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

Authors:
Sergio Arévalo¹,²,³,⁴, Daniel Pérez Rico¹, María Dolores Abarca¹, Laura W. Dijkhuizen¹, Peter Lindblad², Enrique Flores³, Sandra Nierzwicki-Bauer⁴ and Henriette Schluepmann¹

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 bottleneck 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.
Introduction

Cyanobacteria are of critical importance for the biogeochemical cycling of carbon and nitrogen and therefore substantially influence climate and primary production on earth (Sánchez-Baracaldo et al., 2022). Recent insights highlight the importance of symbioses in these cycles (Zehr and Capone, 2020). Moreover, cyanobacteria are capable of forming complex communities; they form symbioses with eukaryotic organisms including algae, plants, fungi, protozoa and invertebrate animals such as sponges and ascidians. This is in part because of their versatile secondary metabolism (Mutalipassi et al., 2021). Filamentous cyanobacteria from the order Nostocales typically differentiate O2-tight heterocysts that can fix dinitrogen in symbioses with plants from all land plant lineages (Rikkinen, 2017). Large sequencing datasets are growing in number confirming and expanding our understanding of those associations. However, understanding of biogeochemical interdependences at the molecular level within these associations has been hampered by the genetic intractability of some particular cyanobacteria, whether in culture or in symbiotic communities.

Genetic alteration in filamentous cyanobacteria has been accomplished mostly in a few species from the Nostoc/Anabaena genus complex. DNA cargo transfer into these filamentous cells was achieved by natural competence, electroporation, E. coli-mediated conjugation and Agrobacterium-mediated transfer (Gutiérrez et al., 2021). Novel approaches for gene transfer have included the use of cell penetrating peptides. Stabilization of the DNA-cargo was further achieved by methylation of the cargo in donor cells (Elhai et al., 1997), but also by engineering the DNA sequence such that the cargo may replicate (using broad spectrum origins of replication) and/or be used as a substrate for homologous recombination. This allowed for the integration into a target locus on either a plasmid or the chromosome. In all cases, integration of the cargo DNA was catalyzed by enzymes from the target cell either by homology directed repair (HDR) or the predominant non-homologous end-joining repair. The efficacy of the HDR, however, is notoriously low and requires cells in the S to G phase of the cell-cycle (Jiang, 2017). Increased efficiencies have been obtained recently in gram-negative bacteria by catalyzing insertion of the cargo DNA using CRISPR-associated transposases. Two systems, the types 1-F and V-K, have been studied in Escherichia coli (Klompe et al., 2019; Strecker et al., 2019). The type I-F from Vibrio cholerae uses a multi-protein effector consisting of the type-characteristic Cas3 protein complexing with CASCADE (complex for antiviral defense)-proteins (Klompe et al., 2019). The type V-K uses the single effector protein Cas12-K which was discovered in filamentous cyanobacteria, including Scytonema hofmannii (Strecker et al., 2019; Saito et al., 2021). The CAST increased the frequency of insertion of the donor DNA up to 80 % without selection. Importantly, it furthermore afforded its guided insertion because the effector protein binds small RNA (including the crRNA encoded by the spacer) that guide the transposition complex to the sequence specified by the crRNA. The mechanism of...
transposition was not thoroughly investigated but both cut- and copy-paste have been reported (Strecker et al., 2019; Vo et al., 2021).

The CASCADE requires more proteins to be expressed than the type V-K, did not transpose DNA cargo over 1 kb in size at high efficiency, and reproducibly inserted it 46-55 bases after the PAM, yet in varying orientations (Klompe et al., 2019). In contrast, the type V-K from *Scytonema hofmanni* inserted cargo DNA up to 10 kb long, 60-66 bp after the PAM in the orientation 5’ left-end (LE) to right end (RE) 3’ at high frequency. The ends of the Tn7-derived type V-K cargo include the 150 bp LE and the 90 bp RE with 3 and 4 transposase binding sites respectively (Peters et al., 2001). The type V-K (CAST) system was plagued with significant off-target insertions (Strecker et al., 2019), and may be affected strongly by transcription (Rodrigo González Linares et al., 2020). Transcripts likely titrated away the guide RNA in a way similar to DNA oligonucleotides or, possibly, the RNApolymerase complex displaces the CAST complex (Bialk et al., 2015; Zhang et al., 2020).

