Testing the retroelement invasion hypothesis for the emergence of the ancestral eukaryotic cell

Gloria Lee[a,b], Nicholas A. Sherer[a,b], Neil H. Kim (김현일)[a,b], Ema Rajic[c], Davneet Kaur[a,b], Niko Urriola[a], K. Michael Martini[a,b,d], Chi Xue[a,b,d,e], Nigel Goldenfeld[a,b,d,e,1], and Thomas E. Kuhlman[a,b,d,f,g,1]

[a]Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]Center for the Physics of Living Cells, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [c]University of Illinois Laboratory High School, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [d]Institute for Universal Biology–NASA Astrobiology Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [e]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [f]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and [g]Department of Physics and Astronomy, University of California, Riverside, CA 92521

Contributed by Nigel Goldenfeld, October 4, 2018 (sent for review May 7, 2018; reviewed by Eugene V. Koonin, Michael Lynch, and Scott William Roy)

Phylogenetic evidence suggests that the invasion and proliferation of retroelements, selfish mobile genetic elements that copy and paste themselves within a host genome, was one of the early evolutionary events in the emergence of eukaryotes. Here we test the effects of this event by determining the pressures retroelements exert on simple genomes. We transferred two retroelements, human LINE-1 and the bacterial group II intron Ll.LtrB, into bacteria, and find that both are functional and detrimental to growth. We find, surprisingly, that retroelement lethality and proliferation are enhanced by the ability to perform eukaryotic-like nonhomologous end-joining (NHEJ) DNA repair. We show that the only stable evolutionary consequence in simple cells is maintenance of retroelements in low numbers, suggesting how retrotransposition rates and costs in early eukaryotes could have been constrained to allow proliferation. Our results suggest that the interplay between NHEJ and retroelements may have played a fundamental and previously unappreciated role in facilitating the proliferation of retroelements, elements of which became the ancestors of the spliceosome components in eukaryotes.

retroelements | LINE-1 | introns | evolution | junk DNA

The complexity of eukaryotes relative to bacteria and archaea is a consequence of the increased connectivity and plasticity of networks and interactions, rather than an increase in the amount of coding DNA (1). Such complexity is mediated by several mechanisms: one is the spliceosome, a complex molecular machine present in eukaryotes that operates on nascent mRNAs to generate mature transcripts. In some animals, for example, the spliceosome can generate multiple mRNAs through alternative splicings of a single primary transcript, allowing access to additional complexity without a concomitant increase in the amount of coding DNA. The spliceosome's primary role is the removal of introns, intervening sequences that disrupt the coding regions of eukaryotic genes and make up, for example, ∼24–37% of the human genome (2). Conversely, bacteria and archaea lack a spliceosome, and intervening sequences are present only in limited numbers as retrotransposable elements called group II introns.

Group II introns are found in only ∼30% of sequenced bacterial species and are generally present in low copy numbers of ∼1–10 per individual in those species where they exist (3). Conversely, retroelements in eukaryotes are vastly more abundant. For example, retrotransposons in humans comprise another ∼45% of the genome in addition to introns and make up the majority of so-called "junk DNA" (2, 4). The human retroelement LINE-1 (or "L1") alone makes up ∼17% of the genome, with ∼500,000 total integrants and ∼80–100 complete and active, or hot (L1H), copies per individual (5, 6). L1 activity contributes significantly to human genetic heterogeneity, disease, development, and evolution (7–10), and its known mechanisms of transposition show significant similarity to those of bacterial group II introns such as Ll.LtrB (11). This motivates their classification together as target-primed retrotransposons (12).

On the basis of manifold sequence, structural, and mechanistic similarities among bacterial group II introns, the spliceosome, eukaryotic spliceosomal introns, and autonomous eukaryotic retrotransposons, it has been hypothesized that an invasion of group II introns from an endosymbiotic eubacterial organelle contributed to the proliferation of introns within eukaryotic genomes before the last eukaryotic common ancestor (13, 14). If so, the resulting disruption to protein coding sequences could be alleviated by, among other contributing factors, consolidation of intron maturase splicing activity within the centralized spliceosome complex (3, 15) and the spatial decoupling of transcription and translation by a nuclear envelope (16, 17), although the order in which these developments occurred remains unclear. However, what enabled the proliferation of retroelements in eukaryotes and the evolutionary pressures and mechanisms limiting proliferation of retroelements in bacteria and archaea remain poorly understood and the subject of speculation (13, 18), particularly in light of the horizontal transfer of proliferative autonomous retroelements from humans to bacteria, as in the case of the recent transfer of L1 to the pathogen *Neisseria gonorrhoeae* (19).

Significance

Phylogenetic evidence suggests that a factor in the emergence of the ancestral eukaryotic cell may have been selection pressure resulting from invasion and proliferation of retroelements. Here we experimentally determine the effects of a retroelement invasion on genetically simple host organisms, and we demonstrate theoretically that the observed effects are sufficient to explain their observed rarity in bacteria. We also show that nonhomologous end-joining (NHEJ), a mechanism of DNA repair found in all extant eukaryotes, but only some bacteria, significantly enhances the efficiency of retrotransposition and the effects of retroelements on the host. We hypothesize that the interplay of NHEJ and retroelements may have played a previously unappreciated role in the evolution of advanced life.

Author contributions: N.A.S., N.H.K., N.G., and T.E.K. designed research; G.L., N.A.S., E.R., D.K., N.U., K.M.M., C.X., N.G., and T.E.K. performed research; N.A.S., K.M.M., C.X., N.G., and T.E.K. analyzed data; and G.L., N.A.S., D.K., K.M.M., C.X., N.G., and T.E.K. wrote the paper.

Reviewers: E.V.K., National Institutes of Health; M.L., Arizona State University; and S.W.R., San Francisco State University.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: nigel@illinois.edu or thomask@ucr.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1807709115/-/DCSupplemental.

Published online November 19, 2018.

To illuminate the changes in cellular machinery and tolerance of retroelements that would have been necessary to go from simple bacterial-like systems to eukaryotic ones, it would be important to understand precisely how retroelements may produce deleterious effects (20), what limits their activity in simple genomes, and what may have enabled their proliferation in eukaryotic genomes. To this end, we have constructed a bacterial version of L1 to quantitatively assess the function and effects of retroelement expression in the bacteria *Escherichia coli* and *Bacillus subtilis*, and we compare its effects with those of the bacterial group II intron Ll.LtrB. We find that L1 is functional in *E. coli*, successfully integrating into its genome. We demonstrate that retroelement expression is severely detrimental to both *E. coli* and *B. subtilis*, with wild-type *B. subtilis* in particular unable to tolerate any retroelement expression. We find that capacity of the host to perform nonhomologous end joining (NHEJ) repair of DNA double breaks increases retrotransposition rates by approximately three orders of magnitude, and that, surprisingly, NHEJ also strongly enhances bacterial sensitivity to the activity of retroelements. We show that these results demonstrate that retroelement activity generally leads to low copy numbers or extinction, as seen in bacteria and archaea, and that proliferation of retroelements in eukaryotes and subsequent addition of complexity to the eukaryotic genome may have been enabled by precise tuning of parameters, leading to suppression of growth defects and enhancement of integration efficiency.

