IS200/IS605 Family-Associated TnpB Increases Transposon Activity and Retention

Davneet Kaur[1], Thomas E. Kuhlman[2,3,4]*

[1]Department of Physics, University of Illinois Urbana Champaign; Urbana, IL, USA.
[2]Department of Physics, University of California Riverside; Riverside, CA, USA.
[3]Biophysics Program, University of California Riverside; Riverside, CA, USA.
[4]Microbiology Program, University of California Riverside; Riverside, CA, USA.

* To whom correspondence should be addressed. Tel: +1 (858) 531-4386, Email: thomas.kuhlman@ucr.edu

ABSTRACT

The IS200/IS605 family are abundant insertion sequences associated with one of the most numerous genes in nature, *tnpB*. Sequence homologies suggest that transposon-borne TnpB may be an evolutionary precursor to CRISPR-Cas system effector proteins, and TnpB has been shown to function as a Cas-like programmable DNA endonuclease. However, the role of TnpB in transposition and its contributions to genome dynamics remain elusive. Using a suite of fluorescent reporters coupled to transposition, we show that TnpB increases IS608-transposon activity and prevents transposon loss from genomes. Analyzing our results through a mathematical model of transposon dynamics, we discuss the multifaceted roles TnpB may play in transposon regulation. The mutually beneficial transposon-TnpB interaction may explain the prevalence of *tnpB*, creating conditions for the proliferation and appropriation of TnpB's RNA-guided endonuclease activity for adaptive immunity.

INTRODUCTION

Transposable elements (TEs) are mobile DNA sequences that can move around or create copies of themselves in an organism's genome (*1*). TEs are major contributors to disease, development, and evolution (*2–13*). As significant sources of mutations, revealing their network of interactions and dynamics is a key factor in understanding the dynamics of evolution. Together with DNA modification systems and auxiliary genes, TEs form a tangled web of interactions that are integral to the proper functioning of organisms (*9*, *14–16*). In some cases, TE activity is even demonstrably necessary for the viability of an organism (*17*). Yet, our understanding of the scope of TE interactions and dynamics remains limited.

IS608 is an autonomous TE from *Helicobacter pylori* that is a representative of the IS200/IS605 family of insertion sequences, which all transpose through similar mechanisms. The IS200/IS605 family is widely distributed, with 153 members spread over 45 genera and 61 species of eubacteria and archaea[1]. IS608 codes for the transposase TnpA that executes recognition, excision, and reintegration (*19–22*) and contains an additional gene of thus far unknown function, *tnpB* (*23*). TnpB proteins are an extremely abundant but poorly characterized family of nucleases that are encoded by many bacterial and archaeal TEs (*24*, *25*) which often contain only *tnpB* (*26*). More than 1 million putative *tnpB* loci were recently identified in bacterial and archaeal genomes, making it one of the most common prokaryotic genes (*27*). TnpB's abundance and conserved association with transposons are indications of its importance in the regulation of TEs and in the

evolution of prokaryotic systems. Yet, the function of TnpB in transposition remains unclear. We address this question in this work.

TnpB contains three major domains: an N-terminal putative helix–turn–helix (Pfam: HTH_OrfB_IS605), a longer and more variable central domain (Pfam: OrfB_IS605) and a C-terminal zinc finger domain (Pfam: OrfB_Zn_ribbon) (*28*). The N-terminal domain contains a RuvC-like motif. TnpB's RuvC endonuclease site has been speculated to play a role in its regulation of transposition (*28*, *29*). Interestingly, the RuvC domain shares sequence similarity to a newly discovered group of non-autonomous transposons, ISC, that are the closest homologues of Cas9, positioning TnpB as a possible ancestor of Cas9 proteins (*24*, *30*, *31*) and providing further possible insight into TnpB's activity. Based on additional phylogenetic work, TnpB has also been proposed to be an ancestor of Cas12 proteins (*27*).

Cas proteins are a part of bacterial defense mechanisms against foreign DNA and RNA (*32–34*). Using a guide RNA, they can locate and cut specific DNA sequences using their endonuclease sites, which include a RuvC endonuclease site. In their mechanisms, Cas9-RuvC and Cas12-RuvC are responsible for nicking the non-target strand of the DNA (*32*). However Cas12b, which contains just the RuvC endonuclease site, has been engineered to make double stranded cuts with only its RuvC site (*35*). Interestingly, TnpB has been determined to be a programmable RNA-guided endonuclease that under different conditions will make either dsDNA or ssDNA breaks (*27*, *29*). This programmability and diversity of function suggests that TnpB may be a promising candidate for development as a genome editing tool.

To reveal the function of TnpB, we created a variety of inducible IS608 that are tagged with fluorescence reporter genes to visualize and quantify TE protein levels and activity (**Figures 1A,S1A**) and introduced them to *E. coli* (*36*). Based on TnpB's endonuclease activity, its phylogenetic connection to Cas systems, and our observations of its effect on transposition of IS608, we discuss the multifaceted roles it may play in the regulation of transposition. Insertion sequences are in general thought to be more damaging than beneficial and only provide a temporary selective advantage to their host (*37*, *38*). We discuss the possibility that TnpB may have been essential in the maintenance of TEs in bacteria by increasing TE retention. Furthermore, the IS200/IS605 family of transposons may have enabled the proliferation of TnpB proteins throughout bacteria, as they often carry only *tnpB*.

MATERIAL AND METHODS

Reagents

Media: Supplemented M63: Ammonium sulphate ((NH4)2SO4, Millipore Sigma, A4418), Potassium phosphate monobasic (KH2PO4, Millipore Sigma, P9791), Iron(II) sulfate heptahydrate (FESO4 7H2O, Millipore Sigma, F8633), Magnesium sulfate (MGSO4, Millipore Sigma, M2643, Thiamine hydrochloride (Millipore Sigma, T1270), Glycerol (Millipore Sigma, G5516), Kanamycin sulfate (Millipore Sigma, 60615), Chloramphenicol (Millipore Sigma, C0378), Anhydrotetracycline hydrochloride (Millipore Sigma, 37919), IPTG (ThermoFisher, 15529019), and Potassium hydroxide (KOH, Millipore Sigma, 221473). Luria-Bertani: Tryptone (ThermoFisher, 211705), Sodium Chloride (NaCl, Millipore Sigma, S98880), and Yeast Extract

(ThermoFisher, 212750). *qPCR Materials*: CFX96 Touch Real Time PCR System (BIO-RAD), SsoAdvanced Universal SYBR Green Supermix (BIO-RAD, 172572), UltraPure Dnase/Rnase-Free Distilled Water (Invitrogen, 10977015), 0.2 ml 8-Tube PCR Strips (BIO-RAD, TBS0201), and 0.2 ml Flat PCR Tube 8-Cap Strips (BIO-RAD, TCS0803). *Plate Reader Materials*: CLARIOStar (BMG Labtech), and Costar 96-well microplate (Corning, 3596). *Molecular Cloning Enzymes*: Phusion Flash Master Mix (ThermoFisher, F548), NheI-HF (New England Biolabs, R3131), XhoI (New England Biolabs, R0146), DpnI (New England Biolabs, R0176), KpnI-HF (New England Biolabs, R3142), I-SceI (New England Biolabs, R0694), SpeI-HF (New England Biolabs, R3133 and DMSO (Millipore Sigma, D8418)

Biological Resources

Strains and Media

Experiments were performed using E. coli K-12 MG1655 Δlac (NCBI:txid511145) (39–43). Molecular cloning and plasmid manipulations were performed using E. coli NEB turbo as a host strain. Cells for measurement of the TE excision response function were grown in M63 minimal medium [100 mM KH2PO4, 15 mM (NH4)2SO4, 1 mM MgSO4, 1.7 μM FeSO4, 5e-5% (wt/vol) thiamine, pH adjusted to 7.0 with KOH] with 0.5% (vol/vol) glycerol as carbon source. Antibiotics were added to the medium as appropriate for plasmid maintenance, and different concentrations of aTc (Sigma Aldrich) and IPTG (ThermoFisher) were added to induce transposase and TnpB production respectively.