Type V-K transposition was reconstructed in vitro; it required TnsB known to join the 3’ ends of Tn7 with target DNA, TnsC an ATP-dependent transposase activator known to form heptameric rings on DNA and TniQ known to recruit TnsC to the target DNA (Stellwagen and Craig, 1997; Peters and Craig, 2001; Park et al., 2021; Querques et al., 2021; Hoffmann et al., 2022). It also needed cas12k which, like Cas9, is related to TnpB from the IS605 transposons (Faure et al., 2019). Unlike TnsA, cas12k did not have the active site known to break the 5’ of the Tn7 ends; the Cas12k required two small RNAs: the constant transactivating RNA that formed a duplex with a part of the otherwise variant crRNA sequence. To ease specifying of the CAST target, the two small RNA binding to Cas12k have been expressed as a single RNA (sgRNA) fusion with a hairpin loop and type two restriction enzyme cutting sites that allow seamless cloning of the variable target sequence; the sgRNA design has been validated and optimized (Strecker et al., 2019).

Here we present the development of synthetic biology vectors with the elements from type V-K CAST designed to study genome editing by way of RNA-guided and catalyzed insertion in filamentous cyanobacteria. We test RNA-guided transposition into expressed gene fusions in mutants of *Anabaena* sp. strain PCC 7120 (*Anabaena*), varying the target loci whilst keeping the sgRNA constant; we further varied the sgRNA sequence, PAM and strand. We characterize the insertions obtained in transconjugants by PCR, confocal microscopy and whole genome sequencing/de novo assembly to determine the mechanism of insertion and unravel eventual off-target effects.

Results

Cloning (Golden Gate) elements to share for studies on RNA-guided transposition in cyanobacteria

Preliminary studies had shown that CAST elements were toxic when *E. coli* containing vectors encoding them were left on plates for over one week. To be able to rapidly test
toxicity and efficacy of combinations of CAST elements in cyanobacteria, we chose to transfer the CAST elements into vectors for Golden Gate cloning. This was also the most efficient approach to eventually share all the individual elements and intermediate vectors in other synthetic biology studies. It further allowed the construction of large vectors systematically that included both, the expression cassettes of CAST elements and the DNA cargo between the LE and RE (Figure 1).

We generated Golden-Gate Level 0 vectors containing the operon tnsB, tnsC and tniQ, the cas12k, and the LE and RE after removal of BsaI and BpiI sites (domestication) by PCR (Figure 1, Level 0). Aiming at restricting the expression of the CAST elements to the cyanobacterial host and further decreasing it with nitrogen in the culture medium of cyanobacteria, we similarly domesticated the glnA promoter from Anabaena (Valladares et al., 2004) and transferred it into a level 0 vector (Figure 1, glnA).

The glnA promoter was only used to generate the level 1 expression cassettes of the CAST transposase proteins, in the first and second position in the final level T vector (Figure 1). The level 1 vector for the sgRNA expression cassette was chosen because its strong promoter would express in the reverse antisense direction and it contained the sgRNA scaffold with Lgul restriction sites allowing for the facile insertion of annealed primers with the spacer sequence specifying the sequence targeted by the sgRNA (Figure 1, level 1 position 3).

Finally, using the Cyano-Gate level T plasmids we assembled conjugative and replicative vectors as well as non-conjugative replicative vectors, and later complemented these with plasmids that were conjugative but not replicative (suicide) (Figure 1, level T). Due to the limited number of positions in the level T assemblies, we fused the LE and RE to the expression cassettes inside the DNA cargo that expressed YFP and erythromycin resistance (Figure 1, Level 1 positions 4 and 5).

The combinations of vectors listed in Supplemental Table S1 proved stable in E. coli. We thus attempted transfer of the conjugative replicative vectors listed in Supplemental Table S2 into Anabaena by triparental conjugation and subsequent antibiotic selection. We concentrated on selection with erythromycin (Em) because this antibiotic is compatible with the plant hosts in the cyanobacterial symbioses of our specific interest, the floating ferns from the genus Azolla.

