https://doi.org/10.1093/nar/gkab752

Orthogonal CRISPR-associated transposases for parallel and multiplexed chromosomal integration

Siqi Yang [1,2,†], Yiwen Zhang[1,2,†], Jiaqi Xu[3], Jiao Zhang[1,2], Jieze Zhang[4], Junjie Yang[1], Yu Jiang[5] and Sheng Yang [1,*]

[1]Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China, [2]College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China, [3]Hangzhou Global Scientific and Technological Innovation Centre, Zhejiang University, Hangzhou, China, [4]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA and [5]Shanghai Taoyusheng Biotechnology Co., Ltd, Shanghai 200032, China

Received May 05, 2021; Revised August 16, 2021; Editorial Decision August 17, 2021; Accepted August 21, 2021

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

ABSTRACT

Cell engineering is commonly limited to the serial manipulation of a single gene or locus. The recently discovered CRISPR-associated transposases (CASTs) could manipulate multiple sets of genes to achieve predetermined cell diversity, with orthogonal CASTs being able to manipulate them in parallel. Here, a novel CAST from *Pseudoalteromonas translucida* KMM520 (PtrCAST) was characterized without a protospacer adjacent motif (PAM) preference which can achieve a high insertion efficiency for larger cargo and multiplexed transposition and tolerate mismatches out of 4-nucleotide seed sequence. More importantly, PtrCAST operates orthogonally with CAST from *Vibrio cholerae* Tn6677 (VchCAST), though both belonging to type I-F3. The two CASTs were exclusively active on their respective mini-Tn substrate with their respective crRNAs that target the corresponding 5 and 2 loci in one *Escherichia coli* cell. The multiplexed orthogonal MUCICAT (MUlticopy Chromosomal Integration using CRISPR-Associated Transposases) is a powerful tool for cell programming and appears promising with applications in synthetic biology.

INTRODUCTION

Cell engineering often involves the interruption of a batch of genes (1,2) and the unfolding of an optimal dose ratio of it (3,4). However, previous genome editing method, at least in *Escherichia coli*, cannot edit more than four target sites with high efficiency (5). Gene integration is especially difficult and results in a large and time-consuming workload when constructing cells. Recently, a variety of CRISPR-associated transposases (CASTs) have been discovered and demonstrated to have the ability of RNA-guided DNA integration on bacterial chromosomes independently of homologous recombination (6–10). MUCICAT based on CAST from *Vibrio cholerae* Tn6677 (VchCAST) has also been demonstrated to have the ability of accelerated construction of optimized strains (2,11).

Currently, Vo *et al.* reported that VchCAST and type V-K CAST from *Scytonema hofmannii* PCC 7110 (ShoCAST) were orthogonal in that they recognize different transposon ends (12), but type V-K CASTs tend to cointegrate its donor plasmid (13–16) due to lack of protein TnsA activity, and have a relatively low integration efficiency on multiple targets (2). Type I-F3 CAST from *Aeromonas salmonicida* S44 (AsaCAST) exhibits RNA-guided DNA transposition activity when introduced to *E. coli,* but the integration efficiency of AsaCAST seems much lower than VchCAST (7). Highly active CASTs capable of MUCICAT need to be identified for the implementation of a parallel and multiplexed genome editing system with VchCAST.

Here, a type I-F3 CAST from *Pseudoalteromonas translucida* KMM520 (PtrCAST) is characterized, which shows no preference in PAM and performs multiplexed integration of large cargo genes with high efficiency. Its orthogonality with VchCAST targeting corresponding five and two loci was proven.

MATERIALS AND METHODS

Strain and medium

Escherichia coli BL21Star™ (DE3) was from Genscript (Nanjing) and cultured in LB (liquid or solid) medium at 37°C.

*To whom correspondence should be addressed. Tel: +86 21 54924173; Email: syang@sibs.ac.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Plasmid construction

The pQCascade/pDCascade and pTns/pTnsABCQ were synthesized by Genscript (Nanjing), and the repeat sequence of crRNA was inserted at the NcoI and BamHI sites of pQCascade/pDCascade, and the interval sequence was inserted at the BsaI site. pDonor with cargo genes (Cm, GFP, etc.) were connected through Gibson-Assembly to construct pDonor-cargo.

To compare the insertion efficiency of ShCAST (from *Scytonema hofmannii*), ShoCAST (from *Scytonema hofmannii* PCC 7110), VchCAST (from *Vibrio cholerae* Tn6677), AsaCAST (from *Aeromonas salmonicida* S44), AvCAST (from *Anabaena variabilis* ATCC 29413), PmcCAST (from *Peltigera membranacea cyanobiont* 210A) and PtrCAST in BL21(DE3). The functional plasmids reported in their respective articles were reconstructed. For ShCAST, the spacer (PSP49 as reported) was added in pHelper_ShCAST_sgRNA (6)(Addgene #127921) through the LguI sites and the origin of pDonor_ShCAST_kanR (6) (Addgene #127924) was replaced with the pSC101 variant. For ShoCAST, pSL1571 (12) was directly used to test the insertion efficiency in the sgRNA-31 targeting site. For VchCAST, pDonorVerA(Vch), pTns(Vch) and pQCascade(Vch)-crRNA4 with the tetracycline (Tet) promoter were used. For AsaCAST, pMTP170 (7)(Addgene #164263) was reconstructed to target the *lacZ* gene (spacer #6 as reported) and pDonorAsa derived from pMTP112 (7) (Addgene #164260) was reconstructed as well. They were used together with pMTP130 (7) (Addgene #162461) and pMTP140 (7) (Addgene #164262). For AvCAST, AvPSP31 was added on pMS4-pHelper (AvCAST)_entry (9) and the *tnsD* gene was deleted, and it was tested with pMS12-pDonor (AvCAST) (9). For PmcCAST, a new target was added on pMS21- pHelper (PmcCAST)_entry (9) and the *tnsD* gene was deleted, and it was tested with pMS29-pDonor (PmcCAST) (9). For PtrCAST, pDonorCm(Ptr), pTns(Ptr) and pQCascade(Ptr)-crRNA4 were used.

The plasmids used in the experiment are shown in Supplementary Table S1. The primers used to construct the plasmids in this experiment are shown in Supplementary Table S2.

Transposition experiments

Transposition experiments for CASTs include the transformation and induction of pDonor, pTns and pQCascade (or variants thereof). First, pDonor and pTns were co-transformed into BL21Star™(DE3) competent cells, and recovered on a double antibiotic LB-agar plate (100 μg/ml ampicillin and 100 μg/ml kanamycin). Then, a single colony was inoculated into liquid LB to make competent cells. pQCascade was introduced by electroporation, and the cells were resuscitated in fresh LB medium in the shaker at 37°C for 1 h. Then, the cells were spread on a triantibiotic LB agar plate (100 μg/ml ampicillin,100 μg/ml kanamycin and 50 μg/ml streptomycin). In the alternative experiment, pDonor and pQCasTns or a single pEffector were transformed into competent cells and grown overnight at 37°C for 16 h. After that, hundreds of colonies were scraped off from the plate. Some of them were re-suspended in fresh LB medium, and then re-spread on a LB-agar plate with 100 ng/ml anhydrotetracycline (aTC) to induce protein expression. Cells were cultured at 37°C for another 16 h, and the films formed were scraped off and resuspended in LB medium. The cells were re-diluted and then spread on a plate containing 1000 ng/ml aTC. The colonies were grown overnight at 37°C and then colony PCR or qPCR was performed for copy number identification.

Similarly, transposition experiments for CASTs also include the transformation and induction of a set of pDonor and pQCasTns (or variants thereof) or a pEffector (or variants thereof). The difference between tri-plasmid and dual-plasmid CASTs is that the latter only requires one transformation and then will be cultured on a LB-agar plate containing their respective antibiotic. Transposition experiments for orthogonal CASTs include the transformation and induction of two sets of pDonor and pQCasTns (or variants thereof).

The methods of transformation and induction are consistent with their respective articles in the experiments comparing the insertion efficiency of ShCAST, ShoCAST, AsaCAST, AvCAST and PmcCAST.