Inducible Transposon Construction

The low copy number plasmid pJK14 (*44*) was used to host the TE in all experiments. pJK14 has a pSC101 replication origin. Plasmid copy number is tightly controlled through the positive feedback of the plasmid-encoded protein RepA (*45*). Additionally, pJK14 is actively segregated to daughter cells through the pSC101 *par* system (*46*). We create an inducible IS608 which is fluorescently tagged and dub it Tn4rev. Plasmid pJK14-Tn4rev was designed using Vector NTI software (Life Technologies) and synthesized *de novo* by GENEWIZ Gene Synthesis Services (GENEWIZ, Inc.) (**Figure S1A.i**). The $P_{Ltet\text{-}01}$ promoter is induced with aTc (*47*) to generate TnpA translationally fused to Venus fluorescent protein. The use of this inducible promoter allows for simple and precise control of TnpA levels within individual cells (*36*). On the plasmid, the TE splits the −10 and −35 sequences of a strong constitutive P_{LacIQ1} promoter (*48*) for the expression of the blue reporter *mCerulean3* (*49*). This promoter generates Cerulean fluorescent protein upon successful TE excision. TnpA then attempts to insert the TE at a new site with the specific sequence TTAC.

Plasmid pZA31- $P_{LlacOid}$-*mCherry-TnpB* was designed using Vector NTI software (Life Technologies) and synthesized *de novo* by GENEWIZ Gene Synthesis Services (GENEWIZ, Inc.). The mCherry-TnpB cassette was inserted into pZA31 at the NheI-HF and XhoI cut sites. As many other TE-associated proteins exhibit strong *cis*-preference (*50*), we introduced *tnpB* to this system in both *trans* and *cis* (**Figures S1A.ii & iii** respectively). For the *trans* combination, $P_{LlacOid}$ is independently induced with IPTG to produce TnpB.

For the *cis* combination, the mCherry-TnpB sequence was amplified from pZA31- $P_{LlacOid}$-mCherry-TnpB and inserted into pJK14-Tn4rev plasmid at the NheI-HF cut site to generate pJK14-Tn4rev-mCherry-*tnpB*. Hence, TnpB is induced simultaneously with the transposon with induction of the $P_{Ltet\text{-}01}$ promoter with aTc (**Figure S1B-C,** see **Supplementary Information SII.i. Introduction of TnpB to the Inducible Transposon**). In **Figure S1A.ii & iii** both strains have *mCherry* (red) translationally fused to *tnpB*, providing a measure of TnpB levels. By removing the excision sites for Tn4rev, RE and LE, we create an immobile version of the TE (**Figure S1A.iv)**, CZ071 pJK14-$P_{Ltet\text{-}01}$-*tnpA* which is introduced *in trans* with either TnpB, pZA31-$P_{LlacOid}$-*mCherry-tnpB* or its negative control plasmid pZA31-$P_{LlacOid}$-SmR. Control strain in **Figure S1A.v** lacks mCherry fusion to confirm that this fusion does not affect the activity of TnpB. We observe identical Venus-TnpA and Cerulean-excision response levels for the strains with and without *mCherry* fusion to the *tnpB* introduced *in cis* with Tn4rev. That the Venus-TnpA fusion does not significantly affect activity of TnpA was confirmed in previous work. All primers used for construction of transposon sequences are listed in **Table S1.** Phusion Flash Master Mix was used for all PCR reactions for plasmid construction and confirmation which are described in further detail in **Supplementary SI.i. Plasmid Construction**. Links to plasmid sequences are listed in **Table S2**.

Fluorescence Measurements

To measure the TE response functions (shown in **Figures 1B-C, 2A, S1B-C, S2, S3A-B, S4**), MG1655 Δ*lac* cells carrying the indicated version of the TE and pZA31 plasmid were grown overnight in LB + 25 µg/mL kanamycin + 34 µg/mL chloramphenicol. The optical density at 600 nm (OD600) of this culture was measured with a Bio-Rad SmartSpec Plus spectrophotometer, and an appropriate volume of the culture was added to 2.5-mL of M63 minimal medium + 0.5% (vol/vol) glycerol + 25 µg/mL kanamycin + 34 µg/mL chloramphenicol in a 20-mm glass test tube to yield a calculated initial OD600 = 0.0008. This tube was grown at 37°C in a New Brunswick C76 water bath shaker with vigorous shaking until it reached an ~OD600 = 0.15–0.20; at this OD, the cells are within the exponential phase of growth and have undergone ~7-8 doublings. The culture was then diluted to OD600 = 0.01 in M63 minimal medium + 0.5% (vol/vol) glycerol + 25 µg/mL kanamycin + 34 µg/mL chloramphenicol and the appropriate concentration of IPTG (0uM, 10uM, 20uM, 50uM, 100uM, 200uM or 2000 uM) was added. The culture was then partitioned into 0.2mL aliquots for each well of a Corning Costar 96-well microplate (Clear, Round well, flat bottom: Well volume: 360uL, Cell growth area: 0.32cm2, TC-Treated, Sterile, Individually wrapped, 3596). Each plate consisted of 8 replicate wells of 12 different aTc concentrations (0ng/mL, 1, 2ng/mL, 3ng/mL, 5ng/mL, 9ng/mL, 20ng/mL, 30ng/mL, 50ng/mL, 80ng/mL, 100ng/mL and 200ng/mL). aTc and IPTG titrated the cells with transposase and TnpB inducers respectively. The plate was loaded into a BioRad Clariostar plate reader for measurements

The cultures were maintained at 37°C in the temperature-controlled environmental chamber of the plate reader with shaking at 300rpm. Optical density measurements and readings in three fluorescent channels, mCherry, Venus, and mCerulean3, were made every 10 minutes. Fluorescent excitation measurements

were made at 561 nm (mCherry excitation, mCherry-TnpB), 514 nm (Venus excitation, Venus-TnpA levels), and 457 nm (mCerulean3 excitation, excision reporting), in that order.

The population-average fluorescence per cell was determined for each well of each experiment by applying a linear fit to the fluorescence vs. the optical density in the exponential phase of growth between ~OD600 = 0.02 to 0.2, between generation 2 to 5, for mCherry, Venus, and mCerulean3. Fluorescence values for each combination of aTc and IPTG concentration combination were averaged over all replicate experiments (3 experiments per inducer combination with 8 replicates per experiment).

Quantification of Fluorescence Per Cell

The Venus, Cerulean and Cherry fluorescences per cell were determined as described for each combination of aTc and IPTG quantity for the strains: 1) Negative Control Strain for background subtraction, MG1655 pJK14 pZA31-$P_{LlacOid}$-SmR, 2) Transposons Tn4rev only, MG1655 pJK14-$P_{Ltet\text{-}01}$-Tn4rev pZA31-$P_{LlacOid}$-SmR, 3) Trans-combination of Tn4rev and TnpB, MG1655 pJK14-$P_{Ltet\text{-}01}$-Tn4rev pZA31-$P_{LlacOid}$-*mCherry-tnpB*, 4) Cis-combination of Tn4rev and TnpB, MG1655 pJK14-$P_{Ltet\text{-}01}$-Tn4rev-*mCherry-tnpB* pZA31-$P_{LlacOid}$-SmR, and 5) Immobilized strain without TnpB, CZ071 pJK14-$P_{Ltet\text{-}01}$-*tnpA* pZA31-$P_{LlacOid}$-SmR and 6) Immobilized strain with TnpB, CZ071 pJK14-$P_{Ltet\text{-}01}$-*tnpA* pZA31-$P_{LlacOid}$-TnpB.

