# 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*

**Abstract**

CRISPR-associated transposons (CASTs) have co-opted CRISPR-Cas proteins and Tn7-family transposons for RNA-guided vertical and horizontal transmission. CASTs encode minimal CRISPR arrays but lack all spacer acquisition genes. Here, we define how different CASTs target new invading mobile elements without updating their own CRISPR arrays. A bioinformatic analysis reveals that all CAST sub-families co-exist with defense-associated CRISPR-Cas systems. Using a quantitative transposition assay, we show that type I-F and I-B CASTs use CRISPR RNAs (crRNAs) from these defense systems for horizontal gene transfer. A high-resolution structure of the type I-F CAST-Cascade in complex with a type III-B crRNA reveals a sequence-independent mechanism for direct repeat recognition. Type I CASTs recognize heterologous CRISPR arrays via a short hairpin in the direct repeat of their crRNA. In contrast, type V CASTs require the Cas12k effector protein but not any crRNA for unguided transposition. This transposition causes random genomic insertions via a copy-and-paste mechanism, even with over-expression of the S15 co-factor. Conversely, a single guide RNA, in concert with S15, increases on-target integration for type V CASTs. These discoveries explain how CASTs horizontally transfer to new hosts without updating their own CRISPR arrays. More broadly, this work will guide further efforts to engineer the activity and specificity of CASTs for gene editing applications.

*Keywords:* gene editing, transposition, bacterial defense, selfish genetic element

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

## Introduction

CRISPR-Cas components have been found associated with multiple systems beyond adaptive immunity (Barrangou and Horvath, 2017; Marraffini, 2015; Mohanraju et al., 2016). For example, CRISPR-associated transposons (CASTs) are an amalgam of a nucleaseinactive CRISPR effector complex and a Tn7-family transposon (Faure et al., 2019a,b; Shen et al., 2022). The ancestral Tn7 transposon consists of five genes, termed *tnsA-E* (Waddell and Craig, 1988; Kubo and Craig, 1990; Craig, 1996; Parks and Peters, 2009; Peters, 2015; Shen et al., 2022). Transposition is catalyzed by *tnsA-C*, whereas *tnsD* and *tnsE* participate in target selection via two distinct mechanisms (Parks and Peters, 2009; Peters, 2015; Shen et al., 2022). CASTs functionally substitute both *tnsD* and *tnsE* with a CRISPR RNA (crRNA)-guided effector complex (Peters et al., 2017; Faure et al., 2019a; Klompe et al., 2019; Strecker et al., 2019). The functions of *tnsD* are substituted by "homing" spacers that target the genomic attachment site for vertical transmission (Saito et al., 2021; Petassi et al., 2020). The mechanism of horizontal transmission, however, remains poorly understood.

CAST systems are organized into two broad categories (Klompe et al., 2019; Saito et al., 2021; Strecker et al., 2019). Type I CASTs, which are highly related to the Tn7 transposase, use a Cascade effector complex to target transposition. By contrast, Type V CASTs evolved from a distinct Tn5053-family transposon and use *cas12k* as the single RNA-guided effector protein (Minakhina et al., 1999; Blackwell et al., 2019). Both CAST subtypes encode CRISPR arrays that are markedly different from defense-associated CRISPR-Cas systems. First, CAST-associated CRISPR arrays are extremely short, generally fewer than three repeats (Klompe et al., 2019; Saito et al., 2021; Strecker et al., 2019). For example, the type I-F3c system only retains a single self-targeting ("homing") spacer, raising the question of how it can also target invading mobile DNA. In addition, type I-C CASTs do not encode any recognizable CRISPR arrays (Rybarski et al., 2021). In contrast, defense-associated CRISPR arrays have tens to hundreds of repeats (Bolotin et al., 2005; Barrangou et al., 2007; Brouns et al., 2008). Second, CASTs do not encode the adaptation genes *cas1* and *cas2*, suggesting that they do not update their own CRISPR arrays (Barrangou et al., 2007; Garneau et al., 2010; Nuñez et al., 2014; Dyda and Hickman, 2015; Lee and Sashital, 2022). Third, CASTs encode an "atypical" repeat that flanks a self-targeting spacer that is only used for vertical transmission (Saito et al., 2021; Petassi et al., 2020). These differences raise the question of how CASTs use these limited CRISPR arrays to target invading mobile elements. Alternatively, there may be one or more novel mechanisms, not previously considered, that CASTS employ during horizontal gene transfer.

Here, we show that type I and type V CASTs use entirely distinct mechanisms for horizontal transmission. A bioinformatic analysis reveals that all CAST subtypes co-occur with CRISPR-Cas defense systems. Type I-F and type I-B CASTs co-opt the spacers from active CRISPR defense systems to mobilize themselves for horizontal dissemination. Mate-out transposition assays demonstrate that both type I-F and I-B CASTs can use crRNAs derived from CRISPR defense systems nearly as efficiently as their own spacers. A cryo-electron microscopy structure of a type I-F TniQ-Cascade complex in complex with a type III-B crRNA shows that Cas6 interacts with the direct repeat (DR) of the crRNA via sequence-independent electrostatic and $\pi$–$\pi$ stacking interactions. Interactions between an evolutionarily conserved Cas6 residue and a nucleotide at the apex of the DR stem-loop is essential for transposition and acts as a molecular ruler for the length of the DR stem. In agreement with this structure, we show that the DR must include a five basepair stem and a five-nucleotide loop for efficient transposition. Because active CRISPR-Cas defense systems include an up-to-date history of invading mobile genetic elements, this mechanism ensures that CASTs can also mobilize into these invading MGEs (McGinn and Marraffini, 2019). In contrast, type V CASTs do not co-opt other CRISPR arrays. Type V CASTs integrate non-specifically via a crRNA-independent copy-and-paste mechanism that requires the Cas12k effector. This process is also independent of the S15 specificity factor. Surprisingly, a single guide RNA (sgRNA) increases on-target specificity relative to the crRNA-tracrRNA pair. Our study resolves the long-standing question of how CASTs can mobilize into novel MGEs without updating their own CRISPR arrays. More broadly, we reveal design principles and potential considerations for optimizing CAST crRNAs for precision gene insertion in diverse organisms.

## Results

### *CASTs co-exist with active CRISPR-Cas defense systems*

We reasoned that CASTs may co-opt other CRISPR arrays that are scattered throughout the host genome for horizontal transmission. To test this hypothesis, we searched for additional CRISPR arrays in the genomes of all CAST-encoding organisms (Figure 1). We identified 921 genomes that encoded a CAST amongst the $\sim$ 1M high-quality assembled genomes in the NCBI reference sequences (RefSeq) database (Figure 1A) (Pruitt et al., 2005; Rybarski et al., 2021). All CASTs encoded very short or undetectable CRISPR arrays (Figure 1B). Next, we searched these CAST-encoding genomes for co-occurring CRISPR-Cas systems and orphaned CRISPR arrays. Defense systems included an active nuclease

(i.e., *cas3*), adaptation genes (i.e., *cas1*, *cas2*, *cas4*), and CRISPR arrays with $\sim$ 10–120 spacers, suggesting active spacer acquisition (Figure 1B) (Camacho et al., 2009; Cock et al., 2009; Skennerton, 2016). We also observed isolated examples of "orphaned" arrays that were not adjacent to a recognizable CRISPR-Cas defense system (Hullahalli et al., 2015; Almendros et al., 2016; Shmakov et al., 2020). Ten percent of genomes that encode a type I-F CAST also encode additional CRISPR-Cas systems and 100% of organisms with a type I-B or type V CAST encode at least one additional CRISPR array (Figure 1C) (Shmakov et al., 2020). 12.5% of type I-B CASTs and 11% of type V CASTs also co-occurred with two or more additional CRISPR-Cas systems (Figure 1C). Type I-F CASTs mainly co-occurred with type III-B, I-F, I-E CRISPR defense systems. In two genomes, the type I-F CAST co-existed with a type II-A defense system (Figure 1D). By contrast, type I-B and V CASTs co-occurred with type III-B and type I-D defense systems (Figure 1D). Below, we test the hypothesis that CASTs can use protospacers from defense-associated CRISPR arrays for horizontal transmission.

*Type I-F CASTs mobilize using heterologous CRISPR arrays*

To determine whether CASTs can co-opt other CRISPR arrays, we first compared the sequences and secondary structures of their direct repeats (DRs) (Lorenz et al., 2011). DRs from the type I-F CAST are structurally identical to defense-associated I-F and III-B CRISPR-Cas systems, with a five nucleotide (nt) loop, five basepair (bp) stem, and five nt 3'-handle (Figure 2A). By contrast, the type I-E DR consists of a four nt loop, seven bp stem, and four nt 3'-handle. The type I-C and II-A DRs are even more divergent from the CAST I-F (Figure S1A).