Results

Description of Constructs. To fully appreciate how human LINE-1 (L1) and bacterial Ll.LtrB molecularly affect their host genomes, we first review their remarkably similar mechanisms of action, likely evincing their shared evolutionary origin. L1 codes for the proteins ORF1p and ORF2p, and Ll.LtrB codes for LtrA. Although ORF1p is thought to bind transcribed L1 mRNA to prevent degradation, ORF2p and LtrA both contain endonuclease and reverse transcriptase domains facilitating replication of the retroelements into new chromosomal loci. After transcription and translation, each protein binds in cis to its encoding RNA, and the resulting ribonucleoprotein particle can then bind and cut a target DNA molecule, using the endonuclease domain. The mRNA 3′ end hybridizes with the cut DNA, which is used by the reverse transcriptase domain as a primer for target-primed reverse transcription (21). This generates a new cDNA copy of the retroelement at a nonspecific location in the genome, a process known as ectopic retrotransposition. L1 retrotransposition rates are poorly quantified in human somatic cells, and in *E. coli*, ectopic retrotransposition of Ll.LtrB occurs with a frequency of ∼1 per 10^9 exposed cells (11, 22, 23). In its native host, *Lactococcus lactis*, Ll.LtrB can also undergo a process called retrohoming, in which integration is targeted to a unique, specific site in the *ltrB* gene with ∼100% efficiency (11, 22, 23).

One author (T.E.K.) extracted the active or hot L1 element (L1H) #4–35 (5) from his own genome and modified it for tunable expression in *E. coli*. PCR was used to add a T7*lac* promoter at the 5′ end and a strong ribosomal binding site (RBS) to drive ORF1p expression (Fig. 1*A*, *Top*). The construct, named TL1H, was ligated into the plasmid pTKIP-neo (24, 25) and transformed into *E. coli* strain BL21(DE3). TL1H expression is tunable via addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). We also synthesized de novo a version of L1H optimized for bacterial expression, EL1H (Fig. 2*A*). This construct uses *E. coli* codon bias, drives both *ORF1* and *ORF2* expression with consensus RBS sequences, and includes a ∼100-bp DNA-encoded poly-A tract at the 3′ end, a feature shown to enhance retrotransposition efficiency (26).

Similarly, Ll.LtrB was transformed into *E. coli* BL21(DE3) on the plasmid pET-TORF/retromobility indicator gene (RIG), a kind gift of the Marlene Belfort laboratory (11, 27). pET-TORF/RIG uses the same pBR322 plasmid backbone as pTKIP, and Ll.LtrB is expressed from the same T7*lac* promoter as employed for L1 expression (Fig. 1*B*). Hence, expression levels of both L1 and Ll.LtrB are comparable between experiments in *E. coli*. In *B. subtilis*, we subcloned TORF/RIG and EL1H into the shuttle vector pHCMC05 under control of the IPTG-inducible hyperspank promoter (28).

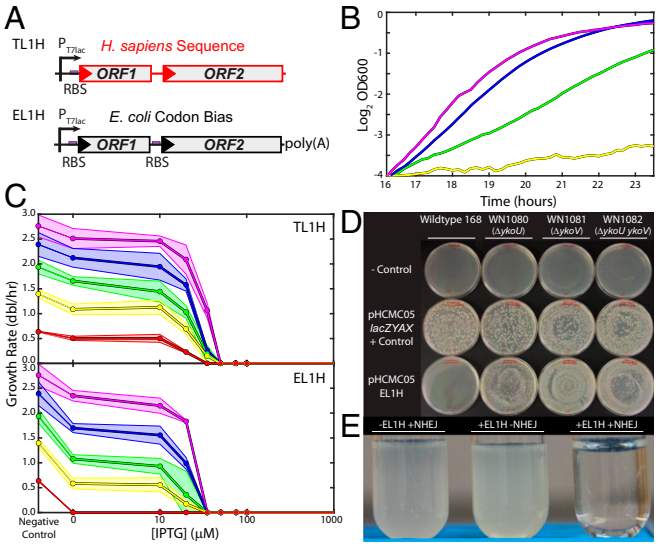

Fig. 1. Bacterial L1 elements and effects on growth. (*A*) L1 constructs used in this study. (*Top*) TL1H has human sequence (indicated by red), and was modified for expression in *E. coli* using a bacterial T7*lac* promoter and a consensus Shine Dalgarno RBS driving *ORF1*. (*Bottom*) EL1H is driven by P_{T7lac} and has consensus RBS for *ORF1* and *ORF2*. EL1H has a 100-bp 3′ poly(A) tract and has *E. coli* codon bias (indicated by black). (*B*) L1 is detrimental to *E. coli* growth. Example growth curves for BL21(DE3) pTKIP-TL1H growing in M63 glucose medium including 0 (magenta), 10 μM (blue), 20 μM (green), and 35 μM (yellow) IPTG. (*C*) Growth response as a function of [IPTG] for BL21 (DE3) pTKIP-TL1H (*Top*) and pTKIP-EL1H (*Bottom*) in various media; magenta, RDM glucose; blue, RDM glycerol; green, cAA glucose; yellow, M63 glucose; red, M63 glycerol. Growth rates were determined using the slope of the best fit regression of the initial linear portion of Log_2(OD600) vs. time, as in *B*. Points are the average of three independent replicates, and shaded regions indicate the SD. (*D*) Wild-type *B. subtilis* cannot survive transformation with EL1H (first column), whereas NHEJ knockouts relieve sensitivity (second column: Δ*ykoU*; third column Δ*ykoV*; fourth column Δ*ykoU* Δ*ykoV*). First row: negative control (TE buffer only); second row: positive control (pHCMC05-*lacZYAX*); third row: pHCMC05-EL1H. We performed transformations in four independent replicates with identical results. (*E*) Example *E. coli* BL21(DE3) cultures in RDM glucose grown for 20 h. (*Left*) pTKIP, pUC57-NHEJ. (*Middle*) pTKIP-EL1H, pUC57. (*Right*) pTKIP-EL1H, pUC57-NHEJ. All cultures contain no IPTG and 100 ng/mL aTc.

Effects of Retroelement Expression on Growth. To assess the effects of L1 expression on bacteria, we first transformed pTKIP-TL1H/EL1H constructs into *E. coli* BL21(DE3), a strain that expresses T7 polymerase (29). A decrease in growth rate in response to increasing L1 expression is immediately apparent in cultures titrated with IPTG (Fig. 1 *B* and *C*). To test the generality of this effect, we next assessed the effects of L1 expression on *B. subtilis*. In contrast to *E. coli*, *B. subtilis* is a Gram-positive bacterium able to repair DNA double-strand breaks through a simple two-protein NHEJ system in a manner similar to eukaryotes (30). Hence, we hypothesized that *B. subtilis* would be more resistant to L1 and cleavage of DNA by ORF2p endonuclease than *E. coli*, which lacks capacity for NHEJ repair. Instead, we find the opposite: wild-type *B. subtilis* 168 cannot survive transformation with pHCMC05-EL1H (Fig. 1*D*). Conversely, we obtain high-yield transformation of EL1H into *B. subtilis* strains with