Background subtractions corresponding to the respective aTc and IPTG concentrations were subtracted from the fluorescence measurements to determine fluorescence per cell for each “color”. The resulting Venus and Cerulean fluorescence values per cell were plotted vs. aTc and fitted to a Hill function of the below form. The resulting mCherry fluorescence values per cell were plotted vs. IPTG and fitted to a Hill function (**SI.ii. Fluorescence Per Cell Fitting to a Hill Function**, **Equation S1.2.1**). The quantitative features of the responses are extracted. We use the inflection point (**Equation S1.2.2**) to compare relative initiation points of the Hill fits. The slope at the inflection point is used as a measure of the sensitivity of the system to induction (**Equation S1.2.4**). The error in the location and slope of the inflection is determined using the 95% confidence interval of the Hill fit variables and their error propagation equations (**Equations S1.1.5 – S1.1.6**).

Calculation of Cumulative Cerulean Fluorescence Per Cell

In previous work, we demonstrate that excision events are followed by spikes of Cerulean fluorescence. Thus, we measure excision rate per cell as the slope of Cerulean fluorescence vs. optical density, as described in the section **Quantification of Fluorescence per Cell**. To determine the total number of excision events from the plasmids in the exponential phase, we integrate Cerulean fluorescence per cell (AU) over time in exponential growth phase (**Figure S7**, see **SI.iii. Calculation of Cumulative Cerulean Fluorescence Per Cell)**

Quantitative-PCR Protocol

qPCR was used to determine relative plasmid copy number, unexcised plasmid-transposon number and total transposon number for cells induced with varying amounts of aTc and IPTG inducer concentrations. qPCR was performed using a Bio-Rad CFX96 Touch Real-Time PCR thermal cycler with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). All four primer pairs listed in **Table S3** were designed to have similar melting temperature values so that the successful annealing temperatures ranges of all reactions overlap. All reactions were performed concurrently on DNA templates from the same batches for uniform concentrations across all qPCR reactions. MG1655 Δ*lac* pJK14-$P_{Ltet\text{-}01}$-Tn4rev pZA31-$P_{LlacOid}$-*mCherry-tnpB* were grown overnight in LB + 25 µg/mL kanamycin + 34 µg/mL chloramphenicol. The optical density at 600 nm (OD600) of this culture was measured with a Bio-Rad SmartSpec Plus spectrophotometer, and an appropriate volume of the culture was added to 2.5-mL of M63 minimal medium + 0.5% (vol/vol) glycerol + 25 µg/mL kanamycin + 34 µg/mL chloramphenicol in a 20-mm glass test tube to yield a calculated initial OD600 = 0.008. The appropriate concentrations of aTc (0 ng/mL or 200ng/mL) and IPTG (0uM or 2000 uM) were added. This tube was grown at 37°C in a New Brunswick C76 water bath shaker with vigorous shaking until it reached an ~OD600 = 0.15–0.20; at this OD, the cells are within the exponential phase of growth and have undergone ~4-5 doublings. From this culture, 0.5mL was centrifuged, the supernatant was discarded and resuspended in UltraPure DNase/RNase-Free Distilled Water (Invitrogen) water. This suspension was used to generate six concentrations of a 5× dilution series of cells for use in qPCR. Amplification reactions were performed in 0.2 ml 8-Tube PCR Strips (Bio-Rad™, clear #TBS0201) with 0.2 ml Flat PCR Tube 8-Cap Strips (Bio-Rad™, optical, ultraclear, #TCS0803). Optimum amplification conditions were determined for each by amplifying six concentrations of a 5× dilution series of MG1655 pJK14-Tn4rev pZA31-*mCherry-tnpB* using two-step amplification with a thermal gradient of 55–75°C. The optimum reaction conditions were determined to be: Annealing temperature: 59°C and Extension time: 30 seconds and used for all 4 reactions.

Primer pair (i) for number of cells: MG1655-*nth* F and MG1655-*nth* R (**Table S3**) generates a 196bp amplicon from the *nth* gene of MG1655 under the optimum conditions with T_m = 86.0°C and efficiency ε = 93.74 ± 1.25% [defined by $N(n) = N_0(1+\varepsilon)^n$, where *n* is cycle number]. Primer pair (ii) for number of plasmids: DK-pJK14-qPCR F and DK-pJK14-qPCR R (**Table S3**) generates a 367bp amplicon from the pJK14-$P_{Ltet\text{-}01}$-Tn4rev plasmid with T_m = 82.0°C and efficiency ε = 96.00 ± 0.76%. Primer pair (iii) for number of unexcised plasmids: pJK14-RE\IP-qPCR F and pJK14-RE\IP-qPCR R (**Table S3**) generates a 226 bp amplicon from pJK14-$P_{Ltet\text{-}01}$-Tn4rev with T_m = 81.0°C and efficiency ε = 95.23 ± 0.66%. Primer pair (iv) for total number of transposons: *tnpA-venus*-qPCR F and *tnpA-venus*-qPCR R (**Table S3**) generates a 148 bp amplicon from pJK14-Tn4rev with T_m = 87.0°C and efficiency ε = 99.15 ± 0.56%.

Statistical Analysis

Statistical analysis and error calculation of qPCR data is described in sections **SI.iv. Quantitative-PCR Analysis** and **SI.v. Error Calculation for Quantitative-PCR** of the Supplementary Information respectively. Two-sided P-values listed in **Tables S7 and S8** were calculated by conducting a Student's T-

Test. For the fluorescence data, standard error or the mean was determined for data for each fluorescence color and induction level of aTc and IPTG. The 95% confidence intervals for the fitted parameters were extracted from the Hill function fit to the induction curves (**Figures 1B,S2,2A,S4,S3A-B,S1B-C**) and used to determine the standard error of the mean for data in **Figures 1D-F,2C-E,S3C-E**.

Sequence Data Resources

Sequences for Tn4rev, PLlacOid-mCherry-TnpB, and PLlacOid-SmR are available from Genbank with accession numbers OP581959, OP581957, and OP581958 respectively.

Data Base Referencing

Links to transposon sequences are included in the Supplementary Information **Table S2**.

RESULTS

TnpB tunes TnpA concentration to increase excision

For strains with active TEs, Venus-TnpA concentration and the number of plasmid-excision events (Cumulative Cerulean, CC) increase with TnpB (**Figures 1B-D,2A-C,S2**) and become more sensitive to induction with aTc (**Figures 1E-F,2D-E,** see Supplementary Sections **SII.ii., SII.iii.** and **SI.iv.** for further analysis). Excision of the TE happens in response to the TnpA concentration in cells. Hence, we quantify the excision response as a function of Venus-TnpA concentration for all concentrations of TnpB (**Figures 2B,S5**, **Table S5**, Supplementary Section **SII.v.**). As the concentration of transposase increases, the number of excision events increases proportionally (**Figure 2B**). The data collapse to a single curve, suggesting that TnpB does not directly affect excision. Rather, TnpB tunes the concentration of TnpA to higher values, which then catalyzes greater excision. A linear regression fit finds a slope of 545.79 CC-AU/V-AU with an R^2 value of 0.913 (**Figure 2B**). Introduction of TnpB to an immobilized version of the transposon does not result in an increase in Venus-TnpA (**Figure 1C**), suggesting that TnpB is affecting transposition and not TnpA stability to achieve higher Venus-TnpA concentrations. We hypothesize that TnpB is increasing the number of transposons per cell, either through better retention of transposons or replication.