We developed a conjugation-based chromosomal transposition assay to determine whether CASTs can exploit these heterologous CRISPR arrays (Figure 2B) (Curtiss, 1969; Petassi et al., 2020). In this assay, the CAST genes, a CRISPR array, and a chloramphenicol (Cm) resistance marker surrounded by left and right inverted repeats are assembled into a conditionally replicative R6K plasmid that only replicates in *pir+* strains (Kolter et al., 1978; Ferrières et al., 2010; Rakowski and Filutowicz, 2013). The *pir+* donor also includes a chromosomally integrated RP4 conjugation system (Bradley and Whelan, 1985; Rakowski and Filutowicz, 2013). Donor cells are auxotrophic for diaminopimelic acid (DAP), allowing for counter-selection on DAP- plates following conjugation with a recipient strain (BAUMAN and DAVIS, 1957; Ferrières et al., 2010). The BL21(DE3) recipient cells support CAST expression and transposition (Strecker et al., 2019; Klompe et al., 2019). Conjugative transfer of the R6K plasmid into the recipient cells and subsequent transposition of the CAST cargo

into the host genome (targeting *lacZ* ) results in chloramphenicol-resistant, $\Delta$*lacZ* recipient cells. The R6K plasmid is lost shortly after conjugation in the recipient cells (*pir*-) and the donor cells are also removed due the absence of DAP (Choi and Schweizer, 2006). Genomic transposition efficiency can be scored quantitatively via the ratio of recipient colonies on standard (DAP-) agar plates vs. CmR cells. Targeting *lacZ* results in white colonies on Cm/X-gal plates; integration outside *lacZ* produces blue colonies on the same plates (Chaffin and Rubens, 1998; UF, 1995). Finally, we also scored the insertion accuracy via both Sanger- and whole-genome long-read sequencing of individual clones.

We first tested this assay with the native and atypical direct repeats from the well-characterized *V. cholerae* HE-45 Type I-F3a system (Figure 2C) (Klompe et al., 2019). This CAST encodes an atypical direct repeat and a homing spacer for site-specific integration into the host's genome. We removed the homing spacer to avoid spurious transposition events (Petassi et al., 2020). Transposition efficiency was scored using a *lacZ*-targeting spacer (Klompe et al., 2019). A scrambled spacer or a scrambled direct repeat served as negative controls. The transposition efficiency was $1.4 \pm 0.2\%$ of all viable recipient cells. This was suppressed below the limit of detection ($< 10^{-6}$ cfus) when either the spacer or the repeat were scrambled. All chloramphenicol-resistant colonies ($n = 395$ across three biological replicates) were white on X-gal plates, suggesting transposition into *lacZ* (Figure S1B). Sanger sequencing of the insertion junctions from 32 colonies showed that the cargo inserted $\sim$ 42–46 bp downstream of the end of target site (Figure S2A). Integration occurred in the forward direction in 91% of all cases and in the reverse direction in the remaining 9%. Whole-genome long-read sequencing indicated a single transposition event at the expected target size. An atypical direct repeat supported a nearly identical transposition efficiency and insertion orientation (Figure S2). The atypical direct repeat maintains the same overall stem-loop structure but has 12 nucleotide substitutions relative to the typical direct repeat (Petassi et al., 2020). Because the typical and atypical direct repeats maintained a high transposition rate, we conclude that the CAST effector complex can tolerate DRs with divergent RNA sequences.

Next, we tested whether this CAST can use DRs from co-occurring CRISPR defense systems (Figure 2C) (Klompe et al., 2019). For this assay, the native CAST array targeted *lacZ* but encoded the DR from defense-associated CRISPR-Cas systems. All other protein and cargo components remained unchanged. Surprisingly, type I-F and III-B DRs supported transposition efficiencies that were comparable to those from the native CAST, despite having no RNA sequence similarity (Figure S1). CRISPR RNAs with type I-E DRs transposed

$\sim 10^3$-fold less efficiently than the native CAST crRNAs (Figure 2F). In all cases, $> 99\%$ of the resulting colonies were white on X-gal plates, indicating targeted transposition into *lacZ* (Figure S1B). Integration occurred in the forward direction in 90% of all cases and in the reverse direction in the remaining 10% (Figure S2B). Integration events for all repeat types occurred $\sim$ 43–45 bp from the 3' end of the target to the integrated transposon (Figure 2E). Long-read sequencing showed that a single copy of the cargo was inserted into *lacZ* (Figure 2D) (Klompe et al., 2019). By contrast, type I-C and II-A direct repeats did not support any transposition activity ($< 10^{-6}$ cfus). The structures of these DRs differ substantially from the I-F DR, indicating that the DR stem loop structure is a major determinant of transposition.

*Cas6 stabilizes direct repeats via sequence-independent electrostatic interactions.*

To investigate the molecular basis for how CASTs exploit heterologous CRISPR arrays, we used cryo-electron microscopy to solve the structure of the *V. cholerae* HE-45 Cascade co-purified with a type III-B crRNA (Figure S3). The crRNA contained a native direct repeat from the type III-B system and a 32 bp spacer. The density for Cascade and the crRNA was refined with a prior model (PDB: 6PIG) (Halpin-Healy et al., 2020; Jia et al., 2020; Li et al., 2020; Park et al., 2022). The overall structure was quite similar to the prior model ($C_A - RMSD = 0.83$Å), indicating the native DR from the type III-B can assemble a functional Cascade (Figure 3).

The type III-B direct repeat engages Cas6 via sequence-independent interactions with the ribose phosphate backbone (Figures 3B–C). The guanidine (G54) at the apex of the stem-loop is flipped out of the plane and enters in a long-range $\pi$–$\pi$ interaction with Cas6(F138). A helix with three arginines (R117, R121, R125) also forms a strong positive pocket to stabilize the crRNA handle. A multiple sequence analysis of I-F Cas6 proteins indicates that these electrostatic interactions are conserved across the entire CAST sub-family (Figure 3D). Thus, Cascade engages diverse direct repeats via crRNA-sequence independent mechanism.

We tested the functional significance of the conserved Cas6 residues using the transposition assay described above (Figure 3D). Mutating any of the arginines to an alanine suppressed transposition below our detection range ($< 10^{-6}$ cfus). Similarly, Cas6(F138A) reduces transposition $> 10^4$-fold, indicating that the $\pi$–$\pi$ interaction is also necessary for stably engaging the DR (Figure 3C). We conclude that Cas6 stabilizes diverse DRs via RNA sequence-independent electrostatic and $\pi$–$\pi$ stacking interactions.

*The Direct Repeat Tunes Transposition Efficiency*

The reduced transposition efficiency with type I-E DRs indicates additional constraints on the CAST crRNA. To test these constraints, we systematically varied the DR sequence and/or structure and assayed the resulting transposition efficiency (Figure 4A). We first scrambled the DR nucleotide sequence but retained the 5 bp stem, the 5 nt loop, the 5 bp 5' handle, and the 8 bp 3' handle of the type I-F CAST. Surprisingly, this crRNA maintained wild type transposition efficiency (Figure 4B). By contrast, scrambling the stem-loop entirely abolished transposition. These results confirm that Cas6-DR contacts are sequence independent but require a structured DR to maintain activity.

Next, we systematically varied the length of the stem, loop, and the 5' and 3' handles to determine the key determinants of efficient transposition. Starting with the CAST I-F DR, changing the stem length by even a single basepair reduced transposition efficiency up to five-fold (Figure 2F). Increasing the length of the stem from five to seven basepairs (as in the type I-E DR) decreased transposition efficiency 500-fold as compared with the type CAST I-F DR (Figure S4A). Decreasing the loop by one nucleotide also reduced transposition efficiency 100-fold (Figure 4B). Changing the length of the 5' and 3' handles modestly reduced transposition efficiency. Consistent with these findings, shortening the type I-E DR stem from seven to five basepairs significantly increased transposition. Adding one nucleotide from the loop to five nucleotides also improved transposition 500-fold relative to the type I-E DR (Figure S4B). These results underscore that the DR structure is the key determinant for assembling a TniQ-Cascade effector complex. The stem must be five basepairs, whereas the loop can tolerate one nucleotide changes from the five-nucleotide native sequence. The structural basis for both effects likely arises from the base stacking interaction with Cas6.

*Type I-B CASTs co-opt co-occurring CRISPR arrays for horizontal transfer*

All type I-B CASTs co-occur with type I-A, I-B, I-D, or III-B defense cas systems or with orphaned CRISPR arrays (Figure 1C). To test whether type I-B CASTs can use these CRISPR arrays, we adapted the mate-out transposition assay to the *Anabaena variabilis* ATCC 29413 type I-B CAST (Figure S5A) (Saito et al., 2021). We first tested transposition with the native CAST DR. Transposition efficiency with the native CRISPR sequence was $\sim 10^3$-fold lower than the type I-F CAST (Figure S5B). This may be due to poor expression in *E. coli* since we did not optimize codon usage, promoters, or translation efficiency. Most of chloramphenicol-resistant colonies where white ($\sim$ 90%, $n = 152$), indicating that *lacZ* was disrupted. Sanger sequencing across the insertion junctions confirmed on-target integration $\sim$ 44–48 bp away from the target site (Figure S5E). Scrambling the crRNA without

preserving the DR structure ablated all transposition activity(Figure S5B). These results indicate that type I-B CASTs are active in the mate-out transposition assay.

