Supplemental Information for

Testing the Retroelement Invasion Hypothesis for the Emergence of the Ancestral Eukaryotic Cell

Gloria Lee, Nicholas A. Sherer, Neil H. Kim, Ema Rajic, Davneet Kaur, Niko Urriola, K. Michael Martini, Chi Xue, Nigel Goldenfeld, and Thomas E. Kuhlman

correspondence to: thomask@ucr.edu

SUPPLEMENTARY METHODS

Plasmid Construction

To create the human-extracted version of L1, TL1H, TEK extracted his genomic DNA by a buccal swab followed by phenol-chloroform extraction and ethanol precipitation (1). Using TEK's genomic DNA, we amplified an L1 element using primers designed to target the highly active L1 element #4-35 identified by Beck et al.(2) between the AccI restriction sites they used for cloning and testing of retrotransposition efficiency. The PCR fragment resulting from TEK's DNA was amplified again using primers containing a T7*lac* promoter with a consensus Shine-Dalgarno RBS that annealed to the 5' end of ORF1 and that flanked the L1 element with ApaI and SalI restriction sites. The resulting PCR product was purified (QIAquick PCR purification kit, QIAGEN), digested with ApaI and SalI, and ligated into similarly prepared pTKIP-*neo* plasmid (3, 4), forming plasmid pTKIP-TL1H.

The *E. coli* optimized L1 element, EL1H, was designed with Vector NTI software (Thermo Fisher Scientific) and synthesized *de novo* and cloned into pUC57-*kan* by GENEWIZ Gene Synthesis (GENEWIZ). The EL1H cassette is flanked by I-SceI restriction sites and LP1 and LP2 sequences for chromosomal integration in future experiments (3-5). BL21(DE3) transformed with pUC57-EL1H do not survive due to leaky L1 expression from the high copy number pUC57 plasmid. We digested pUC57-EL1H with I-SceI (New England Biosciences) and gel purified the resulting EL1H fragment (QIAquick gel extraction kit, QIAGEN). We ligated EL1H into pTKIP-*neo* plasmid that had been I-SceI digested, dephosphorylated with Antarctic Phosphatase (NEB), and gel purified. The resulting plasmid, pTKIP-EL1H, is medium copy number and was used for all experiments described here.

To generate pTKIP-EL1HID, the mTFP1ID (Integration Detection) cassette sequence was designed with Vector NTI software and synthesized and cloned into pUC57-*kan* by GENEWIZ. We non-directionally subcloned mTFP1ID into an XhoI restriction site designed into pTKIP-EL1H between ORF2 and the poly(A) tract, and the correct orientation was determined by screening resulting clones by restriction fragment length analysis following digestion with PvuII (NEB).

To create plasmids pHCMC05-EL1H and pHCMC05-TORF/RIG, EL1H was amplified from pTKIP-EL1H and TORF/RIG was amplified from pET-TORF/RIG using primers including AatII and XmaI restriction sites. The resulting products were digested, gel purified, and ligated into the AatII and XmaI restrictions sites within the MCS of pHCMC05. pHCMC05-*lacZYAX* was made by amplifying the *lacZYA* operon from *E. coli* MG1655 by colony PCR, including some sequence of the *cynX* gene downstream of *lacA*, using primers containing XbaI and XmaI restrictions sites. This PCR fragment was digested, PCR purified, and ligated into the XbaI and XmaI restriction sites of pHCMC05. pHCMC05-*lacZYAX* was designed to be exactly the same

www.pnas.org/cgi/doi/10.1073/pnas.1807709115

size as pHCMC05-EL1H (13,941 bp) and express *lacZ* from the same promoter and ribosomal binding site as EL1H and TORF/RIG in pHCMC05.

To assess the effects of NHEJ enzymes in *E. coli*, we designed a cassette expressing the *B. subtilis* NHEJ genes *ykoU* and *ykoV* from the synthetic promoter P_{Ltet01}(6) using VectorNTI software. We optimized codon usage of *ykoU* and *ykoV* for expression in *E. coli* and included in the cassette the gene encoding tet repressor expressed from a strong constitutive P_{lacIQ} promoter. The cassette was synthesized *de novo* by GENEWIZ and cloned into the plasmid pUC57-*kan*. To generate pUC57-cat-NHEJ, we exactly replaced the kanamycin resistance gene neo with chloramphenicol acetyltransferase, *cat*, by recombineering in strain SW102 (7). To create plasmid pZA31-NHEJ, the NHEJ cassette was subcloned into pZA31 by cutting pUC57-kan-NHEJ with I-SceI and ligating the cassette into the pZA31 backbone amplified from pZA31-*luc* using primers including I-SceI restriction sites. The control plasmid pZA31-*tetR* was created in a similar fashion by amplifying the *tetR* gene from pUC57-NHEJ using primers including I-SceI restriction sites and ligating into pZA31.

All Sanger sequencing was performed by ACGT, Inc and the UIUC Core Sequencing Facility.

B. subtilis Transformation

Colonies of each *B. subtilis* strain were picked from nonselective Lysogeny Broth (LB) plates and used to inoculate cultures containing 2 ml per transformation reaction MC transformation medium (5.36 g K_2HPO_4, 26.2 g KH_2PO_4, 10 g D-glucose, 0.5 g casamino acids, 1 g L-glutamate, 5 ml 300 mM sodium citrate, 0.5 ml 22 mg/ml ferric ammonium citrate, 1.7 ml 1 M $MgSO_4$, 2.5 ml 10 mg/ml L-tryptophan, and 2.5 ml 10 mg/ml L-phenylalanine per 500 ml medium). These cultures were grown for 5 – 6 hours in a 37ºC shaking water bath until entering stationary phase growth, at which point 400 µl of each strain was added to 20 µl of each miniprepped plasmid in a 5 ml round bottom polypropylene tube (Falcon Corning). The tubes were placed back in the 37ºC shaking water bath for two hours, after which the entire mixture was spread on LB agar plates containing 5 µg/ml chloramphenicol.