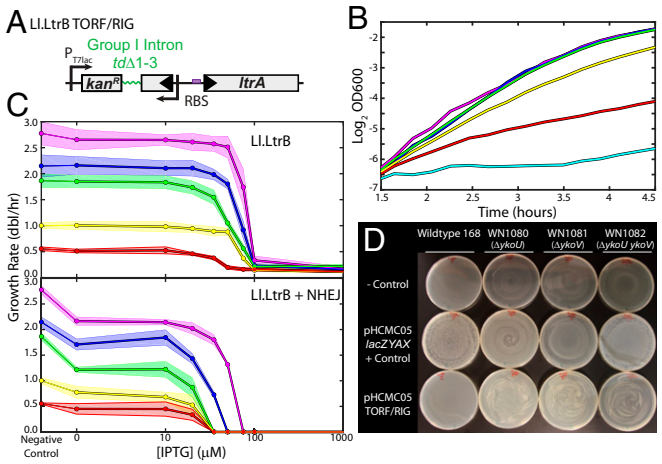

Fig. 2. Effects of Ll.LtrB on bacterial growth. (*A*) The Ll.LtrB construct TORF/RIG. TORF/RIG drives the expression of the Ll.LtrB group II intron, with the *ltrA* coding sequence toward the 3′ end of the intron driven by a strong RBS. TORF/RIG includes a kanamycin resistance gene encoded in the opposite orientation whose coding sequence is disrupted by the group I intron *td*Δ1–3 for determination of retrotransposition frequencies. (*B*) Expression of TORF/RIG is detrimental to *E. coli* growth. Example growth curves for BL21(DE3) pET-TORF/RIG growing in M63 glucose medium including 0 (magenta), 10 μM (blue), 20 μM (green), 35 μM (yellow), 50 μM (red), and 100 μM (cyan) IPTG. (*C*) Growth response as a function of [IPTG] for BL21(DE3) pET-TORF/RIG pZA31-*tetR* (*Top*) and pET-TORF/RIG pZA31-NHEJ (*Bottom*) in various media; magenta, RDM glucose; blue, RDM glycerol; green, cAA glucose; yellow, M63 glucose; red, M63 glycerol. Growth rates were determined using the slope of the best fit linear regression line of Log2(OD600) vs. time, as in *B*. Points are the average of three independent replicates, and shaded regions indicate the SD. (*D*) Wild-type *B. subtilis* cannot survive transformation with pHCMC05-TORF/RIG (first column), whereas NHEJ knockouts somewhat relieve sensitivity (second column: Δ*ykoU*; third column: Δ*ykoV*; fourth column: Δ*ykoU* Δ*ykoV*). First row: negative control (TE buffer only); second row: positive control (pHCMC05-*lacZYAX*); third row: pHCMC05-TORF/RIG. We performed transformations in four independent replicates with identical results.

knockouts of the individual NHEJ repair enzymes Ku (*ykoV*) or LigD (*ykoU*), as well as with both Ku and LigD knocked out (31). A Miller assay of expression level from the positive control plasmid pHCMC05-*lacZYAX* expressing *E. coli*'s metabolic *lac* enzymes from the hyper-spank promoter shows that expression is weak but leaky in *B. subtilis* (*SI Appendix*, Fig. S1). We conclude that wild-type *B. subtilis* is extremely sensitive to even low levels of L1H expression, and that this growth defect is enhanced by NHEJ repair.

We next cloned and expressed the *B. subtilis* NHEJ enzymes (BsKu and BsLigD) in *E. coli* under control of the aTc inducible P_{Ltet01} promoter (32). We verified that BsKu and BsLigD were functional in *E. coli* by ensuring their ability to rescue strains where we induced the homing endonuclease I-SceI to create double-stranded chromosomal breaks at chromosomally integrated I-SceI recognition sites (*SI Appendix*, Fig. S2) (24, 25, 33). We then verified the enhancement of lethality of L1 by NHEJ by cotransformation of BL21(DE3) with plasmids expressing L1 and NHEJ enzymes. We find that even low leakage expression of EL1H without addition of IPTG is lethal to *E. coli* with concomitant induction of NHEJ enzymes with 100 ng/mL aTc (Fig. 1*E*).

To quantify the effect of L1 and Ll.LtrB RIG expression on *E. coli* growth, we measured the growth rate as a function of expression level by titration with IPTG and periodic measurement of optical density in a variety of growth media (Fig. 1 *B* and *C* for L1 and Fig. 2 *B* and *C* for Ll.LtrB). Even with no induction, leaky expression of L1 significantly reduces the growth rate relative to the parent strain carrying an empty plasmid, and complete growth arrest occurs at IPTG concentrations of 35–50 μM (Fig. 1*C*).

We measured the transcriptional response function of the T7*lac* promoter by qRT-PCR (*SI Appendix*, Fig. S3 *A–D*) of L1 mRNA extracted from bacteria grown at those IPTG concentrations at which cultures survive (*SI Appendix*, Fig. S3*E*). The resulting dose-responses as a function of L1 RNAs and Ll.LtrB RNAs per cell are shown in Fig. 3. In Fig. 3*A*, data from TL1H are plotted as blue points, EL1H as red points, and EL1H+NHEJ as black points. In Fig. 3*B*, data from Ll.LtrB are plotted as red points, and Ll.LtrB+NHEJ as black points. The normalized growth rate decreases exponentially with increasing numbers of retroelement RNAs, and growth conditions do not affect this response. Solid lines in Fig. 3 correspond to fits to the exponential function $\exp[-bL]$, where L is the average number of L1 or Ll.LtrB RNAs per cell and the parameter b quantifies the growth defect and sensitivity to retroelement expression. We find that, on average, each L1 transcript yields a decrease in *E. coli*'s

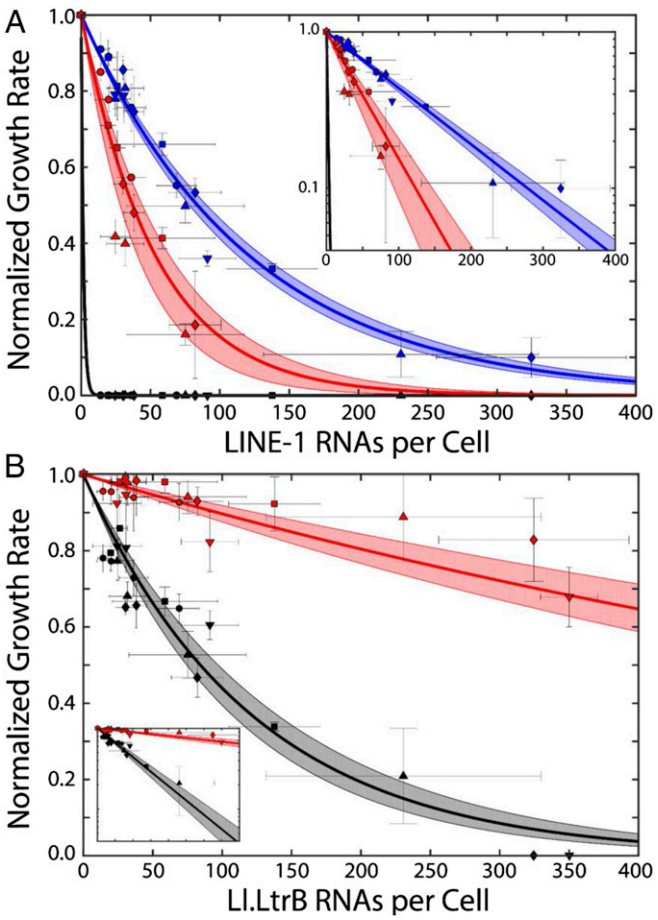

Fig. 3. Quantification of physiological effects of retroelement expression. (*A*) Normalized growth rate as a function of L1 expression on *E. coli* growth in a variety of media. ●, RDM glucose; ■, RDM glycerol; ◇, cAA glucose; ▲, M63 glucose; ▼, M63 glycerol. Blue points: TL1H; red points: EL1H; black points: EL1H and TL1H+NHEJ. Each point corresponds to the mean of three growth and four qRT-PCR measurements; error bars: SEM. Solid lines: fits to $\exp[-b^*L]$, yielding $b = 0.0083 \pm 0.0006$ (TL1H), $b = 0.019 \pm 0.006$ (EL1H), and $b = 0.600 \pm 0.031$ (TL1H and EL1H+NHEJ). Fit errors are 95% CI (shaded regions). (*Inset*) Same, with log *y* axis. (*B*) Same as *A*, quantifying effects of pET-TORF/RIG pZA31-*tetR* (red) and pET-TORF/RIG pZA31-NHEJ (black). (*Inset*) Scales are identical to *A*. Exponential fits yield $b = 0.0011 \pm 0.0002$ (−NHEJ), $b = 0.0082 \pm 0.0011$ (+NHEJ).