RNA-guided transposition efficiently targets expressed genes irrespective of locus position in the chromosome

Reference vectors (Supplemental Table S1, pAzUT.3 and .4) containing only the cargo YFP and erythromycin or spectinomycin/streptomycin resistance expression cassettes yielded clones in all the Anabaena strains tested: the wild-type PCC7120, the derived strains containing the GFP fused to the ammonium transporter protein Amt1 (amt1::gfp, CSVT15; (Merino-Puerto et al., 2010) or the septal-protein SepJ (sepJ::gfp, CSAM137; (Flores et al., 2007).
Vectors containing all of the CAST and the cargo DNA but inactive sgRNA, containing the sgRNA in isolation, or containing all the CAST elements and active sgRNA targeting the GFP sequence yielded transconjugants on the nitrogen-rich BG11 medium (Supplemental Table S1). Expression of the CAST components, therefore, was well enough controlled behind the P\text{glnA} or the inherent toxicity of the CAST proteins was low enough for the transconjugant plasmids to be maintained. Strain CSAM137 did not grow on BG11₀ medium without nitrogen, suggesting that the SepJ::Gfp could not completely substitute for SepJ function under these conditions, although the strain did not show the characteristic filament fragmentation of sepJ inactivation mutants in BG11 (not shown, CSAM; Flores et al., 2007). The strain survived the short periods of induction for P\text{glnA}-dependent CAST-expression on BG11₀ as described below. Nevertheless, using strain CSVT15 the effect of the plasmids pAzUT3, 7 and 9 on the growth of the transconjugants could be tested (Supplemental Figure S1). In BG11 medium when plasmids were selected with erythromycin, all transconjugants grew equally well (Supplemental Figure S1 8d). In contrast, in BG11₀ medium when P\text{glnA} strongly drives the expression of the Cas12K from pAzUT7, transconjugant growth was inhibited (Supplemental Figure S1 28 d, pAzUT7). Expression of the all the CAST components in transconjugants with pAzUT9 did not seem to affect growth as in the case when Cas12k was expressed with the sgRNA scaffold only. Absence of sgRNA target is known to promote high-frequency off-target transposition which could be toxic (Xiao et al., 2021); a toxic effect of Cas12K in the absence of TnsB, TnsC and TniQ was not reported.

Homogenous clones cured of the donor plasmids were obtained at high frequency after sonication and growth on BG11 medium under erythromycin selection when using pAzUT14 in either the CSVT15 or CSAM137 strains. pAzUT14 expressed the sgRNA1 targeting the sense strand of the GFP 163 bp into the 563 bp protein (Supplemental Figure S2). Genotyping of the clones obtained at the amt1::gfp locus of CSVT15 was carried out by PCR amplification spanning either ends of the transposon (Figure 2A, PCR1, PCR2), the region targeted by the sgRNA (PCR3) and the pAzUT14 backbone (PCR4): in all three clones the expected size fragments were amplified at either ends and the total size of the inserted transposons was consistent with a single 5’LE to RE3’transposon insertion (Figure 2B, PCR1,2,3). The backbone of the donor vector could not be detected confirming that the clones were cured of it (Figure 2B, PCR4). Genotyping the clones obtained at the sepJ::gfp locus after curing pAzUT14 from transconjugants of CSAM137 gave a similar result (Figure 3A): amplicons spanning either ends (Figure 3B, PCR1, PCR2), amplicon spanning the entire length of the insertion (Figure 3B, PCR3) and absence of the backbone (Figure 3B, PCR4).

Using the vector pAzUT18 which encodes an sgRNA targeting the sense strand at the wild-type locus alr3727, all transconjugant colonies tested by PCR, as described in Supplemental Figure S3A, had traces of the cargo transposon insertion (Supplemental Figure S3B) unlike the parental wild-type strain (Figure 3B, WT), which confirms efficacy of the CAST on a wild-type gene.
Therefore, PCR-genotyping demonstrated that CAST was able to accurately guide the cargo transposon insertion into the same location, specified by the same sgRNA, independently of the Anabaena loci chosen.

Integration of the cargo DNA inactivated the targeted gene functions

Whilst confocal viewing was not ideal for screening because it was not able to distinguish whether the fluorescence detected from the expression of GFP was encoded in the plasmid or the DNA cargo transposon insertion in a chromosomal locus. In strains cured of the plasmids however, it verified strong YFP expression: this was seen as yellow fluorescence in the cytosol of all cells of the clones UU1 and UU2 (Figure 4 YFP, BF+YFP) compared to the respective parents CSVT15 and CSAM137 which expressed the GFP fusions. This confirmed that the single copy insertion of RNA-guided transposon cargo into the chromosome loci may be used to tag Anabaena, which is known to be polyploid (average number of chromosome 8.2, Hu et al., 2007).

The filaments from strain CSVT15 differed little from the wild-type Anabaena (Figure 4 Anabaena). In contrast, the filaments were much shorter and the cells less regularly spaced in the UU1 clone (Figure 4, UU1) compared with its parental strain (CSVT15), consistent with disruption of a function in the amt1::gfp locus under the conditions tested here.

