Figure S1

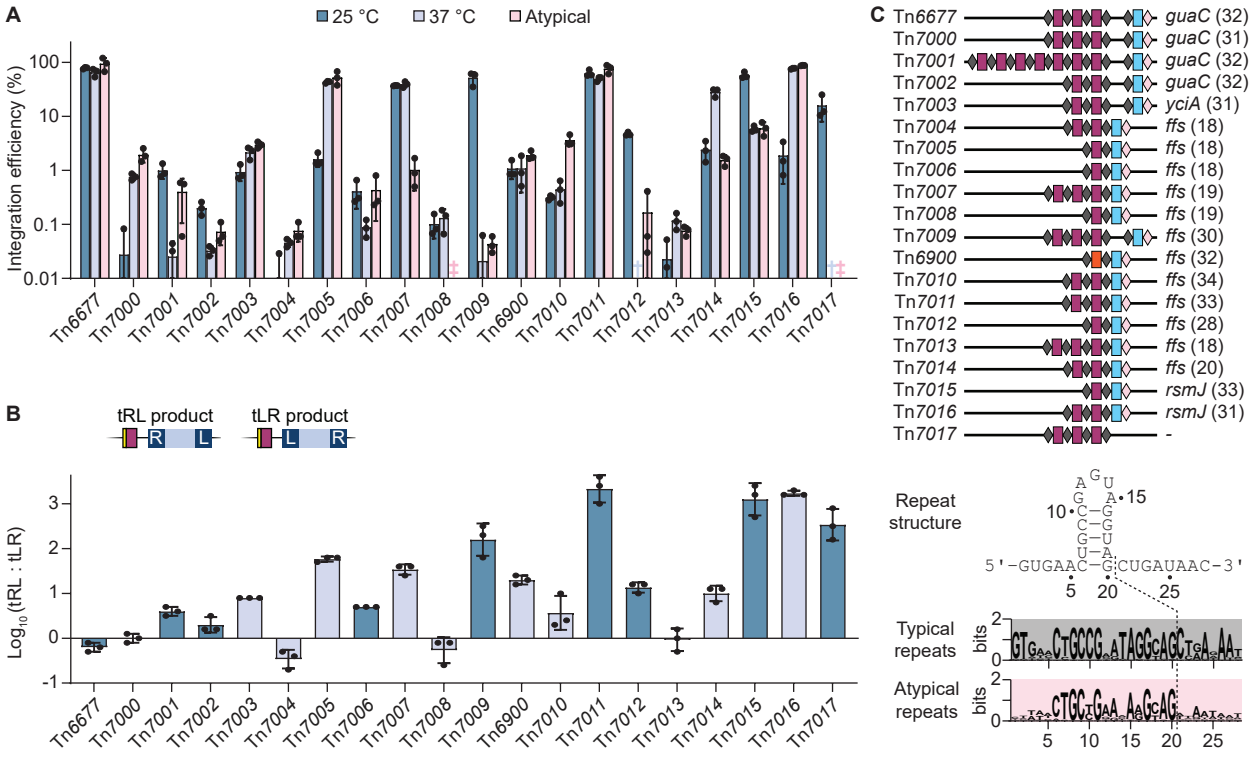

Figure S1: Transposition activity of type I-F3 CRISPR-Tn under different conditions, related to Figure 1
(A) Integration efficiency for all homologs with regular CRISPR repeats at either 25 °C or 37 °C incubation, and atypical repeats (with perfectly matching 32-nt spacers) at 37 °C incubation, measured by qPCR. †, not detected. ‡, not tested. (B) Possible mini-Tn integration orientations (top), and the observed bias (tRL:tLR) for each CRISPR-Tn system under the temperature conditions as presented in (A), determined from qPCR measurements. Integration orientation data may be skewed for low efficiency systems because of detection limitations. (C) Layout of typical (dark grey) and atypical (light pink) repeats within the native CRISPR arrays. Atypical spacers (light blue) and their target genes (*guaC*, *yciA*, *ffs*, and *rsmJ*) are indicated. The orange spacer for Tn*6900* targets a plasmid-encoded protospacer found adjacent to the integration site of Tn*6900*. The bracketed number indicates the length of the atypical spacer. Consensus logos of typical and atypical repeats are shown (bottom), revealing loss of conservation for the last 8 bp of the atypical repeats. Data in (A) and (B) are shown as mean ± s.d. for n = 3 biologically independent samples.