EVOLUTION

growth rate of ∼0.83 ± 0.06% (TL1H) or 1.9 ± 0.6% (EL1H) in the absence of NHEJ, and ≥45 ± 1.6% with NHEJ. Each Ll.LtrB transcript reduces the growth rate by 0.11 ± 0.02% in the absence of NHEJ and 0.82 ± 0.11% with NHEJ. As might be expected because of the ability of LtrA maturase to excise Ll.LtrB from interrupted genes, the growth defect resulting from Ll.LtrB is weaker than that from L1.

The Ll.LtrB growth defect is also evident in plating assays to determine retrotransposition efficiency. Induction of Ll.LtrB expression with 100 μM IPTG reduces the number of viable colony forming units (cfus) per milliliter per OD by ∼10×. Simultaneous induction of Ll.LtrB with 100 μM IPTG and induction of NHEJ enzymes with 100 ng/mL anhydrotetracycline reduces viable cfus/OD/mL by ∼100×, whereas induction of expression of NHEJ enzymes alone has no detectable effect.

Finally, we attempted to transform Ll.LtrB into *B. subtilis* as the plasmid pHCMC05-TORF/RIG, with Ll.LtrB under control of the *lacI*-regulated hyper-spank promoter. As with L1, we find that wild-type *B. subtilis* 168 cannot survive transformation with Ll.LtrB, whereas knockouts for the NHEJ genes *ykoU*, *ykoV*, and both *ykoU* and *ykoV* are transformed with high yield (Fig. 2*D*).

L1 and Ll.LtrB Successfully Integrate in *E. coli* Chromosome. Several lines of evidence demonstrate that both Ll.LtrB and L1 successfully retrotranspose into the bacterial chromosome. *E. coli* carrying the pTKIP-EL1H plasmid was induced to express EL1H for several generations. Surviving cells were transformed with the plasmid pTKRED, which expresses the homing endonuclease I-SceI (24, 25, 33), to digest pTKIP-EL1H in vivo. Colony PCR and gel electrophoresis (Fig. 4*A*) show that cells no longer carrying pTKIP-EL1H still contain EL1H, demonstrating successful chromosomal integration. Colony PCR was also used to determine whether any surviving cells acquired the entire active EL1H sequence, using primers that amplified a 500-bp portion near the 5′ end. A positive signal was detected in 3 of 80 screened colonies, and was verified via sequencing (*SI Appendix*, Fig. S4).

As another phenotypic test, we synthesized the construct EL1HID (Fig. 4*B*) to report EL1H integration via fluorescence. EL1HID contains an mTFP1 gene expressed from a strong promoter whose −10 and −35 sequences are separated by the group I intron *td*Δ1–3 (34). After transcription, *td*Δ1–3 catalyzes its own excision from the transcribed mRNA, which reconstitutes the mTFP1 promoter, and allows expression of teal fluorescent protein on successful retrotransposition. When EL1HID was transformed into *E. coli* and weakly induced, ∼1% of cells exhibited a total fluorescence >10× brighter than any cells from control strains. With simultaneous weak induction of NHEJ enzymes, the fluorescent population increased to ∼80% (Fig. 4 *C*–*F*).

Using a similar RIG in Ll.LtrB (11), we found that NHEJ also enhances the rate of Ll.LtrB ectopic retrotransposition. The RIG is composed of a kanamycin resistance gene, the sequence of which is interrupted by *td*Δ1–3 (Fig. 2*A*). After growing cultures of *E. coli* expressing Ll.LtrB and plating on selective media containing kanamycin, we determined the frequency of successful ectopic retrotransposition to be $3.0 \pm 0.9 \times 10^{-9}$, consistent with measurements by Coros et al. (11). For cells simultaneously expressing NHEJ enzymes, the efficiency increased approximately three orders of magnitude to $4.6 \pm 0.4 \times 10^{-6}$.

Discussion

That both human L1H and bacterial Ll.LtrB expression results in exponential decrease in growth rate suggests a simple universal underlying mechanism: each retroelement mRNA transcript has a probability of integrating and disrupting essential genes affecting growth. In the simplest model of this type, the probability

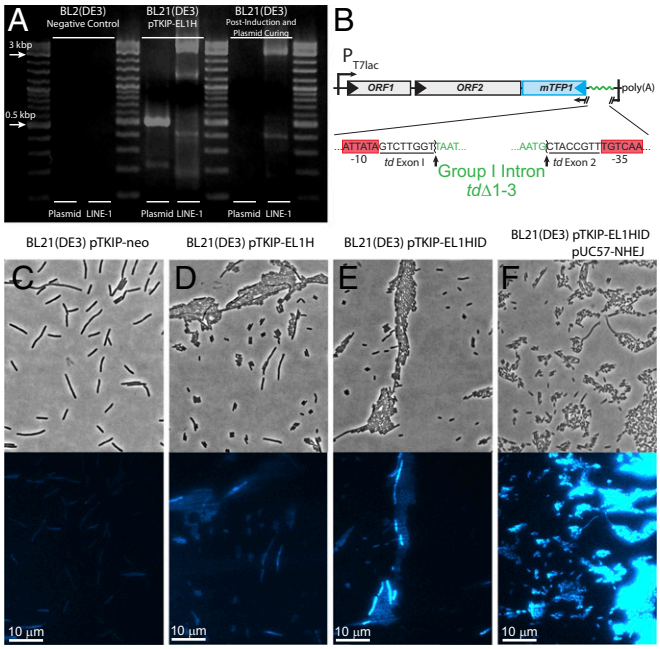

Fig. 4. L1 integrates into the *E. coli* genome. (*A*) Nonclonal colony PCR to detect EL1H (LINE-1 lanes) and pTKIP (plasmid lanes). (*Left*) BL21(DE3) negative control. (*Middle*) BL21(DE3) pTKIP-EL1H positive control. (*Right*) Strain post EL1H exposure and plasmid curing. (*B*) EL1HID, a construct for detecting successful retrotransposition of EL1H in individual cells by fluorescence. The integration detection cassette (ID) consists of *mTFP1* with consensus σ^{70} promoter and RBS. −10 and −35 core promoter sequences are split by the group I intron *td*Δ1–3 (sequences shown below). Upon successful retrotransposition, the cell fluoresces blue. (*C*–*F*) Phase contrast (*Top*) and fluorescence microscopy (*Bottom*) of induced (20 μM IPTG) (*C*) BL21(DE3) pTKIP-neo negative control, (*D*) BL21(DE3) pTKIP-EL1H, (*E*) BL21(DE3) pTKIP-EL1HID, and (*F*) BL21(DE3) pTKIP-EL1HID pUC57-NHEJ (0 IPTG, 5 ng/mL aTc).

that a cell will survive is described by a binomial distribution with zero disruptive integration events, leading to an exponential decrease in growth rate with transcript number; including variable integration rates and physiological responses does not significantly affect the resulting behavior (*SI Appendix, Supplementary Analysis*). As a consequence, in bacteria, the growth defect is a monotonically increasing function of the integration rate. To further understand how retrotransposons will proliferate within a host genome, we constructed a simple model of retroelement activity, motivated by the existing body of work on retroelement activity (20, 35–41), and analyzed its dynamics (*SI Appendix, Supplementary Analysis*). Populations of asexually multiplying cells were simulated on the basis of measured integration rates and growth defects, and allowed to evolve over 10,000 generations. The resulting phase diagrams are shown in Fig. 5 for retrohoming (reflective boundary conditions) and retrotransposition (absorbing boundary conditions), respectively. We find that retrohoming generally leads to low but stable numbers of retroelements, whereas the parameters with which retrotransposition occurs must be finely tuned to achieve long-lived states with proliferation of retrotransposons in the host.