Growth Rate Determination

To measure the effect of retroelement expression on *E. coli*'s growth rate shown in Figs. 1-2, starter cultures were prepared by inoculating LB + 100 µg/ml ampicillin with glycerol stocks of BL21(DE3) carrying the indicated retrotransposon or control plasmid(s). This starter culture was grown at 37ºC in a shaking water bath (New Brunswick C76). Once OD600 of this culture reached ~0.4 – 0.5, 1 µl of the starter culture was added to 50 ml of the experimental medium + 100 µg/ml ampicillin, pre-warmed to 37ºC and thoroughly mixed. 2 ml of this medium was then added to each well of a microplate (Thermo Fisher Scientific Nunclon Delta Surface). Appropriate concentrations of IPTG were added to each well, such that each induction condition was performed in triplicate. The plate was then loaded into a Tecan Infinite f200 plate reader pre-warmed to 37ºC. Measurements of OD600 were performed every 10 minutes over the course of ~24 hours with the temperature maintained at 37ºC and with continuous shaking. The growth rates reported in Figs. 1-2 are averages of the doubling time determined as the slope of the logarithm, base 2, of the background subtracted OD600 versus time in the regime of exponential growth.

Microscopy

To perform fluorescence microscopy, 50 µl samples of culture were spread onto 1% agarose pads prepared on glass slides (Fisher Scientific Premium, 3" × 1" × 1 mm), covered with a #1.5 glass cover slip (VWR, 22 × 30 mm). The slide was placed on a Nikon Eclipse Ti-E fully automated inverted microscope with Perfect Focus System (PFS) automated focus correction. Images were taken using a Nikon CFI Apo TIRF 100x oil immersion objective (NA = 1.49) and captured using an Andor iXon Ultra 897 EMCCD camera with 100 ms exposure. Fluorescent excitation was performed using highly inclined and laminated optical sheet (HILO) laser illumination (8) at 457 nm. 457 nm excitation was provided by a 40 mW Argon laser (CVI Milles Griot). Filter set used was Z457/10x ET485/30m (Chroma).

Quantitative RT-PCR

In vitro transcription of TL1H RNA was performed to generate absolute standards for qRT-PCR. BL21(DE3) pTKIP-TL1H was grown to stationary phase in PDM (plasmid DNA medium), mini-prepped (QIAprep Spin Miniprep kit, Qiagen) to extract the plasmid, digested with I-SceI to linearize the plasmid, and PCR-purified. MegaScript T7 High Yield Transcription Kit (Thermo Scientific) was used to transcribe retrotransposon RNA from the linearized plasmid in vitro. The RNA was then digested with Turbo DNase (Ambion) and purified via LiCl precipitation. The concentration of the RNA was measured with a Nanodrop 2000c spectrophotometer (Thermo Scientific), then serially diluted to obtain five samples with concentrations ranging from 10^{-3}-10^{-7} of the original concentration.

To extract RNA from cells expressing the retrotransposon, BL21(DE3) pTKIP-TL1H was grown in LB until exponential phase (OD_{600}~0.2), then inoculated into flasks containing 10mL of the specified medium that had been titrated with IPTG. Media used were RDM + 0.5% glucose, RDM + 0.5% glycerol, M63 + 0.5% glucose + 0.1% casamino acids (cAA glucose), M63 + 0.5% glucose, and M63 + 0.5% glycerol. Each medium was titrated using IPTG concentrations of 0 µM, 10 µM, 20 µM, 35 µM, and 50 µM. After 6 doublings to reach exponential phase (OD_{600}~0.2), RNA was extracted using the Trizol Max Bacterial RNA Isolation Kit (Thermo Fisher Scientific), digested with Turbo DNase (Thermo Fisher Scientific), and the resulting RNA concentration was measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific).

Melt curves showed a sharp peak at 84.5°C, indicating specific amplification of the desired product. Negative controls of RT^- RNA extracted from cells crossed the cycle threshold number much later than cDNA samples, verifying that initial plasmid DNA was successfully digested by the DNAse and did not contribute to qPCR measurements. To determine the average number of RNAs per cell, RNA counts were normalized by the number of bacteria added to each reaction, determined by colony counts derived from plating 20 independent experimental replicates of each growth condition.

Ll.LtrB Retrotransposition Frequency Assays

To determine the retrotransposition efficiency of Ll.LtrB with and without NHEJ expression, we followed the protocol of Coros et al., 2005, with slight modifications. We created strains with all possible combinations of pET-TORF/RIG, empty pTKIP (as negative control), pUC57-*cat*-NHEJ, pUC57-*cat*, pZA31-NHEJ, and pZA31-*tetR*, for a total of eight strains. From frozen glycerol stocks, we inoculated each strain into 2 ml of LB medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and allowed these cultures to grow to OD_{600} ~1.0 in a 37 ºC shaking water bath. At this point, a sample from each tube was used to inoculate 10 ml LB medium + 100 µg/ml ampicillin and 34 µg/ml chloramphenicol in 50 ml baffled Erlenmeyer flasks

at an initial OD_{600} of 0.01. These cultures were allowed to grow until OD_{600} ~0.2, at which point Ll.LtrB expression was induced by the addition of 100 µM IPTG. The cultures were allowed to grow for another 3 hours in the 37 ºC shaking water bath, at which time a 100 µl sample was taken to measure OD_{600}, and a 10 µl sample was taken to generate 10-fold serial dilutions in phosphate buffered saline (PBS). 10^{-4}, 10^{-5}, and 10^{-6} fold dilutions were plated on plates containing LB agar + 100 µg/ml ampicillin and 34 µg/ml chloramphenicol to determine the total number of bacteria in the cultures. The remainder of the 10 ml culture was collected by centrifugation and plated on LB + 40 µg/ml kanamycin plates. The reported efficiencies are the number of resulting kanamycin resistant colonies divided by the total number of bacteria spread on the plate and are the averages and standard deviations of at least three independent replicates.