Figure S2

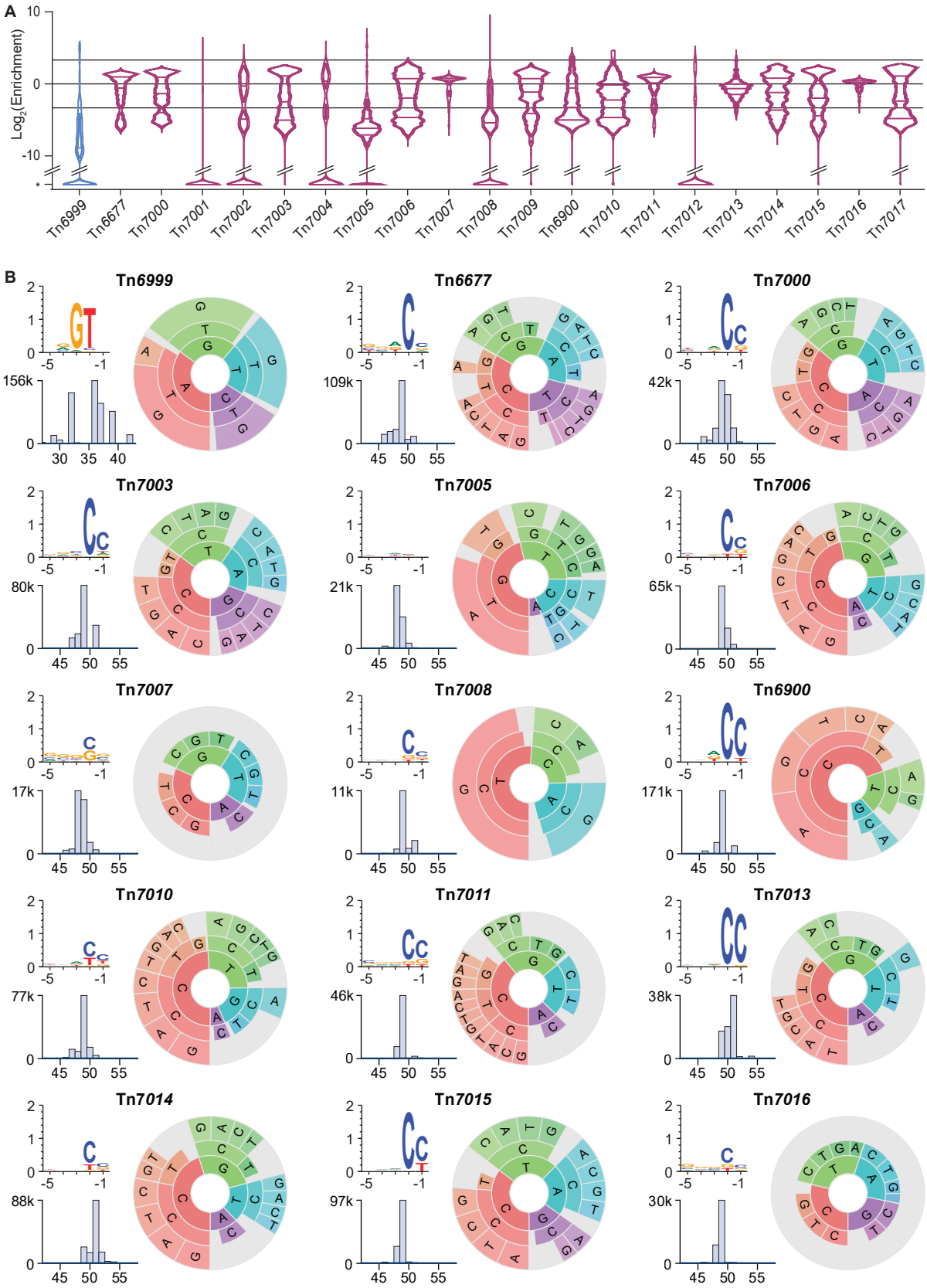
A
Log2(Enrichment)
10
0
-10
*
Tn6999
Tn6677
Tn7000
Tn7001
Tn7002
Tn7003
Tn7004
Tn7005
Tn7006
Tn7007
Tn7008
Tn7009
Tn6900
Tn7010
Tn7011
Tn7012
Tn7013
Tn7014
Tn7015
Tn7016
Tn7017
B
Tn6999
156k
Tn6677
109k
Tn7000
42k
Tn7003
80k
Tn7005
21k
Tn7006
65k
Tn7007
17k
Tn7008
11k
Tn6900
171k
Tn7010
77k
Tn7011
46k
Tn7013
38k
Tn7014
88k
Tn7015
97k
Tn7016
30k