The phase portrait in Fig. 5*B* shows that there exists a small set of parameter values (low growth defect, b, of less than 0.01 and high integration rate, μ, of $\sim 10^{-3}$ retrotransposon^{-1}·cell^{-1}·generation^{-1}), where retrotransposons can proliferate to high numbers. Coupling of the integration rate and growth defect implies that increases in the integration rate inexorably push bacteria toward the upper right of the phase diagram, and thus toward extinction. Hence, the bacterial phase space is highly

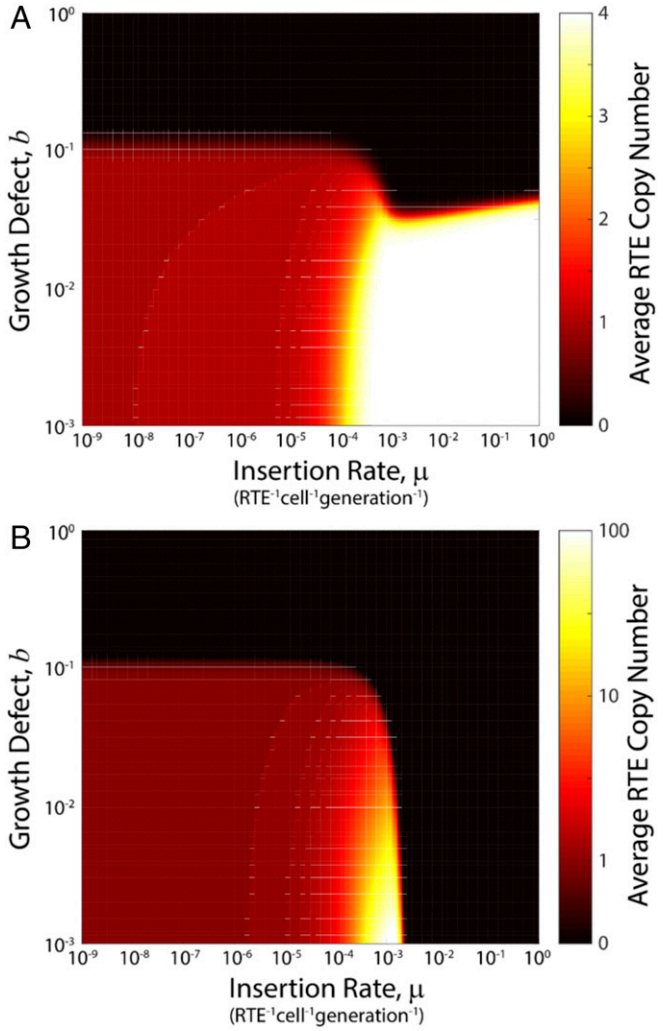

Fig. 5. Phase diagram of retrotransposon dynamics. We simulated the model of retrotransposon dynamics, *SI Appendix*, Eq. 2.7 (*SI Appendix*, *Supplementary Analysis*), using a total system size [defined as the number of available empty sites in the environment plus (effective) number of individuals in the population] of $\Omega = 10^9$, with an initial population of $\psi_1 = 0.1$ and all other states empty. This initial state was allowed to evolve for 10,000 generations with $\Delta = 10^{-8}$ retrotransposon^{-1}·cell^{-1}·generation^{-1} and $\beta = 10^{-2}$ cell^{-1}·generation^{-1}, at the conclusion of which we calculated the average number of retrotransposons per cell over the extant population. Results are shown for (*A*) reflecting boundary conditions with $x_{max} = 4$ and (*B*) absorbing boundary conditions with $x_{max} = -\ln(0.1)/b$.

constrained, and they are unlikely to be found within this small proliferative regime.

To demonstrate this, we performed simulations using absorbing boundary conditions across parameter values, and for each, 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 more than ~7 orders of magnitude, depending on its dynamics and effects. For those parameter regimes corresponding to the aggressive autonomous retrotransposon L1 ($b \geq 10^{-2}$, $\mu \geq 10^{-2}$ retrotransposon^{-1}·cell^{-1}·generation^{-1}), extinction of retroelements 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 also see that the small parameter regime in which 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.

Hence, this simple model suggests that for retroelements to proliferate to high numbers within asexual populations, the coupling of integration rate and growth defect must be weakened. In addition, increases in retrotransposition efficiency by NHEJ, present in all extant eukaryotes, must also be compensated for by suppression of the growth defect to enable proliferation. Indeed, it is hypothesized that many eukaryotic features arose specifically to mitigate the effects of retroelements (3, 13, 16, 17, 42, 43). For example, the nuclear membrane allows the spliceosome to complete intron excision before nuclear export and translation (16, 17). Furthermore, important spliceosomal components are derived from group II introns, and consolidation of splicing activity into the spliceosomal complex may facilitate efficient intron removal (3, 13). With the spliceosome, further complexity added to the eukaryotic genome by retroelements could then be exploited for benefit through, for example, alternative splicing by exon-skipping in some eukaryotes. In summary, proliferation of retroelements plays a dual role. On the one hand, group II introns create genome instability and negative physiological effects. On the other hand, by duplicating themselves, copies of group II introns are free to diversify and become the ancestors of both spliceosome and spliceosomal introns (13, 14).

We hypothesize that NHEJ enhances retrotransposition by directly joining the newly reverse-transcribed retroelement with the remaining free end of the endonuclease-induced break. Without NHEJ, this break can only be repaired through homologous recombination, generally leading to removal of the integrant and apparent low retrotransposition efficiencies, as observed in NHEJ-deficient *E. coli*. However, it is surprising that minimal, two-protein bacterial NHEJ systems interact with and enhance human L1 retrotransposition efficiency. Intriguingly,

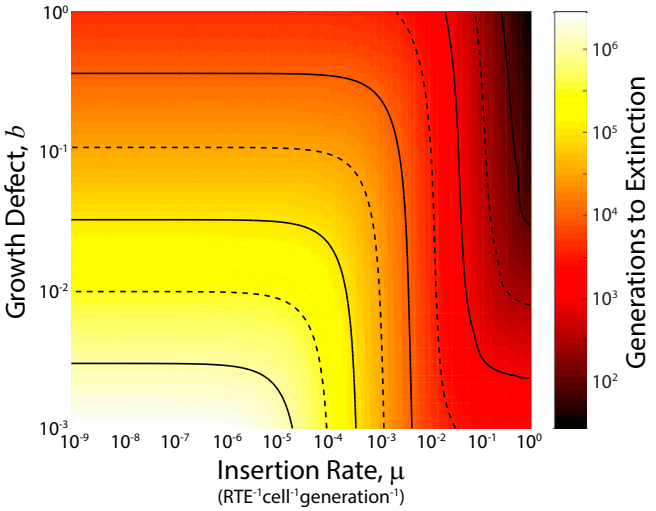

Fig. 6. Time to extinction of retrotransposons in a bacterial population. Simulations of the model *SI Appendix*, Eq. 2.7 (*SI Appendix*, *Supplementary Analysis*), with absorbing boundary condition at $x_{max} = -\ln(0.1)/b$, system size of $\Omega = 10^9$, $\Delta = 10^{-8}$ retrotransposon^{-1}·cell^{-1}·generation^{-1}, $\beta = 10^{-2}$ cell^{-1}·generation^{-1} and initial population of $\psi_1 = 0.1$ with all other states empty. Color indicates the number of generations required for the average number of retrotransposons per cell to drop below 1/Ω. Solid contour lines indicate major decade divisions; dashed contour lines indicate half-decade divisions.