Limitations of this assay should be kept in mind when interpreting the apparently low observed retrotransposition rates. First, to detect retrotransposition by the RIG kanamycin-resistance cassette, bacteria must survive the integration and grow as colonies. Given the lethality of Ll.LtrB as evidenced by bulk growth rate measurement and plating assays, it is possible that the actual retrotransposition rate is significantly higher but results in cell death in a large fraction of instances. Secondly, TORF/RIG is constructed such that the RIG cassette is in the middle of the intron, in contrast to our LINE-1 mTFP1 ID detection cassette which is at the immediate 3' end. Consequently, ~4 kbp of Ll.LtrB must be successfully reverse transcribed before cells become resistant to kanamycin. The vast majority of LINE-1 elements in the human genome are 5' truncations, where reverse transcription aborts before completion. The precise mechanistic details of L1 truncation remain unclear, but it has been suggested that NHEJ may contribute in humans (9). Similarly, an artifactually low retrotransposition efficiency of the Ll.LtrB RIG cassette would be obtained if NHEJ results in premature abortion of reverse transcription before completion of the kanamycin resistance gene. Furthermore, as both LINE-1 and Ll.LtrB in these experiments and those of Coros et al., 2005 are expressed using T7 polymerase, which is neither native to bacteria nor essential for their survival, the simplest way for bacteria to escape the negative effects of retroelement expression is to stop producing T7 polymerase, through mutation, excision of the λ(DE3) prophage, or other means. We frequently find bacteria which no longer express T7 polymerase dominate in cultures induced for retroelement expression and allowed to grow for long times (>24 hours), and such mutants would contribute to an artifactually low estimate of retrotransposition frequency.

SUPPLEMENTARY ANALYSIS

Exponential Growth Defect Arises as a Direct Consequence of Genomic Integration

The observed exponential decay in normalized growth rate can be explained by a simple model where we consider the effect that integrations will have by disrupting essential chromosomal genes and thus cell viability. In the simplest model of this kind, we consider that there are two sub-populations of cells: those that grow normally, and those with retroelement integrations disrupting all growth. In this binary model, there are L transcripts, each with a probability w of integrating and disrupting growth per generation, and the probability q of a cell having no integrations affecting growth per generation given by a binomial distribution evaluated at zero:

$$q = \binom{L}{0} w^0 (1-w)^L = \exp\left[-\ln\left(\frac{1}{1-w}\right)L\right]. \quad (1.1)$$

In our growth experiments, an exponentially growing individual cell, in the absence of

integrations, will produce $g_0 dt$ new individuals in a time interval dt. This leads to a simple model of exponential growth of the form $\frac{dx}{dt} = g_0 x$. If we consider a binary model with a population x of normal cells and a population y of cells with no growth due to integrations, an individual of x will still produce $g_0 dt$ new individuals but only a fraction q of these will be able to grow. This leads to the population model (10):

$$\frac{dx}{dt} = q g_0 x, \quad \frac{dy}{dt} = (1-q) g_0 x. \tag{1.2}$$

The total population of cells in this model grows as $x_0 + y_0 + \frac{x_0}{q}\left[\exp(q g_0 t) - 1\right]$. Thus the measured growth rate would be $q g_0$ and the normalized growth rate is just q. We fit eq. (1.1) to the form $\exp[-bL]$ and make the identification $b = -\ln[1-w]$, which means $b \approx w$ for $w \ll 1$. That is, b is approximately equal to the probability of a retroelement transcript integrating and disrupting growth. Moreover, we expect that the rate of obtaining integrations affecting growth, w, is proportional to the overall rate of integration, μ. Consequently, this simple binary model recapitulates the exponential dependence of the growth rate on the number of transcripts and demonstrates that the exponential dependence implies that the growth defect, b, is expected to be directly coupled to the integration rate, μ.

More complex models of the impact of transposable element integration can be developed, with more than two sub-populations and more nuanced assumptions about the physiological effects. But we find that the dynamics of these models reduce to that of the two rate model presented above, with renormalized parameters. An example of one such model is as follows. Let the number of cells with no chromosomal integrations harming their growth be N_0, the population of cells with one integrant be N_1, and so forth. Then a set of differential equations describing the population dynamics in exponential growth with growth rate g_0 is

$$\frac{dN_x}{dt} = g_0 f(x)(1-\mu) N_x + g_0 f(x-1) \mu N_{x-1}, \tag{1.3}$$

where $f(x)$ is a monotonically decreasing function describing the inhibition of cell growth due to gene disruption by integrations, μ is the mutation rate, and the index x runs from 0 to some integer $x_{\max}$ where the number of integrants is so high the cell cannot function and dies. Making the substitution $(1-\mu) = q$,

$$\frac{dN_x}{dt} = g_0 f(x) q N_x + g_0 f(x-1)(1-q) N_{x-1}. \tag{1.4}$$

This is a lower triangular system of equations whose eigenvalues are the diagonals. After many generations, the largest eigenvalue will dominate and correspond approximately to the measured growth rate. Since $f(x)$ is a monotonically decreasing function, this means the growth rate is $g_0 f(0) q$. $f(0) = 1$ and thus the growth rate is $q g_0$ and the normalized growth rate is q. This is the same result as the binary model discussed above.

Retrotransposon Dynamics in Bacteria Lead to Low Numbers or Extinction

Next, we wish to understand how retrotransposons will proliferate within an asexual host genome given the experimentally measured integration rates and growth defects. A substantial body of work has already been performed studying the population genetics of mobile elements (11-18). Note that the asexuality of this population makes this model distinct from that of Hickey and colleagues (19, 20), where the inferred effects of retroelements and their proliferation is contingent upon sexual reproduction and outbreeding. Here, we construct a simple model of the mean behavior of retroelement activity based upon our experimental measurements and analyze its dynamics. We find that retrohoming generally will lead to low but stable numbers of retroelements, while the parameters with which retrotransposition occurs must be finely tuned in order to get long-lived states with significant proliferation of retrotransposons in the host.