Next, we tested whether the type I-B CAST can use DRs from a co-occurring CRISPR defense system. Surprisingly, the predicted structure of the co-occurring CRISPR defense system is divergent from that of the CAST (Figure S5B). The orphaned DR supported low transposition, $\sim$ 20-fold lower than the native DR (Figure S5C). We did not detect integration for any other DRs ($< 10^{-6}$ cfus). We conclude that type I-B and I-F CASTs can both co-opt heterologous CRISPR arrays, so long as the crRNA DRs can be structurally accommodated within the Cascade effector complex.

*Type V CASTs transpose via a CRISPR RNA-independent mechanism*

All type V CASTs co-exist with either type I or III defense-associated CRISPR systems (Figure 1). Therefore, we assayed whether the *S. hofmannii* (Sh) type V CAST can use spacers from these CRISPR arrays for horizontal transfer (Strecker et al., 2019). As before, we removed the CAST's homing spacer and targeted a single guide RNA (sgRNA) or the native tracr/crRNA to *lacZ*. We also tested whether the *E. coli* or *S. hofmannii* small ribosomal subunit protein S15 aides transposition (Park et al., 2022; Schmitz et al., 2022). Transposition efficiency was scored by measuring the colony forming units on chloramphenicol-resistant plates. On-/off-target events were confirmed via Sanger and long-read whole genome sequencing.

Transposition activity remained high with the native crRNA, the sgRNA, or crRNAs with type I or III direct repeats (Figure 5A). Surprisingly, deleting the entire crRNA ($\Delta$crRNA) did not diminish this activity. Transposition was dependent on *cas12k*, as its ablation dropped transposition below our detection limit ($< 10^{-7}$ cfus). Recent structural studies identified the small ribosomal protein S15 as a core CAST subunit that increases the on-target transposition rate (Park et al., 2022; Schmitz et al., 2022). Surprisingly, expressing either *E. coli* or *S. hofmannii* S15 (EcS15 or ShS15) reduced transposition $\sim$ 50–250 fold, respectively. Blue/white screening and Sanger sequencing of the insertion junction of individual chloramphenicol-resistant clones indicated different integration modes for CASTs assembled with either a sgRNA or a crRNA.

Next, we analyzed the transposition fidelity and mechanism using pooled long-read whole-genome sequencing (Jain et al., 2016). CASTs assembled with a sgRNA transposed at the target site most frequently, especially in recipient cells over-expressing shS15 (50% on-target, $n = 84$) (Figure 5B) but dropped to 5% ($n = 166$) when S15 was not over-expressed (Figure S7). Two genomic hotspots accounted for 12% of the off-target transposition events

(Figure S7). These off-target sites harbored a putrescine utilization protein cluster, phage shock protein cluster and sugar transporter protein cluster. Although we inspected the genomic sequence adjacent to the insertion sites closely, we could not identify any off-target sites that harbored partial homology with the crRNA. With S15, 88% of transposons inserted via a copy-and-paste mechanism (i.e., duplication of the cargo and integration the entire plasmid into the genome). The remaining 12% of transposition events inserted via a cut-and-paste mechanism (Figure 5E). On-target transposition was $> 95\%$ in the L/R orientation ($n = 46$ insertion events). Long-read sequencing of the integration site revealed that the cargo DNA inserted $\sim$ 36–41 bp or 24 bp, away from the end of the *lacZ* target site (Figure 5C).

In contrast to the sgRNA, all crRNA-tracrRNA combinations, as well as $\Delta$crRNA negative control, showed 100% off-target transposition (Figure 5D). Importantly, type I-D and III-D DRs did not improve on-target integration, even with the recipient cells over-expressed ShS15. The CAST randomly dispersed its cargo throughout the genome. Again, we did not find off-target sites that harbored partial homology with the crRNA. 68% of these off-target insertions occurred via a copy-and-paste mechanism (Figure 5E). To determine the number of transposition events per cell, we individually long-read sequenced three clones. All of the genomic off-target transposition occurred via a cut-and-paste insertion mechanism that is indicative of a functional shCAST. Moreover, we observed one clone with up to three insertions at the same genomic site. These results are consistent with our earlier bioinformatic observation of multi-site insertion in other type V systems (Figure S6) (Rybarski et al., 2021). We conclude that type V CASTs do not require any CRISPR array to mobilize via unguided transposition (Querques et al., 2021; Xiao et al., 2021; Wang et al., 2022; Tenjo-Castaño et al., 2022; Park et al., 2022). Taken together, these results highlight that type V CASTs are much more permissive than either type I-F or I-B systems. S15 partially suppresses random integration, protecting the host's genome from deleterious CAST hyperactivity. More broadly, these systems can disperse via unguided transposition without co-opting defense-associated crRNAs.

## Discussion

Here, we show that type I and type V CASTs employ two distinct strategies for horizontal transmission, even though they share the same mechanism (a homing spacer) for vertical transmission (Figure 6). For horizontal transmission, type I CASTs can co-opt defense associated CRISPR arrays already present in the host. These arrays are updated by defense-

associated Cas1-Cas2 (Lee and Sashital, 2022). By contrast, type V CASTs integrate non-specifically via a crRNA-independent mechanism. Type V CASTs are exclusively found in cyanobacteria, suggesting limited horizontal transmission (Strecker et al., 2019; Faure et al., 2019a). This may be due to the dependence on S15 for on-target integration, the limited range of the homing spacer, and the possible host toxicity associated with random integration. Analagously to Tn7, type V CASTs may also retain a *tnsE*-like mechanism to target replicating MGEs (Kubo and Craig, 1990; Craig, 1996; Parks and Peters, 2009; Peters, 2015; Shen et al., 2022). Additional studies will be required to understand and minimize random integration by type V systems for gene editing applications.

Type I CASTs use the information in heterologous CRISPR arrays to direct their own transposition. Despite the large sequence divergence between the direct repeats of these systems, a type I-F CAST reconstituted with type I-E, I-F, and III-B crRNAs mobilizes via on-target transposition with high integration efficiency. Prior studies have shown that type I-F Cascades have a flexible PAM requirement and can tolerate variable crRNA lengths by adjusting the number of Cas7 repeats in the Cascade assembly (Kuznedelov et al., 2016; Gleditzsch et al., 2016; Inga et al., 2019; Tuminauskaite et al., 2020; Wimmer et al., 2022). We conjecture that Cascades from type I-F CASTs may also assemble for a variable number of *cas7* subunits, especially on heterologous crRNAs. Similarly, a type I-B CAST can co-opt spacers from other CRISPR defense systems. This is an elegant solution because active defense systems will continuously update their own CRISPR arrays with a record of prior infections. The most recent mobile genetic elements are inserted proximal to the leader of the CRISPR array and are expressed at the highest levels, providing an ample source of crRNAs for the CAST to use for horizontal gene transfer (McGinn and Marraffini, 2019). Thus, Cascade supports high transposition activity from heterologous crRNAs, opening new opportunities for crRNA engineering.

Our bioinformatic analysis shows that all type I-B systems co-exist with at least one active CRISPR defense system that can be co-opted by the CAST for horizontal gene transfer. However, we only detected additional CRISPR arrays in $\sim 10\%$ of genomes that harbor a type I-F CAST. There are several possibilities to explain this observation. First, CASTs may recognize invading DNA by interacting with replisome associated DNA structures (e.g., the replication fork) or other replisome components (e.g., the sliding clamp). For example, Tn7 encodes *tnsE*, a gene that directs transposition adjacent to replisomes (Nancy, 1991; Shen et al., 2022). The interactions between CASTs and host proteins are a promising area for future research. Second, plasmids and phages also encode their own CRISPR arrays and

even full CRISPR systems (Pinilla-Redondo et al., 2022). These CRISPR arrays can be co-opted by CASTs for mobilization. Third, we may be oversampling type I-F systems from a small group of cultivatable organisms that are over-represented in the NCBI database. Additional genome-resolved metagenomic studies will shed light on the co-occurrence of CASTs and defense-associated CRISPR arrays in the same host genomes.

Defense-associated CRISPR sub-types can also share crRNAs. For example, type III systems lack *cas1* and *cas2* and co-occur in genomes containing type I CRISPR-Cas loci (Makarova et al., 2015). Type III systems can use the pre-processed crRNAs from type I-F systems, acting as secondary defenses that counteract viral escape (Silas et al., 2017; Vink et al., 2021). Another example is the type VI-B system of *Flavobacterium columnare*, which is also acquisition-deficient. This system can acquire spacers in *trans* from a type II-C system that is encoded in the same genome (Hoikkala et al., 2021). CASTs may also use heterologous *cas1* and *cas2* for spacer acquisition into their own arrays. However, CAST CRISPR arrays are extremely short, suggesting that their expansion is not a major mechanism for horizontal transmission. The plasticity of spacer acquisition between CRISPR sub-types suggests that other *cas1/cas2*-deficient systems may use similar mechanisms to target viral pathogens.