NHEJ proteins also heavily associate with telomeres and are required for proper telomere length regulation and end protection (44, 45). Furthermore, the reverse transcriptase activity of telomerase likely shares a common ancestor with group II introns, and in some organisms (e.g., *Drosophila*), telomere maintenance is performed by retroelements rather than telomerase (13). Combined with our results, we conjecture that NHEJ systems, together with retroelement proliferation, were implicated in the unexplained evolutionary transition from generally circular bacterial chromosomes to linear eukaryotic chromosomes (13, 42, 45).

Methods

Strains and Media. Manipulation of constructs was performed with *E. coli* strain NEBTurbo (New England Biosciences). Experiments assaying effects of retroelement expression in *E. coli* were performed in the strain BL21(DE3). *B. subtilis* experiments were performed with strain 168, as well as Δ*ykoU* (WN1080/BFS1845), Δ*ykoV* (WN1081/BFS1846), and Δ*ykoU* Δ*ykoV* (WN1082/BFS1847) knockout strains (31).

Plasmid Construction. See *SI Appendix* for descriptions of plasmid constructs.

***B. subtilis* Transformation.** *B. subtilis* transformation was performed as described in ref. 46, with modifications (*SI Appendix, Supplementary Methods*).

LacZ Measurements. *B. subtilis* 168 pHCMC05-*lacZYAX* was inoculated into RDM glucose and, when OD600 of the culture reached ∼0.3–0.5, 0.5 mL culture was added to 0.5 mL Z-buffer + 0.1% SDS with 100 μL toluene. This mixture was vortexed and incubated in a 37 °C water bath for 30 min. The LacZ assay was then performed as previously described (*SI Appendix*, Fig. S1) (47, 48).

Growth Rate Determination. Detailed methods of growth rate determination can be found in the *SI Appendix*.

Microscopy. To perform fluorescence microscopy, 50 μL samples of culture were spread onto 1% agarose pads prepared on glass slides, covered with a #1.5 glass coverslip and imaged; see *SI Appendix* for details.

Quantitative RT-PCR. Methods for qRT-PCR can be found in *SI Appendix*.

Ll.LtrB Retrotransposition Frequency Assays. Retrotransposition efficiency of Ll.LtrB with and without NHEJ expression was determined by the protocol of ref. 11, with modifications; see *SI Appendix, Supplementary Methods*.

ACKNOWLEDGMENTS. We thank Prof. Douglas Mitchell (University of Illinois Urbana–Champaign) for the gift of *B. subtilis* 168 and plasmids, Wayne L. Nicholson (University of Florida) for the gift of *B. subtilis* NHEJ knockout strains, and Marlene Belfort (University of Albany, State University of New York) for the gift of Ll.LtrB constructs and sequence information. This work was supported by the NSF Center for the Physics of Living Cells (Grant PHY 1430124), the Alfred P. Sloan Foundation (Award FG-2015-65532), and the Institute for Universal Biology, through partial support by the NASA Astrobiology Institute (NAI) under Cooperative Agreement No. NNA13AA91A issued through the Science Mission Directorate. G.L. is supported by the National Science Foundation Graduate Research Fellowship Program under Grant DGE-1144245. All work was reviewed and approved by the University of Illinois Urbana–Champaign Institutional Review Board and Institutional Biosafety Committee.

1. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA).
2. Lander ES, et al.; International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921, and erratum (2001) 412:565.
3. Lambowitz AM, Belfort M (2015) Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol Spectr* 3:MDNA3-0050-2014.
4. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7:e1002384.
5. Beck CR, et al. (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141: 1159–1170.
6. Richardson SR, et al. (2015) The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol Spectr* 3:MDNA3-0061-2014.
7. Goodier JL (2014) Retrotransposition in tumors and brains. *Mob DNA* 5:11.
8. Baillie JK, et al. (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479:534–537.
9. Coufal NG, et al. (2009) L1 retrotransposition in human neural progenitor cells. *Nature* 460:1127–1131.
10. Kano H, et al. (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev* 23:1303–1312.
11. Coros CJ, et al. (2005) Retrotransposition strategies of the Lactococcus lactis Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol* 56:509–524.
12. Beauregard A, Curcio MJ, Belfort M (2008) The take and give between retrotransposable elements and their hosts. *Annu Rev Genet* 42:587–617.
13. Novikova O, Belfort M (2017) Mobile group II introns as ancestral eukaryotic elements. *Trends Genet* 33:773–783.
14. Irimia M, Roy SW (2014) Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol* 6:a016071.
15. Lambowitz AM, Zimmerly S (2011) Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 3:a003616.
16. Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440:41–45.
17. Doolittle WF (2014) The trouble with (group II) introns. *Proc Natl Acad Sci USA* 111: 6536–6537.
18. Boeke JD (2003) The unusual phylogenetic distribution of retrotransposons: A hypothesis. *Genome Res* 13:1975–1983.
19. Anderson MT, Seifert HS (2011) Opportunity and means: Horizontal gene transfer from the human host to a bacterial pathogen. *MBio* 2:e00005-11.
20. Iranzo J, Cuesta JA, Manrubia S, Katsnelson MI, Koonin EV (2017) Disentangling the effects of selection and loss bias on gene dynamics. *Proc Natl Acad Sci USA* 114: E5616–E5624.
21. Moran JV, et al. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87:917–927.
22. Cousineau B, Lawrence S, Smith D, Belfort M (2000) Retrotransposition of a bacterial group II intron. *Nature* 404:1018–1021.
23. Ichiyanagi K, et al. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol* 46:1259–1272.
24. Kuhlman TE, Cox EC (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res* 38:e92.
25. Tas H, Nguyen CT, Patel R, Kim NH, Kuhlman TE (2015) An integrated system for precise genome modification in *Escherichia coli*. *PLoS One* 10:e0136963.
26. Doucet AJ, Wilusz JE, Miyoshi T, Liu Y, Moran JV (2015) A 3′ poly(A) tract is required for LINE-1 retrotransposition. *Mol Cell* 60:728–741.
27. Beauregard A, Chalamcharla VR, Piazza CL, Belfort M, Coros CJ (2006) Bipolar localization of the group II intron Ll.LtrB is maintained in Escherichia coli deficient in nucleoid condensation, chromosome partitioning and DNA replication. *Mol Microbiol* 62:709–722.
28. Nguyen HD, et al. (2005) Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* 54:241–248.
29. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189:113–130.
30. Bowater R, Doherty AJ (2006) Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet* 2:e8.
31. Moeller R, et al. (2007) Role of DNA repair by nonhomologous-end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *J Bacteriol* 189:3306–3311.
32. 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* 25:1203–1210.
33. Kuhlman TE, Cox EC (2010) A place for everything: Chromosomal integration of large constructs. *Bioeng Bugs* 1:296–299.
34. Belfort M, Chandry PS, Pedersen-Lane J (1987) Genetic delineation of functional components of the group I intron in the phage T4 *td* gene. *Cold Spring Harb Symp Quant Biol* 52:181–192.
35. Charlesworth B, Charlesworth D (1983) The population dynamics of transposable elements. *Genet Res* 42:1–27.
36. Charlesworth B, Langley CH (1986) The evolution of self-regulated transposition of transposable elements. *Genetics* 112:359–383.
37. Dolgin ES, Charlesworth B (2006) The fate of transposable elements in asexual populations. *Genetics* 174:817–827.
38. Langley CH, Brookfield JFY, Kaplan N (1983) Transposable elements in mendelian populations. I. A theory. *Genetics* 104:457–471.
39. Brookfield JFY (2005) The ecology of the genome: Mobile DNA elements and their hosts. *Nat Rev Genet* 6:128–136.
40. Hellen EHB, Brookfield JFY (2013) Transposable element invasions. *Mob Genet Elements* 3:e23920.
41. Lynch M, Bürger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84:339–344.
42. Koonin EV (2016) Viruses andd mobile elements as drivers of evolutionary transitions. *Philos Trans R Soc B* 371:20150442.
43. Brodt A, Lurie-Weinberger MN, Gophna U (2011) CRISPR loci reveal networks of gene exchange in archaea. *Biol Direct* 6:65.
44. Riha K, Heacock ML, Shippen DE (2006) The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology. *Annu Rev Genet* 40:237–277.
45. de Lange T (2015) A loopy view of telomere evolution. *Front Genet* 6:321.
46. Sysoeva TA, et al. (2015) Structural characterization of the late competence protein ComFB from Bacillus subtilis. *Biosci Rep* 35:e00183.
47. Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
48. Kuhlman T, Zhang Z, Saier MH, Jr, Hwa T (2007) Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci USA* 104:6043–6048.