First, to introduce direct competition for resources such that extinction is a possible outcome, we construct the model with a limited system size Ω. Within the system, we place N_x cells carrying x copies of the retrotransposon, leaving E empty space. Normalizing by Ω, the mean behavior of the system is described by the equations

$$\begin{aligned} 1 &= \varepsilon + \sum_{x=0}^{\infty} \psi_x; \quad \varepsilon \equiv \frac{E}{\Omega}, \ \psi_x \equiv \frac{N_x}{\Omega} \\ \frac{\partial \psi_x}{\partial \tau} &= \varepsilon e^{-bx} \psi_x - \beta(1-\varepsilon)\psi_x + \mu(x-1)\psi_{x-1} - \mu x \psi_x + \Delta(x+1)\psi_{x+1} - \Delta x \psi_x \\ \frac{\partial \varepsilon}{\partial \tau} &= \beta(1-\varepsilon)\sum_{x=0}^{\infty} \psi_x - \varepsilon \sum_{x=0}^{\infty} e^{-bx} \psi_x, \end{aligned} \tag{2.1}$$

where τ is the generation time, β is the death rate per generation [~10^{-2} – 10^{-3} cell^{-1} generation^{-1} (21)], Δ is the mutation rate per retrotransposon per cell per generation resulting in inactivation of a copy of the retroelement [~10^{-8} retrotransposon^{-1} cell^{-1} generation^{-1} (22)], b is the growth defect, and μ is the transposition rate per retrotransposon per cell per generation. As we have demonstrated above, the values of μ and b will depend on the retroelement in question and the presence or absence of NHEJ, with μ ~10^{-2} – 1 and b ~10^{-2} – 0.6 for LINE-1, and μ ~10^{-9} – 10^{-6} and b ~10^{-3} – 10^{-2} for Ll.LtrB. Here we assume that each element will contribute an equal amount to the growth defect, while, in reality, the effects of each insertion will be drawn from some distribution of possible fitness effects. However, we expect that this simple mean-field model will allow insights into the average behavior of the system.

To determine non-trivial stationary states, we set time derivatives to zero, and the ψ_x equations yield a set of recursion relations:

$$\psi_x^* = \frac{\beta + (\mu+\Delta)(x-1) - \varepsilon^* \left(\beta + e^{-b(x-1)}\right)}{\Delta x} \psi_{x-1}^* - \frac{\mu}{\Delta} \frac{x-2}{x} \psi_{x-2}^*. \tag{2.2}$$

For example,

$$\psi_1^* = \frac{\beta - \varepsilon^*(\beta+1)}{\Delta} \psi_0^*, \tag{2.3}$$

which is only non-negative when

$$\varepsilon^* \le \frac{\beta}{\beta+1}. \tag{2.4}$$

Inspecting the equation for ε, we find

$$\varepsilon^* = \frac{\beta}{\beta + \sum_{x=0}^{\infty} e^{-bx}\psi_x^* \Big/ \sum_{x=0}^{\infty}\psi_x^*} \geq \frac{\beta}{\beta+1}, \tag{2.5}$$

with equivalence only for $b = 0$. Hence, the only internally consistent nontrivial stationary state is

$$\begin{aligned} \varepsilon^* &= \frac{\beta}{\beta+1} \\ \psi_0^* &= 1 - \frac{\beta}{\beta+1} \\ \psi_x^* &= 0 \qquad \forall x > 0, \end{aligned} \tag{2.6}$$

i.e., extinction of the retrotransposon. It should be noted that extinction as the sole stationary state is a consequence of the absorbing nature of the wildtype state, ψ_0. Once cells lose all retrotransposons and enter ψ_0, there is no way to leave. One possible way to avoid this is by including the possibility of horizontal transfer. However, because the cells in our experiments do not undergo horizontal transfer and the rates of horizontal transfer in the wild are poorly quantified, we do not include this possibility in our modeling.

It is possible there exist interesting non-stationary states or other states that, while not truly stationary, are extremely long lived. We therefore simulated the model to determine the phase portrait of possible states as a function of b and μ for the initial conditions beginning with one copy of retrotransposon per cell ($\psi_1 = 0.1$ and $\varepsilon = 0.9$). For the simulations to be tractable, we set a boundary at some maximum number of retrotransposons per cell, $x_{\max}$. We consider setting such a boundary in two ways. First, we set a small fixed number of available insertion sites; once occupied, no further insertions are possible (*i.e.*, reflecting boundary conditions). We suggest that such conditions would correspond to the retrohoming of group II introns. Next, from our experimental data we find that when the growth rate has decreased below ~10% of the nominal value, cells cannot survive and cease to grow. Hence, as a second approach in our simulations we set a dynamic boundary by $x_{\max} = -\ln(0.1)/b$, and where insertions beyond this maximum number result in cell death (*i.e.*, absorbing boundary conditions). We suggest that these conditions would correspond to the nonspecific retrotransposition and amplification of retroelements.

Phase diagrams of simulations with populations of cells allowed to evolve over 10,000 generations are shown in Fig. 5A and Fig. 5B for reflecting and absorbing boundary conditions, respectively. For both conditions, the majority of parameter values quickly lead to extinction. With reflecting boundary conditions, Fig. 5A, a high insertion rate allows saturation of all available integration locations. This corresponds to retrohoming, where insertion rates correspond to ~1 per intron per cell per generation (23), but with low growth defect. As we now demonstrate, this saturated regime is approximately stable and will persist for extremely long times.