TnpB increases transposon excision and retention

E. coli with TnpB introduced *in trans* with Tn4rev were grown with and without TnpA induction (0 or 200ng/mL aTc) and with and without TnpB induction (0 or 2000μM IPTG). qPCR reactions (described in **Quantitative-PCR Protocol**) were performed to measure relative quantities per cell of plasmid numbers, total transposon numbers, and plasmid-borne transposon numbers. We find that the total plasmid number per cell is consistent for all concentrations of inducers (**Figure 2F**). The total transposon number drops when cells are grown with TnpA induction alone relative to when transposition is not induced (**Figure 2G**,

left set of bars). Simultaneous induction with TnpB improves transposon retention at values closer to when transposition is not induced (**Figure 2G**, right set of bars). Induction of TnpB alone has no significant effect on transposon count (**Figure 2G, Table S7**). Finally, the number of transposons remaining in the original plasmid number drops when cells are grown with TnpA induction alone relative to when transposition is not induced (**Figure 2H**, left set of bars). Simultaneous induction with TnpB further reduced plasmid-transposon numbers, indicating an increase in excision with TnpB (**Figure 2H**, right set of bars). Induction of TnpB alone has no significant effect on plasmid-transposon count (**Figure 2H, Table S8**).

TnpB increases the effects on growth from each transposition event

TnpB is unstable in the absence of other transposon features (*29*): TnpB in the absence of Tn4rev (**Figure 3A**, red) achieves significantly lower mCherry-TnpB concentrations compared to the strain with TnpB introduced *in trans* with Tn4rev for the same range of induction (**Figure 3A**, blue). Consequently, lower concentrations of TnpB do not cause significant growth defects (**Figure 3A**). Upon the co-expression of TnpA and TnpB for the immobile TE strain, there is greater growth defect than for the immobile TE without TnpB expression for the same concentrations of TnpA (**Figure 3B**), suggesting that TnpA and TnpB have an undetermined collaborative interaction that is independent of the execution of transposition.

To determine the growth defect contributed by an excision event, we fit the growth rate for each strain to an exponential decay function with CC(aTc) for each [IPTG] (**SII.vii. TnpB Increases Growth Rate Defect from Transposon Excision Events, Figure 3C**, **Table S9)**. The coefficient of exponential decay (*b*) increases as a function of TnpB (**Figure 3D**), suggesting that a combined effect of TnpA and TnpB is causing the defect. We propose that TnpB co-expression is increasing the insertion efficiency of the TE and that *b* is a measure of the successful insertion rate (**SII.viii. Exponential Growth Defect Arises as a Direct Consequence of Genomic Integration**). As the insertion rate improves, the excision events that result in mutational damage increase, leading to the greater growth decay.

Modeling

To understand how TnpB affects transposon dynamics, we analyzed a mean field theory in which the population averaged transposition rates, number of excised and unexcised plasmids and number of transposons are considered for a representative of the population (see **SIII. Supplementary Modeling** for details) which is evolved over time.

In our model, the representative starts with an experimentally determined average number of transposons, $X_o = 12$. The excision (χ, $0 \leq \chi \leq 1$) and insertion (μ) rates of each TE are proportional to the concentration of transposase molecules, which are generated by the TEs themselves. The single stranded transposon excises and then insert primarily into the lagging strand of the target during replication ($0 \leq \mu \leq 1$) (*1*, *5*) which results in a maintenance or loss of TE numbers and an average of at most half the daughter chromosomes containing a TE at the new site. We propose that TnpB extends the range of insertion rate, possibly doubling it by nicking the leading strand during insertion, resulting in the introduction of a TE in

both daughter chromosomes. TnpB-assisted homing is considered by a rate of reintroduction of transposons to excision sites, C_{hom} (TE^{-1} $[TnpB]^{-1}$ $generation^{-1}$) (*29*). The final numbers and activity levels of the transposon in cells evolved (**Figure 4A**) over experimental time scales and plotted in **Figure 4B-E**. Our model recapitulates our experimental observations of an increase and earlier initiation of TnpA-Venus, total transposons, number of excision events and excised plasmids, with an increase in TnpB concentration.

We find that co-expression of TnpA and TnpB results in growth defects, even in the absence of transposition. This suggests an interaction where the function of TnpB-RuvC may be analogous to RuvC's function in Cas9 and Cas12. We propose that TnpA-TnpB may interact to make dsDNA cuts and, within this mechanism, TnpB may nick the complementary strand of the TnpA-target site. This break would introduce the TE to both the target and its complementary strand, resulting in both daughter strands carrying TEs at this new site, potentially doubling the insertion rate. Simultaneously, TnpB on its own may make dsDNA cuts at the excision sites to execute TE homing (*29*). Thus, we propose that TnpB has a dual effect on transposition—aiding in TnpA-executed transposition and providing a homing mechanism.

DISCUSSION

Transposition is a "leaky" process in which TEs are lost (**Figures 2G,S6,** see **SII.vi. Decay of Transposase numbers as a Function of TnpA and TnpB Concentrations)**. This inefficiency compounded with the mutational damage of insertional sequence (IS) proliferation, makes IS extinction within hosts inevitable(*37*, *38*). By mitigating the decay of TE numbers, TnpB can prolong TE lifetimes within hosts. We demonstrate that TnpB expression in conjunction with TnpA increases TE retention (**Figure 2G**). TE retention further increases TnpA levels in cells (**Figures 1B-D,S6)**, which induces greater excision numbers (**Figures 2A-C,2H**). However, this increased TE retention comes at the cost of higher damage and resulting detrimental growth effects (**Figures 3C-D**). Considering TnpB's conservative function and abundance in bacterial and archaeal genomes, TnpB must have played an important role in the proliferation and maintenance of TEs throughout these genomes. Despite its cost to fitness, TnpB is maintained perhaps in part due to its role in the evolution of adaptive immunity (*25*, *30*, *31*, *51–53*). Additionally, TnpB's proliferation via IS200/IS605 transposons and positive impact on TE proliferation perhaps explain its abundance, shedding light on key dynamics leading to the evolutionary transition of the emergence of adaptive immunity in bacteria.

We considered two mechanisms of TnpB interaction that allow it to increase TE retention: (i) TnpB allows an additional "homing" mechanism that introduces a copy of the TE back into its original excision site (*29*), and (ii) TnpB increases the insertion efficiency of the TE into its target site, reducing TE loss. Our qPCR results confirm that target site insertion increases with the introduction of TnpB. However, the increase in the number of plasmid-excision events with maximal TnpB vs. no TnpB induction (ratio of 1.67±0.22, **Figures 2B**) is larger than the corresponding increase in the number of excised plasmids (ratio of 1.25±0.23, **Figure 2E**), suggesting that every excision event does not result in a proportionate number of excised plasmids with a p-value of 0.028. This implies that TnpB is also aiding in homing of the TE back into its original site on the plasmid. Indeed, our model can only account for the large range of increase of

excision events with increased TnpB only if we include TnpB-induced homing of the TE back into its original site on the plasmid (**Figure 4D**). Importantly, increased insertional efficiency is required to account for the overall increase in the number of excised plasmids (**Figure 4E**) and cannot be achieved with homing alone.