Transposition efficiency analysis by colony PCR

The single colony on the plate was directly picked up and resuspended in 10 μl sterile water as a template. The PCR reaction consisted of 200 μM dNTP and 0.5 μM primers. 30 cycles of thermal cycling were done, and the annealing step was at 57°C. Paired primers contained either upstream and downstream primers of the target site, or one genome specific primer and one transposon specific primer to detect all possible integration directions. The PCR amplification products were separated by 2% agarose gel electrophoresis, and the negative control samples were always analyzed in parallel with the experimental samples to identify non-specific bands.

The primers used for verification in this experiment are shown in Supplementary Table S2.

Transposition efficiency analysis by qPCR

Specific primers pairs were designed for each crRNA, with one located in the upstream of protospacer and the other in the interior of left end (LE) or right end (RE) to amplify an ∼200 bp fragment of possible integration in both orientations. A pair of genome-specific primers were designed to amplify the *E.coli* reference gene *rssA* for normalization purposes. qPCR reaction (20 μl) contained 10 μl 2 × SYBR Green qPCR Mix (With ROX) (SparkJade), 3 μl H_2O, 2 μl 10 mM primer and 1 μl genomic DNA (diluted to 1 ng/μl) extracted from scraped colonies using a TIANamp Bacteria DNA kit (TIANGEN) with RNase digestion. Reactions were prepared in a 96-well white PCR plate (BioRad) and measured on the CFX384 Real-Time PCR Detection System (BioRad) using the following thermal cycling parameters: polymerase activation and DNA denaturation (95°C for 2 min), 40 amplification cycles (95°C for 10 s, 55°C for 20 s) and the final melt-curve analysis (59–95°C in increments of 0.5°C/5 s).

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

By testing our primer pairs with each of these samples diluted across three orders of magnitude, and then determining the resulting Cq values and PCR efficiencies, the experimental and reference amplicons were amplified with similar efficiencies, and the primer pairs selectively amplified the intended transposition product (Supplementary Figure S1).

Each biological sample was analyzed by three parallel reactions: one reaction containing a primer pair for the *E. coli* reference gene, a second reaction containing a primer pair for detecting integration in the LR orientation, a third reaction containing a primer pair for detecting integration in the RL orientation.

Transposition efficiency for each orientation is then calculated as $2^{\Delta Cq}$, in which ΔCq is the Cq difference between the experimental reaction and the control reaction (8). Total transposition efficiency for a given experiment is calculated as the sum of transposition efficiencies for both orientations. All measurements presented in the text and figures were determined from three independent biological replicates.

The primers used for verification in this experiment are shown in Supplementary Table S2.

PAM screens

A randomized target PAM library was generated using synthesized ssDNA oligonucleotides (Tsingke) with six randomized bases upstream of the protospacer (Supplementary Table S3). Oligonucleotides were used to generate a PCR product for subsequent Gibson assembly into pTarget_CAST (Addgene #127926). Gibson products were transformed into DH5α competent cells, recovered for 1 h, and spread on chloramphenicol plates. Cells were harvested 12 h later and verified. Plasmid DNA was extracted using a ToloPrep kit (Tolobio). 500 ng of the library target DNA was co-transformed with 500 ng of both pTns/pTnsABCQ and pDonor into BL21Star™(DE3) competent cells. Cells were culture for 1 h and spread on plates containing 100 μg/ml ampicillin, 100 μg/ml kanamycin and 25 μg/ml chloramphenicol. Cells were harvested 12 h later and made into competent cells. 500 ng of pQCascade/pDCascade was transformed into the cells, recovered for 1 h and spread on plates with 100 μg/ml ampicillin, 100 μg/ml kanamycin, 25 μg/ml chloramphenicol, and 50 μg/ml streptomycin. Insertion products containing the randomized PAM sequence were amplified using primers listed in Supplementary Table S2 and sequenced using a NovaSeq 6000 (Illumina).

The 6 bp random sequence between the plasmid backbone sequence 'TAGTATCTACGATACGTAGTATCT ACGATACGTAG' and crRNA sequence 'GAGAAGTC ATTTAATAAGGCCACTGTTAAACG' from the Illumina reads were picked up using seqkit (17) with custom scripts.

The sequence logo drawing was made by WebLogo: https://weblogo.berkeley.edu/logo.cgi (18).

To obtain the distance distribution between the target site and transposon integration site, the sequences between crRNA sequence 'GAGAAGTCATTTAATAAGGCCACT GTTAAACG' and transposon end sequence 'TGTTGT TTGAAGTATAAGTTGGCATAAGTACAAACGA' or 'TGTTGTTTGAAGTATAAGTTGACATATC' from the Illumina reads were picked up, and the length distribution of the sequences were counted. The counts of each PAM are listed in Supplementary Table S4.

Next-generation sequencing, data analysis

To determine the integration sites and their distribution from the whole genome sequencing (WGS) data, split-read mapping approaches were performed (using custom Bash scripts) as previously reported (2).

RESULTS

The highly efficient RNA-guided transposition of PtrCAST

Among the five types of CASTs, the type IV, and I-E CASTs are missing related transposition proteins (19,20), while V-K CASTs generate cointegrate products of the donor plasmid (13–16). Therefore, CASTs of interest with complete components and high efficiency were screened from type I-B and I-F CASTs for activity characterization (21,22) (sequences and accession numbers are listed in Supplementary Table S5). Among them, the highly active CASTs were further tested for orthogonality with VchCAST.

The type I-F3 CAST elements from *Vibrio natriegens* ATCC 14048 was amplified from genomic DNA. The type I-F3 CAST elements derived from *Pseudoaltermonas translucida* KMM520, the type I-B1 CAST elements derived from *Anabaena variabilis* ATCC 29413 and type I-B2 CAST elements derived from *Cyanothece* sp. PCC 8801 were synthesized and optimized according to the codon bias of *E. coli*. The type I-F3 CASTs were constructed on three plasmids pTns, pQCascade, and pDonor (Figure 1A, Supplementary Table S1): pTns expressed Tn7-like transposon encoded *tnsA*, *tnsB* and *tnsC* genes with the tetracycline inducible (Tet) promoters; pQCascade expressed the *tniQ* gene and CRISPR-related elements, including a crRNA or a CRISPR array (unique spacers interspaced between repeats), *cas6*, *cas7* and *cas5/8* genes with two Tet promoters. pDonor contains the transposon ends (Left end (LE) and Right end (RE)) and the cargo gene (Figure 1A). The type I-B CASTs were constructed on three plasmids in a similar form (Figure 1B, Supplementary Table S1). Since type I-B CAST has Tn7-like *tnsD* and *tniQ* genes, *tniQ* gene was also expressed on pTnsABCQ in addition to *tnsA*, *tnsB* and *tnsC* genes. *tnsD* gene, crRNA or CRISPR array, *cas5*, *cas6*, *cas7* and *cas8* genes were constructed on pDCascade (Figure 1B).