With a boundary set at $x_{\max}$, the model becomes

$$
\begin{aligned}
&1=\varepsilon+\sum_{x=0}^{\infty}\psi_x;\qquad \varepsilon\equiv\frac{E}{\Omega},\ \psi_x\equiv\frac{N_x}{\Omega}\\
&\frac{\partial\psi_{x<x_{\max}}}{\partial\tau}=\varepsilon e^{-bx}\psi_x-\beta(1-\varepsilon)\psi_x+\mu(x-1)\psi_{x-1}-\mu x\psi_x+\Delta(x+1)\psi_{x+1}-\Delta x\psi_x\\
&\frac{\partial\psi_{x_{\max}}}{\partial\tau}=\varepsilon e^{-bx_{\max}}\psi_{x_{\max}}-\beta(1-\varepsilon)\psi_{x_{\max}}+\mu(x_{\max}-1)\psi_{x_{\max}-1}-\Delta x_{\max}\psi_{x_{\max}}-\left[\mu x_{\max}\psi_{x_{\max}}\right]\\
&\frac{\partial\varepsilon}{\partial\tau}=\beta(1-\varepsilon)\sum_{x=0}^{x_{\max}}\psi_x-\varepsilon\sum_{x=0}^{x_{\max}}e^{-bx}\psi_x+\left[\mu x_{\max}\psi_{x_{\max}}\right],
\end{aligned}
\tag{2.7}
$$

with the terms in square brackets present only for absorbing boundary conditions. In this case, the ψ_x equations can be manipulated to yield recursion relations in terms of the $\psi_{x_{\max}}$ state. In particular, for reflecting boundary conditions we find

$$
\psi^*_{x_{\max}-1}=\frac{\beta-\varepsilon^*\left(\beta+e^{-bx_{\max}}\right)+\Delta x_{\max}}{\mu\left(x_{\max}-1\right)}\psi^*_{x_{\max}},
\tag{2.8}
$$

similar to the condition eq. (2.3) above. To avoid populating lower states and again running afoul of the conditions eqs. (2.4) and (2.5), we demand eq. (2.8) equal zero, yielding:

$$
\varepsilon^*=\frac{\beta+\Delta x_{\max}}{\beta+e^{-bx_{\max}}}\simeq\frac{\beta}{\beta+e^{-bx_{\max}}}
\tag{2.9}
$$

since $\Delta x_{\max}$ is small and approximately negligible. Therefore, only if $\Delta x_{\max}$ is negligible, a meta-stable and extremely long-lived state similar to eq. (2.6) and consistent with eq. (2.5) is possible,

$$
\begin{aligned}
&\varepsilon^*=\frac{\beta}{\beta+e^{-bx_{\max}}}\\
&\psi^*_{x_{\max}}=1-\frac{\beta}{\beta+e^{-bx_{\max}}}\\
&\psi^*_x=0\qquad\forall x<x_{\max}.
\end{aligned}
\tag{2.10}
$$

This demonstrates that the retrohoming strategy allows for low numbers of retrotransposons that are approximately stable and can persist for extremely long times.

For absorbing boundary conditions, the appropriate recursion relation relative to the state with the maximum number of retrotransposons is

$$
\psi^*_{x_{\max}-1}=\frac{\beta-\varepsilon^*\left(\beta+e^{-bx_{\max}}\right)+(\Delta+\mu)x_{\max}}{\mu\left(x_{\max}-1\right)}\psi^*_{x_{\max}}.
\tag{2.11}
$$

In contrast with the argument for retrohoming, the non-negligible factor of $\mu x_{\max}$ in the numerator renders the $\psi_{x_{\max}}$ state unstable. Hence, while the phase portrait Fig. 5B shows that there exists a small set of parameter values ($b < 0.01$ and $\mu \sim 10^{-3}$ retrotransposon^{-1} cell^{-1} generation^{-1}) where the retroelement is able to proliferate to high numbers, these states will eventually go extinct. Thus, the phase portrait with absorbing boundary conditions changes with time, and the result shown in Fig. 5B depends upon the interval over which the simulation is allowed to run. To determine the lifetime of these states, we performed simulations using absorbing boundary conditions for a variety of values of b and μ, and where we recorded the number of generations required for the retrotransposon to go extinct. The result is shown in Fig. 6. From this analysis, we see that the time required for a retrotransposon to go extinct can vary over ~7 orders of magnitude, depending upon

its dynamics and effects. For those parameter regimes corresponding to the aggressive autonomous retrotransposon LINE-1 ($b \geq 10^{-2}$, $\mu \geq 10^{-2}$ retrotransposon^{-1} cell^{-1} generation^{-1}), extinction is rapid, occurring in ~100 – 10,000 generations. Conversely, parameter regimes corresponding to the group II intron Ll.LtrB ($10^{-3} \leq b \leq 10^{-2}$, $10^{-9} \leq \mu \leq 10^{-6}$ retrotransposon^{-1} cell^{-1} generation^{-1}) can persist in low copy numbers (~1 per cell) for millions to tens of millions of generations. We additionally see that the small parameter regime where retrotransposons can proliferate to high copy numbers ($b \leq 10^{-2}$, $\mu \sim 10^{-3}$ - 10^{-4} retrotransposon^{-1} cell^{-1} generation^{-1}) persists for hundreds of thousands to millions of generations, and could be maintained longer with the inclusion of horizontal gene transfer.