We find that co-expression of TnpA and TnpB results in growth defects, even in the absence of transposition. This suggests an interaction where the function of TnpB-RuvC may be analogous to RuvC's function in Cas9 and Cas12. We propose that TnpA-TnpB may interact to collectively make dsDNA cuts and, within this mechanism, TnpB may nick the complementary strand of the TnpA-target site. This break would introduce the TE to both the target and its complementary strand, resulting in both daughter strands carrying TEs at this new site, potentially doubling the insertion rate. Simultaneously, TnpB on its own may make dsDNA cuts at the original excision site to execute TE homing (*29*). Thus, we propose that TnpB has a dual effect on transposition—aiding in TnpA-executed transposition and providing an additional homing mechanism.

DATA AVAILABILITY

Further information and requests for resources, reagents, plasmids and strains should be directed to and will be fulfilled by the lead contact, Thomas E. Kuhlman (thomas.kuhlman@ucr.edu). All data reported in this paper will be shared by the lead contact upon request. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACCESSION NUMBERS

Genetic sequences for the transposon constructs used in this study have been deposited with the NCBI GenBank under accession number OP581959, OP581957, and OP581958.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

Conceptualization, T.E.K.; Methodology, D.K. and T.E.K.; Investigation, D.K., Formal Analysis, D.K.; Writing – Original Draft, D.K. and T.E.K.; Writing – Review and Editing, D.K. and T.E.K.; Funding Acquisition, T.E.K.; Resources, D.K. and T.E.K.; Supervision, T.E.K.

FUNDING

This work was supported by the NSF Center for the Physics of Living Cells (PHY 1430124) and startup funds from the University of California.

CONFLICT OF INTEREST

The authors declare no competing interests.

REFERENCES

1. B. McClintock, The Origin and Behavior of Mutable Loci in Maize. *Proc Natl Acad Sci U S A*. **36**, 344–355 (1950).

2. V. P. Belancio, P. L. Deininger, A. M. Roy-Engel, LINE dancing in the human genome: transposable elements and disease. *Genome Medicine*. **1**, 97 (2009).

3. P. A. Callinan, M. A. Batzer, "Retrotransposable Elements and Human Disease" in *Genome Dynamics*, J.-N. Volff, Ed. (KARGER, Basel, 2006; https://www.karger.com/Article/FullText/92503), pp. 104–115.

4. J.-M. Chen, P. D. Stenson, D. N. Cooper, C. Férec, A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum Genet*. **117**, 411–427 (2005).

5. P. L. Deininger, M. A. Batzer, Alu Repeats and Human Disease. *Molecular Genetics and Metabolism*. **67**, 183–193 (1999).

6. H. H. Kazazian, C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, S. E. Antonarakis, Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*. **332**, 164–166 (1988).

7. K. A. O'Donnell, K. H. Burns, Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mobile DNA*. **1**, 21 (2010).

8. B. Bodega, V. Orlando, Repetitive elements dynamics in cell identity programming, maintenance and disease. *Current Opinion in Cell Biology*. **31**, 67–73 (2014).

9. N. G. Coufal, J. L. Garcia-Perez, G. E. Peng, G. W. Yeo, Y. Mu, M. T. Lovci, M. Morell, K. S. O'Shea, J. V. Moran, F. H. Gage, L1 Retrotransposition in Human Neural Progenitor Cells. *Nature*. **460**, 1127–1131 (2009).

10. S. L. Anwar, W. Wulaningsih, U. Lehmann, Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *Int J Mol Sci*. **18** (2017), doi:10.3390/ijms18050974.

11. K. H. Burns, Transposable elements in cancer. *Nature Reviews Cancer*. **17**, 415–424 (2017).

12. X. Liu, M. Irfan, X. Xu, C.-Y. Tay, B. J. Marshall, Helicobacter pylori infection induced genome instability and gastric cancer. *GENOME INSTAB. DIS.* **1**, 129–142 (2020).

13. R. K. Pradhan, W. Ramakrishna, Transposons: Unexpected players in cancer. *Gene*. **808**, 145975 (2022).

14. H. Kano, I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag, H. H. Kazazian, L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev*. **23**, 1303–1312 (2009).

15. D. Schneider, R. E. Lenski, Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology*. **155**, 319–327 (2004).

16. L. Chao, C. Vargas, B. B. Spear, E. C. Cox, Transposable elements as mutator genes in evolution. *Nature*. **303**, 633–635 (1983).

17. A. J. Modzelewski, W. Shao, J. Chen, A. Lee, X. Qi, M. Noon, K. Tjokro, G. Sales, A. Biton, A. Anand, T. P. Speed, Z. Xuan, T. Wang, D. Risso, L. He, A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*. **184**, 5541-5558.e22 (2021).

18. S. He, A. Corneloup, C. Guynet, L. Lavatine, A. Caumont-Sarcos, P. Siguier, B. Marty, F. Dyda, M. Chandler, B. Ton Hoang, The IS200/IS605 Family and "Peel and Paste" Single-strand Transposition Mechanism. *Microbiol Spectr*. **3** (2015), doi:10.1128/microbiolspec.MDNA3-0039-2014.

19. C. Guynet, A. B. Hickman, O. Barabas, F. Dyda, M. Chandler, B. Ton-Hoang, In Vitro Reconstitution of a Single-Stranded Transposition Mechanism of IS608. *Molecular Cell*. **29**, 302–312 (2008).

20. B. Ton-Hoang, C. Guynet, D. R. Ronning, B. Cointin-Marty, F. Dyda, M. Chandler, Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. *EMBO J*. **24**, 3325–3338 (2005).

21. O. Barabas, D. R. Ronning, C. Guynet, A. B. Hickman, B. Ton-Hoang, M. Chandler, F. Dyda, Mechanism of IS200/IS605 Family DNA Transposases: Activation and Transposon-Directed Target Site Selection. *Cell*. **132**, 208–220 (2008).

22. S. He, C. Guynet, P. Siguier, A. B. Hickman, F. Dyda, M. Chandler, B. Ton-Hoang, IS200/IS605 family single-strand transposition: mechanism of IS608 strand transfer. *Nucleic Acids Res.* **41**, 3302–3313 (2013).

23. D. Kersulyte, B. Velapatiño, G. Dailide, A. K. Mukhopadhyay, Y. Ito, L. Cahuayme, A. J. Parkinson, R. H. Gilman, D. E. Berg, Transposable Element ISHp608 of Helicobacter pylori: Nonrandom Geographic Distribution, Functional Organization, and Insertion Specificity. *J Bacteriol*. **184**, 992–1002 (2002).

24. S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, K. Severinov, F. Zhang, E. V. Koonin, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol*. **15**, 169–182 (2017).

25. W. Bao, J. Jurka, Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mob DNA*. **4**, 12 (2013).

26. P. Siguier, E. Gourbeyre, M. Chandler, Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev*. **38**, 865–891 (2014).

27. H. Altae-Tran, S. Kannan, F. E. Demircioglu, R. Oshiro, S. P. Nety, L. J. McKay, M. Dlakić, W. P. Inskeep, K. S. Makarova, R. K. Macrae, E. V. Koonin, F. Zhang, The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*. **374**, 57–65 (2021).

28. C. Pasternak, R. Dulermo, B. Ton-Hoang, R. Debuchy, P. Siguier, G. Coste, M. Chandler, S. Sommer, ISDra2 transposition in Deinococcus radiodurans is downregulated by TnpB. *Molecular Microbiology*. **88**, 443–455 (2013).

29. T. Karvelis, G. Druteika, G. Bigelyte, K. Budre, R. Zedaveinyte, A. Silanskas, D. Kazlauskas, Č. Venclovas, V. Siksnys, Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*. **599**, 692–696 (2021).

