HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2020 January 05.

Published in final edited form as:

Science. 2019 July 05; 365(6448): 48–53. doi:10.1126/science.aax9181.

RNA-guided DNA insertion with CRISPR-associated transposases

Jonathan Strecker[1,2,3,4], **Alim Ladha**[1,2,3,4], **Zachary Gardner**[1,2,3,4], **Jonathan Schmid-Burgk**[1,2,3,4], **Kira S. Makarova**[5], **Eugene V. Koonin**[5], **Feng Zhang**[1,2,3,4,†]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Abstract

CRISPR-Cas nucleases are powerful tools to manipulate nucleic acids; however, targeted insertion of DNA remains a challenge as it requires host cell repair machinery. Here we characterize a CRISPR-associated transposase (CAST) from cyanobacteria *Scytonema hofmanni* which consists of Tn7-like transposase subunits and the type V-K CRISPR effector (Cas12k). ShCAST catalyzes RNA-guided DNA transposition by unidirectionally inserting segments of DNA 60–66 bp downstream of the protospacer. ShCAST integrates DNA into unique sites in the *E. coli* genome with frequencies of up to 80% without positive selection. This work expands our understanding of the functional diversity of CRISPR-Cas systems and establishes a paradigm for precision DNA insertion.

†Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Author contributions: J.S., and F.Z. conceived the project. J.S., A.L., and Z.G. performed bacterial experiments. J.S. purified CAST proteins and performed in vitro reactions. A.L. performed genome targeting experiments, and insertion specificity analysis with help from J.S.B. K.S. and E.V.K. identified CAST loci and performed bioinformatics analysis. F.Z. supervised the research and experimental design. J.S. and F.Z. wrote and revised the manuscript with input from all authors.

Competing interests: J.S. and F.Z. are co-inventors on US provisional patent application no. 62/780,658 filed by the Broad Institute, relating to CRISPR-associated transposases. F.Z. is a co-founder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences.

Data and materials availability: Expression plasmids are available from Addgene under UBMTA; support forums and computational tools are available via the Zhang lab website (zhanglab.bio).

Supplementary Materials
Materials and Methods
Figs. S1 to S12
Tables S1 to S6
References (35–38)

Prokaryotic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated proteins (Cas) systems provide adaptive immunity against foreign genetic elements via guide-RNA dependent DNA or RNA nuclease activity (1–3). CRISPR effectors, such as Cas9 and Cas12, have been harnessed for genome editing (4–9) and create targeted DNA double-strand breaks in the genome, which are then repaired by endogenous DNA damage repair pathways. Although it is possible to achieve precise integration of new DNA following Cas9 cleavage either through homologous recombination (10) or non-homologous end-joining (11, 12), these processes are inefficient and vary greatly depending on cell type. Homologous recombination repair is also tied to active cell division making it unsuitable for post-mitotic cells. Recently, an alternative approach to make point mutations on DNA has been developed that relies on using dead Cas9 (13) to recruit cytidine or adenine deaminases to achieve base editing of genomic DNA (14–16). However, base editing is restricted to nucleotide substitutions, and thus efficient and targeted integration of DNA into the genome remains a major challenge.

To overcome these limitations, we sought to leverage self-sufficient DNA insertion mechanisms, such as transposons. We explored bioengineering approaches of CRISPR-Cas effectors to facilitate DNA transposition (Fig. S1). Cas9 binding to DNA generates an R-loop structure exposing a substrate for enzymes that act on single-stranded DNA (ssDNA). By tethering nickase Cas9(D10A) to the ssDNA transposase TnpA from *Helicobacter pylori* IS608 (17, 18) we observe targeted DNA insertions *in vitro* and in *E. coli* that are dependent on TnpA transposase activity, Cas9 sgRNA, and the presence of an insertion site within the ssDNA. However, the requirement of ssDNA donor will require continued development for efficient synthesis and delivery to cells.

A number of CRISPR-Cas systems lacking active nuclease domains have been identified previously, including minimal type I loci lacking the Cas3 helicase-nuclease (19) and type V loci containing a Cas12 effector with a naturally inactivated RuvC-like nuclease domain (20). The absence of nuclease domains raises questions as to the biological function of these CRISPR-Cas systems which can only bind but not cleave DNA. Recently, an association between Tn7-like transposons and subtype I-F, subtype I-B, or subtype V-K (formerly, V-U5) CRISPR-Cas systems was reported (21, 22). The CRISPR-Cas associated Tn7-like transposons contain *tnsA*, *tnsB*, *tnsC*, and *tniQ* genes (21), similar to the canonical Tn7 heterotrimeric TnsABC complex (23, 24). Tn7 is targeted to DNA via two alternative pathways that are mediated respectively by TnsD, a sequence-specific DNA binding protein which recognizes the Tn7 attachment site (25, 26), and TnsE, which facilitates transposition into conjugal plasmids and replicating DNA (27).

The association between Tn7-like transposons and CRISPR-Cas systems suggests that the transposons might have hijacked CRISPR effectors to generate R-loops in target sites and facilitate the spread of transposons via plasmids and phages (21). In the case of subtype V-K, the position of the CRISPR-Cas locus is frequently conserved in predicted transposons, suggesting that CRISPR-Cas is linked with transposition (22). However, since canonical Tn7 transposons often carry cargo genes with defense functions that are beneficial to the host cell (24), it is also possible that CRISPR-Cas may be cargo genes. To date, no functional data on transposon-encoded CRISPR-Cas systems have been reported. Here, we show that Tn7-like

transposons can be directed to target sites via crRNA-guided targeting and elucidate the mechanism of crRNA-guided Tn7 transposition. We further demonstrate that Tn7 transposition can be reprogrammed to insert DNA into the genome of *E. coli*, highlighting the potential of using RNA-guided Tn7-like transposons for genome editing.