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

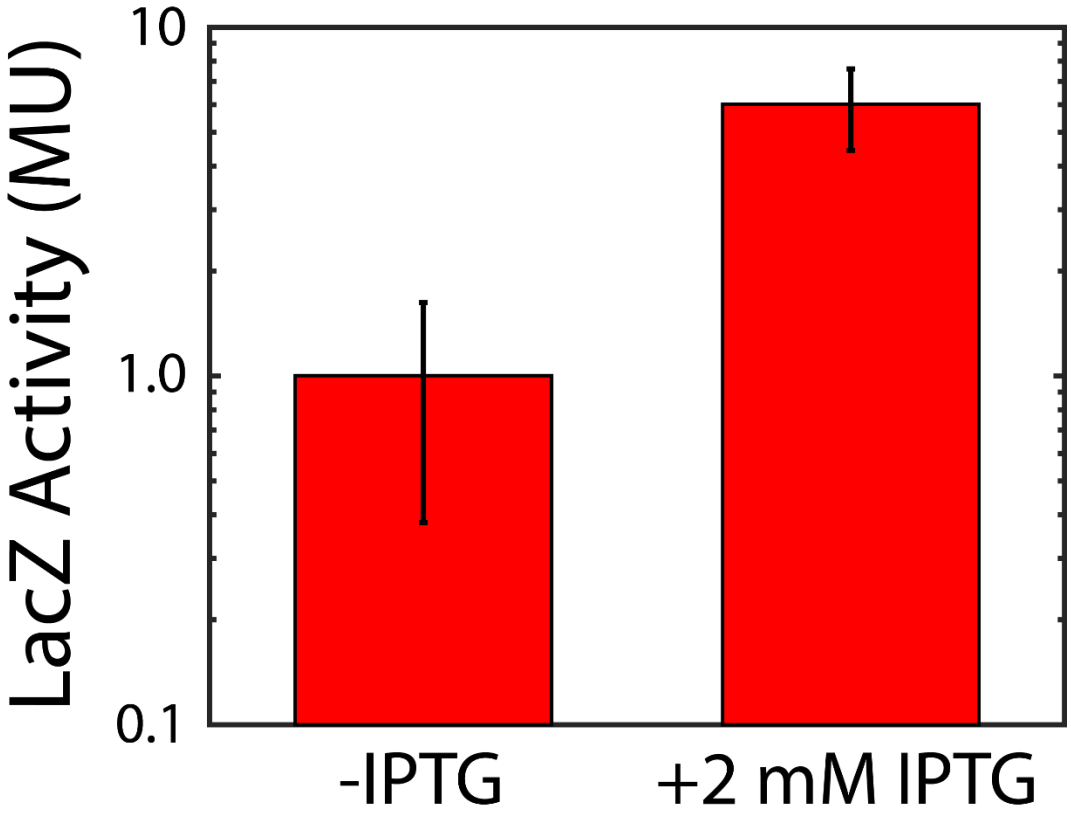

Figure S1. Expression from the hyper-spank promoter of pHCMC05 in *Bacillus subtilis.* LacZ activity of uninduced (left) and induced (right) *B. subtilis* 168 transformed with pHCMC05-*lacZYAX* was measured with a Miller assay (24). Bars are the mean of six independent replicates and error bars are the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