Type V CASTs have a distinct transposition mechanism and method of horizontal transmission. Transposition does not require a CRISPR array but is dependent on Cas12k. The small ribosomal protein S15 suppresses, but does not completely abrogate, random transposition. Mechanistically, S15 forms a complex with Cas12k and TniQ to stabilize the R-loop (Park et al., 2022; Schmitz et al., 2022). In agreement with earlier work, S15 also increases on-target insertion in the mate-out transposition assay (Park et al., 2022; Schmitz et al., 2022). However, these assays retain high off-target integration rates. We note that our assays are conducted *in vivo* and target the *E. coli* genome, which provides ample off-target sites and structures. Additional host factors may further participate in transposition *in vivo* as compared to the purified *in vitro* system (Park et al., 2022). Interestingly, over-expressing S15 suppressed overall integration via an unknown mechanism. Preventing this off-target transposition activity while boosting on-target integration will be increasingly important for domesticating type V CASTs in heterologous organisms.

## Supplemental Information

Supplemental information includes seven figures and two tables.

Bioinformatics data available: https://github.com/KuangHu/CAST-crosstalking-repo

## Declarations

*Author Contributions.* K.H. and I.J.F. conceived the project. K.H., C.-W.C., and Z.Y. performed all experiments and analyzed the data. K.H. and C.-W.C. prepared the figures. I.J.F. and C.O.W. secured the funding. I.J.F. and C.O.W. supervised the project. K.H., C.-W.C., C.O.W. and I.J.F. wrote the manuscript with input from all co-authors.

*Funding.* This work was supported by NIGMS grants R01GM124141 (to I.J.F.) and R01GM088344 (to C.O.W.), the Welch Foundation grant F-1808 (to I.J.F.), and the College of Natural Sciences Catalyst Award for seed funding.

*Declaration of Interests.* K.H., C.O.W. and I.J.F. have filed a patent application relating to CRISPR-associated transposons.

## Materials and Methods

*Bioinformatic analysis of CAST co-occurrence with other CRISPR systems*

Genomes containing CAST systems were collected from NCBI genomic databases (Pruitt et al., 2005). We searched for CRISPR-Cas systems in these genomes using Opfi, a Python library to search DNA sequencing data for putative CRISPR systems (Hill et al., 2021). First, we located all regions containing a CRISPR array that was not associated with a CAST. Within those regions, we next searched for *cas* genes located no more than 25 kilobase pairs away from CRISPR array using BLAST and a previously-developed database of diverse *cas* genes (Camacho et al., 2009; Rybarski et al., 2021). We subtyped CRISPC-cas systems based on signature genes (Makarova et al., 2015; Makarova and Koonin, 2015).

*Proteins and nucleic acids*

Oligonucleotides were purchased from IDT. Gene blocks for CRISPR arrays were purchased from Twist Biosciences. The R6K plasmid for mate-out transposition assays was obtained from Addgene (#64968) (Choi et al., 2005). The type I-F *Vibrio cholerae* HE-45 CAST was subcloned from Addgene (#130637 and #130633) (Klompe et al., 2019). The type V *Scytonema hofmanni* (Sh)CAST was obtained from Addgene (#127922) (Strecker et al., 2019). The type I-B *Anabaena variabilis* ATCC 29413 CAST was subcloned from Addgene (#168137) (Saito et al., 2021). For mate-out transposition assays, each of these systems was PCR amplified and cloned into pTNS2 to replace the parental mini-Tn7 (Addgene #64968) by Golden Gate assembly. The repeat, spacer, chloramphenicol resistance cargo, and left and right inverted repeats were synthesized by IDT and cloned into the same plasmid. Full plasmids information can be found in Table s1.

*Cascade purification*

Plasmids for type I-F CASTs Cascade over-expression were constructed by subcloning the individual genes into pRSFDuet1 (Addgene #126878) to create pIF1008. Type I-F Cascade was co-expressed with 6xHis-MBP-TEV-TniQ and a type III-B crRNA in NiCo21 cells (NEB). Cells were then induced with 0.5 mM isopropyl at 18 ℃ for another 18–20 hours before harvesting. Cells were centrifuged and re-solubilized in lysis buffer containing 25 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, and 1 mM DTT. Cascade complexes were purified via the N-terminal maltose binding protein (MBP) tag using amylose beads (NEB) and eluted with lysis buffer containing 10 mM maltose. MBP was removed using TEV protease at 4 ℃overnight. The sample was further diluted to 100 mM NaCl and developed over an anion exchange column (5 mL Q column HP). After loading Cascade, the column was washed extensively with buffer A (25 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM DTT). The complex was eluted with a 25 column volume gradient of buffer B (25mM Tris pH 7.5, 1 M NaCl, 5% glycerol, and 1 mM DTT.) Cascade was further purified by size exclusion chromatography using a Superose 6 increase column (GE healthcare) in SEC buffer (25 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, and 1 mM DTT). Fractions were further pooled and concentrated to 0.25 mg/ml and stored in the −80℃ freezer.

*Cryo-electron microscopy (cryo-EM)*

*Sample preparation and data collection*

Purified TniQ-Cascade complexes was diluted to a concentration of 0.25 mg/ml in 25 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, and 1 mM DTT. Samples were deposited on an Ultra Au foil R 1.2/1.3 grid (Quantifoil) that was plasma-cleaned for 1.5 min (Gatan Solarus 950). Excess liquid was blotted away for 4 s in a Vitrobot Mark IV (FEI) operating at 4 ℃ and 100% humidity before being plunge-frozen into liquid ethane. Data were collected on a Glacios cryo-transmission electron microscope (TEM; Thermo Fisher Scientific) operating at 200 kV, equipped with a Falcon IV direct electron detector camera (Thermo Fisher Scientific). Movies were collected using SerialEM at a pixel size of 0.94Å with a total exposure dose of $40e^{-}/\text{Å}^{2}$.

*Data processing and model building*

Motion correction, contrast transfer function (CTF) estimation, and particle picking were all performed on cryo SPARC live and further transferred to cryoSPARC for two-dimensional (2D) classification, ab initio 3D reconstruction calculation, 3D classification, and nonuniform refinement (Punjani et al., 2017). Because of the flexibility of TniQ and

Cas6, particle subtraction and focused refinement were also performed in cryoSPARC. A full description of the cryo-EM data processing workflows can be found in Figure S3. A published Cascade structure (PDB: 6PIG) was docked into cryo-EM density maps using Chimera before being refined in Coot, ISOLDE, and PHENIX (Emsley et al., 2010; Croll, 2018; Adams et al., 2010; Halpin-Healy et al., 2020). Full cryo-EM data collection and refinement statistics can be found in Table S2.

*Conjugation-based transposition assays*

CASTs were cloned into a conditionally replicative R6k plasmid (Addgene #64968). The CAST I-F system's proteins, CRISPR array, and inverted repeats were subcloned from Addgene plasmids #130637, #130634, and #130633 to generate pIF1001. CAST V and inverted repeat constructs were subcloned from Addgene plasmids #127922 and #127924 to generate pIF1005. CAST I-B system's proteins and inverted repeat constructs were subcloned from Addgene plasmids #168137 and #168146 to generate pIF1003.

For transposition, the R6k plasmid was transformed into MFDpir cells, which contain the genomically-integrated RP4-based transfer machinery, termed the donor strain (Ferrières et al., 2010). All growth steps were conducted at 37℃. The donor strain was grown with 0.3 mM diaminopimelic acid (DAP) and appropriate antibiotics. The recipient strain was grown in lysogeny broth (LB). The donor and recipient cells were gently washed four timtes in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$) by spinning and resuspending 1 mL cultures. The cell density was estimated by taking an optical density reading after resuspension and the donor and recipient cells were combined in a 3:1 ratio. This mixture was plated on a non-selective plate containing DAP (0.3 mM) for conjugation. The conjugation plate was incubated overnight. The conjugation mixture was collected and washed by mixing with 1 mL PBS, vortexed, and gently spun down four times. Multiple ten-fold dilutions of this mixture were plated onto selective (LB+12 $\mu$l/ml chloramphenicol) and non-selective plates. The cfu/ml was calculated by counting the colonies on plates with 50-500 colonies. The integration efficiency equal to the cfu/ml on selection plate divide by cfu/ml on non-selection plates.

*DNA sequencing of transposition products*

*Sanger sequencing*

Individual colonies were resuspended in LB, pelleted by centrifugation at 16,000 g for 1 min and resuspended in 80 $\mu$l of $H_2O$, before being lysed by incubating at 98℃ for 10 min in a thermal cycler. Cell debris was pelleted by centrifugation at 16,000 g for 1 min, and the

supernatant was removed and serially diluted with 90 $\mu$l of $H_2O$ to generate lysate dilutions for PCR analysis. PCR products were generated with Q5 Hot Start High-Fidelity DNA Polymerase (NEB) using 1 $\mu l$ of the diluted lysate per 10 $\mu$l reaction volume. Reactions contained 200 $\mu$M dNTPs and 0.5 $\mu$M primers and were subjected to 30 thermal cycles. PCR amplicons were resolved by 1% agarose gel electrophoresis and visualized by staining with ethidium bromide (Thermo Scientific). To map integration sites by Sanger sequencing, bands were excised after separation by gel electrophoresis, DNA was gel-extracted (Qiagen), and samples were submitted to Sanger sequencing (Eton).

*High-throughput long-read whole genome Sequencing*

Colonies from plate-based transposition reactions were washed off and diluted to an $OD_{600} \sim 0.5$ using LB. The liquid culture was then grown for two hours at 37°C. Genomic DNA was extracted (ProMega Wizard Genomic DNA kit) and barcoded with a Nanopore technologies rapid barcoding kit. The barcoded DNA was sequenced on a MinION nanopore sequencer using the manufacturer-suggested protocol. Output reads were analyzed using *seqkit* (Shen et al., 2016). First, the reads were processed by *seqkit* to collect adjacent target-DNA sequences from all the reads containing 40 bp of the shCAST left end sequence. Then these sequences were mapped onto the BL21(DE3) genome using BLAST+ with a 95% sequence identity cutoff (Camacho et al., 2009). The position of the transposition reaction was defined as the number of basepairs between the end of the PAM and the beginning of the shCAST left-end sequence.

*Single-colony whole genome sequencing*

Single colonies were grown overnight in LB with 17 ng/$\mu$l chloramphenicol at 37°C. 1 ml of the overnight culture was pelleted by centrifugation at 16,000 g for 1 min and the gDNA was extracted as described above. Genomic DNA samples were separated and barcoded in units of 12 per batch using the MinION rapid barcoding kit. Samples were loaded into a MinION flowcell (FLO-MIN106D) and sequenced with a MinION Mk1B device. The raw read fastq files were assembled with flye (Kolmogorov et al., 2020).

## References

R. Barrangou, P. Horvath, A decade of discovery: CRISPR functions and applications, Nature microbiology 2 (2017) 1–9.

L. A. Marraffini, Crispr-cas immunity in prokaryotes, Nature 526 (2015) 55–61.

P. Mohanraju, K. S. Makarova, B. Zetsche, F. Zhang, E. V. Koonin, J. Van der Oost, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems, Science 353 (2016) aad5147.

G. Faure, S. A. Shmakov, W. X. Yan, D. R. Cheng, D. A. Scott, J. E. Peters, K. S. Makarova, E. V. Koonin, CRISPR–Cas in mobile genetic elements: counter-defence and beyond, Nature Rev. Microbiol. 17 (2019a) 513–525.

G. Faure, K. S. Makarova, E. V. Koonin, Crispr–cas: complex functional networks and multiple roles beyond adaptive immunity, Journal of molecular biology 431 (2019b) 3–20.

Y. Shen, J. Gomez-Blanco, M. T. Petassi, J. E. Peters, J. Ortega, A. Guarné, Structural basis for DNA targeting by the tn7 transposon, Nature Structural & Molecular Biology 29 (2022) 143–151.

C. S. Waddell, N. L. Craig, Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes., Genes & development 2 (1988) 137–149.

K. M. Kubo, N. L. Craig, Bacterial transposon Tn7 utilizes two different classes of target sites, Journal of bacteriology 172 (1990) 2774–2778.

N. Craig, Transposon Tn7, Transposable Elements (1996) 27–48.

A. R. Parks, J. E. Peters, Tn7 elements: engendering diversity from chromosomes to episomes, Plasmid 61 (2009) 1–14.

J. E. Peters, Tn7, Mobile DNA III (2015) 647–667.

J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons, Proc. Natl. Acad. Sci. USA 114 (2017) 7358–7366.

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.

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.

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.

M. T. Petassi, S.-C. Hsieh, J. E. Peters, Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons, Cell 183 (2020) 1757–1771.

S. Minakhina, G. Kholodii, S. Mindlin, O. Yurieva, V. Nikiforov, Tn5053 family transposons are res site hunters sensing plasmidal res sites occupied by cognate resolvases, Molecular microbiology 33 (1999) 1059–1068.

G. Blackwell, Z. Iqbal, N. Thomson, Evolution and spread of bacterial transposons, Microbiology society (2019).

J. R. Rybarski, K. Hu, A. M. Hill, C. O. Wilke, I. J. Finkelstein, Metagenomic discovery of CRISPR-associated transposons, Proceedings of the National Academy of Sciences 118 (2021) e2112279118.

A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology 151 (2005) 2551–2561.

R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, Crispr provides acquired resistance against viruses in prokaryotes, Science 315 (2007) 1709–1712.

S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, J. Van Der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes, Science 321 (2008) 960–964.

J. E. Garneau, M. Ève Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán, S. Moineau, The CRISPR/Cas bacterial immune system cleaves bacteriophage and

plasmid DNA, Nature 468 (2010) 67–71.

J. K. Nuñez, P. J. Kranzusch, J. Noeske, A. V. Wright, C. W. Davies, J. A. Doudna, Cas1-cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity, Nature structural & molecular biology 21 (2014) 528–534.

F. Dyda, A. B. Hickman, Mechanism of spacer integration links the CRISPR/Cas system to transposition as a form of mobile dna, Mobile DNA 6 (2015) 1–5.

H. Lee, D. G. Sashital, Creating memories: molecular mechanisms of CRISPR adaptation, Trends in Biochemical Sciences (2022).

J. McGinn, L. A. Marraffini, Molecular mechanisms of CRISPR-Cas spacer acquisition, Nature Reviews Microbiology 17 (2019) 7–12.

K. D. Pruitt, T. Tatusova, D. R. Maglott, Ncbi reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic acids research 33 (2005) D501–D504.

C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, Blast+: architecture and applications, BMC bioinformatics 10 (2009) 1–9.

P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, et al., Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics 25 (2009) 1422–1423.

C. Skennerton, Minced—mining CRISPRs in environmental datasets, git (2016).

K. Hullahalli, M. Rodrigues, B. D. Schmidt, X. Li, P. Bhardwaj, K. L. Palmer, Comparative analysis of the orphan CRISPR2 locus in 242 *Enterococcus faecalis* Strains, PloS 10 (2015) e0138890.

C. Almendros, N. M. Guzmán, J. García-Martínez, F. J. M. Mojica, Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems, Nature Microbiology 1 (2016) 1–8.

S. A. Shmakov, I. Utkina, Y. I. Wolf, K. S. Makarova, K. V. Severinov, E. V. Koonin, Crispr arrays away from cas genes, The CRISPR journal 3 (2020) 535–549.

R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, Viennarna package 2.0, Algorithms for molecular biology 6 (2011) 1–14.

R. Curtiss, Bacterial conjugation, Annual review of microbiology 23 (1969) 69–136.

R. Kolter, M. Inuzuka, D. R. Helinski, Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K, Cell 15 (1978) 1199–1208.

L. Ferrières, G. Hémery, T. Nham, A.-M. Guérout, D. Mazel, C. Beloin, J.-M. Ghigo, Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery, Journal of bacteriology 192 (2010) 6418–6427.

S. A. Rakowski, M. Filutowicz, Plasmid R6K replication control, Plasmid 69 (2013) 231–242.

D. E. Bradley, J. Whelan, Conjugation systems of IncT plasmids, Microbiology 131 (1985) 2665–2671.

N. BAUMAN, B. D. DAVIS, Selection of auxotrophic bacterial mutants through diaminopimelic acid or thymine deprival, Science 126 (1957) 170–170.

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

D. Chaffin, C. Rubens, Blue/white screening of recombinant plasmids in Gram-positive bacteria by interruption of alkaline phosphatase gene (phoZ) expression, Gene 219 (1998) 91–99.

W. UF, New multifunctional Escherichia coli-Streptomyces shuttle vectors allowing blue-white screening on XGal plates, Gene 165 (1995) 149–150.

T. S. Halpin-Healy, S. E. Klompe, S. H. Sternberg, I. S. Fernández, Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system, Nature 577 (2020) 271–274.

N. Jia, W. Xie, M. J. de la Cruz, E. T. Eng, D. J. Patel, Structure-function insights into the initial step of dna integration by a CRISPR-Cas-Transposon complex, Cell Research 30 (2020) 182–184.

Z. Li, H. Zhang, R. Xiao, L. Chang, Cryo-em structure of a type IF CRISPR RNA guided surveillance complex bound to transposition protein TniQ, Cell Research 30 (2020) 179–181.

J.-U. Park, A. W.-L. Tsai, A. N. Rizo, V. H. Truong, T. X. Wellner, R. D. Schargel, E. H. Kellogg, Structures of the holo CRISPR RNA-guided transposon integration complex, Nature (2022) 1–3.

M. Schmitz, I. Querques, S. Oberli, C. Chanez, M. Jinek, Structural basis for RNA-mediated assembly of type V CRISPR-associated transposons, bioRxiv (2022) 2022–06.

M. Jain, H. E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community, Genome biology 17 (2016) 1–11.

I. Querques, M. Schmitz, S. Oberli, C. Chanez, M. Jinek, Target site selection and remodelling by type V CRISPR-transposon systems, Nature 599 (2021) 497–502.

R. Xiao, S. Wang, R. Han, Z. Li, C. Gabel, I. A. Mukherjee, L. Chang, Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition, Molecular Cell 81 (2021) 4457–4466.