30. V. V. Kapitonov, K. S. Makarova, E. V. Koonin, ISC, a Novel Group of Bacterial and Archaeal DNA Transposons That Encode Cas9 Homologs. *Journal of Bacteriology*. **198**, 797–807 (2016).

31. S. Shmakov, O. O. Abudayyeh, K. S. Makarova, Y. I. Wolf, J. S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, F. Zhang, E. V. Koonin, Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. *Mol Cell*. **60**, 385–397 (2015).

32. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. **337**, 816–821 (2012).

33. Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, A. Nakata, Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol*. **169**, 5429–5433 (1987).

34. F. J. Mojica, C. Díez-Villaseñor, E. Soria, G. Juez, Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*. **36**, 244–246 (2000).

35. J. Strecker, S. Jones, B. Koopal, J. Schmid-Burgk, B. Zetsche, L. Gao, K. S. Makarova, E. V. Koonin, F. Zhang, Engineering of CRISPR-Cas12b for human genome editing. *Nat Commun*. **10**, 212 (2019).

36. Kim Neil H., Lee Gloria, Sherer Nicholas A., Martini K. Michael, Goldenfeld Nigel, Kuhlman Thomas E., Real-time transposable element activity in individual live cells. *Proceedings of the National Academy of Sciences*. **113**, 7278–7283 (2016).

37. N. Cerveau, S. Leclercq, E. Leroy, D. Bouchon, R. Cordaux, Short- and Long-term Evolutionary Dynamics of Bacterial Insertion Sequences: Insights from Wolbachia Endosymbionts. *Genome Biology and Evolution*. **3**, 1175–1186 (2011).

38. A. Wagner, Periodic Extinctions of Transposable Elements in Bacterial Lineages: Evidence from Intragenomic Variation in Multiple Genomes. *Molecular Biology and Evolution*. **23**, 723–733 (2006).

39. Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular Systems Biology*. **8**, 610 (2012).

40. T. E. Kuhlman, E. C. Cox, DNA-binding-protein inhomogeneity in $E.$ coli modeled as biphasic facilitated diffusion. *Phys. Rev. E*. **88**, 022701 (2013).

41. T. E. Kuhlman, E. C. Cox, Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Research*. **38**, e92 (2010).

42. T. E. Kuhlman, E. C. Cox, A place for everything. *Bioeng Bugs*. **1**, 296–299 (2010).

43. H. Tas, C. T. Nguyen, R. Patel, N. H. Kim, T. E. Kuhlman, An Integrated System for Precise Genome Modification in Escherichia coli. *PLoS One*. **10**, e0136963 (2015).

44. J. B. Kinney, A. Murugan, C. G. Callan, E. C. Cox, Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences*. **107**, 9158–9163 (2010).

45. G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, R. Díaz-Orejas, Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev*. **62**, 434–464 (1998).

46. W. T. Tucker, C. A. Miller, S. N. Cohen, Structural and functional analysis of the par region of the pSC 10 1 plasmid. *Cell*. **38**, 191–201 (1984).

47. R. Lutz, H. Bujard, Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research*. **25**, 1203–1210 (1997).

48. M. P. Calos, J. H. Miller, The DNA sequence change resulting from the IQ1mutation, which greatly increases promoter strength. *Molec. Gen. Genet.* **183**, 559–560 (1981).

49. M. L. Markwardt, G.-J. Kremers, C. A. Kraft, K. Ray, P. J. C. Cranfill, K. A. Wilson, R. N. Day, R. M. Wachter, M. W. Davidson, M. A. Rizzo, An Improved Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching. *PLoS One*. **6**, e17896 (2011).

50. M. D. Weinreich, A. Gasch, W. S. Reznikoff, Evidence that the cis preference of the Tn5 transposase is caused by nonproductive multimerization. *Genes Dev.* **8**, 2363–2374 (1994).

51. E. V. Koonin, K. S. Makarova, Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biol Evol*. **9**, 2812–2825 (2017).

52. M. Krupovic, K. S. Makarova, P. Forterre, D. Prangishvili, E. V. Koonin, Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biology*. **12**, 36 (2014).

53. E. V. Koonin, M. Krupovic, Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet*. **16**, 184–192 (2015).

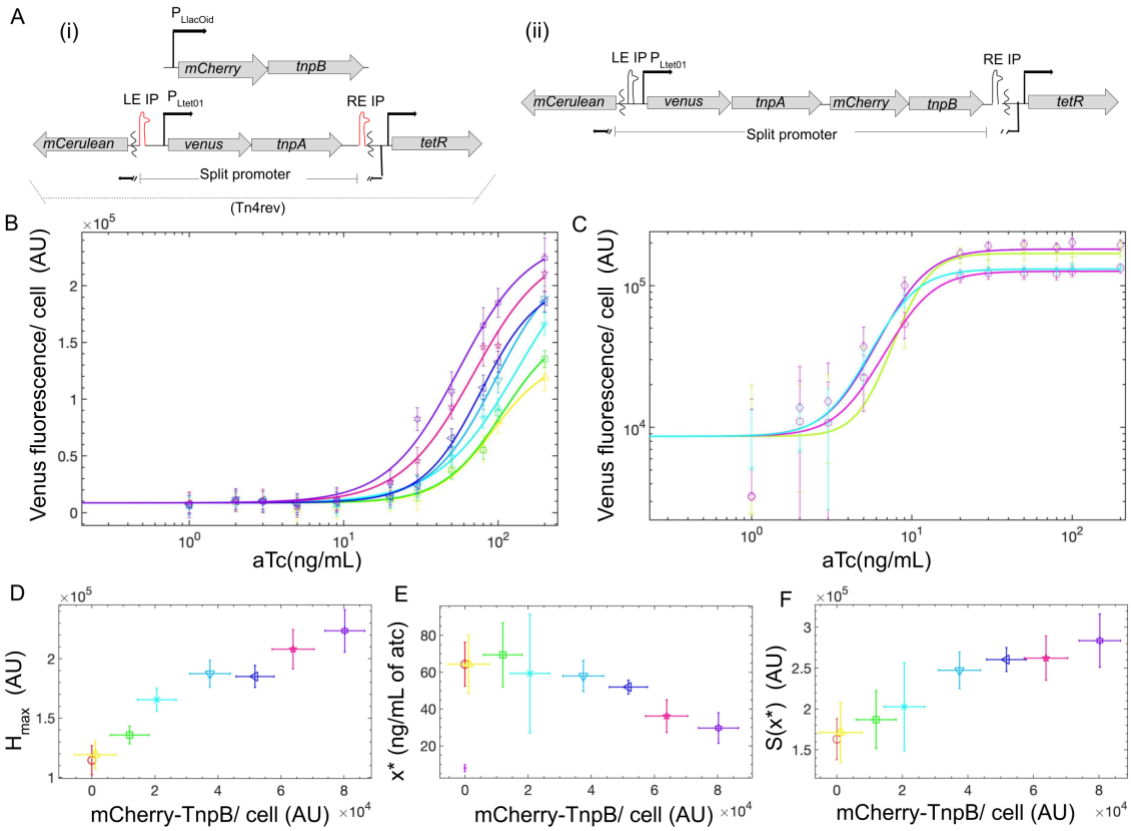

Fig. 1: (A) Constructs for Transposition Quantification: TnpB introduced with Tn4rev (i) *in trans* and (ii) *in cis*. **(B)** Venus fluorescence as a function of [aTc] = {0ng/mL, 1ng/mL, 2ng/mL, 3ng/mL, 5ng/mL, 9ng/mL, 20ng/mL, 30ng/mL, 50ng/mL, 80ng/mL, 100ng/mL and 200ng/mL} for TnpB *in trans* for [IPTG] = {0μM (yellow, ∆), 10μM (green, ☐), 20μM (turquoise, ✳), 50μM (teal, ▽), 100μM (navy, ◁), 200μM (magenta, ✋) & 2000μM (violet, ✳)}. **(C)** (i) Immobilized transposon, [IPTG] = 0 (purple, ◊), (ii) Immobilized transposon, [IPTG] = 2000μM (lizard green, +), (iii) Immobilized transposon with TnpB, [IPTG] = 0 (fuchsia, ○) and (iv) Immobilized transposon with TnpB, with [IPTG] = 2000μM (arctic, X). Solid lines, same color: Hill function fit; see **Supplementary Information SII.i**). Fitted parameters in **Table S4** and plotted in **D-F** using corresponding color and symbols. Also included is Tn4rev only strain (red, ○) **(D)** TnpB induction results in higher saturating TnpA-Venus (H_{max}). **(E)** TnpB results in earlier TnpA-Venus expression [inflection point of curves in (**B**), x^*]. **(F)** TnpB increases TnpA-Venus sensitivity to aTc [slope at inflection point $S(x^*)$].

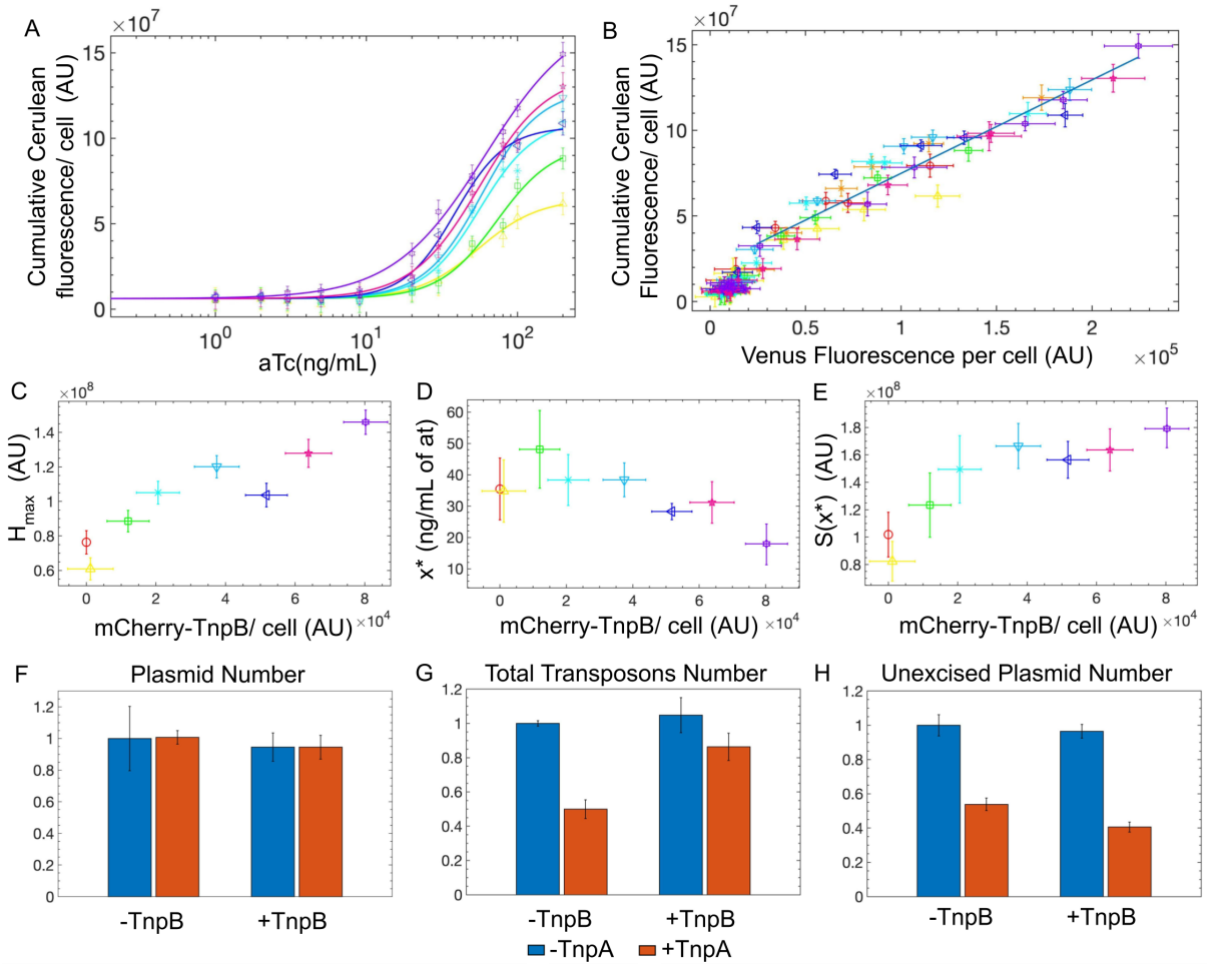

Fig. 2: Excision as Function of TnpA: (A) Cumulative Cerulean (CC) fluorescence as a function of [aTc] = {0ng/mL, 1ng/mL, 2ng/mL, 3ng/mL, 5ng/mL, 9ng/mL, 20ng/mL, 30ng/mL, 50ng/mL, 80ng/mL, 100ng/mL and 200ng/mL} for TnpB *in trans* for [IPTG] = {0μM (yellow, △), 10μM (green, ☐), 20μM (turquoise, ✳), 50μM (teal, ▽), 100μM (navy, ◁), 200μM (magenta, ✋) & 2000μM (violet, ✳)}. Solid lines, same color: Hill function fit. Fitted parameters are plotted in **C-E** and **Table S6. (B)** CC vs. Venus-transposase concentration plotted for Tn4rev only strain (red, ○), TnpB *in trans* for all [IPTG] (same color and symbols as **A**), and TnpB *in cis* (orange, X). Solid line: linear regression fit. **(C)** TnpB induction results in higher saturating CC levels (H_{max}). **(D)** TnpB induction results in earlier excision [inflection point of curves in (**A**), x^*]. **(E)** TnpB increases excision sensitivity to aTc concentration [slope at inflection point $S(x^*)$]. **(F-H)** qPCR determining relative (**F**) plasmid, (**G**) total transposon, and (**H**) unexcised plasmid counts. P-values given in **Tables S7 and S8**. See **Table S3** for primer sequences.