Similarly, clone UU2 with the cargo transposon inserted in the sepJ::gfp exhibits shorter filaments, irregularly shaped cells and cells dividing in an envelope resembling that of heterocysts a phenotype characteristic for loss of function SepJ protein (Flores et al., 2007). We conclude therefore that the RNA-guided transposition in Anabaena may serve to generate tagged inactivation at expressed loci specified by the sgRNA.

The sgRNA targeting the sense strand of expressed GFP and containing the GTT PAM was effective

We next explored how the sgRNA sequences affect the specificity and efficacy of RNA-guided transposition (Supplemental Figure S2). The sgRNA from pAzUT14 efficiently targeted the sense-strand of the expressed gfp fusions and contained in its 5 prime the PAM GTT 162 bases into the gfp sequence (Figure 2, 3). Including the PAM in the sgRNA sequence for strong binding of the sgRNA to Cas12k was attempted because the sgRNA locus on the donor plasmid is protected by the proximity of the LE of the transposon from DNA cargo insertions and thus inactivation. Insertions of the cargo transposon were observed when sgRNA was used from pAzUT10 that contained the PAM GGTT and targeted the sense strand of gfp, 271 bp into the coding sequence of gfp, but these were recovered at much lower frequency (Supplemental Figure S2, and data not shown). No insertions were observed for pAzUT12 where the sgRNA contained the PAM GGTT and targeted the antisense strand of gfp 563 bp into the gfp sequence (Supplemental Figure S2, and data not shown).
The cargo transposon inserted mostly 63 b after the PAM and led to 2-5 base duplications at the insertion sites

PCR-results suggested that, in all cases, the cargo transposon was inserted in the 5’ LE to RE 3’ orientation, but PCR could not distinguish whether the insertions had been a result of cut- or copy-paste mechanisms. For accurate sequence information on the integration loci, DNA was extracted from the six independently obtained clones exhibiting RNA-guided transposition (Figure 3 and 4), their respective parental strains, and from the wild-type Anabaena. The DNA was then sequenced using Minion long-reads with a minimum 50 x coverage which allowed assemblies to be resolved into a single full-length chromosome with high-accuracy sequence information.

The sequences immediately adjacent to the cargo transposon were extracted and aligned for comparison (Figure 5). In four of the six genomes analyzed, the LE inserted exactly 63 bases behind the PAM. When comparing sequences at the LE with RE, duplications of the bases at the vicinity of the insertion due to the resolution of transposase complex were two- to four bases long, and in five out of six cases they were exactly five bases (Figure 5, duplications highlighted in purple). RNA-guided cargo transposon insertions were therefore very reproducible.

RNA-guided transposition was unidirectional 5’ LE to RE 3’ and single copy insertion without co-integration of donor plasmid

To inspect the overall structure of the loci targeted by the cargo transposon in each clone sequenced, 13 kb regions of the consensus assembly sequences spanning the insertions were viewed in IGV along with aligned reads from the clone and its parent. A typical result is shown for the clone UU1.4 in Figure 6. Alignment of reads obtained from the parental strain CSV15 revealed some single nucleotide polymorphisms between the clone and the reference strain, yet the foremost difference was the 2,918 bp insertion in the clone UU1.4 corresponding in size with the cargo transposon. Automatic annotation (Figure 6, Annotation (prokka)) identified amt1 as amtB based on its homology to amtB from E. coli. Individual alignments using known sequences (Figure 6, BLAT alignments of known sequences) identified the start of the GFP, the gap caused by the insertion (line with arrows), and the remainder of gfp; it furthermore identified the yfp (VECTOR_GFPLIKE) and erythromycin resistance. Additionally, downstream of the cargo transposon insertion, it identified the vector sequences used to generate the amt1::gfp fusion in the parental strain CSV15: these vector sequences inserted through a single cross-over homologous recombination event. Results obtained from the analyses of the other six clones investigated revealed an identical mechanism of insertion of the RNA-guided cargo transposon. From the restricted analyses carried out here we conclude, therefore, that cargo transposon insertions guided by CAST were unidirectional from 5’LE to RE 3’ with a precise resolution of the 5’and 3’end with cut.
No sign of off-target insertions or remobilization of endogenous mobile elements

We next examined whether exposure to the CAST machinery led to the remobilization of mobile genetic elements (MGE) already present within the genome of the parental strains. The automatic detection of indels and deletions in genomes of the transconjugants compared to their respective parental strain gave the results listed in Supplemental Table S3. Manual inspection of the regions corresponding to these events using evidence from the aligned long-reads, however, could not verify any changes but instead assigned the indels to computation artefacts. We therefore conclude that exposure to the CAST machinery did not cause the remobilization of endogenous mobile genetic elements in the six cases that were analyzed in this study.

