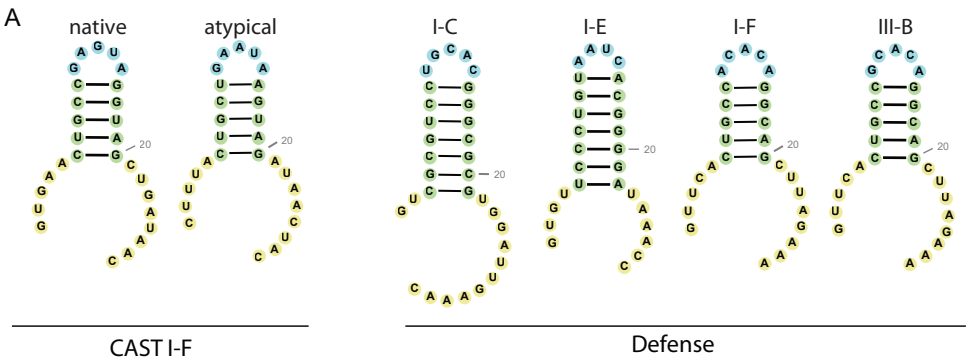
# Supplemental Information for: Distinct horizontal transfer mechanisms for type I and type V CRISPR-associated transposons

Kuang Hu[a,*], Chia-Wei Chou[a], Claus O. Wilke[b], Ilya J. Finkelstein[a,**]

[a]*Department of Molecular Biosciences, The University of Texas at Austin, 100 East 24th St., Stop A5000, Austin, 78712, Texas, USA*
[b]*Department of Integrative Biology, The University of Texas at Austin, 2415 Speedway, Stop C0930, Austin, 78712, Texas, USA*

[*]Corresponding author: kh36969@utexas.edu
[**]Corresponding author: ilya@finkelsteinlab.org

*Preprint submitted to Elsevier* *March 2, 2023*




B

| | | | Defense | | | |
|---|---|---|---|---|---|---|
| | native | atypical | I-C | I-E | I-F | III-B |
| blue | 0 | 0 | N.A. | 2 | 0 | 0 |
| white | 395 | 181 | N.A. | 43 | 167 | 401 |

Figure S1: **A type I-F CAST catalyzes on-target transposition using heterologous CRISPR arrays.** (A) The structure of native type I-F CAST direct repeats (DRs) compared to (B) the structures of DRs from co-occurring CRISPR-Cas systems. Colors indicate different structural elements in the DR. Blue: loop; green: stem; yellow: handles. (C) On-target transposition into *lacZ* results in white colonies on LB+CmR+X-gal plates. In this example, the type I-F CAST and its native crRNA are guided to *lacZ* in the recipient cell genome. Integration is confirmed via junction PCR and Sanger sequencing. (D) On-target transposition remains at $\sim$ 100% with DRs from the defense-associated type I-E, I-F, and III-B CRISPR arrays. The type I-C DR is not active (N.A.).