Figure S2: PAM requirements and integration site variation for CRISPR-Tn homologs, related to Figure 2
(A) Violin plots displaying the enrichment of 1042 individual N5 PAM variants after RNA-guided transposition with different CRISPR-Tn homologs. Note that CRISPR-Tn with low overall integration efficiencies at 37°C may show skewed data due to artefactual bottlenecking of the PAM diversity. (B) For each homolog with >0.5% total integration efficiency at 37 °C we show: Top left - a WebLogo for the top 5% enriched PAM sequences. The y-axis shows the information content of a sequence position, in bits. The x-axis shows the base position relative to the target site, with -1 representing the base immediately adjacent to the protospacer. Low sequence conservation represents the absence of sequence restraints and therefore more flexible PAM requirements. Right – a PAM wheel for visualizing PAM sequence conservation including positional information. The inner ring represents the -1 position, the middle ring the -2 position and the outer ring the -3 position of the PAM sequence. Bottom left - integration site distribution for 'CC' (Type I-F3) or 'GTN' (Type V-K) PAMs obtained from the PAM library dataset. The x-axis shows the distance in bp from the 3' end of the target site to the integrated transposon (and therefore includes the 5-bp target site duplication).

Figure S3

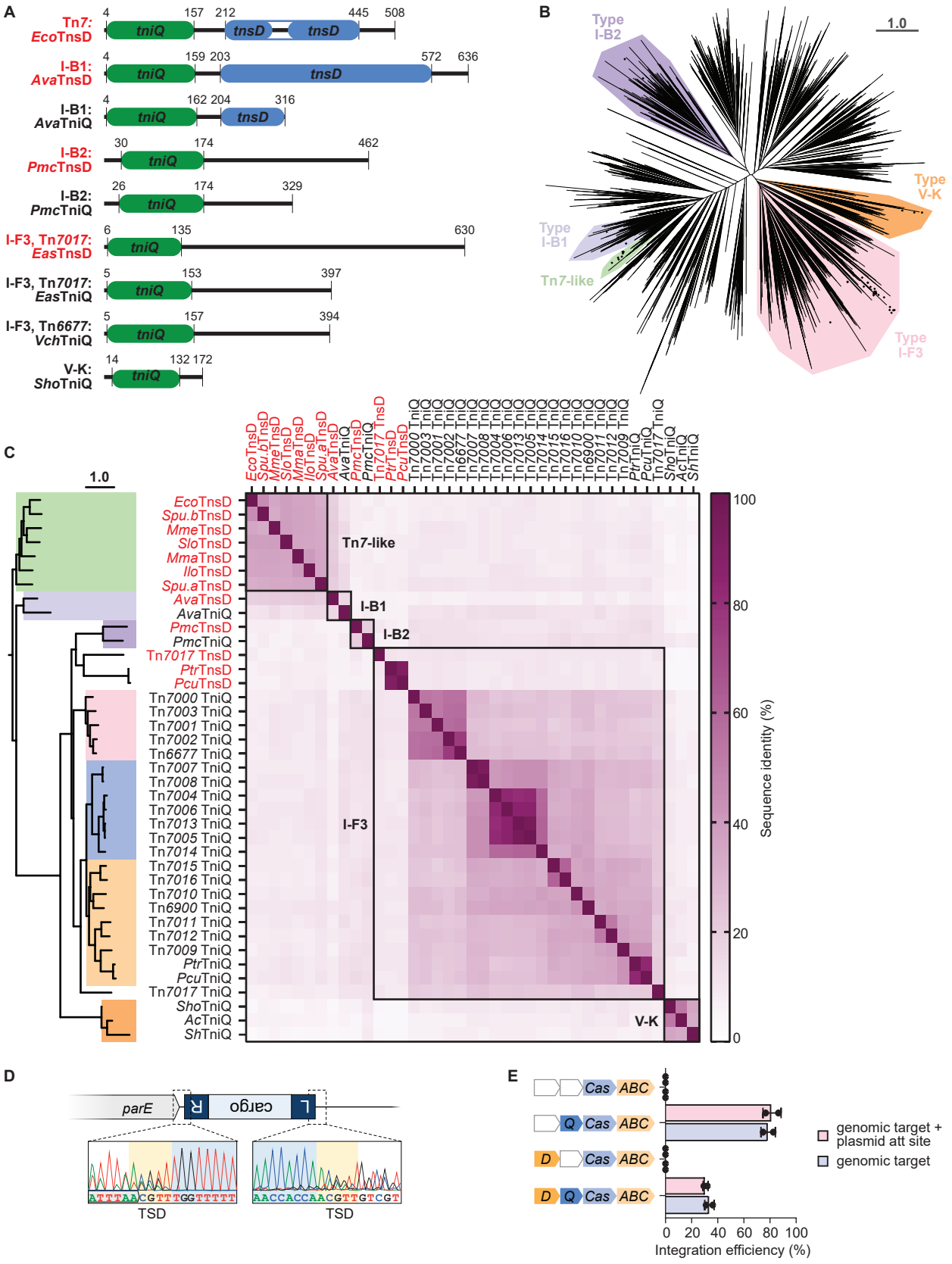
A
Tn7: EcoTnsD
I-B1: AvaTnsD
I-B1: AvaTniQ
I-B2: PmcTnsD
I-B2: PmcTniQ
I-F3, Tn7017: EasTnsD
I-F3, Tn7017: EasTniQ
I-F3, Tn6677: VchTniQ
V-K: ShoTniQ
tniQ
tnsD
B
Type I-B2
Type V-K
Type I-B1
Tn7-like
Type I-F3
1.0
C
Tn7-like
I-B1
I-B2
I-F3
V-K
Sequence identity (%)
D
parE
R
cargo
L
TSD
E
Cas ABC
Q Cas ABC
D Cas ABC
D Q Cas ABC
genomic target + plasmid att site
genomic target
Integration efficiency (%)

Figure S3: Analysis of TniQ-family proteins associated with CRISPR-Tn, related to Figure 3
(A) Size and domain organization of representative TniQ family proteins from characterized CRISPR-associated transposons and Tn*7*. Predicted TniQ (PF06527) and TnsD (PF15978) domains are shown. TniQ proteins involved in RNA-guided transposition are labelled in black while TnsD proteins involved in protein-mediated homing are shown in red. (B) Phylogenetic tree of a large diversity of TniQ proteins based on the TniQ domain as predicted and aligned by hmmersearch. This dataset shows the evolutionary relationship of different Tn*7*-like transposons and highlights the vast diversity of TniQ proteins that remains unexplored (C) Dendrogram of selected TniQ/TnsD proteins (left) and sequence identity matrix (right). Ptr and Pcu represent two type I-F CRISPR-Tn systems from *Parashewanella tropica* and *Parashewanella curva* C51, respectively, and were shown to encode two TniQ family proteins of which the TnsD variants are expected to home to *parE*. Although Tn*7017* TnsD and *Ptr*TnsD/*Pcu*TnsD are predicted to direct transposition to the same site, the proteins show very low sequence conservation. (D) Sanger sequencing data verifying integration at a plasmid-encoded *parE* sequence. Heterogeneity in the sequencing chromatogram at the transposon-target boundary is the result of amplification from a bacterial culture rather than individual clones and indicates flexibility in the exact integration site. (E) Integration efficiency at the genomic protospacer with or without pTarget present, under different gene deletion conditions. Data are shown as mean ± s.d. for n = 2 biologically independent samples.