Figure S2

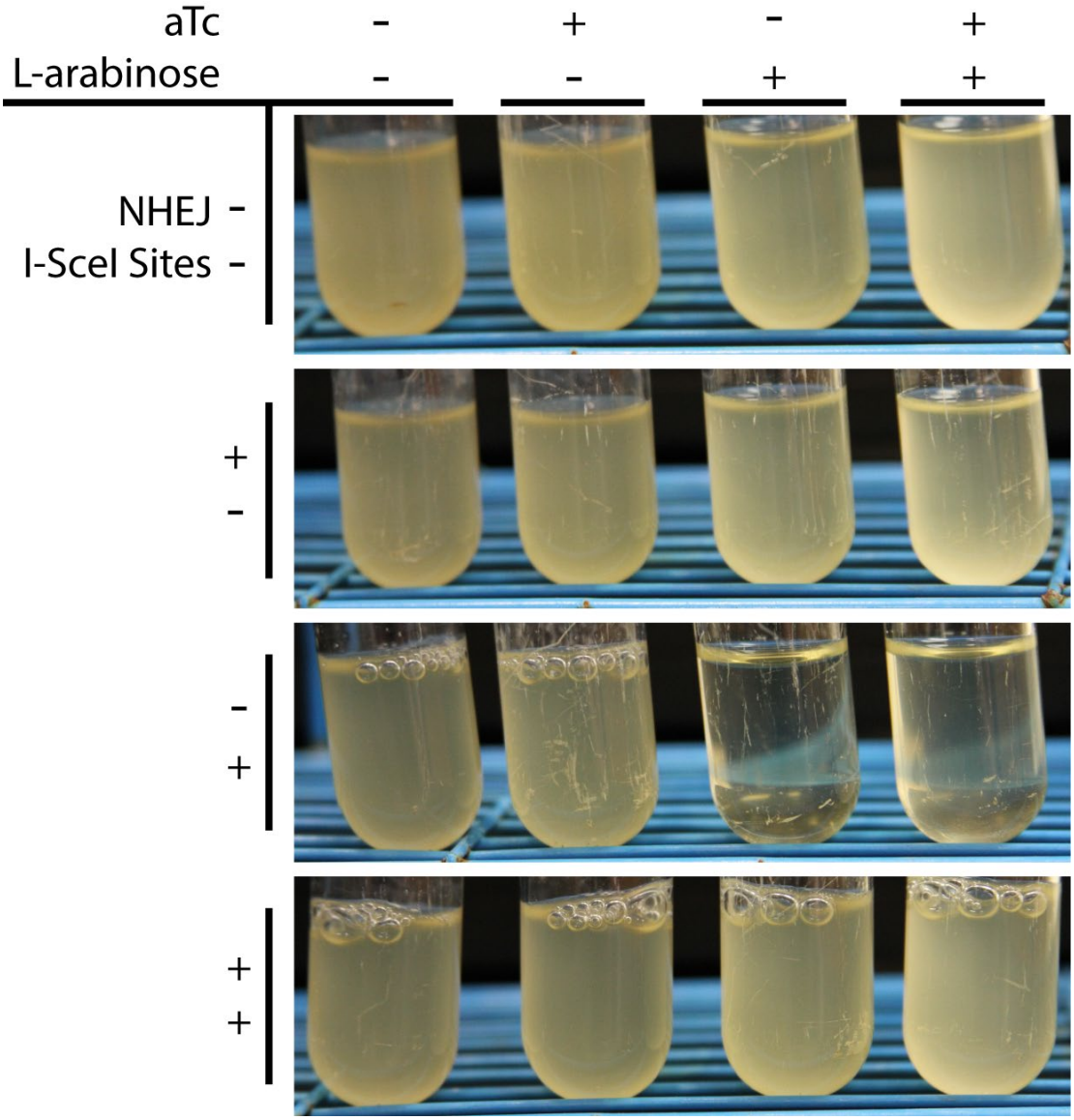

Figure S2. *B. subtilis* NHEJ enzymes function in *E. coli.* Turbidity of cultures grown for ~36 hours at 30 ºC after inoculation with identical amounts of cells. Bacterial strains are MG1655 Δ*lac* carrying the plasmid pTKRED, which expresses the homing endonuclease I-SceI when induced with L-arabinose. 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

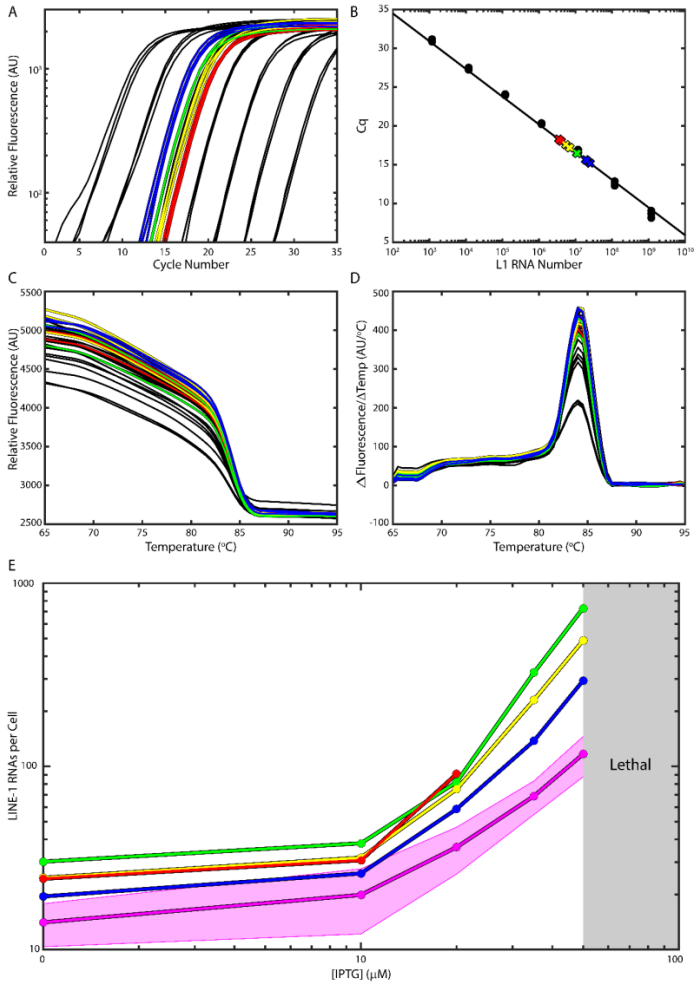

Figure S3. Quantitative RT-PCR to determine T7*lac* promoter response function. (A) Amplification curves of reverse transcribed serial 10x dilutions of *in vitro* transcribed TL1H RNA as an absolute standard (black), along with reverse transcribed RNA extracted from BL21(DE3) pTKIP-TL1H grown in M63 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

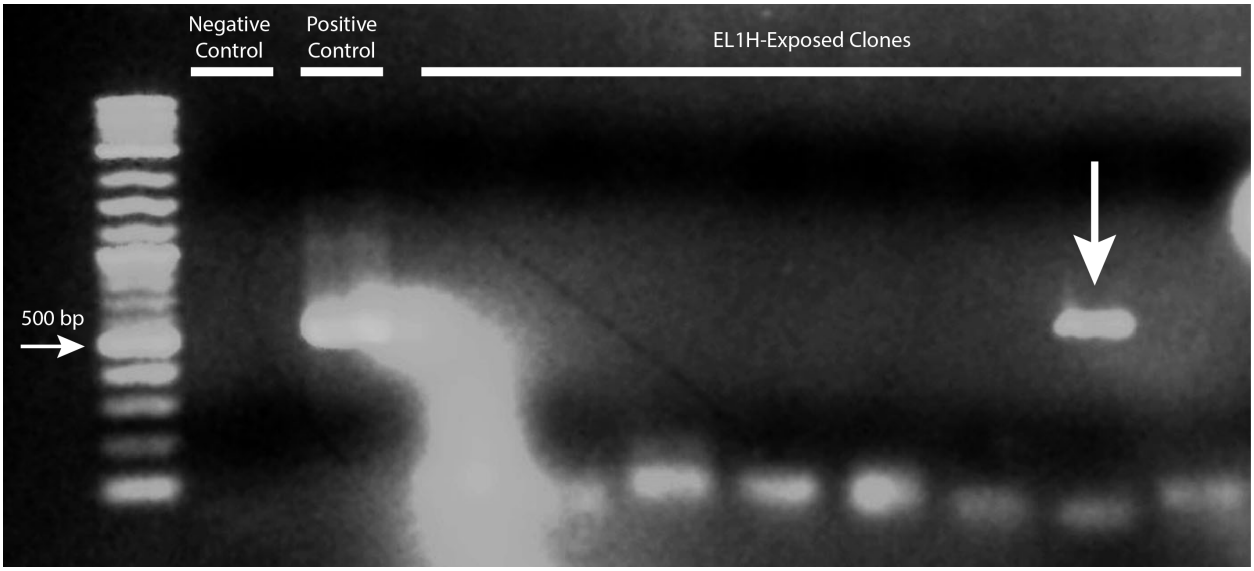

Figure S4. Detection of Full-Length EL1H Genomic Integrants. Representative 2% agarose gel electrophoresis of colony PCR of eight isolated colonies of BL21(DE3) that had been exposed to EL1H and cured of pTKIP-EL1H using primers that anneal to the 5' end of EL1H and produce a 500 bp amplicon. BL21(DE3) 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).