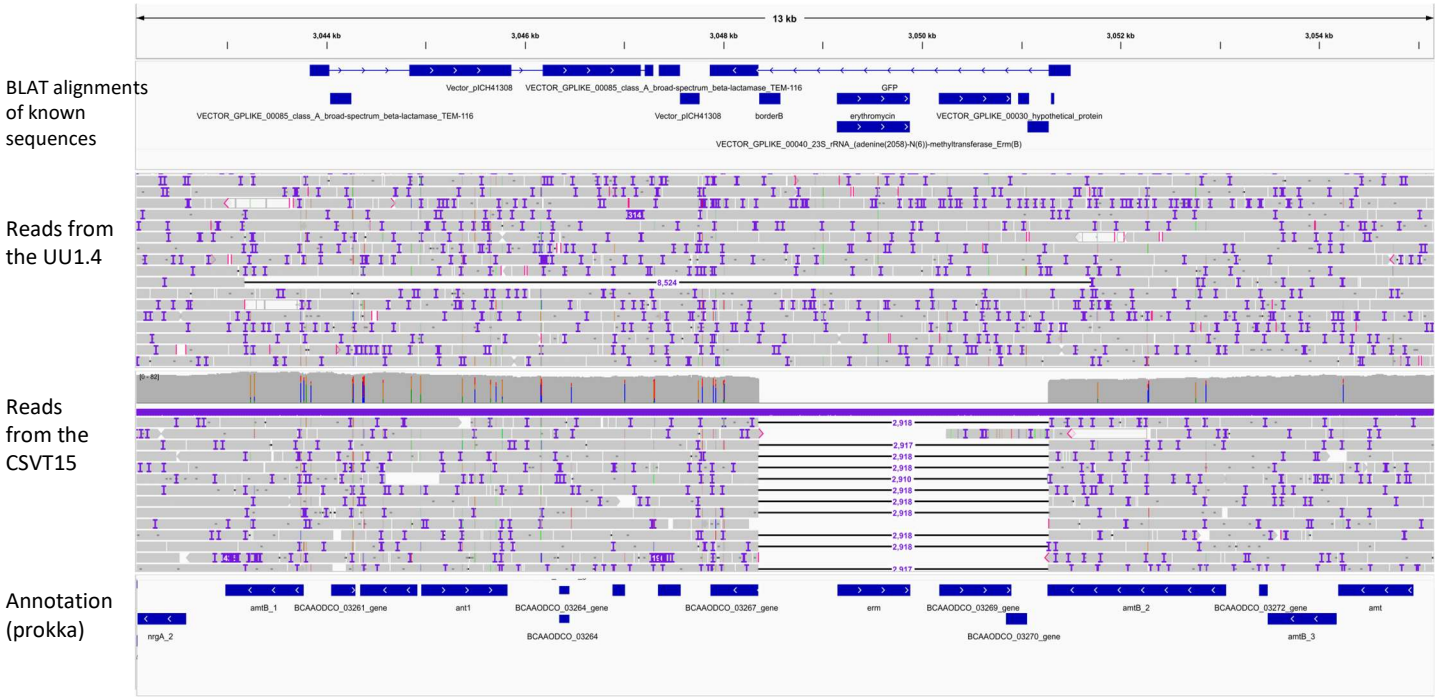

Figure 6 The *amt1:gfp* locus in the strain UU1.4 compared to the parental strain CSVT15. Strains were sequenced using MinIon with minimally 50 times coverage and their genomes assembled *de novo*. The assembly was automatically annotated with prokka, and known sequences were aligned using the BLAT-aligner to the *amt1::gfp* locus. In addition, MinIon reads obtained from sequencing strain UU1.4 and the parental strain CSV15 were aligned to the assembled genome of UU1.4. The assembly, automated and manual annotation, and the alignments were then visualized in Integral Genome Viewer (IGV) in the 13 kbp region spanning the 2918 bp cargo transposon.