Discussion

A Golden Gate toolbox extension for genome engineering by RNA-guided transposition in filamentous cyanobacteria

The discovery of type II restriction enzymes with three- to four-base overhangs opened the way to seamless cloning of many fragments simultaneously. Yet the most important advance towards synthetic biology was their use in a cloning system with a controlled “vocabulary” in which overhang sequences specify the position and orientation in an expression cassette, or cloning assembly of multiple cassettes or non-coding elements: this allowed sharing of the individual elements in plasmids between laboratories and new combinations to be generated at great speed. Here we used the vocabulary from the “Golden-Gate” system expanding on an existing set of vectors named CyanoGate (Vasudevan et al., 2019): we added a conjugative suicide vector, all the CAST elements and the P_glnA promoter for nitrogen-regulated expression in cyanobacteria under nitrogen starvation.

We combined CAST elements and the cargo DNA in a single conjugative vector (Figure 1), but unlike in Rubin et al. (2021), we inserted the sgRNA cassette in the reverse complement orientation such that RNA polymerase on the strongly expressed sgRNA gene would not affect the transposase binding to the LE. In addition, we designed 34 nt-long sgRNA spacers, well over the minimum required for specific targeting of RNA-guided transposition in vitro (Xiao et al., 2021). Also, the PAM is not cleaved by the CAST as in the case of Cas9 (Heler et al., 2015) and insertions near the PAM of the sgRNA would be suppressed by the LE and RE borders of the donor plasmid (Strecker et al., 2019). We therefore included the PAM in the sgRNA spacer to test whether the binding of the sgRNA PAM sequence to the N-terminal groove of Cas12k would lead to tighter binding and therefore repression of random transposition (Xiao et al., 2021). As presented in the results, this configuration resulted in the specific RNA-guided integration at either of the loci targeted by the sgRNA in Anabaena (Figure 2 and 3).
Accurate targeting of the RNA-guided transposon cargo in *Anabaena*

In all six transconjugant *Anabaena* clones analyzed in depth, no off-target effects were detected (Supplemental Table S3). These effects could have been off-target insertions of cargo DNA, or deletions and remobilization of endogenous mobile elements inside the *Anabaena* genome. Absence of off-target effects contrasted with previous reports on poor target specificity in a variety of gram-negative bacteria (Strecker et al., 2019; Rubin et al., 2021). Results from Xiao et al. (2021) showed that when Cas12 lacks the sgRNA, RNA transposition occurs at a high rate and in random locations, therefore, the authors suggested that sgRNA binding suppresses the random transposition. In our system, sgRNA was thus expressed at a sufficiently high level such that when bound to the Cas12k it suppressed random transposon insertions.

It is recognized that, higher throughput analysis will be required to investigate off-target effects of the engineered CAST system used here for *Anabaena*. We observed poor growth of clones on BG110 medium, when the P_{glnA} directs optimum expression of the Cas12k and an sgRNA scaffold, suggesting that the constitutive expression of these CAST components may be toxic to the *Anabaena* cells (Supplemental Figure S1). We observed toxicity of the CAST system in *E. coli* (results not shown). The relative growth handicap imparted by the plasmids encoding the CAST in BG110 medium was exploited to cure the replicative plasmid containing the CAST away from inside the clones exhibiting transposon integration (Figure 2, 3).

We used the unmodified LE sequence which is suspected to direct homing via delocalized CRISPR RNA to the tRNA-leu in *Scytonema hofmanii (sh)*: it contains a 17 bp motif matching the shtRNA-Leu gene. However, in our construct the CRISPR direct repeat found upstream of the 17-bp motif in the shCAST locus is missing (Saito et al., 2021). A more thorough understanding of the LE sequence is critically needed in many respects, but most importantly for the design of intron boundaries in the LE and cargo DNA so as to engineer protein fusions in coding-loci targeted by the RNA guided transposition. TnsB binding sites are the characteristic features of the LE and RE and TnsB was found to bind the backbone of the DNA only and to have a not very strict DNA-sequence preference upon binding (Kaczmaraska et al., 2022).

RNA-guided transposition was unidirectional from LE to RE, reflected a copy and paste mechanism, and was likely influenced by RNA

Unidirectional insertion of the cargo (Figure 5) is consistent with the unidirectional formation of the polymeric TnsC complex with the target DNA and Cas12k serving to recruit the TnsB transposase at the target site (Park et al., 2021; Querques et al., 2021). Co-integration of the plasmid sequences supplied along with the transposon cargo was reported to occur at frequencies ranging from some 85% (Vo et al., 2020; Vo et al., 2021) to 19.4%, and 0.6% in the specific case of the Cas12k-homing endonuclease fusion (Tou et al., 2022). An explanation for co-integration was also proposed: the CAST lacks the TnsA required to excise the transposon and the Cas12k lacks the endonuclease activity.
to substitute for the TnsA. In *Anabaena*, we recovered only single copies of the cargo transposon and there were no plasmid sequence co-integrates (Figure 6). This may be due to the presence of accessory proteins in the cyanobacteria in which the Cas12k has evolved to be devoid of the nickase activity typically found in TnsA (Xiao et al. 2021 and ref therein). Our results are of insufficient throughput to conclude definitely. The sgRNA targeting the antisense strand of the *gfp*-gene fusions at either loci was not effective. This may have been due to the sense transcript RNA duplexing efficiently with the sgRNA and consequently less efficient RNA/DNA hetero-duplexing at the target location. For example, the gRNA depletion with DNA oligonucleotides was reported to be effective in the case of Cas9 bound gRNA (Bialk et al., 2015). The lesser efficacy of the sgRNA with a spacer containing GTT PAM may stem from a suboptimal fit at the T287 from the WED domain and the R421 from the PI domain (Xiao et al., 2021).

RNA-guided transposition to understand genetic features in complex communities such as symbioses

RNA-guided transposition may be used to target a locus in a specific organism in a mixture effectively: this has been recently demonstrated in bacterial communities of various complexities (Rubin et al., 2021). The advantage of this approach does not only reside in the catalysis, and therefore the efficiency with which the cargo DNA is inserted into the DNA, but also in the insertion of tags at precise locations resulting in engineered loci that may be selected or followed visually in complex mixtures. We used YFP to trace whether the colonies were homogenous genetically after sonication and prolonged selection (Figure 4). The method lends itself to study fitness of the cells with engineered loci in mixed communities so as to study the role of the original versus engineered loci in microbe interactions.

We used a replicative plasmid in this initial study to be assured of sufficient expression of the CAST elements and sufficient cargo substrate for transposition. A replicative plasmid is not desired for the precise engineering of loci in microbes within complex communities because this may lead to a larger proportion of off-target insertions of the cargo. Because curing the replicative plasmid used for the delivery of the CAST and cargo is cumbersome, we tested a conjugative suicide vector (pAzUT.17) for delivery of the cargo DNA in a shortened protocol: after consecutive conjugation and sonication (to reduce the length of the filaments), cells were immediately transferred onto BG11\(_0\) medium for two days, and then on BG11-medium and selection. RNA-guided cargo DNA transposition events from conjugation with suicide plasmids were detected in clones that had no trace of the suicide plasmid (Supplemental Figure S4).

We conclude that filamentous cyanobacteria, such as *Anabaena*, are amenable to genome engineering using RNA-guided transposition with the Cas12k-based CAST system. We next will need to test the approach in other species and using alternative methods of DNA-transfer. We urgently will test the method on symbiotic species, such
as *Nostoc punctiforme* and *Nostoc azollae*, especially when present in complex microbial consortia.

### Materials and Methods

#### Bacterial strains and growth conditions

*Anabaena* sp. (also known as *Nostoc* sp.) strain PCC 7120 (Merino-Puerto et al., 2010) and CSAM137 (Flores et al., 2007) were cultured photoautotrophically in BG11 or BG11₀ (without NaNO₃) medium (Rippka et al., 1974) at 30°C, under constant white light (35 μE·m⁻²·s⁻¹) and shaking. For solid cultures, 1% (w/v) agar (Bacto-Agar, Difco) was added. When required, media were supplemented with 5 μg/ml (solid medium) or 2.5 μg/ml (liquid medium) streptomycin sulfate (Sm), spectinomycin (Sp) or erythromycin (Em).

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

#### DNA tools and vectors

*Scytonema hofmannii* Tn7-like transposase components including shCas12k, the operon encoding TnsB, TnsC and TniQ and the optimized sgRNA scaffold were obtained from pHelper_ShCAST; the left- and right-end (LE and RE) sequences of the Tn7-like transposon were obtained from pDonor_ShCAST (Strecker et al., 2019b). The pJ23119 promoter and T7Te terminator were PCR amplified from pHelper_ShCAST together with the optimized single guide (sg) RNA-scaffold sequence that contained the lgu1 sites for insertion of the target-specific spacer. The *P_glnA* was amplified from the pRL3845 plasmid (Valladares et al., 2004). All other promoters and terminators were obtained from the CyanoGate system (Vasudevan et al., 2019). The spacer part of the sgRNA was assembled by hybridization of two partially complementary synthetic oligonucleotides (Integrate DNA Technologies; Figure S2). The coding sequences for the erythromycin and Sm/Sp resistance genes were obtained from the CyanoGate system (Vasudevan et al., 2019).

Vectors to assemble the plasmids to test CAST in cyanobacteria were obtained from the MoClo Plant Tool kit following the pipeline described in (Engler et al., 2014), where level 0 plasmids contain individual components (promoters, coding regions, terminators, etc.) and expression cassettes are assembled in level 1 plasmids. The final level T plasmids, containing all the expression cassettes, were assembled using the replicative and conjugative vector pCAT.000 or the replicative but not conjugative vector pCAT.334 form the CyanoGate system (Vasudevan et al., 2019). In addition, cyanobacterial replication encoded by OriT in pCAT.000 was replaced with the ColE1 from pEERM3
(Englund et al., 2015) to obtain a T-level backbone allowing conjugation but not replication (committing suicide) in the cyanobacterial host.

**Plasmid construct and bacterial colony screening**

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

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

Level 1 plasmids were assembled by digestion with BsaI and ligation, following the MoClo cloning protocol (Weber et al., 2011). To obtain pAzU1.3.3, the annealed oligonucleotide spacers were introduced in plasmid pAzU1.3 containing the sgRNA scaffold by LguI digestion and ligation (Figure 1). To obtain level T plasmids, level 1 inserts were introduced in pCAT.000 or pCAT.334, together with an end-linker (L), by BpiI digestion and ligation. All restriction enzymes and T4-DNA ligase were from ThermoFisher.

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

**Cyanobacterial transformation and sonication of exconjugants**

*E. coli* strains ED8634 (containing pRL443 encoding the conjugation machinery) and HB101 (containing pRL623 and the cargo plasmid) were used for tri-parental conjugation as described in (Elhai and Wolk, 1988). The mixture of *E. coli* and cyanobacteria was spread and cultured on filters deposited on solid BG11 + 5% LB plates for 24 hours and then transferred to BG11 medium for after another 24 hours. Filters were moved to BG11 (or BG11r for the rapid conjugation protocol) plates supplemented with Sm, Sp and Em for 48 hours, then transferred back to BG11 medium supplemented with the corresponding antibiotics. Potential positive colonies growing after two weeks were re-streaked and confirmed by PCR amplification followed by sequencing, as described above.

To generate clonal strains of exconjugants, exconjugants that had integrated the eYFP-encoding sequence into the genomic DNA were grown in 25 ml of BG11 medium to an optical density (OD) of 1 at 750 nm. Then, under sterile conditions, 1 ml of culture was
removed and placed in a sterile plastic tube for sonication. During the sonication step, the filament-length was monitored until most of the filaments were broken down to 2-3 cells. Finally, serial dilutions were made and 50 μl of each dilution was plated on solid BG11 medium supplemented with the appropriate antibiotic, and allowed to grow under the conditions described above. The genotypes of the clonal colonies thus obtained were analyzed by PCR (Supplemental Table S4).

To test relative growth rates, cultures grown in liquid BG11 medium with the corresponding antibiotic for 1 week were washed with BG110 medium, inoculated with the chlorophyll equivalents indicated and incubated in the light at 30°C for 8, 12 and 28 days.

**Confocal microscopy**

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

**DNA extraction and sequencing**

Cyanobacterial genomic DNA was isolated using the GeneJET Genomic DNA Purification Kit (ThermoFisher) following the manufacturer’s instructions. DNA quality and concentration were analyzed by UV absorption, q-bit and on gel. Minion sequencing libraries were generated using 0.3-3 μg DNA with the SQK-LSK109 kit following instructions by Nanopore Technologies (version NBE_0905_v109_revAK_14Aug2019); in case of multiplexing the EXP-NBD104 extension was combined with NEB Blunt/TA Ligase Master Mix (M0367, New England BioLabs (NEB)). Briefly, the DNA was first repaired (NEBNextFFPE Repair mix (M6630) and NEBNext Ultra II End repair/dA-tailing Module (E7546)), then bound on AMPure XP beads (Beckman & Coulter) and cleaned twice with 75% v/v ethanol, before elution in water at 50°C for 10 min. Subsequently, when barcoded, the barcode adapters were ligated and the DNA cleaned once more using the AMPure XP beads and 75% v/v ethanol. Finally, the DNA was ligated to the sequencing primers in the presence of the tether, and washed with Short Fragment Buffer (SQK-LSK109 kit, Nanopore Technologies) bound once more on the AMPure XP beads, before elution in elution buffer at 50°C for 10 min. Priming and loading of the recycled MinION flowcell (R9.4.1 Nanopore Technologies, starting with 600 active pores) was as per the manufacturers
instruction using reagents from EXP-FLP002 (Nanopore Technologies). The flowcell was washed between the loading of different libraries using the EXP-WSH004 (Nanopore Technologies) reagents that included DNase.