Characterization of a transposon associated with a type-V CRISPR system

Among the transposon-encoded CRISPR-Cas variants, the subtype V-K are the simplest because they contain a single-protein CRISPR-Cas effector (20, 21, 28), Cas12k (formerly, C2c5). Subtype V-K systems are so far limited to cyanobacteria and the latest non-redundant set includes 63 loci that, in the phylogenetic tree of Cas12k, split into 4 major branches, covering a broad taxonomic range of Cyanobacteria (22). All V-K systems are embedded within predicted Tn7-like transposable elements with no additional *cas* genes, suggesting that, if they are active CRISPR-Cas systems, they might rely on adaptation modules supplied *in trans*. Of the 560 analyzed V-K spacers, only 6 protospacer matches were identified: 3 from cyanobacterial plasmids, and 3 from single-stranded transposons of IS200 or IS650 families (22). These findings suggest the possibility that V-K systems provide a biological advantage for the host transposons by directing integration into other mobile genetic elements, to enhance transposon mobility, and to minimize the damage to the host.

For experimental characterization, we selected two Tn7-like transposons encoding subtype V-K CRISPR-Cas systems (hereafter, CAST, CRISPR-associated Transposase). The selected CAST loci were 20–25 kb in length and contained Tn7-like transposase genes at one end of the transposon with a CRISPR array and Cas12k on the other end, flanking internal cargo genes (Fig. 1A, Fig. S2A, B). We first cultured the native organisms *Scytonema hofmanni* (UTEX B 2349; Fig. 1B), and *Anabaena cylindrica* (PCC 7122) and performed small RNA-sequencing to determine if the CRISPR-Cas systems are expressed and active. For both loci, we identified a long putative tracrRNA that mapped to the region between Cas12k and the CRISPR array, and in the case of *S. hofmanni* (ShCAST) we detected crRNAs 28–34 nt long, consisting of 11–14 nt of direct repeat (DR) sequence with 17–20 nt of spacer (Fig. 1C, Fig. S2C).

To investigate whether ShCAST and AcCAST function as RNA-guided transposases, we cloned the four CAST genes (tnsB, tnsC, tniQ, and Cas12k) into a helper plasmid (pHelper) along with the endogenous tracrRNA region and a crRNA targeting a synthetic protospacer (PSP1). We predicted ends of the transposons by searching for TGTACA-like terminal repeats surrounded by a duplicated insertion site (21) and constructed donor plasmids (pDonor) containing the kanamycin resistance gene flanked by the transposon left end (LE) and right end (RE). Given that CRISPR-Cas effectors require a protospacer adjacent motif (PAM) to recognize target DNA (29), we generated a target plasmid (pTarget) library containing the PSP1 sequence flanked by a 6N motif upstream of the protospacer. We co-electroporated pHelper, pDonor, and pTarget into *E. coli* and extracted plasmid DNA after 16 h (Fig. 1D). We detected insertions into the target plasmid by PCR for both ShCAST and AcCAST and deep sequencing confirmed the insertion of the LE into pTarget. Analysis of PAM sequences in pInsert plasmids revealed a preference for GTN PAMs for both ShCAST and AcCAST systems, suggesting that these insertions result from Cas12k targeting (Fig.

1E, Fig. S3A,B). We next examined the position of the donor in pInsert products relative to the protospacer. Insertions were detected within a small window 60–66 bp downstream from the PAM for ShCAST and 49–56 bp from the PAM for AcCAST (Fig. 1F). No insertions were detected in the opposite orientation for either system, indicating that CAST functions unidirectionally. Although DNA insertions could potentially arise from genetic recombination in *E. coli*, the discovery of an associated PAM sequence and the constrained position of insertions argues against this possibility.

To validate these findings, we transformed *E. coli* with ShCAST pHelper and pDonor plasmids along with target plasmids containing a GGTT PAM, an AACC PAM, and a scrambled non-target sequence. We assessed insertion events by quantitative droplet digital PCR (ddPCR), which revealed insertions of the donor only in the presence of pHelper and a pTarget containing a GGTT PAM and crRNA-matching protospacer sequence (Fig. 1G). Additional experiments with 16 PAM sequences confirmed a preference for NGTN motifs (Fig. S3C). As further validation, we recovered pInsert products and performed Sanger sequencing of both LE and RE junctions. All sequenced insertions were located 60–66 bp from the PAM and contained a 5-bp duplicated insertion motif flanking the inserted DNA (Fig. S4), consistent with the staggered DNA breaks generated by Tn7 (30). As Tn7 inserts into a CCCGC motif downstream of its attachment site, we hypothesized that the sequence within the insertion window might also be important for CAST function. We generated a second target library with an 8N motif located 55 bp from the PAM and again co-transformed the library into *E. coli* with ShCAST pHelper and pDonor followed by deep sequencing (Fig. S5A). We observed only a minor sequence preference upstream of the LE in pInsert, with a slight T/A preference 3 bases upstream of the insertion site (Fig. S5B–D). ShCAST can therefore target a wide range of DNA sequences with minimal targeting rules. Together these results indicate that AcCAST and ShCAST catalyze DNA insertion in a heterologous host and that these insertions are dependent on a targeting protospacer and a distinct PAM sequence.

Genetic requirements for RNA-guided insertions

We next sought to determine the genetic requirements for ShCAST insertions in *E. coli* and constructed a series of pHelper plasmids with deletions of each element. Insertions into pTarget required all four CAST proteins and the tracrRNA region (Fig. 2A). To better characterize the tracrRNA sequence, we complemented pHelper$_{\Delta tracrRNA}$ with various tracrRNA driven by the pJ23119 promoter. Expression of the 216-nt tracrRNA variant 6 alone was sufficient to restore DNA transposition (Fig. 2B). The 3’ end of the tracrRNA is predicted to hybridize with a crRNA containing 14 nt of the DR sequence and we designed single guide RNAs (sgRNA) testing two linkers between the tracrRNA and crRNA sequences. Both designs supported insertion activity in the context of the tracrRNA variant 6 (Fig. 2C). We observed that expression of tracrRNA or sgRNA with the pJ23119 promoter resulted in a 5-fold increase in the insertion activity compared to the natural locus, suggesting that RNA was rate-limiting during heterologous expression.

As ShCAST does not destroy the protospacer upon DNA insertion, we asked whether multiple insertions could occur in pTarget, or if these are inhibited as with canonical Tn7

(31, 32). We generated target plasmids containing LE+RE, or LE alone, and measured ShCAST transposition activity at 6 nearby protospacers. We observed a strong inhibitory effect on transposition at a protospacer 62 bp from the LE (less than 1% of relative activity to pTarget), and only 5.7% relative activity 542 bp from the LE (Fig. 2D), indicating that CAST transposon ends act in cis to prevent multiple insertions. The presence of LE alone resulted in a weaker inhibitory effect and we observed 61.1% of activity at 542 bp away from the transposon end (Fig. S6A, B).

Our original pDonor contained 2.2 kb of cargo DNA, and we next tested the effect of donor length on ShCAST activity ranging from 500 bp to 10 kb. We observed a 2-fold higher insertion rate with a 500 bp donor, and a similar rate of insertions with 10 kb of payload compared to the original pDonor (Fig. S6C). We were unable to detect re-joined pDonor backbone during transposition in *E. coli* (Fig. S6D, E), suggesting that a linear donor backbone is formed, and not a rejoined product, consistent with the known reaction products of canonical Tn7 (30, 33). Finally, we investigated the requirement of the LE and RE transposon ends sequences contained in pDonor for transposition. Removal of all flanking genomic sequence or the 5 bp duplicated target sites had little effect on insertion frequency, and ShCAST tolerated truncations of LE and RE to 113 bp and 155 bp, respectively (Fig. S7A). Removal of additional donor sequence completely abolished transposase activity, consistent with the loss of predicted Tn7 TnsB-like binding motifs (Fig. S7B, C).

In vitro reconstitution of ShCAST

Although our data strongly suggested that ShCAST mediates RNA-guided DNA insertion, to exclude the requirement of additional host factors, we next sought to reconstitute the reaction in vitro. We purified all four ShCAST proteins (Fig. S8A) and performed in vitro reactions using pDonor, pTarget, and purified RNA (Fig. 3A). Addition of all four protein components, crRNA, and tracrRNA resulted in DNA insertions detected by both LE and RE junction PCRs, as did reactions containing the four protein components and sgRNA (Fig. 3B). The truncated tracrRNA variant 5 was also able to support DNA-insertion in vitro, in contrast with the activity observed in *E. coli*. ShCAST-catalyzed transposition in vitro occurred between 37–50°C and depended on ATP and Mg^{2+} (Fig. S8B, C). To confirm that in vitro insertions are in fact targeted, we performed reactions with target plasmids containing a GGTT PAM, an AACC PAM, and a scrambled non-target sequence, and could only detect DNA insertions into the GGTT PAM substrate with the target sequence (Fig. 3C). In vitro DNA transposition depended on all four CAST proteins, although we identified weak but detectable insertions in the absence of tniQ (Fig. 3D).

Consistent with the predicted lack of nuclease activity of Cas12k, we were unable to detect DNA cleavage in the presence of Cas12k and sgRNA across a range of buffer conditions (Fig. S8D). To determine whether other CRISPR-Cas effectors could also stimulate DNA transposition, we performed reactions with tnsB, tnsC, and tniQ, along with dCas9 and a sgRNA targeting the same GGTT PAM substrate. We were unable to detect any insertions following dCas9 incubation (Fig. 3E), indicating that the function of Cas12k is not merely DNA binding, and that DNA transposition by CAST does not simply occur at R-loop structures. As final validation, we transformed in vitro reaction products into *E. coli* and

performed Sanger sequencing to determine the LE and RE junctions. All sequenced donors were located in pTarget, 60–66 bp from the PAM, and containing duplicated 5-bp insertion sites, demonstrating complete reconstitution of ShCAST with purified components.

ShCAST mediates efficient and precise genome insertions in *E. coli*

To test whether ShCAST could be reprogrammed as a DNA insertion tool, we selected 48 targets in the *E. coli* genome and co-transformed pDonor and pHelper plasmids expressing targeting sgRNAs (Fig. 4A). We detected insertions by PCR at 29 out of the 48 sites (60.4%) and selected 10 sites for additional validation (Fig. S9A). We performed ddPCR to quantitate insertion frequency after 16 h and measured rates up to 80% at PSP42 and PSP49 (Fig. 4B). This high efficiency of insertion was surprising given that insertion events were not selected for by antibiotic resistance, so we performed PCR of target sites to confirm. We detected the 2.5 kb insertion product in the transformed population (Fig. 4C). Re-streaking transformed *E. coli* yielded pure single colonies, the majority of which contained the targeted insertion (Fig. S9B) and the high efficiency of integration was maintained with a variety of donor DNA lengths (Fig. S9C). We analyzed the position of genome insertions by targeted deep sequencing of the LE and RE junctions and observed insertions within the 60–66 bp window at all 10 sites (Fig. 4D, Fig. S10A).

We next assayed the specificity of RNA-guided DNA transposition. We performed unbiased sequencing of donor insertion sites following Tn5 tagmentation of gDNA. We observed one prominent insertion site in each sample, which mapped to the target site, and contained more than 50% of the total insertion reads (Fig. 4E). The remaining off-target reads were scattered across the genome and analysis of the top off-target sites revealed strong overlap between samples revealing that these events are independent of the guide sequence (Fig. S10B, Table S5). Top off-target sites were located near highly expressed loci such as ribosomal genes, serine-tRNA ligase, and enolase, although insertion frequency in these regions were all less than 1% of the on-target site (Table S5). We identified one potential RNA-guided off-target following targeting of PSP42 which contains 4 mismatches to the guide sequence (Fig. S10C). Together, these results indicate that ShCAST robustly and precisely inserts DNA into the target site.

Discussion

Here we demonstrate that CRISPR-Cas systems associated with Tn7-like transposon mediate RNA-guided DNA transposition and elucidate its mechanism. ShCAST mediates unidirectional insertions in a narrow window downstream of the target and inhibits repeated insertions into a single target site (Fig. 5). Although ShCAST and AcCAST exhibit similar PAM preferences, one notable difference is that their respective positions of insertion, relative to the PAM, differ by 10–11 bp, which roughly corresponds to one turn of DNA. Deeper exploration of microbial genomes is expected to uncover CAST systems with a range of diverse properties including targeting preference and activity across different conditions.

Targeted DNA insertion by ShCAST results in the incorporation of LE and RE elements and is therefore not a scarless integration method. One potential generalizable strategy for the use of CAST in the therapeutic context would be to insert corrected exons into the intron before the mutated exon (Fig. S11). CAST could also be used to insert transgenes into "safe harbor" loci (34) or downstream of endogenous promoters so that the expression of transgenes of interest can benefit from endogenous gene regulation.

Further studies should improve our understanding of the function of each transposase subunit in the CAST complex, notably, TniQ, which contains a predicted DNA-binding domain. We originally hypothesized that TniQ is analogous to the site-specific DNA-binding protein TnsD of Tn7, and therefore, might be dispensable for RNA-guided insertions; however, we observed that TniQ is required for RNA-guided insertions in *E. coli*. The observation that *in vitro* transposition can occur to a limited extent in the absence of TniQ is compatible with a model in which TniQ facilitates the formation of the CAST complex and is not essential for catalytic function, therefore, it might be possible to engineer simplified versions of CAST systems without TniQ or with fragments of TniQ.

Our analysis indicated that ShCAST is specific, but under overexpression conditions can integrate at non-targeted sites in the *E. coli* genome via Cas12k-independent mechanisms, and this guide-independent integration seems to favor highly expressed genes. We also observed non-targeted insertions into pHelper in *E. coli* which was independent of Cas12k (Fig. S12) and reminiscent of TnsE-mediated Tn7 insertions into conjugal plasmids and replicating DNA (27). Future protein engineering of the transposase components could improve the targeting specificity of CAST systems.

In summary, this work identifies a function for CRISPR-Cas systems beyond adaptive immunity that does not require Cas nuclease activity and provides a strategy for targeted insertion of DNA without engaging homologous recombination pathways, with a particularly exciting potential for genome editing in eukaryotic cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank R. Macrae for critical reading of the manuscript, and the entire Zhang laboratory for support and advice. We thank F. Chen for imaging *Scytonema hofmanni*.

Funding: J.S. is supported by the Human Frontier Science Program. F.Z. is a New York Stem Cell Foundation–Robertson Investigator. F.Z. is supported by NIH grants (1R01-HG009761, 1R01-MH110049, and 1DP1-HL141201); the Howard Hughes Medical Institute; the New York Stem Cell and Mathers Foundations; the Poitras Center for Affective Disorders Research at MIT; the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT; and J. and P. Poitras.

References

1. Barrangou R, Horvath P, A decade of discovery: CRISPR functions and applications. Nat Microbiol 2, 17092 (2017). [PubMed: 28581505]

2. Mohanraju P et al., Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353, aad5147 (2016). [PubMed: 27493190]

3. Marraffini LA, CRISPR-Cas immunity in prokaryotes. Nature 526, 55–61 (2015). [PubMed: 26432244]

4. Cong L et al., Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823 (2013). [PubMed: 23287718]

5. Mali P et al., RNA-Guided Human Genome Engineering via Cas9. Science 339, 823–826 (2013). [PubMed: 23287722]

6. Zetsche B et al., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015). [PubMed: 26422227]

7. Strecker J et al., Engineering of CRISPR-Cas12b for human genome editing. Nat Commun 10, 212 (2019). [PubMed: 30670702]

8. Teng F et al., Repurposing CRISPR-Cas12b for mammalian genome engineering. Cell discovery 4, 63 (2018). [PubMed: 30510770]

9. Liu J-J et al., CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature 566, 218–223 (2019). [PubMed: 30718774]

10. Jasin M, Rothstein R, Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol 5, a012740 (2013). [PubMed: 24097900]

11. Schmid-Burgk JL, Honing K, Ebert TS, Hornung V, CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat Commun 7, 12338 (2016). [PubMed: 27465542]

12. Suzuki K et al., In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature 540, 144–149 (2016). [PubMed: 27851729]

13. Qi LS et al., Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152, 1173–1183 (2013). [PubMed: 23452860]

14. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424 (2016). [PubMed: 27096365]

15. Gaudelli NM et al., Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). [PubMed: 29160308]

16. Nishida K et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, aaf8729–aaf8729 (2016). [PubMed: 27492474]

17. Guynet C et al., In vitro reconstitution of a single-stranded transposition mechanism of IS608. Mol Cell 29, 302–312 (2008). [PubMed: 18280236]

18. Barabas O et al., Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. Cell 132, 208–220 (2008). [PubMed: 18243097]

19. Makarova KS et al., An updated evolutionary classification of CRISPR-Cas systems. Nature Reviews Microbiology 13, 722–736 (2015). [PubMed: 26411297]

20. Shmakov S et al., Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15, 169–182 (2017). [PubMed: 28111461]

21. Peters JE, Makarova KS, Shmakov S, Koonin EV, Recruitment of CRISPR-Cas systems by Tn7-like transposons. P Natl Acad Sci USA 114, E7358–E7366 (2017).

22. Faure G et al., CRISPR–Cas in mobile genetic elements: counter-defense and beyond. Nat Rev Microbiol in press, (2019).

23. Sarnovsky RJ, May EW, Craig NL, The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J 15, 6348–6361 (1996). [PubMed: 8947057]

24. Peters JE, Craig NL, Tn7: smarter than we thought. Nat Rev Mol Cell Biol 2, 806–814 (2001). [PubMed: 11715047]

25. Waddell CS, Craig NL, Tn7 transposition: recognition of the attTn7 target sequence. Proc Natl Acad Sci U S A 86, 3958–3962 (1989). [PubMed: 2542960]

26. Waddell CS, Craig NL, Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes Dev 2, 137–149 (1988). [PubMed: 2834269]

27. Peters JE, Craig NL, Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. Genes Dev 15, 737–747 (2001). [PubMed: 11274058]

28. Hou S et al., CRISPR-Cas systems in multicellular cyanobacteria. RNA Biol 16, 518–529 (2019). [PubMed: 29995583]

29. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740 (2009). [PubMed: 19246744]

30. Bainton R, Gamas P, Craig NL, Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell 65, 805–816 (1991). [PubMed: 1645619]

31. Skelding Z, Queen-Baker J, Craig NL, Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. EMBO J 22, 5904–5917 (2003). [PubMed: 14592987]

32. Stellwagen AE, Craig NL, Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. EMBO J 16, 6823–6834 (1997). [PubMed: 9362496]

33. Biery MC, Stewart FJ, Stellwagen AE, Raleigh EA, Craig NL, A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic Acids Res 28, 1067–1077 (2000). [PubMed: 10666445]

34. Sadelain M, Papapetrou EP, Bushman FD, Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer 12, 51–58 (2011). [PubMed: 22129804]

35. Li H, Durbin R, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]

36. Leenay RT et al., Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Molecular Cell 62, 137–147 (2016). [PubMed: 27041224]

37. Bainton RJ, Kubo KM, Feng JN, Craig NL, Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell 72, 931–943 (1993). [PubMed: 8384534]

38. Ton-Hoang B et al., Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. EMBO J 24, 3325–3338 (2005). [PubMed: 16163392]

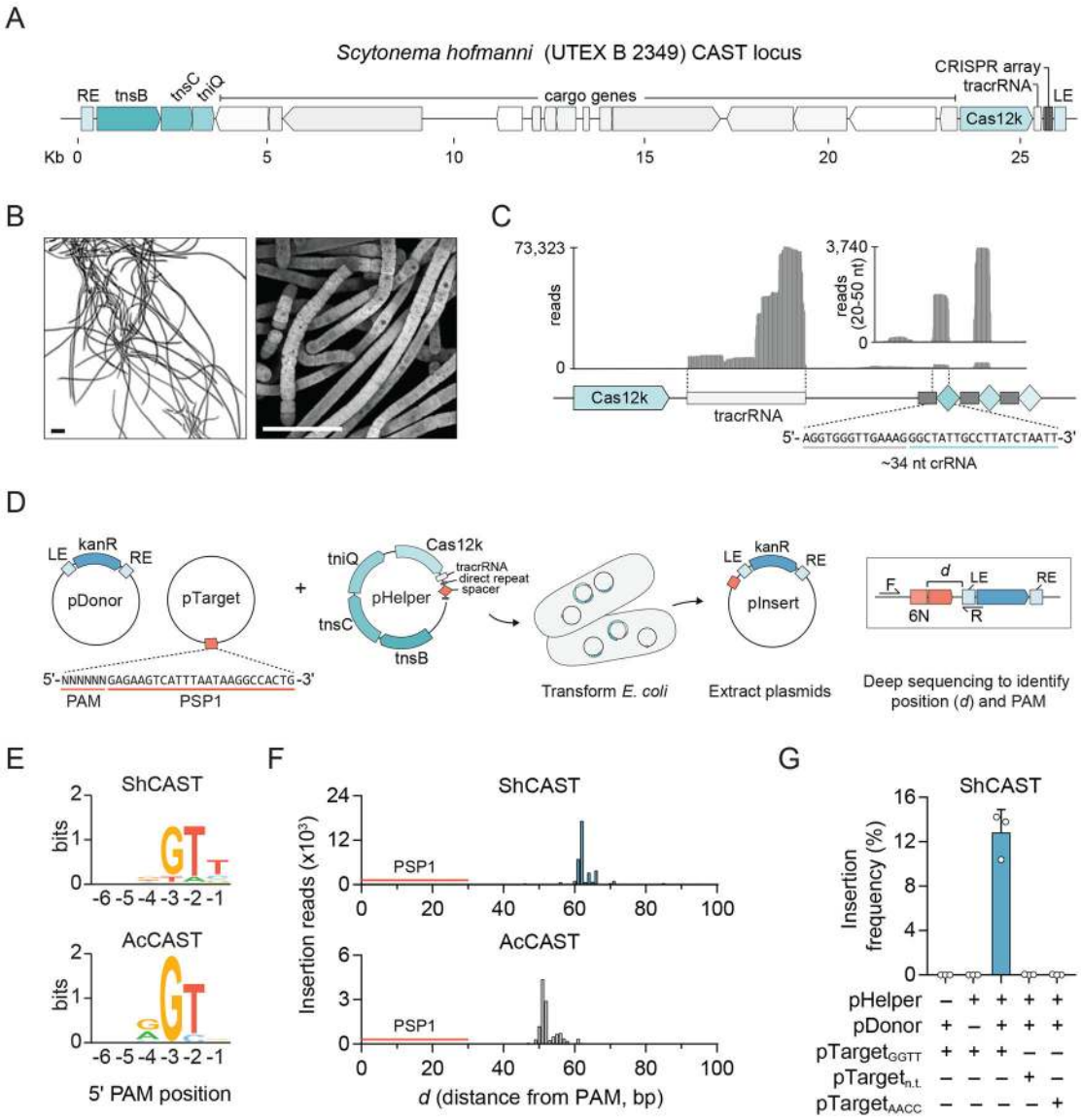

Figure 1. Targeting requirements for CRISPR-associated transposase (CAST) systems

A. Schematic of the *Scytonema hofmanni* CAST locus containing Tn7-like proteins, the CRISPR-Cas effector Cas12k, and a CRISPR array.

B. Fluorescent micrograph of the cyanobacteria *S. hofmanni*. Scale bar, 40 uM.

C. Alignment of small RNA-Seq reads from *S. hofmanni*. The location of the putative tracrRNA is marked.

D. Schematic of experiment to test CAST system activity in *E. coli*.

E. PAM motifs for insertions mediated by ShCAST and AcCAST.

F. ShCAST and AcCAST insertion positions identified by deep sequencing.

G. Insertion frequency of ShCAST system in *E. coli* with pTarget substrates as determined by ddPCR. Error bars represent s.d. from n=3 replicates.

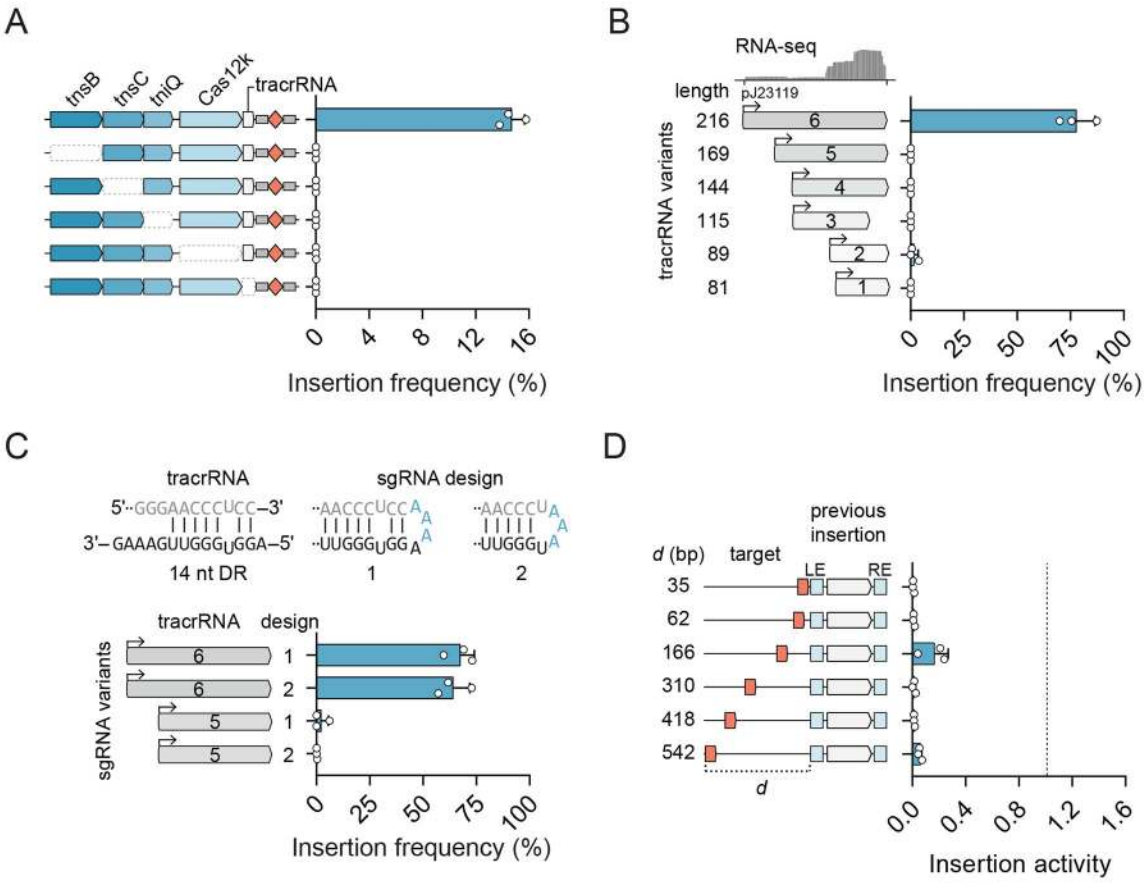

Figure 2. Genetic requirements for RNA-guided insertions

A. Genetic requirement of tnsB, tnsC, tniQ, Cas12k, and tracrRNA on insertion activity. Deleted components are indicated by a dashed outline.

B. Insertion activity of 6 tracrRNA variants expressed with the pJ23119 promoter.

C. Schematic of tracrRNA and crRNA base pairing and two sgRNA designs highlighting the linker sequence (blue).

D. Insertion activity into pTarget containing ShCAST transposon ends relative to activity into pTarget without previous insertion.

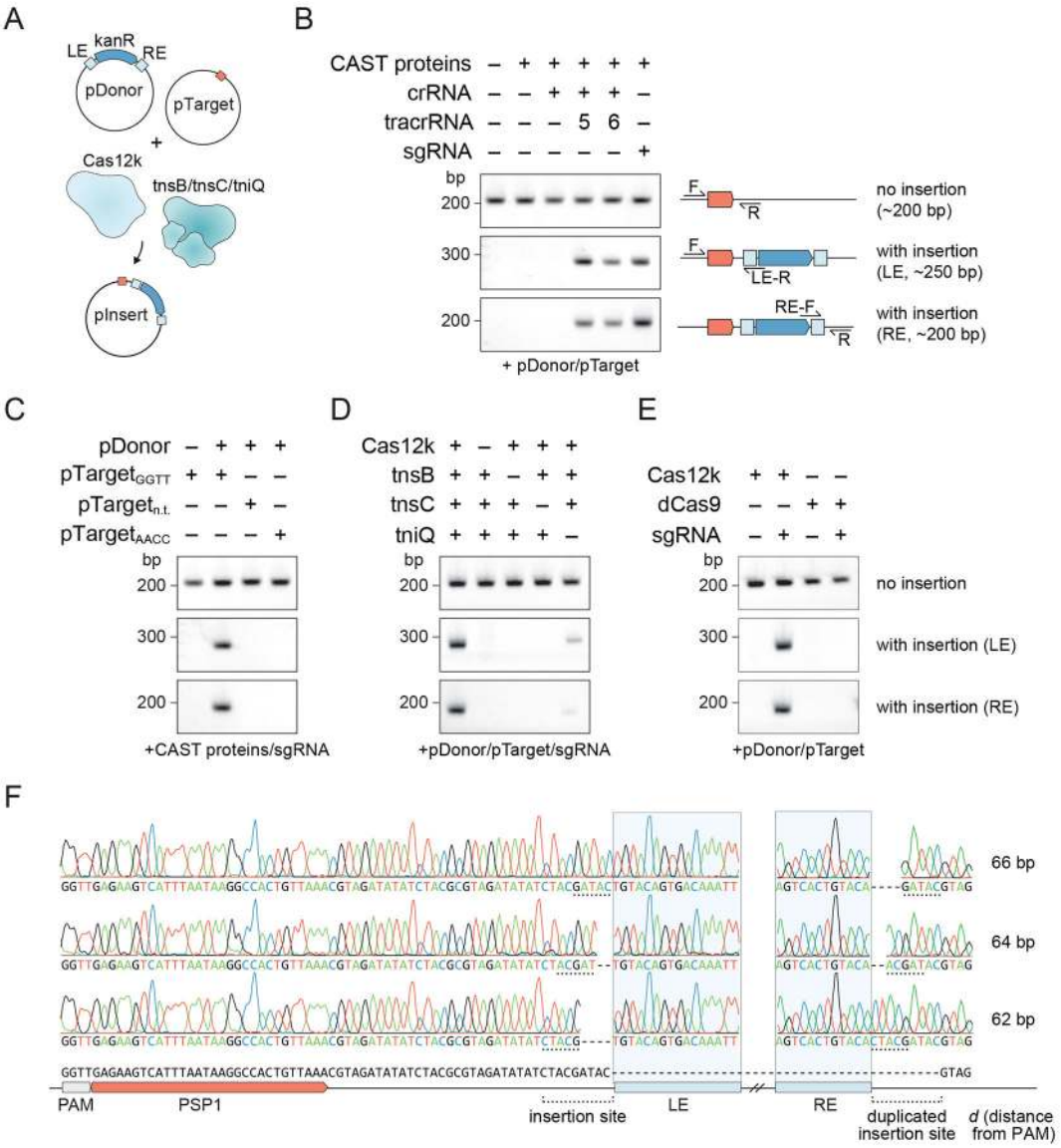

Figure 3. In vitro reconstitution of an RNA-guided transposase.

A. Schematic of in vitro transposition reactions with purified ShCAST proteins and plasmid donor and targets.

B. RNA requirements for in vitro transposition. pInsert was detected by PCR for LE and RE junctions. All reactions contained pDonor and pTarget. Schematics indicate the location of primers and the expected product sizes for all reactions.

C. Targeting specificity of ShCAST in vitro. All reactions contained ShCAST proteins and sgRNA.

D. Protein requirements for in vitro transposition. All reactions contained pDonor, pTarget, and sgRNA.

E. CRISPR-Cas effector requirements for in vitro transposition. All reactions contained ShCAST proteins, pDonor, and pTarget.

F. Chromatograms of pInsert reaction products following transformation and extraction from *E. coli*. LE and RE elements are highlight and the duplicated insertion sites denoted.
G. For all panels, ShCAST proteins were used at a final concentration of 50 nM, and n=3 replicates for all reactions were performed with a representative image shown.

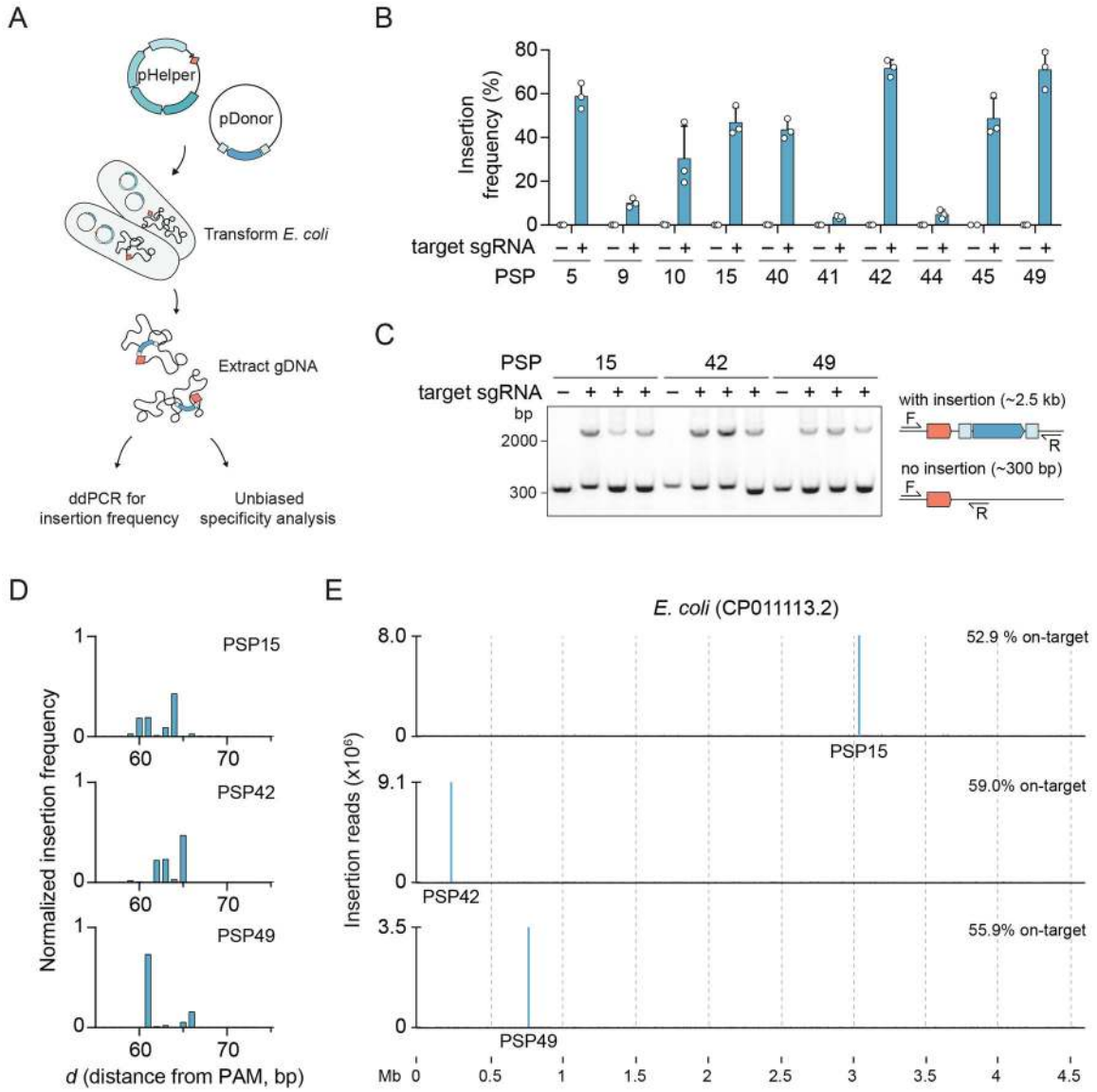

Figure 4. ShCAST mediates genome insertions in *E. coli*

A. Schematic of experiment to test for genome insertions in *E. coli*.

B. Insertion frequency at 10 tested protospacers following ShCAST transformation. Insertion frequency was determined by ddPCR on extracted genomic DNA. Error bars represent s.d. from n=3 replicates.

C. Flanking PCR of 3 tested protospacers in a population of *E. coli* following ShCAST transformation. Schematics indicate the location of primers and the expected product sizes.

D. Insertion site position as determined by deep sequencing following ShCAST transformation.

E. Insertion positions determined by unbiased donor detection. The location of each protospacer is annotated along with the percent of total donor reads that map to the target.

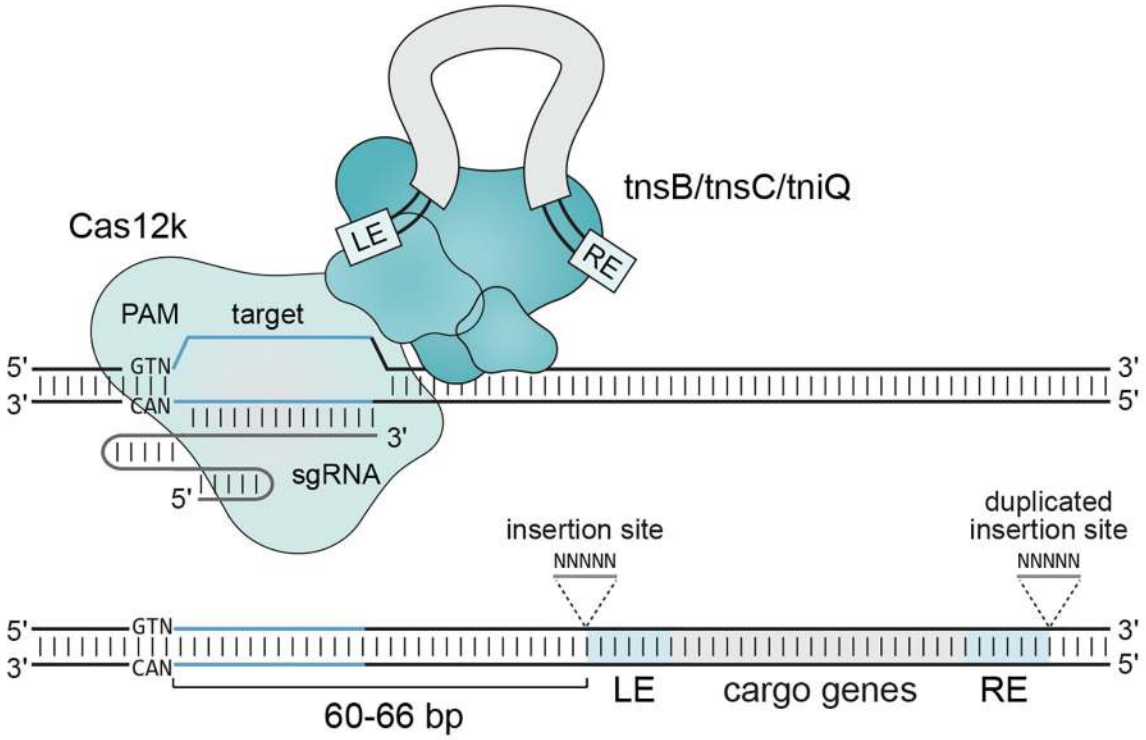

Figure 5. Model for RNA-guided DNA transposition

The ShCAST complex that consists of Cas12k, TnsB, TnsC, and TniQ mediates insertion of DNA 60–66 bp downstream of the PAM. Transposon LE and RE sequences along with any additional cargo genes are inserted into DNA resulting in the duplication of 5 bp insertion sites.