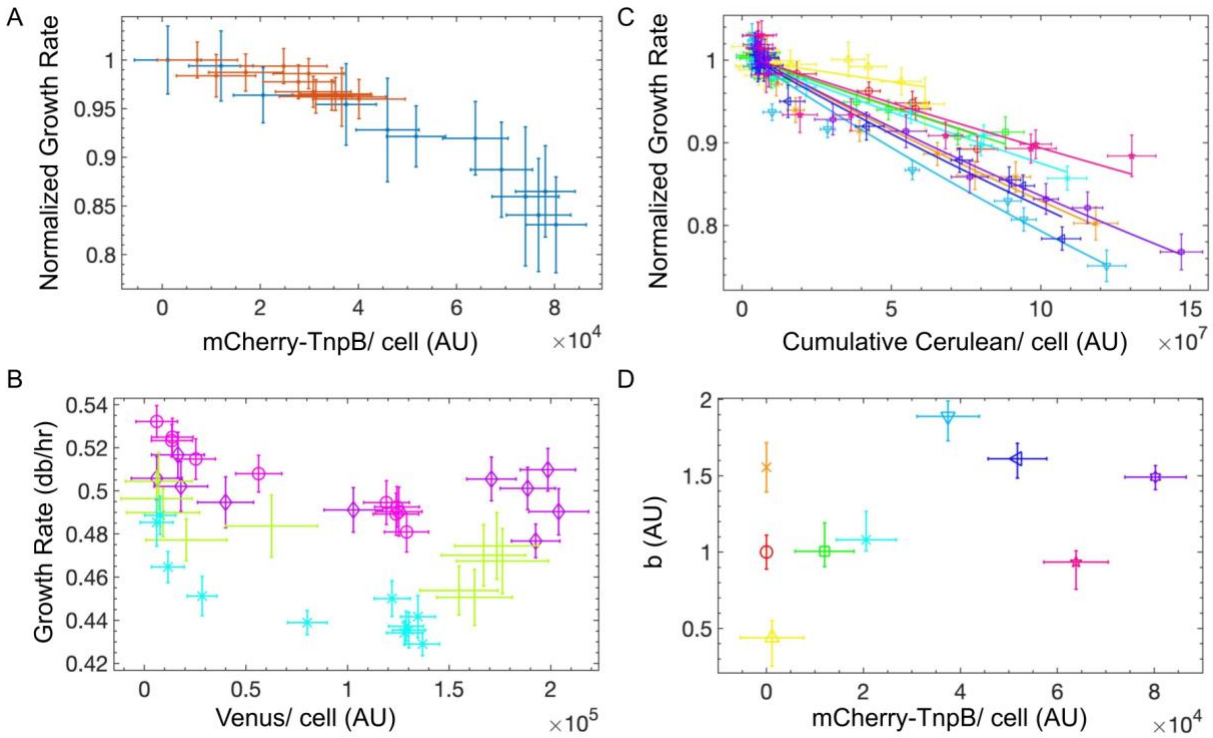

Fig. 3: Effects of TnpB on Growth: (A) TnpB expression with: (i) active TE (blue) and (ii) without active TE (red), for full range of [IPTG] = {0µM,10µM, 20µM, 50µM, 100µM, 200µM & 2000µM}. **(B)** Growth rate of immobilized strain (i) Immobilized transposon, [IPTG] = 0 (purple, ◊), (ii) Immobilized transposon, [IPTG] = 2000µM (lizard green, +), (iii) Immobilized transposon with TnpB, [IPTG] = 0 (fuchsia, ○) and (iv) Immobilized transposon with TnpB, with [IPTG] = 2000µM (arctic, X) versus Venus fluorescence per cell for all [aTc]. **(C)** Growth rate is fit to exponential decay function of CC to determine decay coefficient, *b* (see **Supplementary Information SII.vii)** for Tn4rev only (red, ○), TnpB *in trans* for [IPTG= {0µM (yellow, ∆), 10µM (green, □), 20µM (turquoise, ✳), 50µM (teal, ▽), 100µM (navy, ◁), 200µM (magenta, ✋) & 2000µM (violet, ✳)} and TnpB *in cis* (orange, X). **(D)** *b* increases with TnpB until it saturates. Same colors and symbols as **C**. Outliers correspond to curves with $R^2 < 0.9$ (**Table S9**). Data for TnpB *in cis* is plotted at zero point for mCherry-TnpB.

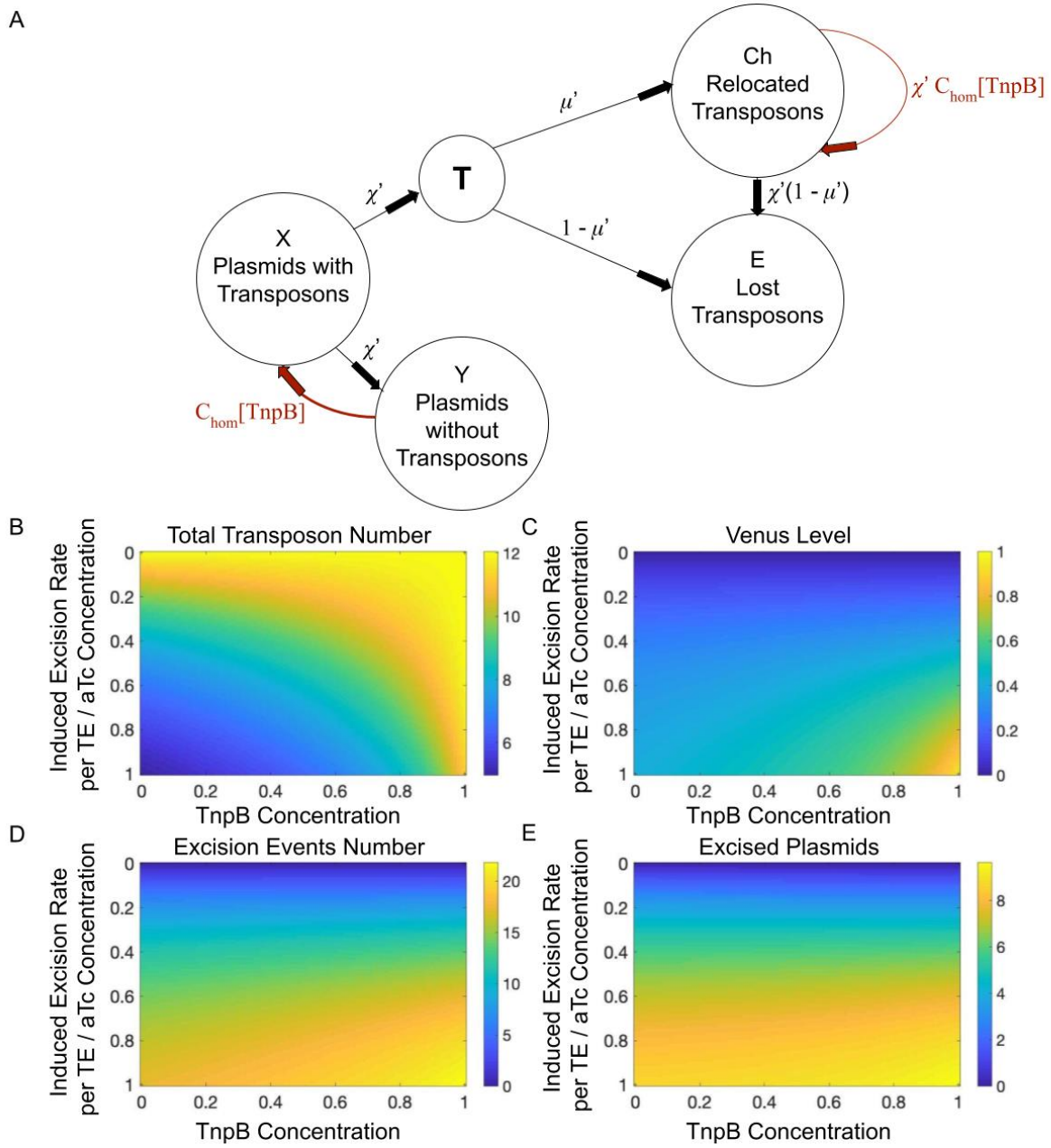

Fig. 4: Modeling of Transposon and TnpB Dynamics: (A) Model Schematic: The rate equations described in **Supplementary SIII.i** are evolved over time and transposon numbers are tracked. **(B-E)** Phase diagrams were generated with the y-axis and x-axis measuring aTc and IPTG induction respectively. Here, the excision rate per TE is a function of [aTc] as it induces TnpA. Tuning [TnpB] from zero to one in the model accounts for increased peak values and earlier initiation of expression that we observe experimentally for **(B)** total transposon number (qPCR data), **(C)** transposase concentration (Venus-TnpA data), **(D)** number of excision events (CC data) and **(E)** total excised plasmids (qPCR data).