FIGURES S1 – S4
Figure S1

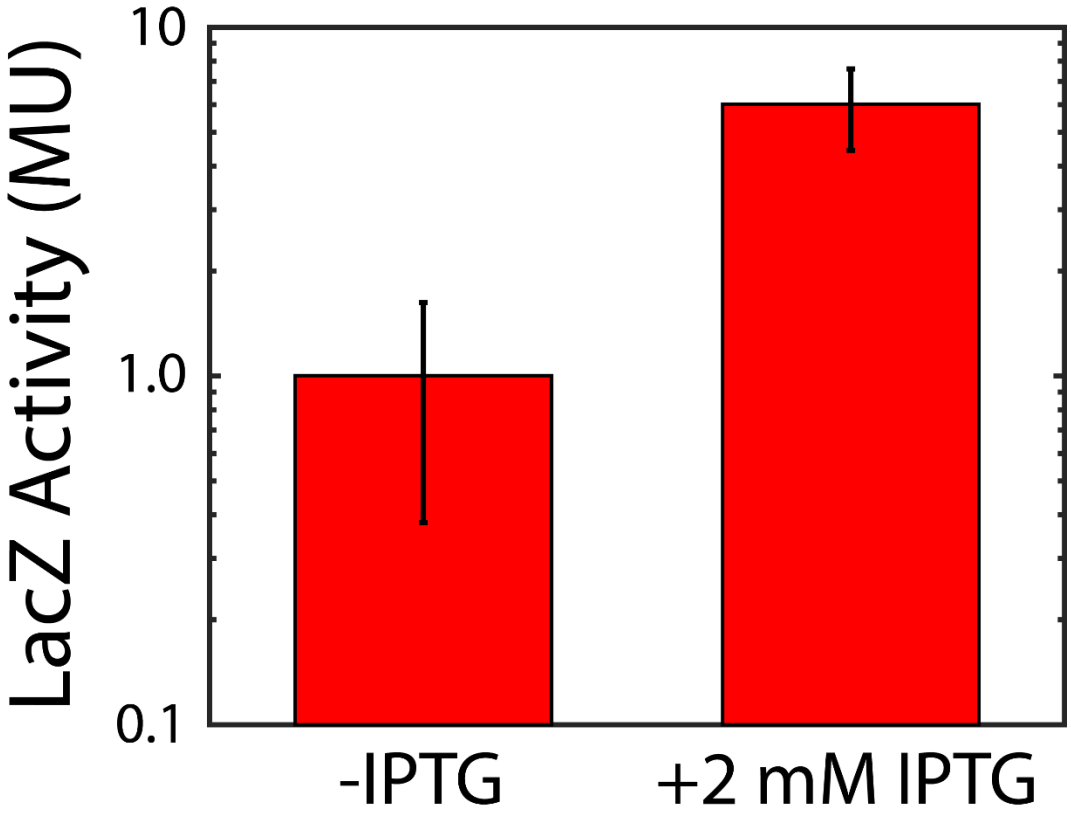

Figure S1. Expression from the hyper-spank promoter of pHCMC05 in *Bacillus subtilis.* LacZ activity of uninduced (left) and induced (right) *B. subtilis* 168 transformed with pHCMC05-*lacZYAX* was measured with a Miller assay (24). Bars are the mean of six independent replicates and error bars are the standard deviation.