**Genome assemblies and annotation**

Data acquisition of the nanopore sequencing used the Minknow program (Oxford Nanopore Technologies) until 50 times genome coverage was achieved for the barcode with the lowest reads. Base calling was carried out separately. Assemblies were computed *de novo* using Flye (Kolmogorov et al., 2019) with default settings, then visualized using Bandage (Wick et al., 2021); polishing using medaka proved not to improve the assemblies; annotation of the assemblies was carried out with bakta (Schwengers et al., 2021) and prokka (Seemann, 2014) for comparison. *Anabaena* genomes, and alignment files of the minion reads aligned to them with minimap2 (Li, 2018), were visualized using Integral Genome Viewer (IGV; Thorvaldsdóttir et al., 2013).

The BLAT function inside the IGV was used to locate YFP, GFP, Amt1, SepJ as well as the left and right end sequences of the Tn7 provided in the original plasmid pAzUT14. Large structural variations between the transconjugant after RNA-guided transposition and the reference genomes were programatically detected using Sniffles, after mapping of the reads using NGMLR (Sedlazeck et al., 2018). The log of the analyses is detailed at https://github.com/lauralwd/anabaena_nanopore_workflow/blob/main/script.sh.

When comparing all of the strains with the transposon targeted into the *amt1::gfp* locus, Sniffles identified 14 insertions, 3 deletions and 6 recombinations (Supplemental Table S2). The insertions were then inspected by extracting the FASTA files of the insertions and deletions, then locating these regions in the assemblies of the transconjugants using BLAT in IGV for further evaluation.

**Material and Data availability**

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

**Accession numbers**

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

**Acknowledgements**

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

RNA-guided transposition for cyanobacteria by Arévalo et al. 2022
**Author contributions**

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


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<sub>glmA</sub> pAzU0.2 for the operon encoding TnsB, TnsC and TniQ (<sub>tnsBCQ</sub>), 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.

**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. Fluorescence 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 Figure 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 Figure 3. The bar corresponds to 10 mm.

**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.
Supplemental Data

The following supplemental data is available in the online version of this article.

Supplemental Figure S1. Relative toxicity of the CAST plasmids in Anabaena.

Supplemental Figure S2. The three different sgRNAs targeting gfp.

Supplemental Figure S3. Efficient targeting at locus alr3727 in wild-type Anabaena.

Supplemental Figure S4. Rapid conjugation protocol for RNA-guided transposition using the suicide plasmid pAzUT.17.

Supplemental Table S1. CASTGATE vectors generated in this study.

Supplemental Table S2. CASTGATE vectors transferred to and tested in Anabaena wild-type, and the CSVT15 and CSAM137 strains in this study.

Supplemental Table S3. Insertions detected by Snifles comparing genome assemblies from the parental strains with those from clones obtained after RNA-guided transposition.

Supplemental Table S4. Primers used for PCR assays and key cloning steps in this study.

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. et 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 \textit{gfp} of the \textit{sepJ::gfp} locus of the parental strain CSAM137, and its detection \textit{a} by PCR. \textbf{A}, Scheme of the \textit{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 \textit{gfp}, also encoded in pAzUT.14. When mobilized with resolution of the transposase complex, the cargo transposon inserts inside the \textit{gfp} of the \textit{sepJ::gfp} fusion. \textbf{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 \textit{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) \textit{Anabaena} sp. PCC 7120; (P) CSAM137; (B) pAzUT.14; (1) AzUU2 clone 3; (2) AzUU2 clone 8; and (3) AzUU2 clone 7. \textit{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. Fluorescence 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.

RNA_Guided_transposition_for_cyanobacteria
Figure 6 The *amt1:gfp* locus in the strain UU1.4 compared to the parental strain CSV15. 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.