J. Y. Wang, P. Pausch, J. A. Doudna, Structural biology of CRISPR-Cas immunity and genome editing enzymes, Nature Reviews Microbiology 20 (2022) 641–656.

F. Tenjo-Castaño, N. Sofos, B. López-Méndez, L. S. Stutzke, A. Fuglsang, S. Stella, G. Montoya, Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon, Nature Communications 13 (2022) 5792.

J.-U. Park, A. W.-L. Tsai, T. H. Chen, E. H. Kellogg, Mechanistic details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM, Proceedings of the National Academy of Sciences 119 (2022).

K. Kuznedelov, V. Mekler, S. Lemak, M. Tokmina-Lukaszewska, K. A. Datsenko, I. Jain, E. Savitskaya, J. Mallon, S. Shmakov, B. Bothner, S. Bailey, A. F. Yakunin, K. Severinov, , E. Semenova, Altered stoichiometry *Escherichia coli* Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation, Nucleic Acids Res (2016) gkw914.

D. Gleditzsch, H. Müller-Esparza, P. Pausch, K. Sharma, S. Dwarakanath, H. Urlaub, G. Bange, L. Randau, Modulating the Cascade architecture of a minimal type I-F CRISPR-Cas system, Nucleic Acids Res 44 (2016) 5872–5882.

S. Inga, R. Marius, S. Tomas, M. Elena, W. Sabine, S. Carla, S. Virginijus, S. Ralf, Decision-making in Cascade complexes harboring crRNAs of altered length, Cell reports 28 (2019) 3157–3166.

D. Tuminauskaite, D. Norkunaite, M. Fiodorovaite, S. Tumas, I. Songailiene, G. Tamulaitiene, T. Sinkunas, Dna interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system, BMC biology 18 (2020) 1–16.

F. Wimmer, I. Mougiakos, F. Englert, C. L. Beisel, Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons, Molecular Cell 82 (2022) 1210–1224.

C. Nancy, Tn7: a target site-specific transposon, Molecular microbiology 5 (1991) 2569–2573.

R. Pinilla-Redondo, J. Russel, D. Mayo-Muñoz, S. A. Shah, R. A. Garrett, J. Nesme, J. S. Madsen, P. C. Fineran, S. J. Sørensen, Crispr-cas systems are widespread accessory elements across bacterial and archaeal plasmids, Nucleic acids research 50 (2022) 4315–4328.

K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, et al., An updated evolutionary classification of CRISPR-Cas systems, Nature Reviews Microbiology 13 (2015) 722–736.

S. Silas, P. Lucas-Elio, S. A. Jackson, A. Aroca-Crevillén, L. L. Hansen, P. C. Fineran, A. Z. Fire, Antonio, Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems, Elife 6 (2017) e27601.

J. N. Vink, J. H. Baijens, S. J. Brouns, Comprehensive PAM prediction for CRISPR-Cas systems reveals evidence for spacer sharing, preferred strand targeting and conserved links with crispr repeats, bioRxiv (2021) 2021–05.

V. Hoikkala, J. Ravantti, C. Díez-Villaseñor, M. Tiirola, R. A. Conrad, M. J. McBride, S. Moineau, L.-R. Sundberg, Cooperation between different CRISPR-Cas types enables adaptation in an RNA-targeting system, Mbio 12 (2021) e03338–20.

A. M. Hill, J. R. Rybarski, K. Hu, I. J. Finkelstein, C. O. Wilke, Opfi: A Python package for identifying gene clusters in large genomics and metagenomics data sets, Journal of open source software 6 (2021).

K. S. Makarova, E. V. Koonin, Annotation and classification of CRISPR-Cas systems, CRISPR: methods and protocols (2015) 47–75.

K.-H. Choi, J. B. Gaynor, K. G. White, C. Lopez, C. M. Bosio, R. R. Karkhoff-Schweizer, H. P. Schweizer, A Tn7-based broad-range bacterial cloning and expression system., Nature methods 2 (2005) 443–448.

A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryosparc: algorithms for rapid unsupervised cryo-EM structure determination, Nature methods 14 (2017) 290–296.

P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallographica Section D: Biological Crystallography 66 (2010) 486–501.

T. I. Croll, Isolde: a physically realistic environment for model building into low-resolution electron-density maps, Acta Crystallographica Section D: Structural Biology 74 (2018) 519–530.

P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, et al., Phenix: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallographica Section D: Biological Crystallography 66 (2010) 213–221.

W. Shen, S. Le, Y. Li, F. Hu, Seqkit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, PloS one 11 (2016) e0163962.

M. Kolmogorov, D. M. Bickhart, B. Behsaz, A. Gurevich, M. Rayko, S. B. Shin, K. Kuhn, J. Yuan, E. Polevikov, T. P. Smith, et al., metaFlye: scalable long-read metagenome assembly using repeat graphs, Nature Methods 17 (2020) 1103–1110.

# Figures

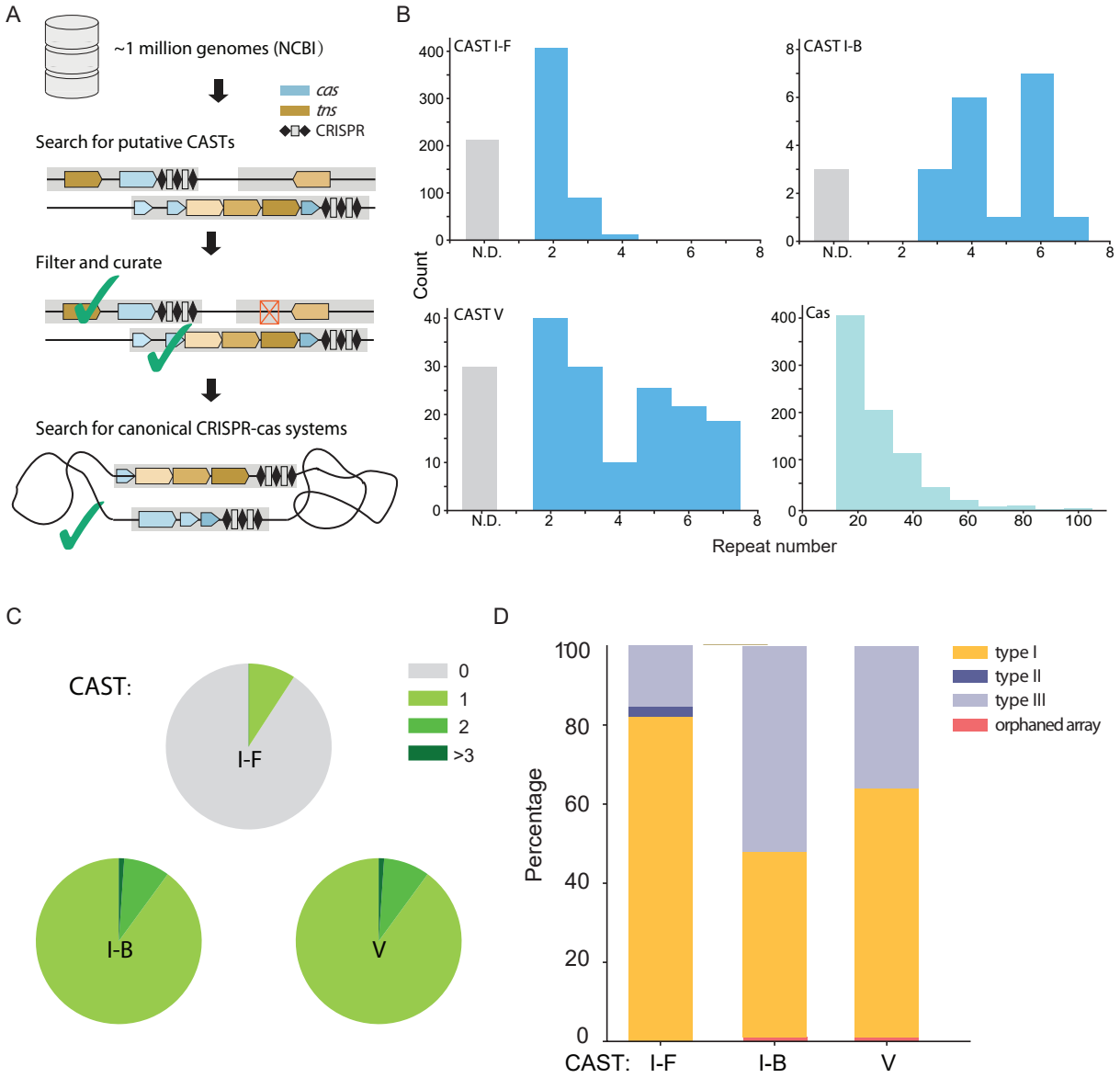


Figure 1: **CASTs co-occur when CRISPR defense systems.** (A) A bioinformatics workflow for annotating CRISPR defense systems that co-occur with CASTs in the same genome. Blue: CRISPR-associated genes; brown: transposase genes. (B) CAST CRISPR arrays are shorter than defense associated CRISPR arrays in the same genomes. (C) CASTs frequently co-exist with one or more additional CRISPR arrays. (D) Defense-associated CRISPR-Cas sub-types that co-exist with CASTs in the NCBI microbial genome database.

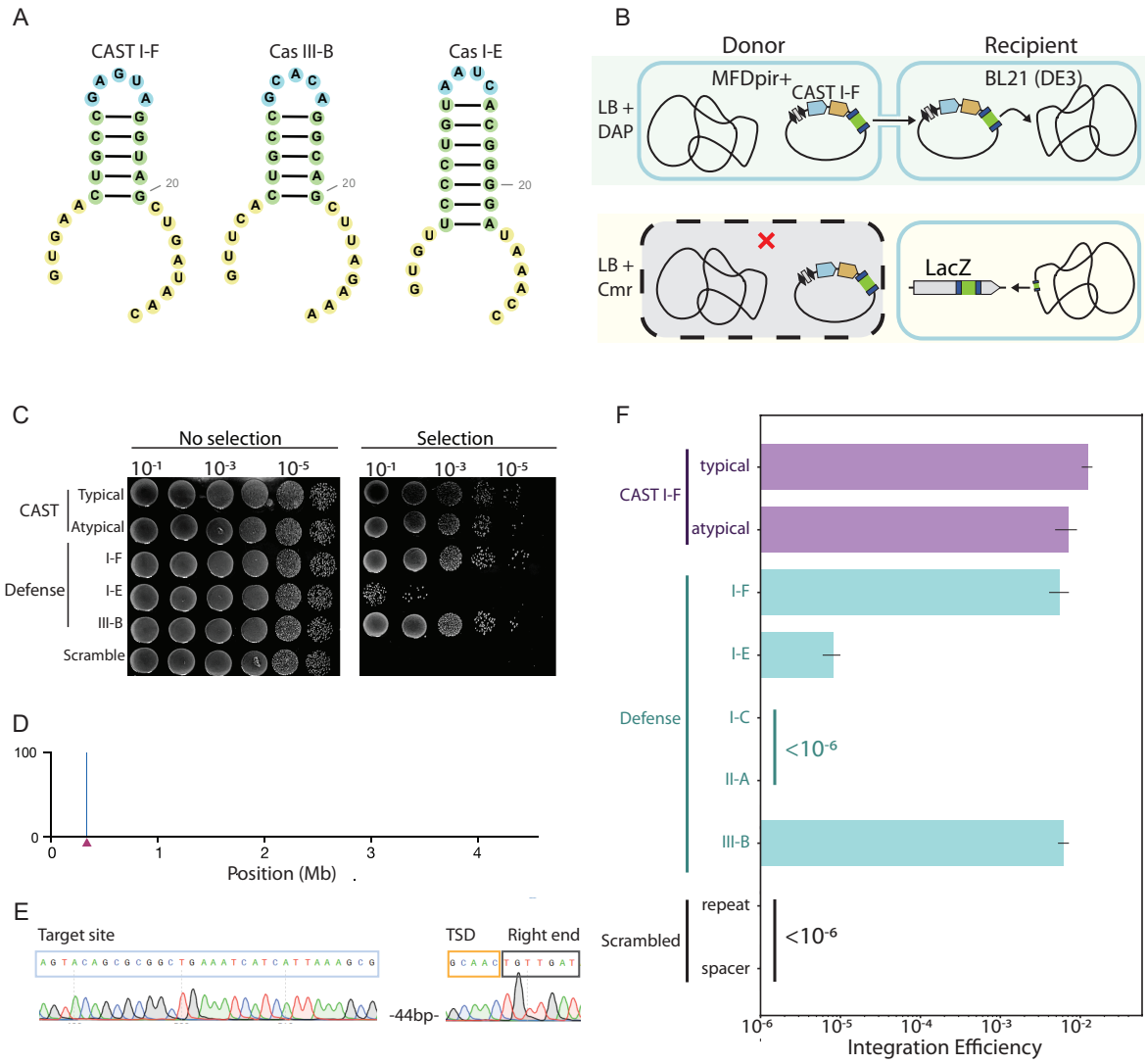


Figure 2: **Type I-F CASTs co-opt defense-associated CRISPR arrays.** (A) The predicted structures of direct repeats (DRs) from a type I-F CAST and co-occurring defense CRISPR-Cas systems. Blue: 4-5 nt loop; green: 5-7 bp stem; yellow: 5'- and 3'-handles. (B) Schematic of a quantitative conjugation-based mate-out transposition assay. A plasmid harboring the CAST, along with the cargo antibiotic resistance (green), and a minimal CRISPR array is conjugated into the recipient strain. Guided transposition into *lacZ* is scored as white, chloramphenicol-resistant clones. The donor strain is removed via counter-selection with diaminopimelic acid (DAP). (C) Direct repeats from the defense associated CRISPR arrays support transposition, but a scrambled direct repeat does not. (D) Colony-resolved long-read sequencing (E) and Sanger sequencing (F) confirms cut-and-paste transposition into *lacZ* (triangle in E). Target site duplication (TSD) is also visible in this data. (F) Quantification of transposition from the native CAST array and co-occurring defense systems. Error bars are the standard deviation across three biological replicates. Scrambling either the repeat or spacer suppressed transposition below our detection limit of $\sim 10^6$ cfus.

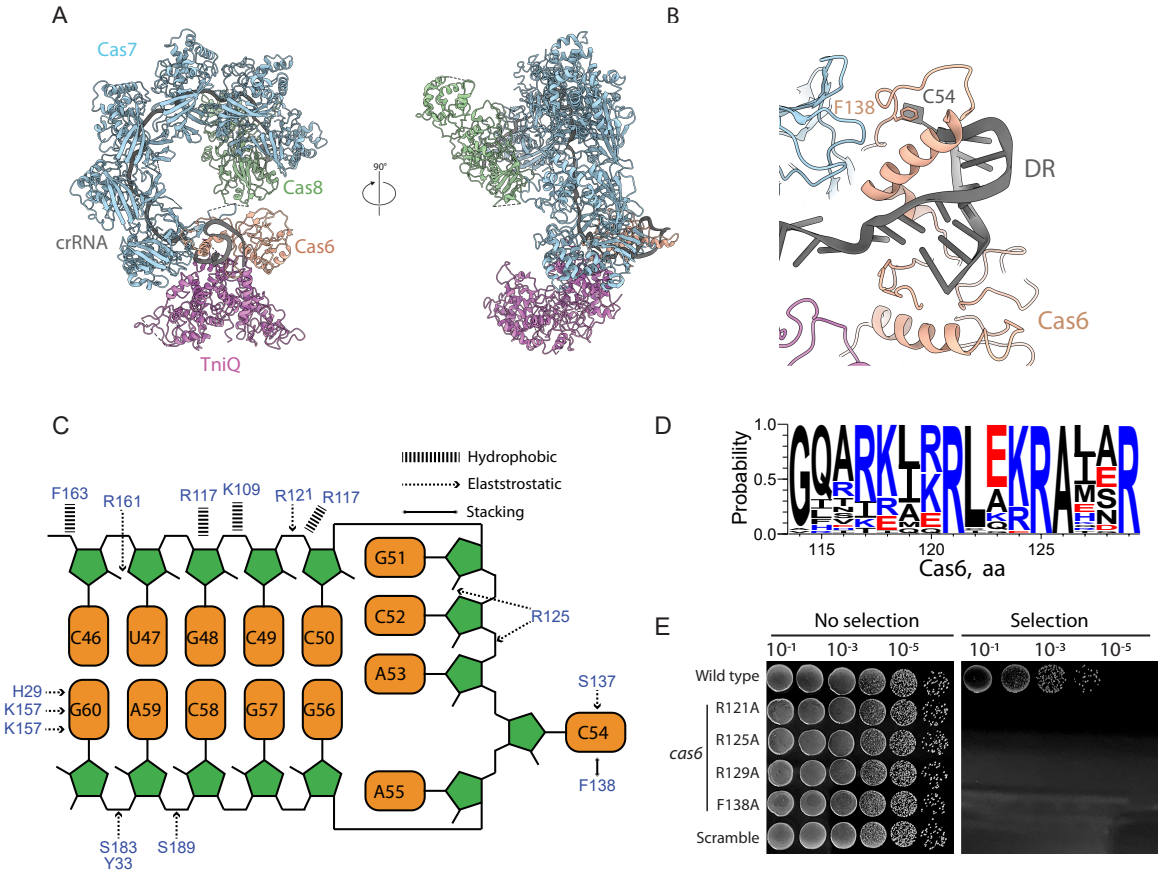


Figure 3: **Cas6 recognizes the crRNA via sequence-independent interactions with the DR.** (A) Structural overview of a type I-F TniQ-Cascade purified with a type III-B crRNA. (B) Close-up view of the direct repeat (gray) and its interactions with Cas6 (salmon). (C) Schematic of the hydrophobic and electrostatic interactions between key Cas6 residues and the direct repeat from the type III-B crRNA. (D) Multiple sequence alignment across all CAST I-F *cas6* genes reveals conserved residues in the arginine-rich helix. (E) Transposition requires Cas6 residues R121, R125, R129, and F138 to coordinate the direct repeat.

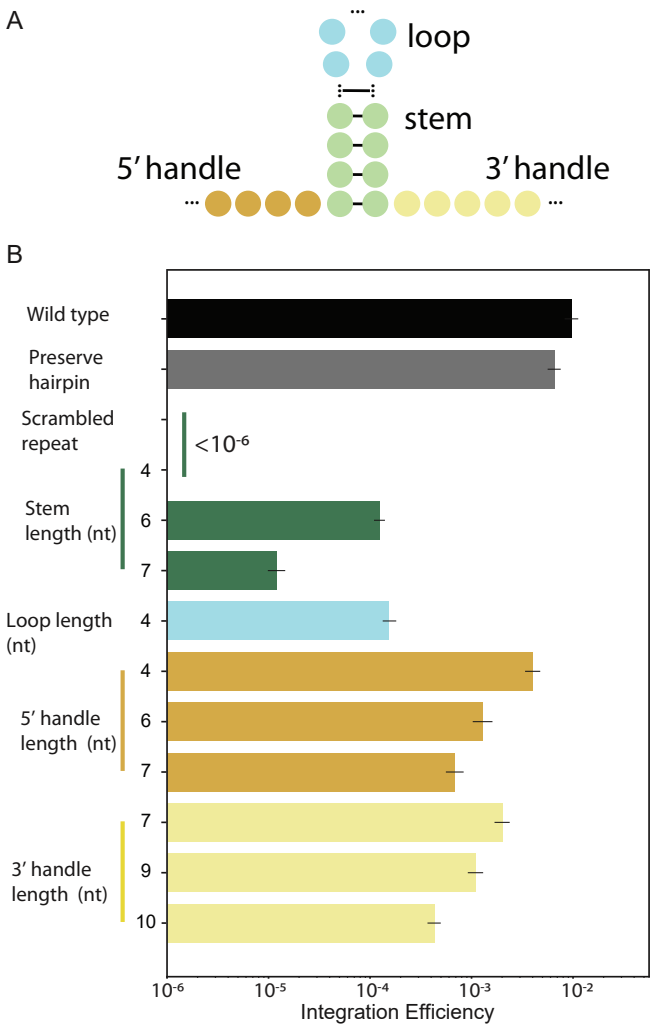


Figure 4: **The crRNA stem-loop length regulates transposition activity.** (A) We tested changes in the direct repeat sequence, stem (green), loop (blue), and handle lengths (5'-orange, 3'-yellow). (B) The effect of each feature on the integration efficiency. Black: DR from the native CAST CRISPR array; gray: a sequence-scrambled DR that preserved the wild type stem loop structure; other colors correspond to the schematic in (A).

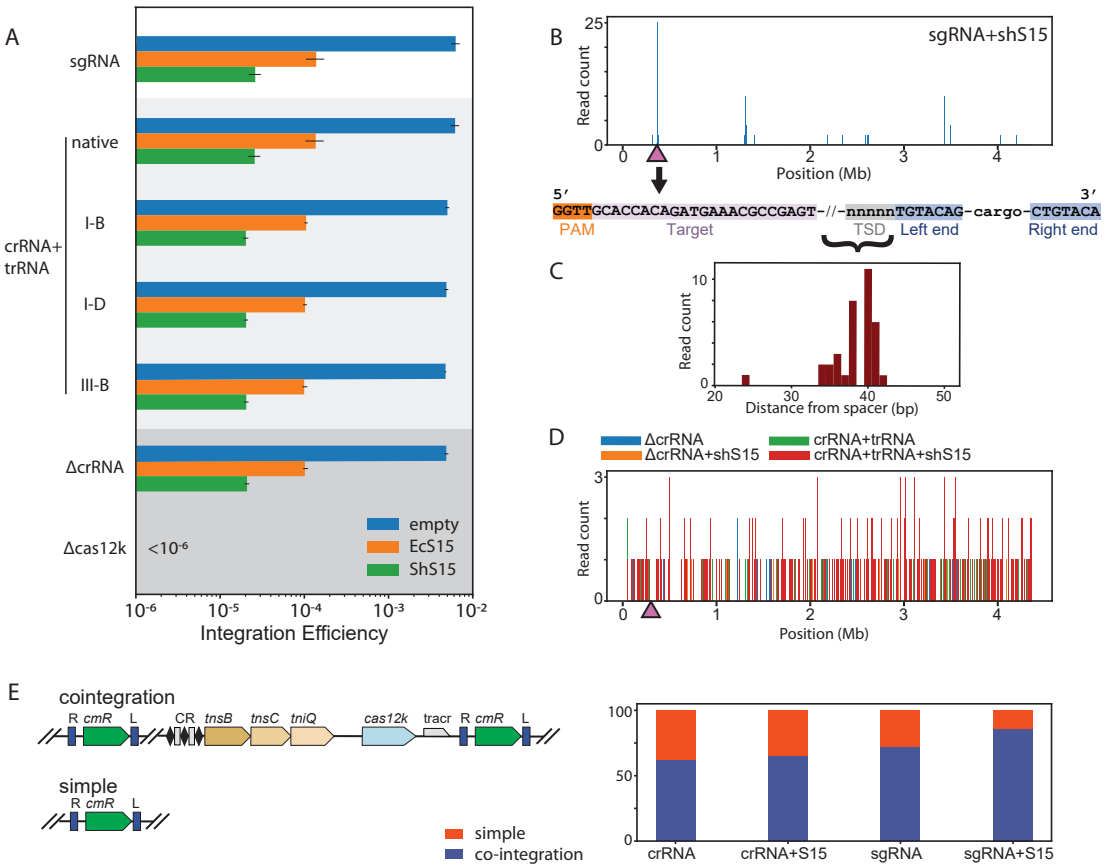


Figure 5: **Unguided integration by type V CASTs.** (A) The small ribosomal protein S15 reduces overall transposition for all crRNA sub-types. (B) (top) Long-read sequencing confirms that most, but not all, insertions are in *lacZ* when ShS15 is over-expressed in the recipient cells. The target site duplication (TSD, gray) and left and right inverted repeats (blue) are also visible at the insertion site. (bottom) (C) The cargo is inserted $\sim$ 35–42 bp away from the target site (purple in B). (D) Substituting the sgRNA with a trac-crRNA results in random transposition through the genome. Adding ShS15 does not increase on-target transposition frequency. (E) Quantification of simple insertion and co-integrate transposition products.

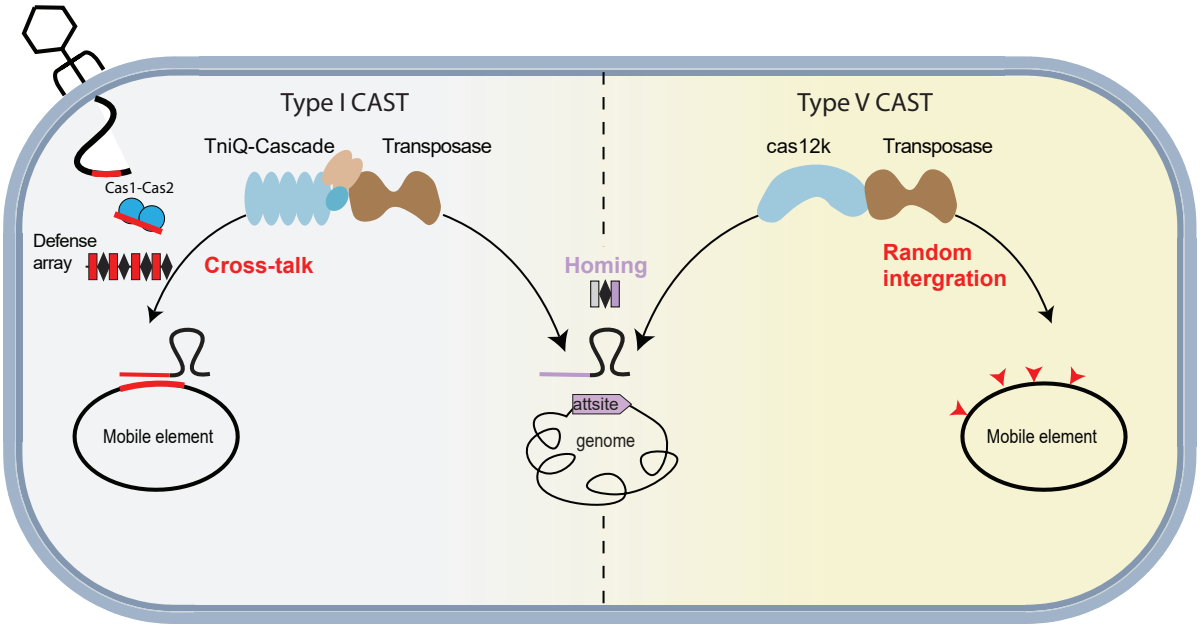


Figure 6: **Dual strategies for horizontal transmission by type I and type V CASTs.** (left) Type I CASTs can co-opt defense associated CRISPR arrays (red) for horizontal transmission. These arrays are updated by defense-associated Cas1-Cas2 integrases. (right) Type V CASTs integrate non-specifically via crRNA-independent transposition. Both systems use a homing spacer for vertical transmission (purple).