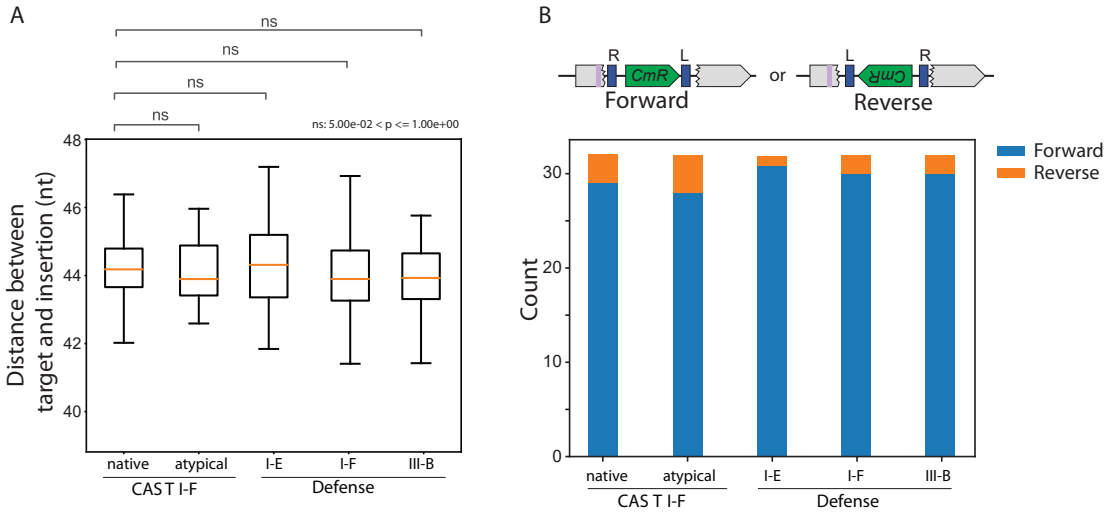



Figure S2: **Defense-associated CRISPR RNAs support on-target integration.** (A) The distance between the target site and the integration site for the indicated crRNAs, as determined via Sanger sequencing of 32 colonies for each condition. n.s.: $p > 0.05$. (B) Schematic of forward and reverse integration orientations (top). The integration orientation was determined via Sanger sequencing for each of the indicated crRNAs (bottom).

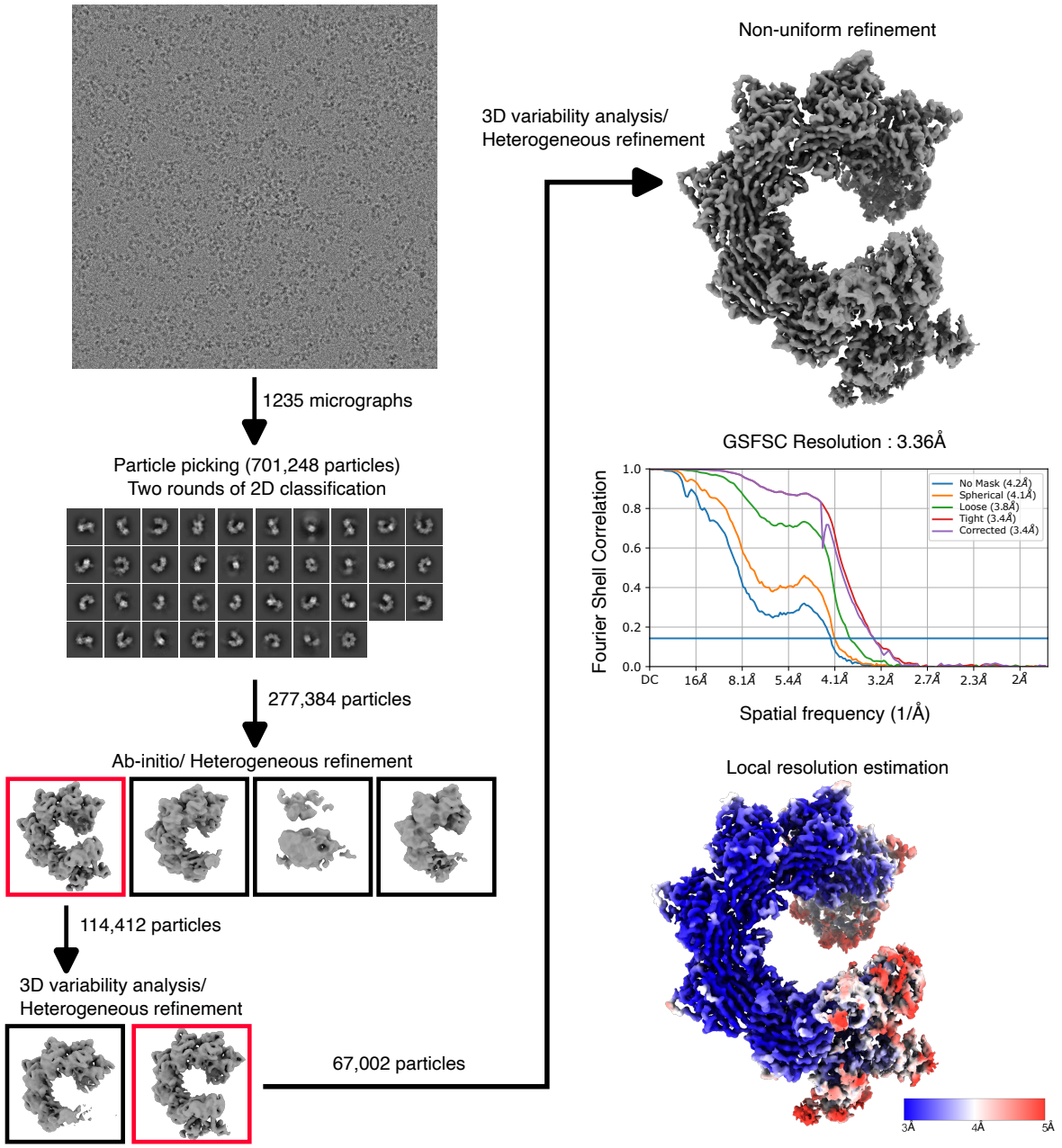





Figure S3: **Cryo-EM image processing workflow for the TniQ-cascade with defense III-B array complex.**

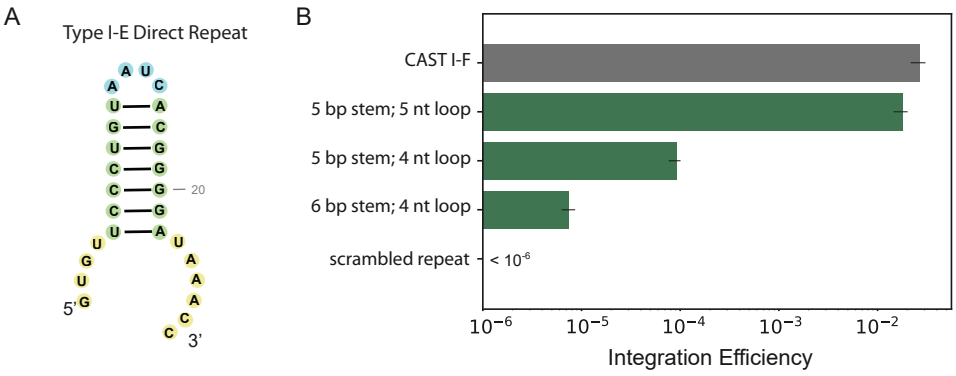



Figure S4: **The length of the direct repeat (DR) stem loop controls transposition.** (A) Predicted structure of the DR from a defense-associated type I-E CRISPR-Cas system. (B) Changing the loop and stem length increases integration efficiency. Error bars: mean of three replicates.

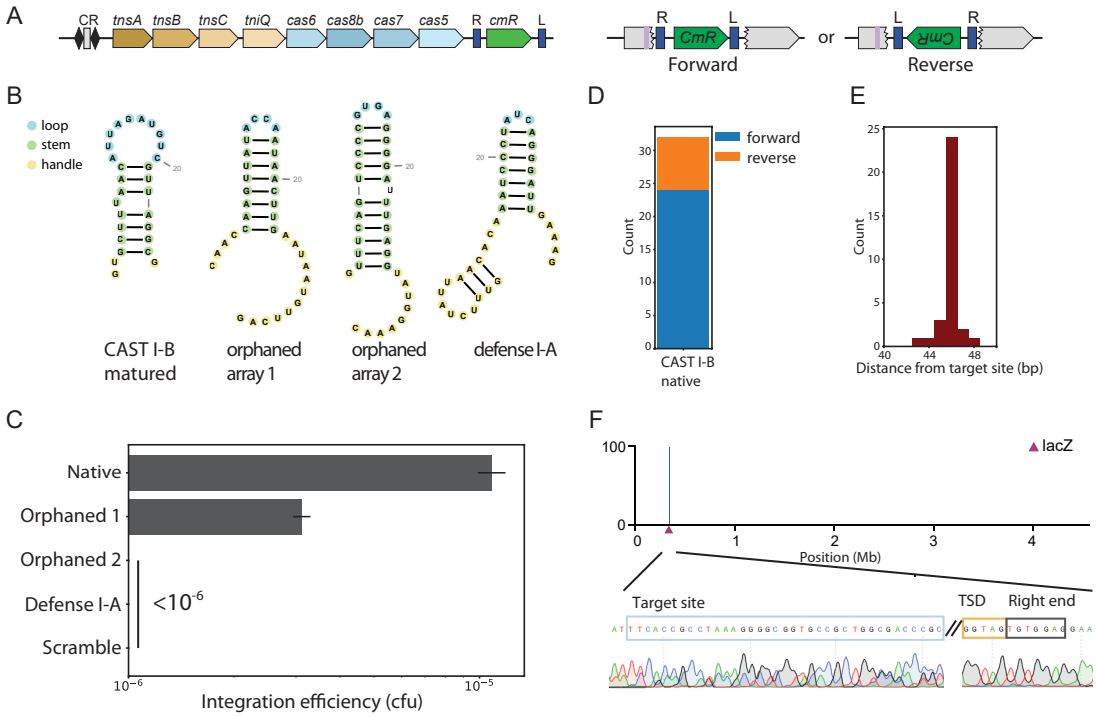



Figure S5: **Type I-B CASTs co-opt spacers from heterologous CRISPR arrays.** (A) Gene architecture of the type I-B *Av*CAST, cloned into the R6k plasmid (pIF1004, Table S1). (B) Predicted structure of the direct repeat (DR) from all CRISPR arrays found in *anabaena variabilis* (ATCC 29413). (C) Quantification of transposition from the native CAST array and co-occurring CRISPR systems. Error bars: standard deviation across three biological replicates. Scrambling the DR suppressed transposition below our detection limit of $\sim 10^6$ cfus. (D) The integration orientation and (E) distance between the target and insertion sites were determined via Sanger sequencing (N=32 clones). (F) Long-read NGS confirms single on-target cut-and-paste transposition into *lacZ*. Target site duplication (TSD) is also visible in this data.

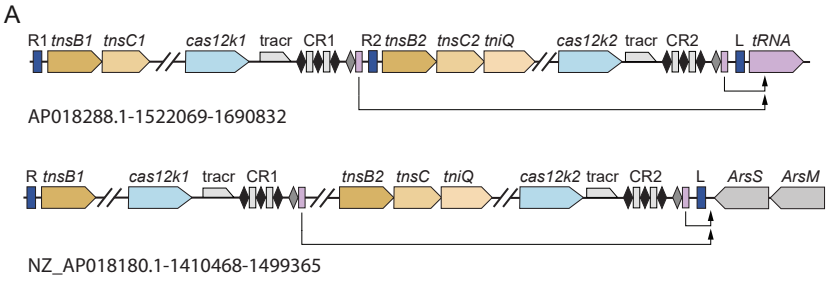



Figure S6: **Examples of metagenomically-encoded type V CASTs that insert twice at the same genomic site.** Assession numbers are indicated for each example.

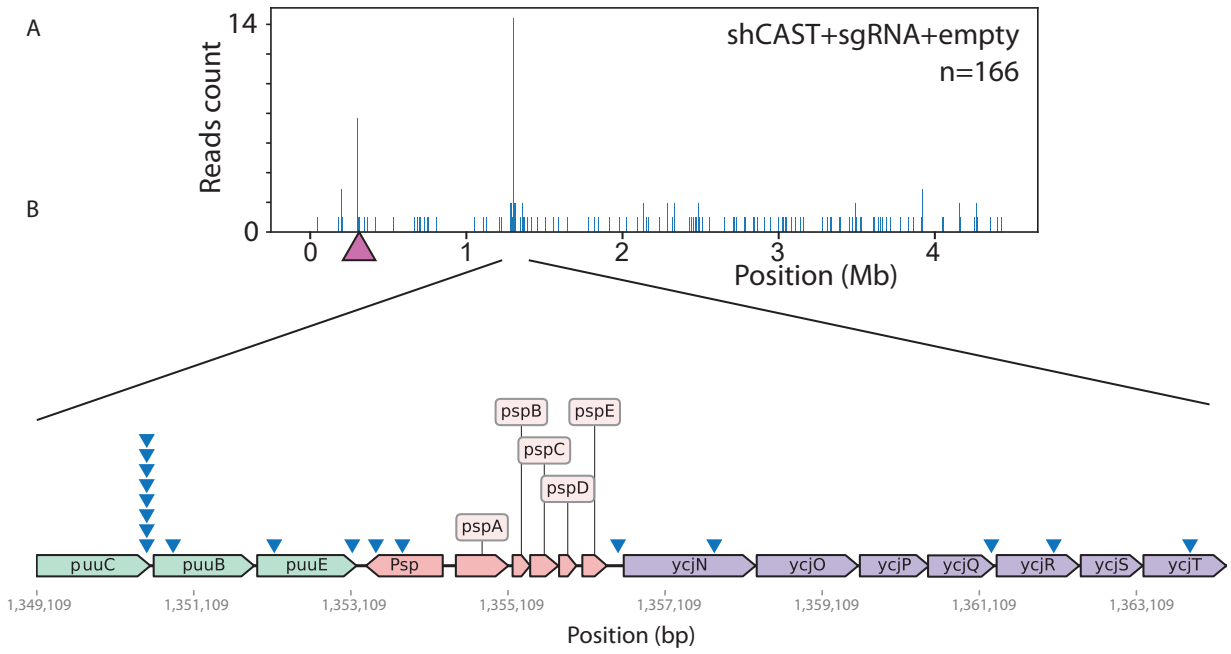



Figure S7: **High resolution view of the off-target genomic hotspot during ShCAST-catalyzed transposition** (A) In the absence of S15, whole-genome long-read sequencing reveals an apparent integration hotspot in the bacterial genome (N=166 insertion events). Diamond: on-target site (*lacZ*). (B) A higher-resolution view of the integration hot spot. Integration occurs over a broad region spanning multiple operons (green, pink, and purple). The events near *puuC* span a $\sim$ 200 bp window without any strong sequence homology to the sgRNA. We conjecture that this hotspot may be due to a structure-specific feature of the bacterial genome.

**Supplemental Tables**

| Plasmid | Description | Source |
|---|---|---|
| pIF1010 | miniTn7-encoding R6K plasmid | Addgene: 64968 [1] |
| pIF1011 | *V. cholerae* TniQ, Cas8, Cas7, and Cas6 | Addgene: 130637 [2] |
| pIF1012 | Kanamycin resistance cassette | Addgene: 130634 [2] |
| pIF1013 | *V. cholerae* TnsA, TnsB, and TnsC | Addgene: 130633 [2] |
| pIF1014 | ShCAST system and its native crRNA | Addgene: 127922 [3] |
| pIF1015 | Kanamycin resistance cassette for ShCAST | Addgene: 127924 [3] |
| pIF1016 | Expression of AvCAST proteins | Addgene: 168137 [4] |
| pIF1017 | AvCAST donor | Addgene: 168145 [4] |
| pIF1008 | R6K backbone with Golden Gate (GG) cloning sites | This paper |
| pIF1001 | R6K plasmid encoding a type I-F CAST but no cargo | This paper |
| pIF1002 | R6K plasmid with a I-F CAST targeting *lacZ* | This paper |
| pIF1003 | R6K plasmid with a I-B CAST but no cargo | This paper |
| pIF1004 | R6K plasmid with a I-B CAST targeting *lacZ* | This paper |
| pIF1005 | R6K plasmid encoding a type V CAST but no cargo | This paper |
| pIF1006 | R6K plasmid with a type V CAST targeting *lacZ* | This paper |
| pIF1008 | Type I-F Cascade over-expression | This paper |
| pIF1009 | Expression of a type III-B defense-associated crRNA | This paper |

Table 1: Plasmids used in this study.

| **Data collection and Processing** | |
|---|---|
| Microscope | Glacios |
| Pixel size (Å) | 0.94 |
| Voltage (kV) | 200 |
| Detector | Falcon IV |
| Exposure ($e^-/Å^2$) | 40 |
| Defocus range ($\mu$m) | -1.2 to -2.2 |
| Final particles | 66,862 |
| Symmetry | C1 |
| Map Resolution (Å) | 3.36 |
| (FSC threshold=0.143) | |
| | |
| **Model refinement and validation** | |
| Initial model used | 6PIG |
| Ramachandran | |
| Flavor (%) | 95.63 |
| Allowed (%) | 4.37 |
| Outlier (%) | 0.00 |
| Rotamer outliers (%) | 0.16 |
| Clash score | 7.88 |
| MolProbity score | 1.73 |
| EM Ringer score | 2.82 |

Table 2: Cryo-EM data collection and processing statistics

## References


[1] K.-H. Choi, H. P. Schweizer, mini-tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*, Nature protocols 1 (2006) 153–161.

[2] S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration, Nature 571 (2019) 219–225.

[3] J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, F. Zhang, Rna-guided DNA insertion with CRISPR-associated transposases, Science 365 (2019) 48–53.

[4] M. Saito, A. Ladha, J. Strecker, G. Faure, E. Neumann, H. Altae-Tran, R. K. Macrae, F. Zhang, Dual modes of CRISPR-associated transposon homing, Cell 184 (2021) 2441–2453.