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

Permalink
https://escholarship.org/uc/item/65v2d4wg

Journal
Proceedings of the National Academy of Sciences of the United States of America, 115(49)

ISSN
1091-6490

Authors
Lee, Gloria
Sherer, Nicholas A
Kim, Neil H
et al.

Publication Date
2018-12-01

DOI
10.1073/pnas.1807709115

Peer reviewed
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 a,b, Ema Rajic a,b, Davneet Kaur a,b, Niko Urriola a, K. Michael Martinii b,d,e, Chi Xue a,b,d,e, Nigel Goldenfeld a,b,d,e,1, and Thomas E. Kuhlman a,b,d,f,1

*Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 1Center for the Physics of Living Cells, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 2University of Illinois Laboratory High School, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 3Institute for Universal Biology–NASA Astrobiology Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 4Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 5Center for Biophysics and Quantitative Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and 6Department 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 splicings of a single primary transcript, allowing access to additional

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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To 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 10^6 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 T7lac promoter at the 5′ end and a strong ribosomal binding site (RBS) to drive ORF1p expression (Fig. 1A, 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-thiogalactopyranoside (IPTG). We also synthesized de novo a version of L1H optimized for bacterial expression, EL1H (Fig. 2A). 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 T7lac promoter as employed for L1 expression (Fig. 1B). 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 hyper-spark promoter (28).

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. 1B 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. 1D). Conversely, we obtain high-yield transformation of EL1H into B. subtilis strains with
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-lacZAYX 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_{LtetO} 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 ELH without addition of IPTG is lethal to *E. coli* with concomitant induction of NHEJ enzymes with 100 ng/mL aTc (Fig. 1E).

To quantify the effect of L1 and L1LtrB 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. 1B and C for L1 and Fig. 2B and C for L1LtrB). 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. 1C).

We measured the transcriptional response function of the T7lac 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. S3E). The resulting dose-responses as a function of L1 RNAs and L1LtrB RNAs per cell are shown in Fig. 3. In Fig. 3A, data from TL1H are plotted as blue points, EL1H as red points, and EL1H+NHEJ as black points. In Fig. 3B, data from L1 LtrB are plotted as red points, and L1LtrB+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 L1LtrB 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
growth rate of $-0.83 \pm 0.06\%$ (TL1H) or $1.9 \pm 0.6\%$ (EL1H) in the absence of NHEJ, and $\geq 45 \pm 1.6\%$ with NHEJ. Each LLlTrB transcript reduces the growth rate by $0.11 \pm 0.02\%$ in the absence of NHEJ and $0.82 \pm 0.11\%$ with NHEJ. As might be expected because of the ability of LtrA maturase to excise LLlTrB from interrupted genes, the growth defect resulting from LLlTrB is weaker than that from L1.

The LLlTrB growth defect is also evident in plating assays to determine retrotransposon efficiency. Induction of LLlTrB expression with $100 \mu M$ IPTG reduces the number of viable colony forming units (cfus) per milliliter on LB by $\sim 10\times$. Simultaneous induction of LLlTrB with $100 \mu M$ IPTG and induction of NHEJ enzymes with $100 \text{ng/mL}$ anhydrotetracycline reduces viable cfus/OD/mL by $\sim 100\times$, whereas induction of expression of NHEJ enzymes alone has no detectable effect.

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

**L1 and LLlTrB Successfully Integrate in E. coli Chromosome.** Several lines of evidence demonstrate that both LLlTrB 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. 4A) 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. 4B) to report EL1H integration via fluorescence. EL1HID contains an mtTFP1 gene expressed from a strong promoter whose $\sim 10$ and $\sim 35$ sequences are separated by the group I intron tda1–3 (34). After transcription, tda1–3 catalyzes its own excision from the transcribed mRNA, which reconstitutes the mtTFP1 promoter, and allows expression of a fluorescent protein on successful retrotransposition. When EL1HID was transformed into E. coli and weakly induced, $\sim 1\%$ of cells exhibited a total fluorescence $\times 10^2$ brighter than any cells from control strains. With simultaneous weak induction of NHEJ enzymes, the fluorescent population increased to $\sim 80\%$ (Fig. 4C–F).

Using a similar RIG in LLlTrB (11), we found that NHEJ also enhances the rate of LLlTrB ectopic retrotransposition. The RIG is composed of a kanamycin resistance gene, the sequence of which is interrupted by tda1–3 (Fig. 24). After growing cultures of E. coli expressing LLlTrB 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 LLlTrB 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 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. 5B shows that there exists a small set of parameter values (low growth defect, b, of less than 0.01 and high integration rate, $\mu$, 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
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 assexual 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, 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 L1.LtrB ($10^{-3} \leq b \leq 10^{-2}$, $10^{-3} \leq \mu \leq 10^{-6}$ retrotransposon$^{-1}$cell$^{-1}$generation$^{-1}$) can persist in low copy numbers (~1 per cell) for millions to tens of millions of generations.
LacZ Measurements. B. subtilis 168 pHCMC05-1azYAX 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. 5I) (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.

LlLTR 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.I.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.