In order to test the transposition activity of type I-F3 CAST, the three plasmids were transformed into *E. coli* BL21Star™(DE3). The cargo on pDonor is a chloramphenicol gene without a promoter, and pQCascade(Ptr) carries the crRNA3 targeting *lacZ* loci on the genome (one is in the *lacZYA* operon, the other is at the λDE3 prophage)(Figure 1C), flanked by a 5′-CC-3′ protospacer adjacent motif (PAM). The crRNA1 targeting *lacZ* on the genome (Supplementary Figure S2A) carried on pQCascade (Vna) is flanked by a 5′-TC-3′ PAM. After being induced by aTC, colony PCR was performed with a/a' and b/b' primers to identify the insertion of PtrCAST, and i/i' primers were used to identify the insertion of VnaCAST.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

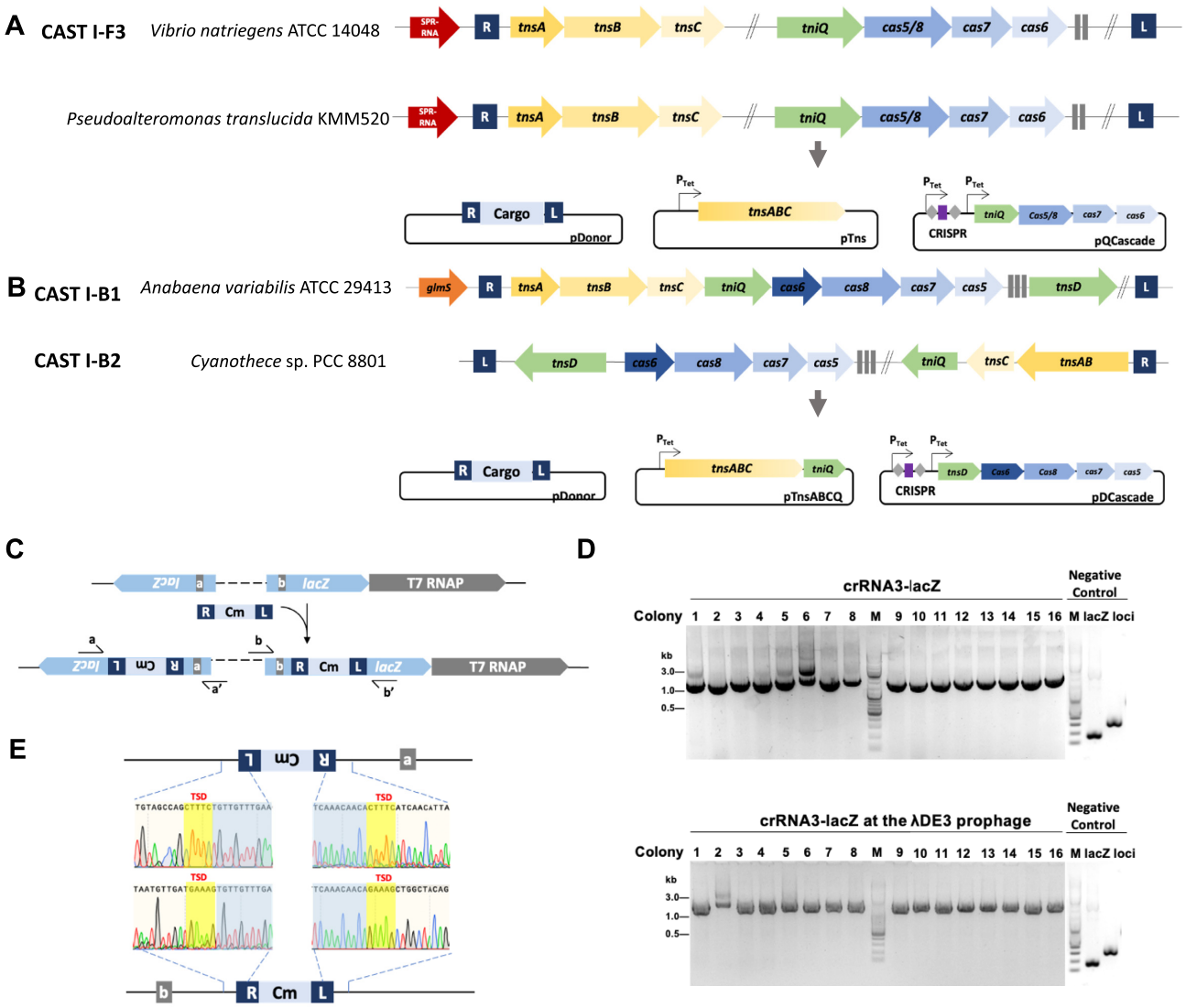


Figure 1. Characterization the RNA-guided DNA integration using CRISPR-associated transposases (CASTs). (**A**) Type I-F3 CASTs from *Pseudoalteromonas translucida* KMM520 and *Vibrio natriegens* ATCC 14048. Plasmid schematics for transposition experiments: pDonor includes RE-cargo-LE; pTns includes Tet promoter, *tnsA*, *tnsB*, and *tnsC* genes; pQCascade includes two Tet promoters, crRNA which is comprised of two repeats (grey diamonds) and a single spacer (purple rectangle), *tniQ*, *cas5/8*, *cas7*, and *cas6* genes. (**B**) Type I-B CASTs from *Anabaena variabilis* ATCC 29413 and *Cyanothece* sp. PCC 8801. Plasmid schematics for transposition experiments: pDonor includes RE-cargo-LE; pTnsABCQ includes Tet promoter, *tnsA*, *tnsB* (or *tnsAB*), *tnsC*, and *tniQ* genes; pDCascade includes two Tet promoters, crRNA which is comprised of two repeats (grey diamonds) and a single spacer (purple rectangle), *tnsD*, *cas8*, *cas7*, *cas6,* and *cas5* genes. (**C**) Genomic loci on BL21Star™(DE3) targeted by crRNA3, and the PCR primer pairs a/a'and b/b'to selectively amplify them. The PAMs and target sites are in grey, together. (**D**) Colony PCR-based analysis of transposition with crRNA3, resolved by agarose gel electrophoresis. The left band in negative control is the *lacZ* locus and the right band is the *lacZ* locus at the λDE3 prophage. (**E**) Sanger sequencing chromatograms for upstream and downstream junctions of genomically integrated transposons from experiments with crRNA3. TSD is the abbreviation for target site duplication. a is the *lacZ* locus and b is the *lacZ* locus at the λDE3 prophage.

Finally, PtrCAST is capable of RNA-guided DNA integration, with 16 positive bands from 16 colonies (Figure 1D) but no insertion colony was detected in VnaCAST (Supplementary Figure S2B). The expected insertion for PtrCAST is a single band (~1.6 kb) compared with the negative control. The lane 6 in Figure 1D (crRNA-lacZ part) has an additional band (~3.2 kb) that represents two copies of the RE-GFP-LE verified in Supplementary Figure S2C.

The inserted colonies of PtrCAST were sequenced next and the sequencing peak diagram at the joint for one of the colonies shows the target site duplication (TSD) on both a and b loci (crRNA3 targeting loci), which also proves RNA-guided DNA transposition activity of PtrCAST because TSD is the feature of Tn7 transposition products (Figure 1E).

Since the PAM of Type I-B CAST is unknown, PAM screens on Type I-B1 CAST from *Anabaena variabilis* ATCC 29413 and Type I-B2 CAST from *Cyanothece* sp. PCC 8801 were performed. A randomized target PAM library was generated using synthesized ssDNA oligonucleotides with 6 randomized bases upstream of the proto-

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

spacer. No insertions were detected through PCR (Supplementary Figure S2D, E).

Characterizing PAM sequences, integration orientation, seed sequence and genome fidelity of PtrCAST

Before utilization of the viable PtrCAST, it's necessary to obtain more knowledge on its features such as PAM preferences, insertion orientation, seed sequence and genome fidelity. To characterize whether PtrCAST has preference to specific PAMs, 21 crRNAs tiled along the *lacZ* gene were tested, covering all 16 PAMs (5′-CA-3′,5′-CG-3′, 5′-AT-3′, 5′-AA-3′, 5′-AG-3′, 5′-GG-3′ were tested twice with two crRNAs and 5′-GC-3′ was tested three times with three crRNAs) (Supplementary Table S3). The insertion efficiency of each crRNA was determined by qPCR, the insertion efficiency of 5′-CC-3′ was normalized to 100% the standardization (LR direction + RL direction), and the relative integration efficiencies of the remaining PAMs were calculated. All 16 kinds of PAMs could be recognized by PtrCAST (Figure 2A), but 4 crRNAs flanking by 5′-AA-3′, 5′-AG-3′, 5′-AT-3′ and 5′-GG-3′ had no integration which may be due to the low binding efficiency of the selected crRNAs (Supplementary Figure S3A). The preference of insertion orientation can be known by testing the insertion efficiency of 16 PAMs determined by qPCR. The insertion efficiency in the LR direction from 16 crRNAs was only 10^{-6}–10^{-1} compared to an efficiency of in the RL direction (Figure 2B).

To avoid the selection bias among crRNAs in the above method, PAM screens with high throughput were performed as follows. The pTarget including a 5′-NNNNNN-3′ sequence as the target for crRNA-guided targeting was constructed and transformed into *E. coli* BL21Star™ (DE3) together with pTns, pQCascade (including the crRNA that targets pTarget), and pDonorCm. All plasmids in the induced *E.coli* were extracted and analyzed by PCR with pTarget specific primer F and RE/LE specific primers R/R'. The PCR products were analyzed by next-generation sequencing (NGS) (Figure 2C). To analyze the processed data, Weblogo (18,23) was used to show the results that PtrCAST has a low degree of PAM conservation and a relative preference for CG PAM in RL orientation (Figure 2D) and obvious conservation for CT PAM in LR orientation (Supplementary Figure S3B). The low conservative degree of PAM sequence in RL orientation is more representative because of the strong insertion preference (Figure 2B) of it, indicating that PtrCAST is PAM-free, and the low conservation of CG PAM is consistent with Figure 2A. CRISPR-associated systems usually recognize specific PAMs, but PtrCAST seems to have no PAM preference which makes target selection more flexible in genome editing like NgAgo (24). From the PAM screens, 86.7% of the integrations are located 48 bp from the 3′ of the target site and 8.7% of the integrations are located 51 bp from the 3′ of the target site (Figure 2E).

To probe the tolerance of transposition to spacer mismatches, consecutive blocks of 4-nucleotide mismatches along the spacer of crRNA4 were tested (Figure 2F). Mismatches within the 4-nucleotide seed sequence severely reduced transposition efficiency, which is different from the 8-nucleotide seed sequence of VchCAST. To characterize the fidelity of PtrCAST on a genome-wide scale, next-generation sequencing was performed and prominent insertion site(s) were observed in each sample and off-target integrations (15.8% for crRNA3 and 9.7% for crRNA4 guided DNA transposition) elsewhere on the genome (Supplementary Figure S3C and Supplementary Table S6), which is more frequent than VchCAST (with less than 1% off-target integration rate) (2,8). Top off-target sites were in *pepQ* gene for crRNA3 (9.8%) and *ldhA* gene for crRNA4 (4.3%). Two potential RNA-guided off-target integrations were observed which contain 13/32 and 11/32 mismatches respectively to the guide sequence (Supplementary Figure S3D). The more off-target events compared with that of VchCAST may be the results of the more flexible PAM and guide sequence tolerance of PtrCAST.

PtrCAST is capable of efficiently large or multiplexed insertions

VchCAST can integrate 0.98–10 kb cargo genes in BL21(DE3), but as the gene size increases, the efficiency gradually decreases (12). After inducing at 37°C, the integration efficiency of 10 kb is less than 25%. The integration efficiency can be increased to about 100% by lowering the temperature to 30°C and extending the integration cycle (12). The insertion efficiencies of 1.3, 4.3 and 15.4 kb cargo genes of PtrCAST in BL21Star™(DE3) after induced at 37°C overnight were all 100% (8 of 8, 8 of 8, 16 of 16 colonies are positive respectively) (Figure 2G, Supplementary Figure S4A–C). To rule the difference between strains out, the same 15.4 kb cargo for VchCAST was constructed and made transposition experiments in BL21Star™(DE3). Consequently, the insertion efficiency of VchCAST for large cargo is similar to PtrCAST in BL21Star™(DE3) (Supplementary Figure S4D) indicating that VchCAST has higher insertion efficiency in BL21Star™(DE3).

To test the ability for PtrCAST targeting multiple sites, a CRISPR array which has eight diverse spacers was constructed on the pQCascade-array8 (Figure 3A). After a round of transposition, 2–8 copies of strains were obtained at one time, of which strains with eight copies accounted for 12.5% of 16 colonies (Supplementary Figure S5). Zhang *et al.* used VchCAST to target eight sites in BL21 (DE3), which requires three passages to obtain strains with eight copies (2). Consequently, the efficiency of PtrCAST in targeting multiple sites is slightly higher than that of VchCAST (Figure 3B, Supplementary Figure S5).

Since the two natural repeat sequences which interspacing the targeting spacer on the chromosome of *Pseudoalteromonas translucida* KMM520 have 3 bp differences, they were named repeat 1(R1) and repeat 2(R2) (Figure 3C). The crossover method of R1 and R2 was used when designing the CRISPR array which has 8 diverse spacers, namely R1-S1-R2-S2-R1-S3-R2-S4-R1-S5-R2-S6-R1-S7-R2-S8-R1 (Figure 3A). Petassi *et al*. have studied the crRNA processing form of type I-F3 CASTs (7), so in accordance with it, PtrCAST can process two crRNA forms that are not identical, named crRNA-AAGUU and crRNA-GAAAU (Figure 3C). Both crRNA-AAGUU and crRNA-GAAAU are active forms confirmed by the CRISPR array which has eight diverse spacers. In addition, the combining

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

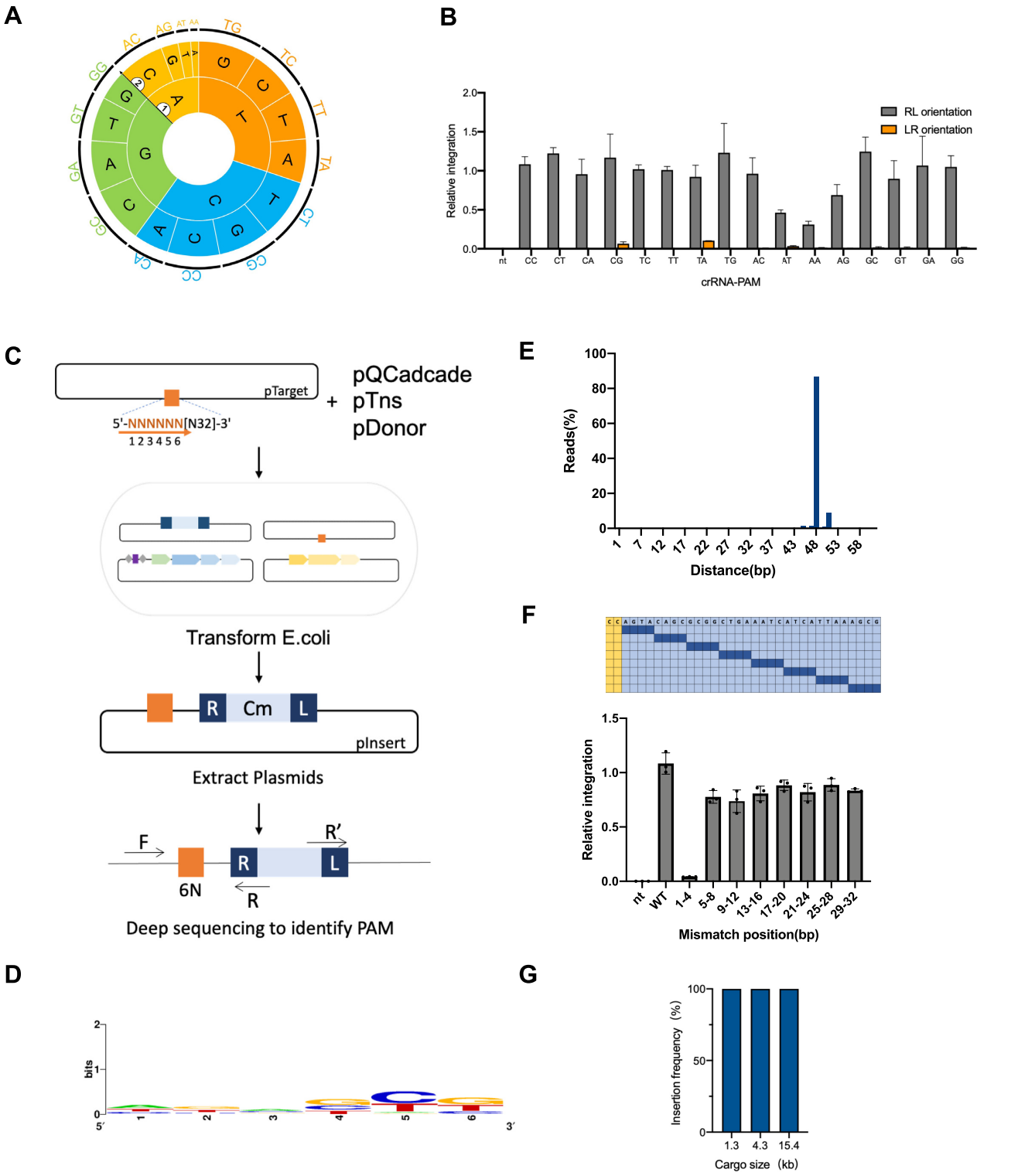


Figure 2. PAMs for type I-F3 CAST from *Pseudoalteromonas translucida* KMM520 are not conservative, and it can insert large cargo. (**A**) PAMs wheels for PtrCAST RNA-guided insertions identified by qPCR. The arrow in the wheel illustrates the orientation of each base. The area of a sector of the ring for one base at one position represents its frequency at this position. The insertion efficiencies of 5′-CA-3′,5′-CG-3′, 5′-AT-3′, 5′-AA-3′, 5′-AG-3′and 5′-GG-3′are the average of two crRNAs; the insertion efficiency of 5′-GC-3′is the average of three crRNAs. (**B**) Alternative integration orientations for 16 PAMs. Efficiencies are detected by qPCR. LR and RL denote transposition products in which the transposon left end and right end are proximal to the target site, respectively. (**C**) Schematic of PAM screens for PtrCAST. The orange arrow indicates each position of the PAM sequence. (**D**) PAMs for PtrCAST RNA-guided insertions in RL orientation identified by Next Generation Sequencing (NGS). Drawing is made by WebLogo: https://weblogo.berkeley.edu/logo.cgi. The position is consistent with the orange arrow in C. (**E**) NGS analysis of the distance between the Cascade target site and transposon integration site, determined by PAM screens. (**F**) crRNAs were mutated in 4-nucleotide blocks to introduce spacer-target DNA mismatches(top), and the resulting integration efficiencies were determined by qPCR (bottom). Data are normalized to *rssA* gene, quantified by three biological replicates. (**G**) Colony PCR-based quantification of insertion efficiency with variable cargo sizes and crRNA4, including 1.3, 4.3 and 15.4 kb.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

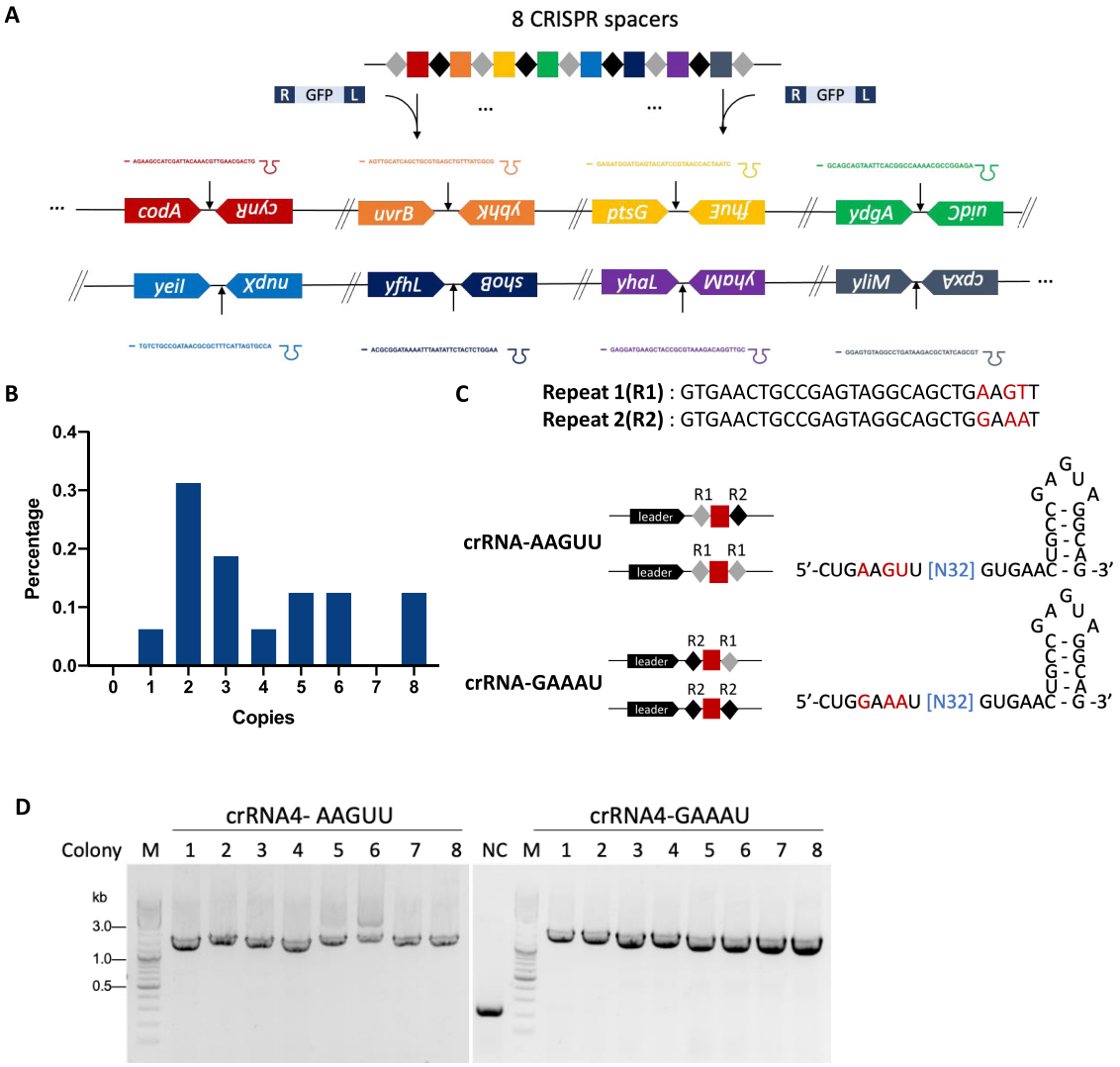


Figure 3. Characterization of multiplexed insertion efficiency and two kinds of crRNAs for PtrCAST. (**A**) Schematic of multiplexed RNA-guided insertion with eight CRISPR spacers targeting eight genomic loci(*codA-cynR*, *uvrB-ybhK*, *ptsG-fhuE*, *ydgA-uidC*, *yeiI-nup, yfhL-shoB*, *yhaL-yhaM*, *yliM-cpxA*) on BL21Star™(DE3). The cargo is GFP. (**B**) Colony PCR-based quantification of insertion efficiency with eight CRISPR spacers, resolved by agarose gel electrophoresis. Data were quantified from 16 colonies. (**C**) Schematic of two kinds of repeat sequences (R1 and R2) and crRNAs (crRNA-AAGUU and crRNA-GAAAU). (**D**) Colony PCR-based quantification of insertion efficiency with crRNA4-AAGUU and crRNA4-GAAAU, resolved by agarose gel electrophoresis.

forms of R1-spacer-R1 and R2-spacer-R2 was constructed, which produced the crRNA forms of crRNA4-AAGUU and crRNA-GAAAU respectively. The transposition efficiencies of the forms of R1-spacer-R1 and R2-spacer-R2 were 100% (8 of 8, 8 of 8 colonies are positive respectively) (Figure 3D).

PtrCAST works orthogonally with VchCAST

First, a test was performed to confirm that PtrCAST and VchCAST can only recognize their respective transposon end sequences. pQCasTns(Vch)-crRNA2 and pDonorGFP(Ptr) targeting the locus between *purT* and *eda* gene, or pQCasTns(Ptr)-crRNA4 and pDonorVerA(Vch) (VerA is the terminator sequence) targeting the *lacZ* locus were co-transformed into *E. coli* BL21Star™(DE3)(Figure 4A). PCR identification was performed after induced by aTC, and it was found that the two combinations did not cross-react, indicating that the two can orthogonally identify different end sequences (Figure 4B).

Next, the orthogonality of PtrCAST and VchCAST was tested, co-transforming four plasmids (Figure 4C) including pQCasTns(Vch)-crRNA2, pQCasTns(Ptr)-crRNA4, pDonorVerA(Vch) and pDonorGFP(Ptr) in *E.coli* BL21Star™ (DE3). The sum size of RE-cargo VerA-LE of pDonorVerA(Vch) was about 0.84 kb, and the sum size of RE-cargo GFP -LE of pDonorGFP(Ptr) was about 2.03 kb (Figure 4D). After being induced by aTC, the insertion efficiency of VchCAST was detected by colony PCR using d/d' primer. The size of the positive band should be 1.50 kb, and the size of the negative band should

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

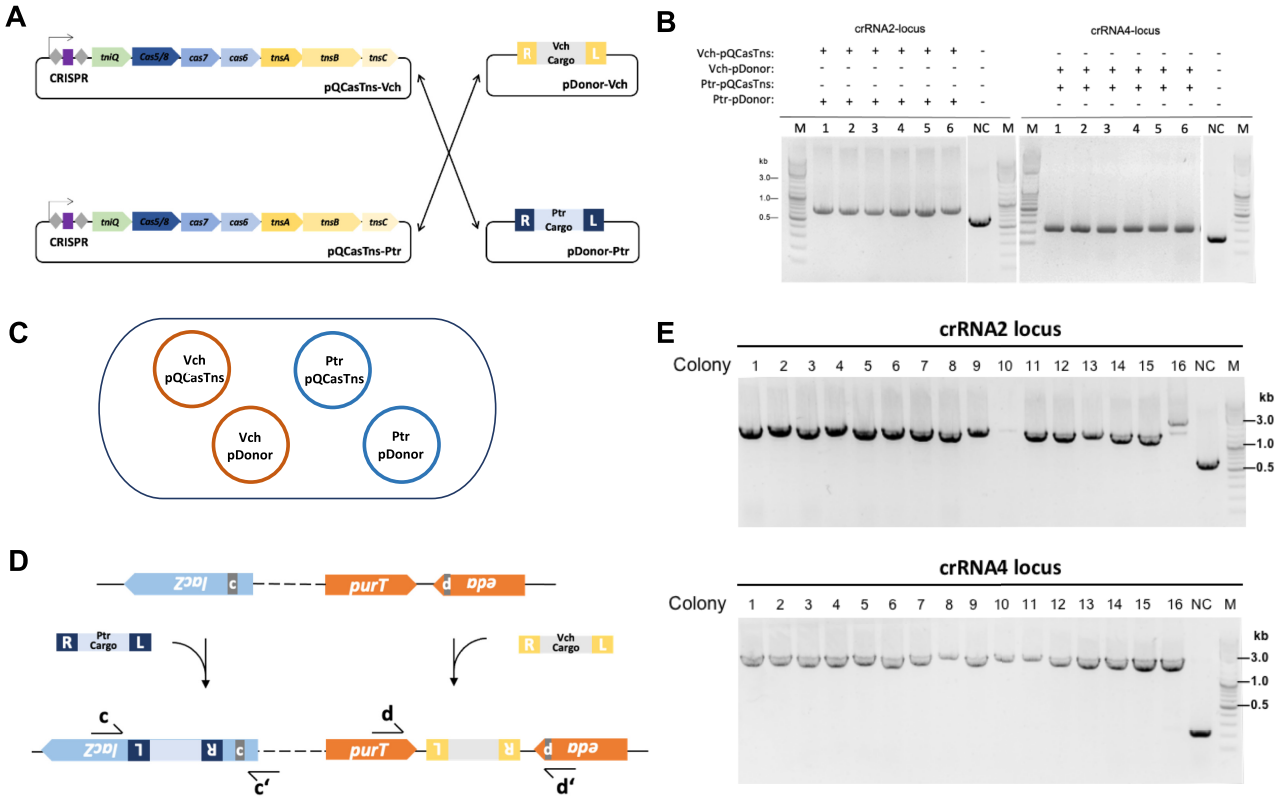


Figure 4. The orthogonality between PtrCAST and VchCAST. (**A**) Schematic of the transposon end recognition between PtrCAST and VchCAST. (**B**) PCR products probing for transposon end recognition at crRNA2 locus or crRNA4 locus with both systems, resolved by gel electrophoresis. Integration proceeds only with a cognate pairing between the expression and donor plasmids. (**C**) Schematic of orthogonal CASTs using four plasmids and the pQCasTns drives protein–RNA expression with a single promoter. (**D**) Genomic loci on BL21Star^{TM}(DE3) targeted by crRNA4 with PtrCAST and crRNA2 with VchCAST, and the PCR primer pairs c/c' and d/d' selectively amplify them. The PAMs and target sites are in grey, together. (**E**) Colony PCR-based analysis of transposition with crRNA4 and crRNA2, resolved by agarose gel electrophoresis.

be 0.66 kb. Using the c/c' primer to detect the insertion efficiency of PtrCAST, where the size of the positive band should be 2.27 kb, and the size of the negative band should be 0.24 kb. VchCAST and PtrCAST target the corresponding sites parallelly and inserted corresponding cargo genes with an efficiency of about 100% (16 of 16 colonies are positive) (Figure 4E). The results have been verified by sanger sequencing (Supplementary Figure S6). As a result, PtrCAST and VchCAST can insert their respective cargo to the corresponding crRNA2 and crRNA4 loci without interfering with each other, achieving 1 (PtrCAST) × 1 (VchCAST) orthogonality.

Building a multiplexed orthogonal CRISPR-associated transposition system using PtrCAST and VchCAST

After proving 1 (PtrCAST) × 1 (VchCAST) orthogonality, multiplexed orthogonality was tested through 2 (PtrCAST) × 1 (VchCAST) orthogonality. VchCAST targeted locus between *purT* and *eda* gene with the terminator cargo, and PtrCAST targeted the *lacZ* loci with the GFP cargo (Figure 5A). After inducing at 37°C, colony PCR with d/d' primers was used to detect the insertion efficiency of VchCAST, and colony PCR with a/a' and b/b' primers was used to detect the insertion efficiencies of PtrCAST. The size of the positive bands should be 2.20 and 2.36 kb respectively, and the size of negative bands should be 0.17 and 0.33 kb respectively. Both VchCAST and PtrCAST target the corresponding sites and insert corresponding cargo genes with an efficiency of about 100% (16 of 16 colonies are positive), resulting in a strain with the desired 2 (PtrCAST) * 1 (VchCAST) insertion (Figure 5A).

Because each CAST uses two plasmids when testing orthogonality, the transformation process was time-consuming. To shorten this process, the pQCasTns and pDonor of each transposition system were merged into pAIO, using one promoter to drive protein expression (Supplementary Figure S7A). In this way, only two plasmids needed to be introduced into one *E. coli* cell, and merging all components of each CAST into one plasmid did not bring any change to insertion efficiency (Supplementary Figure S7B).

The next experiments increased the number of targets targeted by each system, targeting *lacZ* loci on the genome with VchCAST which carrying the cargo VerA, and designing a CRISPR array that has four spacers targeting *nagB*, *nagE* and *manX* genes with PtrCAST which carrying the cargo GFP. There are two spacers of the four spacers in one CRISPR array 480 bp apart that both targeting *manX* gene (Figure 5B). After inducing at 37°C, colony PCR with a/a' and b/b' primers was used to detect the insertion efficiencies of VchCAST. The size of the positive bands should be

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

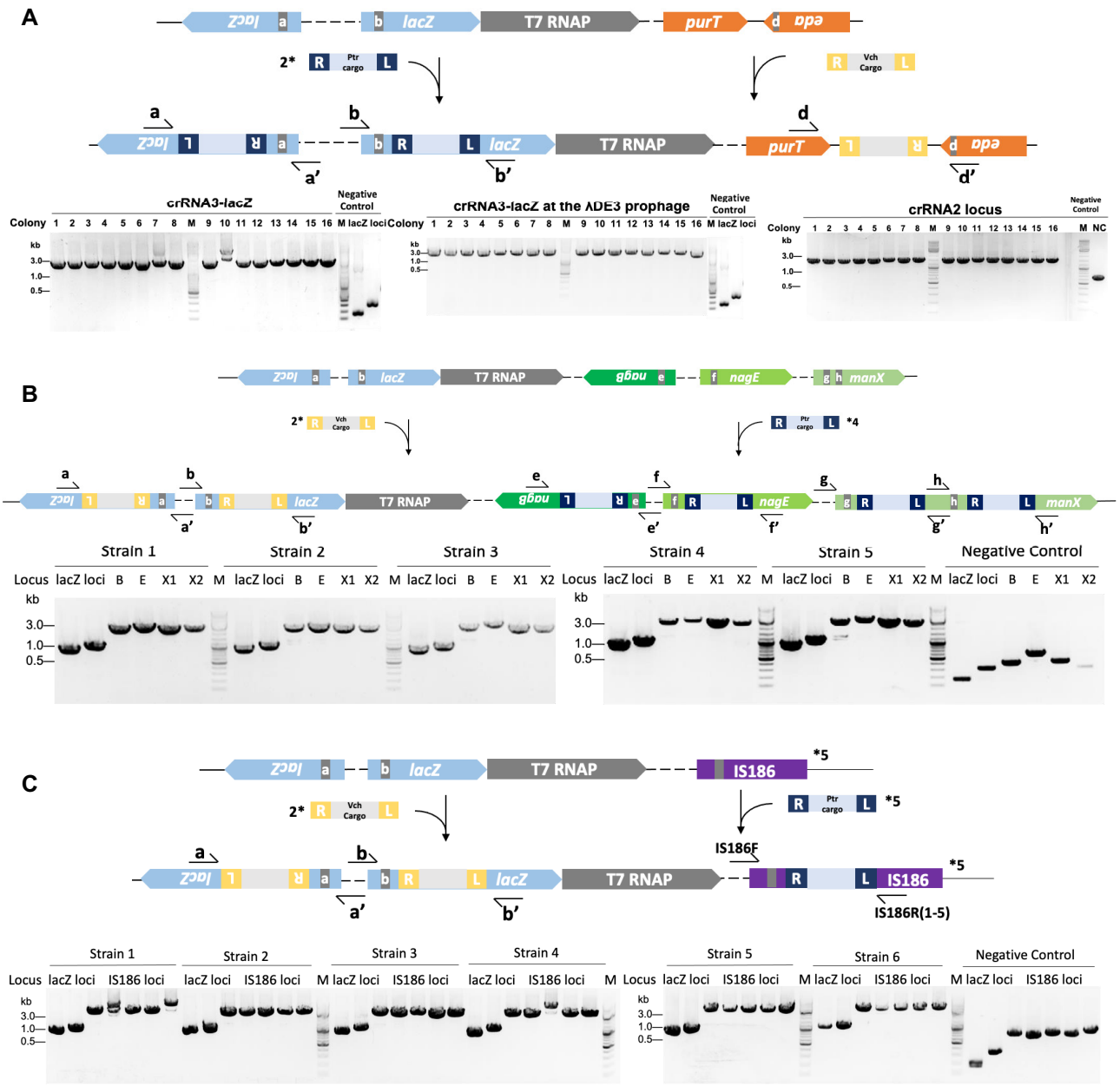


Figure 5. The multiplexed orthogonal transposition using PtrCAST and VchCAST. (**A**) Top: Genomic loci on BL21Star™(DE3) targeted by crRNA3 with PtrCAST and crRNA2 with VchCAST, and the PCR primer pairs a/a', b/b' and d/d' to selectively amplify them. The PAMs and target sites are in grey, together. Bottom: Colony PCR-based analysis of transposition with crRNA3 and crRNA2, resolved by agarose gel electrophoresis. The left band in negative control is the *lacZ* locus and the right band is the *lacZ* locus at the λDE3 prophage. (**B**) Top: Genomic loci on BL21Star™(DE3) targeted by four CRISPR spacers with PtrCAST and crRNA3 with VchCAST, and the PCR primer pairs e/e', f/f', g/g', h/h', a/a' and b/b' to selectively amplify them. The PAMs and target sites are in grey, together. Bottom: Colony PCR-based analysis of transposition with four CRISPR spacers and crRNA3, resolved by agarose gel electrophoresis. The left band in negative control is the *lacZ* locus and the right band is the *lacZ* locus at the λDE3 prophage. (**C**) Top: Genomic loci on BL21Star™(DE3) targeted by crRNA-IS186 with PtrCAST and crRNA3 with VchCAST, and the PCR primer pairs IS186F/IS186R (1-5), a/a' and b/b' to selectively amplify them. The PAMs and target sites are in grey, together. Bottom: Colony PCR-based analysis of transposition with crRNA-IS186 and crRNA3, resolved by agarose gel electrophoresis. The left band in negative control is the *lacZ* locus and the right band is the *lacZ* locus at the λDE3 prophage.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

1.0 and 1.17 kb, and the size of the negative bands should be 0.17 and 0.33 kb respectively. The upstream and downstream primers e/e', f/f', g/g' and h/h' were used to detect the insertion efficiencies of *nagB*, *nagE* and *manX*. The size of the positive bands was about 2.47, 2.70, 2.49 and 2.36 kb, and the size of negative bands should be 0.43, 0.66, 0.46 and 0.33 kb. In the end, PtrCAST and VchCAST targeted the corresponding 4 and 2 loci without interfering with each other, and strains with the desired 2 (VchCAST) × 4 (PtrCAST) insertion were obtained, proving 2 (VchCAST) × 4 (PtrCAST) orthogonality, and the efficiency was about 100% (5 of 5 colonies are positive) (Figure 5B).

Then, a test was performed on more targets to 2 (VchCAST) × 5 (PtrCAST), still targeting the *lacZ* loci on the genome carrying the cargo VerA and replaced the CRISPR array of PtrCAST with a crRNA targeting five IS186 sites, carrying the cargo GFP (Figure 5C). After inducing at 37°C, colony PCR with a/a' and b/b' primers was used to detect the insertion efficiencies of VchCAST. The size of the positive bands should be 1.0 and 1.17 kb, and the size of the negative bands should be 0.17 and 0.33 kb respectively. Primers IS186F/IS186R (1-5) were used to detect the insertion efficiencies of the five sites targeted by PtrCAST. The size of the positive bands was between 2.0 and 2.1 kb, and the size of the negative bands should be 0.78, 0.73, 0.79, 0.73 and 0.85 kb, respectively. A strain with the desired 2 (VchCAST) × 5 (PtrCAST) insertion was obtained, indicating that 2 (VchCAST) × 5 (PtrCAST) orthogonality could be achieved (Figure 5C). In conclusion, VchCAST and PtrCAST can insert corresponding cargo genes into their respective 2 and 5 loci at an efficiency of around 100% (6 of 6 colonies are positive) without interfering with each other.

DISCUSSION

In this study, a CAST that showed high activity in RNA-guided DNA transposition with no preference to PAMs and tolerant to crRNA mismatches was identified. PtrCAST and VchCAST can efficiently insert their respective cargoes into the corresponding five and two loci without interfering with each other.

CASTs need to meet three conditions to achieve complete orthogonality. First, the RE/LE sequences need to be different. Second, the PAMs to be recognized must be different. Finally, the repeat sequences need to be different. The RE/LE sequences of PtrCAST and VchCAST are different, and they can orthogonally identify end sequences in our experiment. However, the flexible PAM of PtrCAST and the similarity of repeat sequences between the two CASTs may lead to the cross reactivity among the other combinations of their spacers. The type I-B CASTs have a longer and dissimilar repeat sequence may be the ideal partners of the type I-F3 CASTs.

Recently, Saito *et al.* detected insertions (less than 1% efficiency) using type I-B1 CAST from *Anabaena variabili* ATCC 29413(AvCAST) which was not detected here (9). There may be two reasons for this. Firstly, outer primers of the target site were used for colony PCR detection and the size of the band was used to judge the occurrence of integration events, while Saito *et al.* used RE-cargo-LE-specific primers and outer primers of the target site for ddPCR detection (9). In principle, the latter method is much more sensitive, and insertions with less than 1% efficiency could be detected in a colony population. Secondly, both *tnsD* and *tniQ* genes of I-B1 CAST were retained in our experiment, and Saito *et al.* found that knocking out *tnsD* gene could improve the efficiency and increase the odds of observing a rare insertion event (9).

The insertion efficiencies of seven CASTs were benchmarked by qPCR to make a directly comparison (Supplementary Figure S8). For most CASTs, the highly efficient spacers from their respectively original articles were used. For PmcCAST, the spacer targeting the *lacZ* gene locus was used because Saito *et al.* only tested one spacer targeting the plasmid to screen PAM (9). Aside from the unexpected low insertion efficiency of ShoCAST in our hands, that of ShCAST, VchCAST, AsaCAST, AvCAST and PmcCAST was consistent with their respective articles. Though the insertion efficiency of PtrCAST is highest among seven CASTs in this trial (Supplementary Figure S8), it is hard to claim that PtrCAST is generally more efficient than other CASTs, since only one target site was tested. Overall, the insertion efficiency of type I-B CASTs should be increased to meet the requirements of constructing strains in cell engineering.

More CASTs will continue to be studied to find those that are orthogonal to both PtrCAST and VchCAST, and the deployment of multiplexed MUCICAT will further accelerate the construction speed of engineered cells.

DATA AVAILABILITY

The raw sequence data reported in this paper including WGS for integration sites, PAM screens and the BL21Star™(DE3) strain have been deposited in deposited in BioProject/SRA: PRJNA747763 and PRJNA747754 and the Genome Sequence Archive (25) National Genomics Data Center (26), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number CRA004516 and CRA004528 that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

The genome differences between BL21(DE3) and BL21 Star™(DE3) have been listed in Supplementary Table S7.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge Professor Samuel H. Sternberg, Professor Joseph E. Peters and Professor Feng Zhang for sharing the plasmids encoding various CASTs. Author contributions: S.Q.Y. and S.Y. designed the overall method and approach. Y. J., J.J.Y. and S.Y. supervised the research. S.Q.Y. constructed the plasmids and strains and performed the transposition experiments. J.J.Y. analyzed the data of NGS. S.Q.Y., Y.W.Z., J.Q.X., J.Z. screened the CAST for experiments. S.Q.Y., Y.W.Z., J.Z.Z. and S.Y. wrote the manuscript. All authors read and commented the manuscript.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021

FUNDING

National Natural Science Foundation of China [21825804, 31921006]; Shanghai Municipal Science and Technology Major Project.
Conflict of interest statement. None declared.

REFERENCES

1. Alper,H., Miyaoku,K. and Stephanopoulos,G. (2005) Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.*, **23**, 612–616.
2. Zhang,Y.W., Sun,X.M., Wang,Q.Z., Xu,J.Q., Dong,F., Yang,S.Q., Yang,J.W., Zhang,Z.X., Qian,Y., Chen,J. *et al.* (2020) Multicopy chromosomal integration using CRISPR-associated transposases. *ACS Synth. Biol.*, **9**, 1998–2008.
3. Ajikumar,P.K., Xiao,W.H., Tyo,K.E.J., Wang,Y., Simeon,F., Leonard,E., Mucha,O., Phon,T.H., Pfeifer,B. and Stephanopoulos,G. (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science (New York, N.Y.)*, **330**, 70–74.
4. Yang,C., Gao,X., Jiang,Y., Sun,B., Gao,F. and Yang,S. (2016) Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in Escherichia coli. *Metab. Eng.*, **37**, 79–91.
5. Feng,X., Zhao,D., Zhang,X., Ding,X. and Bi,C. (2018) CRISPR/Cas9 assisted multiplex genome editing technique in *Escherichia coli*. *Biotechnol. J.*, **13**, e1700604.
6. Strecker,J., Ladha,A., Gardner,Z., Schmid-Burgk,J.L., Makarova,K.S., Koonin,E.V. and Zhang,F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, **365**, 48–53.
7. Petassi,M.T., Hsieh,S.C. and Peters,J.E. (2020) Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons. *Cell*, **183**, 1757–+.
8. Klompe,S.E., Vo,P.L.H., Halpin-Healy,T.S. and Sternberg,S.H. (2019) Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, **571**, 219–225.
9. Saito,M., Ladha,A., Strecker,J., Faure,G., Neumann,E., Altae-Tran,H., Macrae,R.K. and Zhang,F. (2021) Dual modes of CRISPR-associated transposon homing. *Cell*, **184**, 2441–2453.
10. Halpin-Healy,T.S., Klompe,S.E., Sternberg,S.H. and Fernandez,I.S. (2020) Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature*, **577**, 271–274.
11. Zhang,Y., Yang,J., Yang,S., Zhang,J., Chen,J., Tao,R., Jiang,Y., Yang,J. and Yang,S. (2021) Programming cells by multicopy chromosomal integration using CRISPR-associated transposases. *The CRISPR journal*, **4**, 350–359.
12. Vo,P.L.H., Ronda,C., Klompe,S.E., Chen,E.E., Acree,C., Wang,H.H. and Sternberg,S.H. (2021) CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.*, **39**, 480–489.
13. Rubin,B.E., Diamond,S., Cress,B.F., Crits-Christoph,A., He,C., Xu,M., Zhou,Z., Smock,D.C., Tang,K., Owens,T.K. *et al.* (2020) Targeted genome editing of bacteria within microbial communities. bioRxiv doi: https://doi.org/10.1101/2020.07.17.209189, 17 July 2020, preprint: not peer reviewed.
14. Vo,P.L.H., Acree,C., Smith,M.L. and Sternberg,S.H. (2021) Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. *Mob. DNA*, **12**, 13.
15. Rice,P.A., Craig,N.L. and Dyda,F. (2020) Comment on 'RNA-guided DNA insertion with CRISPR-associated transposases'. *Science*, **368**, eabb2022.
16. Strecker,J., Ladha,A., Makarova,K.S., Koonin,E.V. and Zhang,F. (2020) Response to comment on 'RNA-guided DNA insertion with CRISPR-associated transposases'. *Science*, **368**, eabb2920.
17. Shen,W., Le,S., Li,Y. and Hu,F.Q. (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, **11**, e0163962.
18. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
19. Faure,G., Shmakov,S.A., Yan,W.X., Cheng,D.R., Scott,D.A., Peters,J.E., Makarova,K.S. and Koonin,E.V. (2019) CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.*, **17**, 513–525.
20. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I. *et al.* (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
21. Peters,J.E. (2019) Targeted transposition with Tn7 elements: safe sites, mobile plasmids, CRISPR/Cas and beyond. *Mol. Microbiol.*, **112**, 1635–1644.
22. Peters,J.E., Makarova,K.S., Shmakov,S. and Koonin,E.V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E7358–E7366.
23. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
24. Qu,J.Y., Xie,Y.L., Guo,Z.Y., Liu,X.T., Jiang,J., Chen,T., Li,K., Hu,Z. and Luo,D.X. (2021) Identification of a novel cleavage site and confirmation of the effectiveness of NgAgo gene editing on RNA targets. *Mol. Biotechnol.*, https://doi.org/10.1007/s12033-021-00372-1.
25. Wang,Y.Q., Song,F.H., Zhu,J.W., Zhang,S.S., Yang,Y.D., Chen,T.T., Tang,B.X., Dong,L.L., Ding,N., Zhang,Q. *et al.* (2017) GSA: genome sequence archive. *Genom Proteom Bioinf*, **15**, 14–18.
26. Zhang,Z., Zhao,W.M., Xiao,J.F., Bao,Y.M., He,S.M., Zhang,G.Q., Li,Y.X., Zhao,G.P., Chen,R.S., Gao,Y. *et al.* (2020) Database resources of the National Genomics Data Center in 2020. *Nucleic Acids Res.*, **48**, D24–D33.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab752/6363765 by guest on 07 September 2021