Figure S2

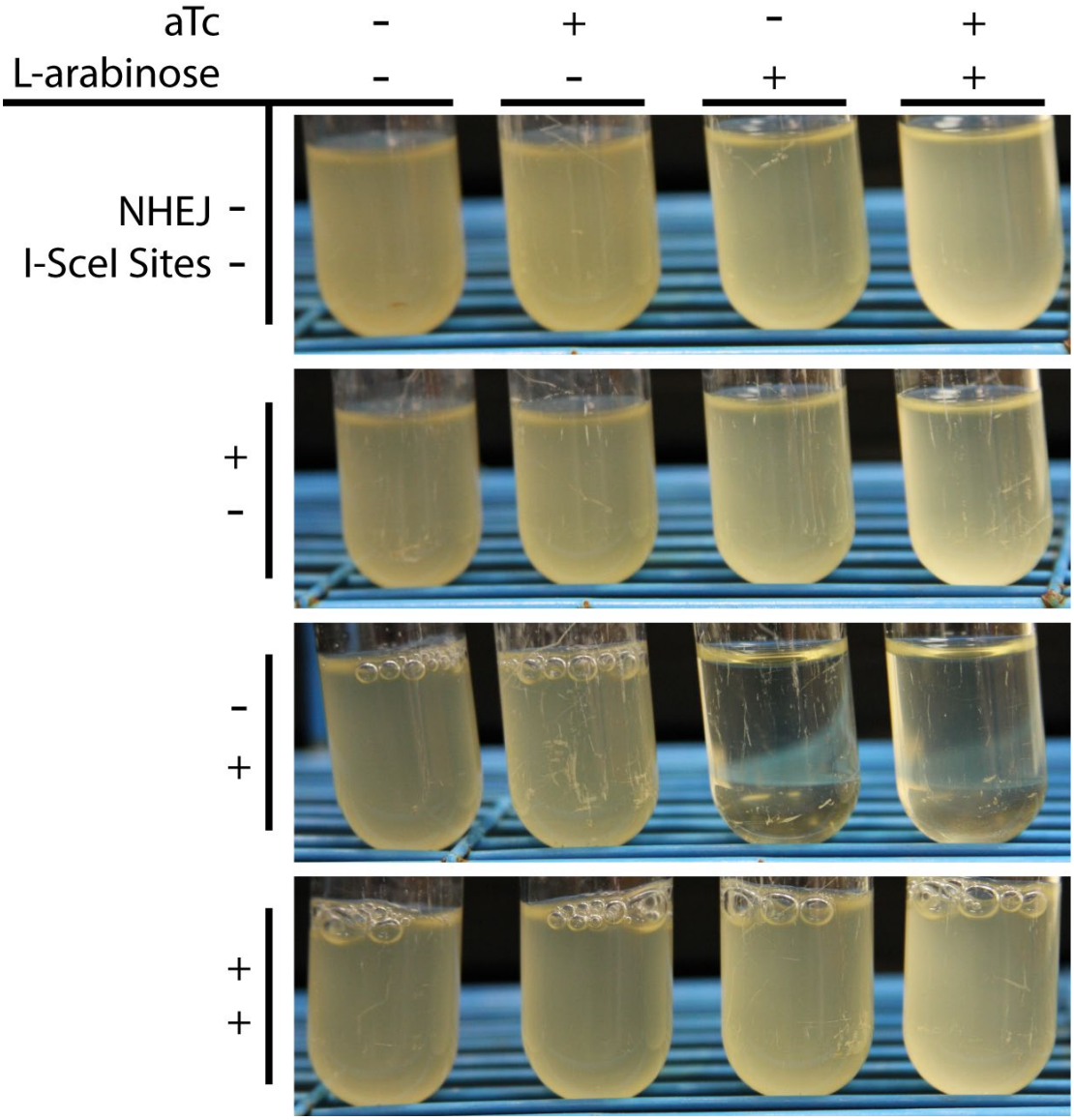

Figure S2. *B. subtilis* NHEJ enzymes function in *E. coli.* Turbidity of cultures grown for ~36 hours at 30 ºC after inoculation with identical amounts of cells. Bacterial strains are MG1655 Δ*lac* carrying the plasmid pTKRED, which expresses the homing endonuclease I-SceI when induced with L-arabinose. Additional plasmids and modifications are, from top to bottom - first row: pUC57-*kan*; second row: pUC57-*kan*-NHEJ; third row: pUC57-*kan* with I-SceI sites integrated at the *atpI* chromosomal locus (3-5); fourth row: pUC57-kan-NHEJ with I-SceI sites integrated at the *atpI* chromosomal locus. Columns correspond to different inducer conditions – first column: 0 aTc, 0 L-arabinose; second column: 100 ng/ml aTc, 0 L-arabinose; third row: 0 aTc, 0.4% w/v L-arabinose; fourth row: 100 ng/ml aTc, 0.4% L-arabinose. Lack of turbidity in row 3, columns 3 and 4 demonstrate that I-SceI double strand breaks are lethal to *E. coli* (3). Recovery of turbidity in row 4, columns 3 and 4 demonstrate that even low, leakage expression of *B. subtilis* NHEJ enzymes rescue *E. coli* growth.

Figure S3

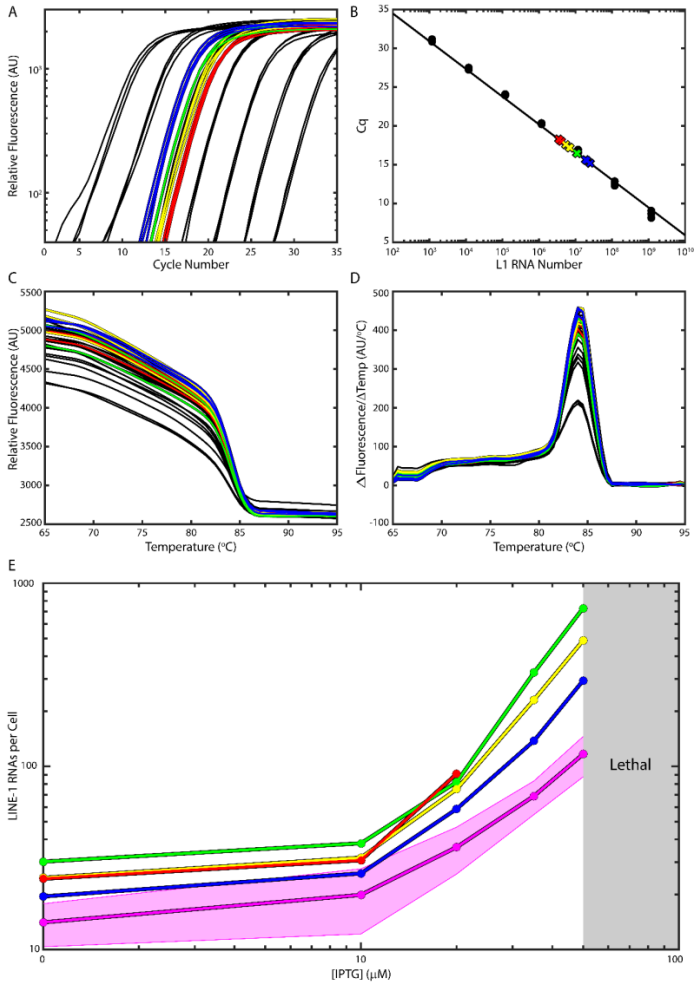

Figure S3. Quantitative RT-PCR to determine T7*lac* promoter response function. (A) Amplification curves of reverse transcribed serial 10x dilutions of *in vitro* transcribed TL1H RNA as an absolute standard (black), along with reverse transcribed RNA extracted from BL21(DE3) pTKIP-TL1H grown in M63 glucose medium with 0 (red), 10 μM (yellow), 20 μM (green), and 50 μM (blue) IPTG. (B) Absolute quantification of TL1H RNA numbers. Black circles are critical cycle numbers (C_q) of the *in vitro* standards from (A), colored crosses are C_qs of BL21(DE3) pTKIP-TL1H RNA with the threshold at ~200 AU. PCR efficiency was 90.5%. (C) Melting curves and their unimodal derivatives (D) resulting from qRT-PCR, demonstrating clean amplification of TL1H cDNA. Melting temp of the amplicon was 84.5 ºC. (E) RNA was extracted from BL21(DE3) pTKIP-TL1H grown in RDM glucose (magenta), RDM glycerol (blue), M63 glucose (yellow), cAA glucose (green), or M63 glycerol (red) with 0, 10, 20, 35, or 50 μM IPTG and quantified through qRT-PCR. Concentrations of IPTG higher than 50 μM were nonviable in all media except M63 glycerol, where concentrations higher than 20-35 μM were generally nonviable. The number of RNAs determined by qRT-PCR was divided by the number of cells added to the reaction, determined by measurement of OD600 and plating performed at the time of harvest. Shaded magenta region shows the standard error of the mean of four experimental replicates for samples prepared in RDM glucose. The standard errors of other samples are similar, but not shown for clarity. The number of LINE-1 RNAs per cell for each growth and induction condition thus obtained were used as the x-axis in main text Figure 3.