Figure S4

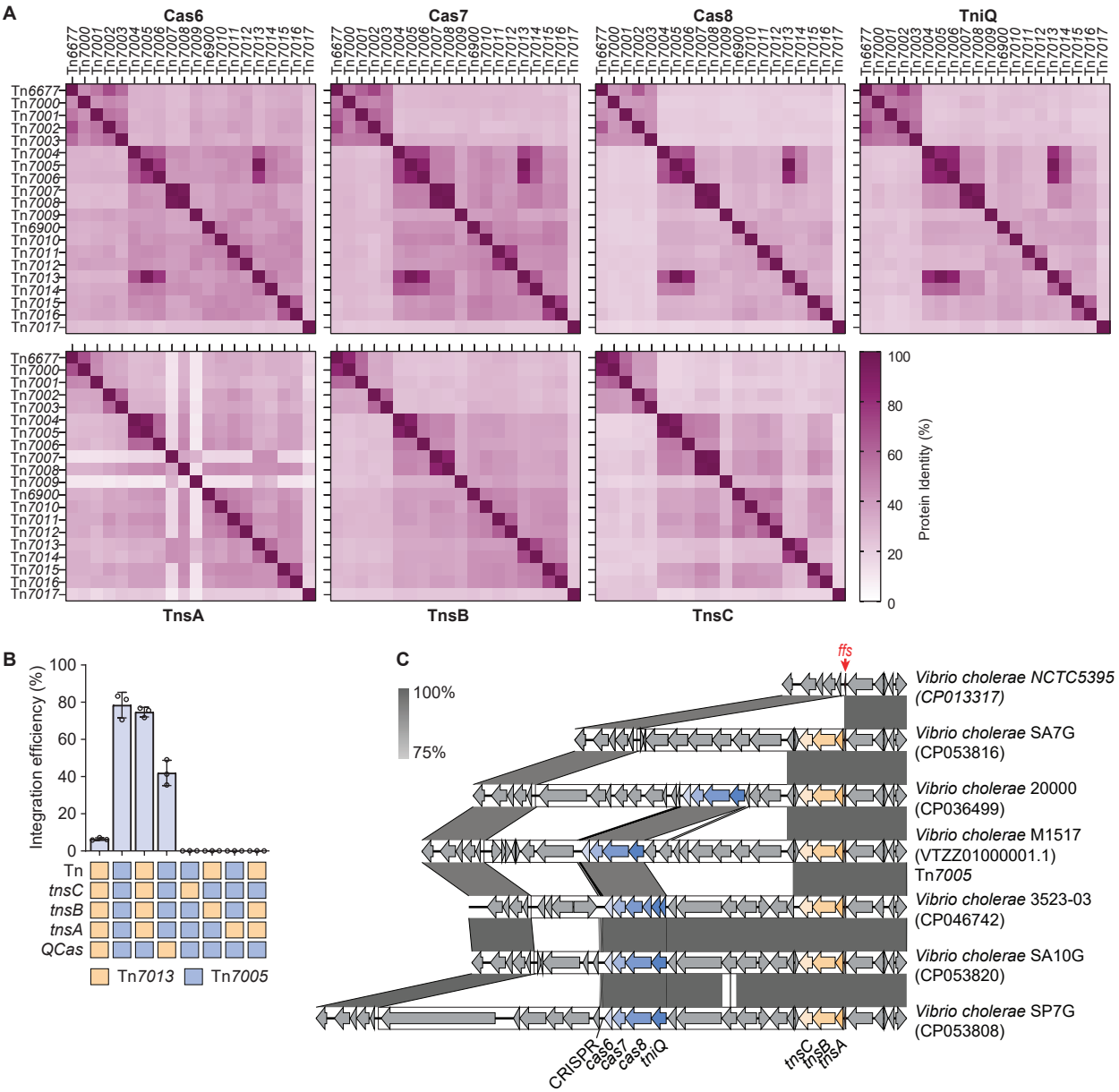

Figure S4: Analysis of co-evolution and recruitment of CRISPR-Tn machinery, related to Figure 4

(A) Sequence comparison of individual CRISPR-Tn protein components involved in RNA-guided transposition. Sequence identity patterns are consistent within the targeting and transposition machinery individually, but show discrepancies when cross-compared. (B) Integration efficiency for different combinations of *tniQ-cas876* (*QCas*) operons and *tnsA*, *tnsB*, and *tnsC* genes or operons from Tn*7005* and Tn*7013*, as measured by qPCR. Data are shown as mean ± s.d. for n = 3 biologically independent samples. (C) Analysis of the *ffs* locus from seven different *V. cholerae* isolates, with syntenic genes highlighted in gray blocks. Five of the loci are occupied by closely related CRISPR-transposons that exhibit similar QCascade and TnsABC operons; the highly dissimilar cargo genes elsewhere in the transposons highlight recombination events that caused payload diversity. Similarly, the transposon of *V. cholerae* SA7G shows high sequence homology of *tnsABC* compared to the other transposons, however, it lacks the CRISPR-Cas machinery altogether. Since each of these transposons likely homed to the *ffs* locus through the conserved atypical crRNAs, this transposon likely lost its CRISPR-Cas module, in concordance with high rates of cargo turnover.

Figure S5

Figure S5: Genomic layout and cargo analysis of native CRISPR-transposons, related to Figure 5
CRISPR-Tn encode numerous cargo genes in addition to the transposase and CRISPR-Cas operons. The native genomic layout of homologs in this study are shown, with the transposon highlighted with a light blue box and shown with the right (R) end on the right side of the figure. Transposition and DNA targeting modules are shown in yellow and blue, respectively. Putative defense systems are highlighted in red when the complete gene set was found. Roman numerals in brackets indicate the predicted subtype of the defense system. A large diversity of defense systems are found, suggesting that CRISPR-Tn systems provide a benefit to bacterial hosts by providing innate immunity from phage predation.

Figure S6

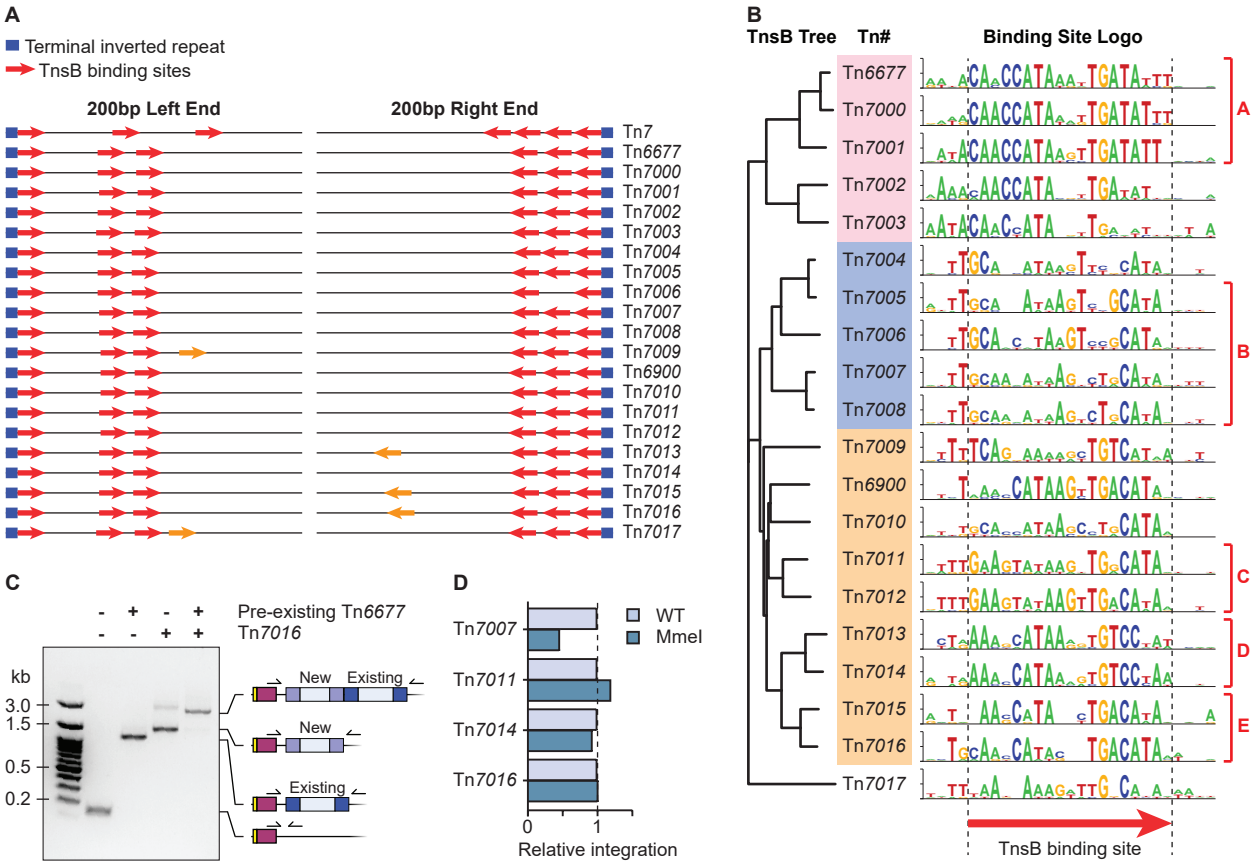

Figure S6: Transposon end sequence analysis and orthogonality of CRISPR-transposons, related to Figure 6
(A) Schematic representation of TnsB binding sites within the transposon ends reveals spatial conservation among different CRISPR-Tn. Blue box, terminal inverted repeat (TIR); red arrow, predicted TnsB binding site; yellow arrow, sequence with some similarity to predicted TnsB binding sites. (B) Dendrogram of TnsB proteins (left) and consensus logos for predicted TnsB binding sites for each CRISPR-Tn (right). Compatibility groups are indicated based on enrichment data shown in Figure 6B, and align with both TnsB phylogeny and predicted TnsB binding site consensus sequences. (C) External-external PCR analysis of integration events with Tn*7016* in *E. coli* BL21(DE3) strains with or without a pre-existing mini-Tn*6677* located downstream of the target site, verifying that orthogonal CRISPR-Tn systems can be used for side-by-side insertions of genetic payloads. (D) qPCR data to verify that mutations within the transposon ends to facilitate MmeI recognition sites for Tn-seq do not drastically affect integration efficiencies of the selected systems.