Figure S4

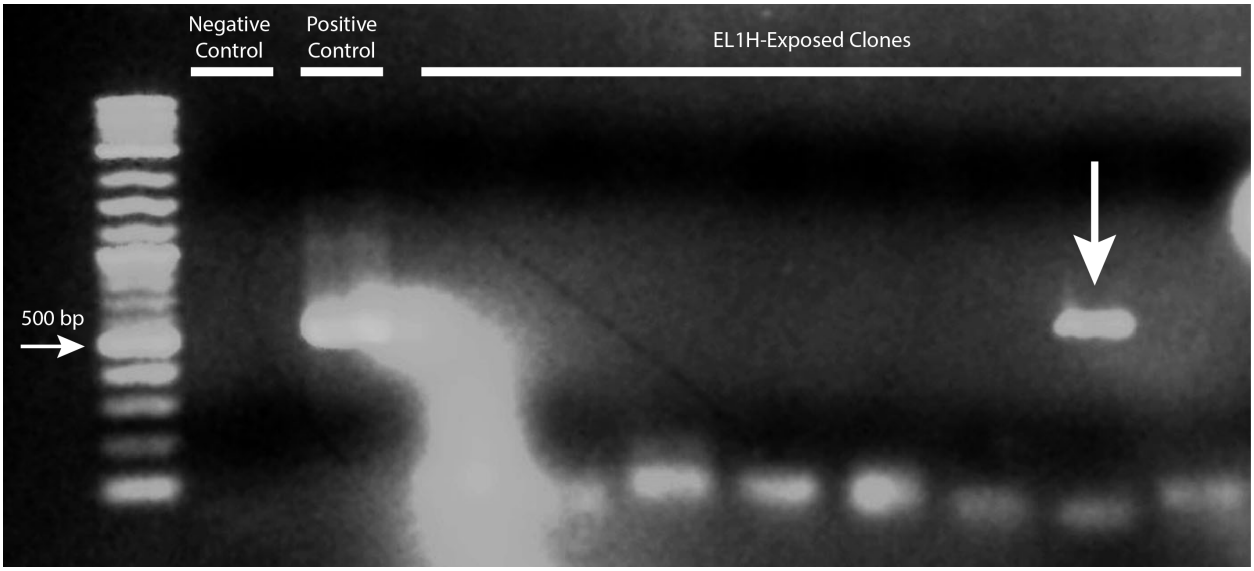

Figure S4. Detection of Full-Length EL1H Genomic Integrants. Representative 2% agarose gel electrophoresis of colony PCR of eight isolated colonies of BL21(DE3) that had been exposed to EL1H and cured of pTKIP-EL1H using primers that anneal to the 5' end of EL1H and produce a 500 bp amplicon. BL21(DE3) was used as a negative control, and BL21(DE3) pTKIP-EL1H as a positive control. The large fluorescent smear near the positive control band was a result of excess ethidium bromide staining. Since EL1H RNA is reverse transcribed and integrated starting from the 3' end, presence of the 5' end indicates complete integration. Out of 80 colonies tested, we found 3 colonies yielding this 500 bp product indicating complete integration of EL1H.

SUPPLEMENTAL REFERENCES

1. Ghatak S, Muthukumaran RB, & Nachimuthu SK (2013) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. *J. Biomol. Tech.* 24(4):224-231.
2. Beck CR*, et al.* (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141(7):1159-1170.
3. Kuhlman TE & Cox EC (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res* 38(6):e92.
4. Kuhlman TE & Cox EC (2010) A place for everything: chromosomal integration of large constructs. *Bioeng Bugs* 1(4):296-299.
5. Tas H, Nguyen CT, Patel R, Kim NH, & Kuhlman TE (2015) An integrated system for precise genome modification in *Escherichia coli*. *PLoS One* 10(9):e0136963.
6. Lutz R & Bujard H (1997) 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 Res* 25(6):1203-1210.
7. Warming S, Costantino N, Court DL, Jenkins NA, & Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Research* 33(4):e36-e36.
8. Landgraf D, Okumus B, Chien P, Baker TA, & Paulsson J (2012) Segregation of molecules at cell division reveals native protein localization. *Nat Meth* 9(5):480-482.
9. Suzuki J*, et al.* (2009) Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet* 5(4):e1000461.
10. Armitage P (1952) The statistical theory of bacterial populations subject to mutation. *Journal of the Royal Statistical Society. Series B (Methodological)*:1-40.
11. Charlesworth B & Charlesworth D (2009) The population dynamics of transposable elements. *Genetical Research* 42(1):1-27.
12. Charlesworth B & Langley CH (1986) The evolution of self-regulated transposition of transposable elements. *Genetics* 112(2):359-383.
13. Dolgin ES & Charlesworth B (2006) The Fate of transposable elements in asexual populations. *Genetics* 174(2):817-827.
14. Langley CH, Brookfield JFY, & Kaplan N (1983) Transposable elements in Mendelian population. I. A theory. *Genetics* 104(3):457-471.
15. Brookfield JFY (2005) The ecology of the genome — mobile DNA elements and their hosts. *Nature Reviews Genetics* 6:128.
16. Hellen EHB & Brookfield JFY (2013) Transposable element invasions. *Mobile Genetic Elements* 3(1):e23920.
17. Iranzo J, Cuesta JA, Manrubia S, Katsnelson MI, & Koonin EV (2017) Disentangling the effects of selection and loss bias on gene dynamics. *Proceedings of the National Academy of Sciences* 114(28):E5616-E5624.
18. Lynch M, Burger R, Butcher D, & Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84(5):339-344.
19. Hickey DA (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* 101(3-4):519-531.
20. Hickey DA (1992) Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. *Genetica* 86(1-3):269-274.
21. Wang P*, et al.* (2010) Robust Growth of Escherichia coli. *Current Biology* 20(12):1099-1103.
22. Hottes AK*, et al.* (2013) Bacterial adaptation through loss of function. *PLoS Genet* 9(7):e1003617.
23. Coros CJ*, et al.* (2005) Retrotransposition strategies of the Lactococcus lactis Ll.LtrB group II intron are dictated by host identity and cellular environment. *Molecular Microbiology* 56(2):509-524.
24. Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratories).