Poxviruses capture host genes by LINE-1 retrotransposition

S. M. Fixsen[1], K.R. Cone[1], S.A. Goldstein[1], T.A. Sasani[1], A.R. Quinlan[1], S. Rothenburg [2], & N.C. Elde[1,3]*

[1] Department of Human Genetics, University of Utah, Salt Lake City, USA
[2] Department of Medical Microbiology and Immunology, University of California, Davis, USA
[3] Howard Hughes Medical Institute, 4000 Jones Bridge Rd, Chevy Chase, MD 20815, USA

*Correspondence to: nelde@genetics.utah.edu

Abstract

Horizontal gene transfer (HGT) provides a major source of genetic variation. Many viruses, including poxviruses, encode genes with crucial functions directly gained by gene transfer from hosts. The mechanism of transfer to poxvirus genomes is unknown. Using genome analysis and experimental screens of infected cells, we discovered a central role for Long Interspersed Nuclear Element-1 (LINE-1) retrotransposition in HGT to virus genomes. The process recapitulates processed pseudogene generation, but with host messenger RNA directed into virus genomes. Intriguingly, hallmark features of retrotransposition appear to favor virus adaption through rapid duplication of captured host genes on arrival. Our study reveals a previously unrecognized conduit of genetic traffic with fundamental implications for the evolution of many virus classes and their hosts.

Summary

Active selfish genetic elements in infected cells aid virus adaptation by catalyzing transfer of host genes to virus genomes.

Introduction

The movement of genetic material from one organism to another by horizontal gene transfer (HGT) bypasses vertical inheritance from parent to offspring (Keeling and Palmer, 2008). HGT can shortcut more gradual mutational processes by providing the opportunity for adaptive leaps through single mutational events where genes can be transferred between genomes across kingdoms of life (Keeling and Palmer, 2008). HGT plays a central role in the diversification of many infectious microbes, which contend with complex environments and host immune defenses to persist and replicate. Horizontal gene transfer is particularly well-described in bacteria, where it can contribute to the emerging crisis of multi-antibiotic resistant pathogens (Wijayawardena *et al*., 2013).

Many classes of viruses acquire genes from the hosts they infect by horizontal transfer (Caprari *et al*., 2015). A conspicuous example is Rous Sarcoma virus, in which recognition of the horizontal transfer of the host c-*Src* gene into the virus genome led to the discovery of oncogenes (Swanstrom et al., 1983). Other co-opted host genes diverge to aid the virus in inhibiting host defenses (Kawagishi-Kobayashi *et al*., 1997, Elde and Malik, 2009). It has become increasingly clear that acquisition of host genes is a common mechanism by which viruses of many classes gain adaptive advantages to propagate (Dolja and Koonin, 2018). Despite the prevalence of HGT in diverse viruses and its importance in virus evolution, the mechanisms by which genes are transferred from host to virus genomes are not well understood.

To investigate mechanisms of transfer we examined poxviruses, which encode a plethora of genes gained by horizontal transfer. Based on phylogenetic analysis, more than 25% of poxvirus genes were acquired by HGT from hosts (Hughes and Friedman, 2005), although similarities not evident in sequence comparisons but revealed by shared protein structures suggest that a higher proportion of genes were acquired by horizontal transfer (Bahar *et al*., 2011). Many captured genes adapt to act as inhibitors of host immune defenses to favor virus replication (Elde and Malik, 2009). Some act as host range factors because their deletion leads to restriction of infectible cell types or species, implying that new acquisitions may aid in host jumps *(*Bratke *et al*., 2013; Haller *et al*., 2014). While the most notorious poxvirus, variola virus which caused smallpox, has been eradicated, other extant poxviruses are capable of infecting humans and pose

the risk of new pandemics, as appears to be the case for monkeypox (Sklenovska and Van Ranst, 2018). Poxviruses also serve as useful model systems for other viruses with large double-stranded (ds) DNA genomes (Deeg *et al*., 2018).

Because poxvirus-encoded copies of host genes lack introns, it was hypothesized that transfer occurs through a spliced mRNA intermediate (Lefkowitz *et al.*, 2006). For integration, mRNA templates need to be reverse transcribed into virus genomes by either a retrovirus or a retrotransposon. The discovery of a provirus encoded in the genome of fowlpox virus (Hertig *et al.*, 1997) and a Short Interspersed Nuclear Element (SINE) retrotransposon found in the taterapox virus genome (Piskurek and Okada, 2007) indicate that either is capable of catalyzing insertions in poxvirus genomes.

Given differences in biochemical activity, it is possible to detect patterns distinguishing mechanisms of retroviral versus retrotransposon-based transfer of host genes into virus genomes. LINE-1 retrotransposons are known to reverse transcribe and integrate host genes in a sequence independent manner, resulting in collections of processed pseudogenes in diverse host genomes (Esnault *et al.*, 2000). LINE-1 transcripts encode two proteins: an RNA binding protein (ORF1) and an endonuclease (EN)/reverse transcriptase (RT) protein (ORF2). ORF2 preferentially nicks one strand of DNA at a TTTT/AA site in the host genome, and reverse transcription is primed from the exposed 3'hydroxl. LINE-1 elements can also integrate at sites of dsDNA breaks in an endonuclease-independent manner (Morrish et al., 2002). LINE-1 RT shows *cis*-preference, meaning it is most likely to transcribe the RNA from which it was translated, but SINE elements like the one found in the taterapox genome can hijack LINE-1 machinery to integrate following replication.

Sometimes LINE-1 enzymes promiscuously reverse transcribe host mRNA, which might facilitate transfers of host genes to virus genomes. It is plausible that the poly(A) tail of either the host mRNA or a poly(A) tract in SINE transcripts initializes LINE-1 reverse transcription, in addition to faithful recognition of LINE-1 transcripts. Short target site duplications (TSDs) of sequences ~8-20bp flank LINE-1 insertions of either LINE-1 or host sequences, likely a product of repairing staggered breaks during integration. The resulting LINE-1 mediated flanking

duplications are copies of host sequences as opposed to specific retrovirus sequences. Additional evidence of LINE-1 action comes from comparisons with closely related species lacking the insertion and instead containing what is termed an "empty site." Together these features can help distinguish whether LINE-1 is involved in the transfer of host genes to virus genomes.

However, given that viruses rapidly mutate, unique signatures of retrovirus or LINE-1 mediated horizontal transfer may quickly degrade and obscure mechanisms of transfer. Therefore, we devised an experimental system to screen for the horizontal transfer of a gene from host to virus genomes using an artificial selection scheme. Combined with observations of viruses encoding recently acquired and highly conserved genes, we reveal LINE-1 elements as molecular accomplices in the transfer of host genes to virus genomes.

Results

We hypothesized that examination of genes recently acquired by HGT might provide clues for revealing mechanisms of gene transfer. Virus homologs of the host Golgi Anti-Apoptotic Protein (GAAP; also called transmembrane Bax inhibitor motif protein family 4 or TMBIM4), found in several orthopoxviruses, are ~75% identical to human GAAP (Gubser *et al*., 2007; Saraiva *et al*., 2013), suggesting a recent acquisition and/or highly conserved function favoring virus replication. Strikingly, we observed nearly identical 21 base pair sequences flanking the open reading frame of virus GAAP (vGAAP), indicative of a TSD, consistent with gene capture resulting from LINE-1 mediated retrotransposition of host mRNA (Fig. 1A and Fig. 1 – supplemental figure 1). The vGAAP gene is situated near one of two identical inverted terminal repeat (ITR) regions of the virus genome, and on opposite ends of the genome in various viruses, suggesting that it was originally transferred into the ITR. A single copy of the putative 21bp TSD at an analogous position near the other ITR in the cowpox genome suggests a pseudo empty site (Fig. 1A). Together with a published account of a non-autonomous SINE in the genome of taterapox virus (Piskurek and Okada, 2007), these observations suggest that host genes can be retrotransposed into poxvirus genomes by LINE-1 activity in a process akin to the generation of processed pseudogenes.

Phylogenetic analysis revealed a single origin for the horizontal transfer of GAAP into poxvirus genomes. In addition, at least two other classes of DNA viruses, herpesviruses and iridoviruses, independently acquired GAAP by horizontal transfer (Fig. 1B) (Bellehumeur *et al*., 2016; de Groof *et al*., 2015). Another study of Squirrel Monkey cytomegalovirus revealed that the S1 gene encodes a protein more than 95% identical to squirrel monkey Signaling lymphocytic activation molecule family 6 (SLAMF6), followed by portions of the 3' untranslated region (UTR) of the host mRNA (Perez-Carmona *et al.*, 2015). Together these findings suggest that many DNA viruses acquire host genes by retrotransposition. However, the majority of host to virus gene transfers lack signatures of retrotransposition, which can quickly degrade through mutation, motivating our development of an experimental strategy for capturing HGT events in real time.

We developed a selection scheme using vaccinia virus (Copenhagen strain), which encodes nearly 200 genes in a 191,737 bp genome (See methods). We focused on K3L, a gene that was presumably acquired from a host by an ancient poxvirus based on sequence identity with human eukaryotic initiation factor 2α (eIF2α; 27.6% identical) (Beattie *et al*., 1991; Elde *et al*., 2009; Elde and Malik, 2009). K3L diverged from eIF2α to encode a protein that blocks the host defense Protein Kinase R (PKR) pathway. PKR is activated by dsRNA produced by viruses in the cytoplasm during transcription of virus genes and virus replication. Upon activation, PKR phosphorylates eIF2α, leading to a block in protein translation and virus replication. K3L acts as a pseudo-substrate of PKR to competitively inhibit PKR from binding eIF2α (Fig. 2 – supplemental figure 1) (Beattie *et al*., 1991; Kawagishi-Kobayashi *et al*., 1997), and can be essential for productive infections depending on the host (Langland and Jacobs, 2002; Park *et al.*, 2019; Park *et al.*, 2021).

Using a mutant virus deleted for K3L and another PKR inhibitor, E3L (Brennan *et al*., 2014), we infected a rabbit kidney cell line (RK13) expressing chromosome-integrated copies of K3L fused to the gene encoding mCherry fluorescent protein (RK13-K3L) (Fig. 2A and see Methods for details). The host copy of mCherry-K3L complements the mutant virus lacking K3L, allowing it to replicate. Viruses grown in RK13-K3L cells were transferred to a non-complementing baby hamster kidney (BHK) cell line. Viruses lacking K3L replicate poorly in BHK cells, providing

strong selection for viruses that "reclaimed" the mCherry-K3L gene by horizontal transfer and/or somehow adapted in the absence of K3L to block host PKR and regain productive replication.

Using this scheme, we screened ~500 million viruses over many experiments and recovered ten virus isolates from mCherry-positive infected cell clones. In the restrictive BHK cells, viruses encoding K3L replicate ~100-fold better than the parent virus deleted for K3L (Fig. 2B), allowing us to identify viruses with improved replication compared to the larger population of K3L-deficient viruses. After plaque purification of candidate mCherry positive isolates, we infected BHK cells with each recovered virus isolate and observed titers on par with viruses encoding K3L (Fig. 2C). By sequencing the genome of each recovered virus isolate, we found that the mCherry-K3L fusion gene was integrated into the virus genome at a unique position in every isolate (Figs. 2D and 3A), providing the opportunity to compare each independent integration event.

All integrations were consistent with a mechanism of LINE-1 retrotransposition-mediated HGT. The sequence of each transferred gene revealed hallmark features of mRNA gene capture, including 5' and 3' UTRs, precise intron splicing, and poly(A) tails (Fig. 3B). For seven of ten isolates, TSDs flanked mCherry-K3L insertions and contained cut site sequences matching the consensus (TTTT/AA) of LINE-1 endonuclease (Fig. 3 and Fig. 3 – supplemental figure 1). Although there were no TSDs flanking mCherry-K3L in isolates 5, 9, and 10, these integrations might result from a endonuclease-independent LINE-1 mechanism at DNA breaks (Morrish *et al.*, 2002). Because our experiments lack any exogenous reverse transcriptase, we tested for endogenous activity in the RK13 cells and detected robust LINE-1 activity (Fig. 3 – supplemental figure 2). These results suggest that LINE-1 mediated retrotransposition may be a common mechanism of HGT into poxvirus genomes. While we do not exclude the possibility that genes can also be transposed into the viral genome by retroviruses, this would require co-infection. LINE-1 mediated retrotransposition may be more likely simply because LINE-1 enzymes are nearly ubiquitous in mammals.

In general, poxvirus genomes are arranged with the essential genes, shared by all poxviruses, in the central region of the genome. Clade- or species-specific genes, which more commonly

exhibit evidence of HGT, are enriched near the ends of the genome (Lefkowitz *et al*., 2006; McLysaght *et al*., 2003). This pattern may reflect that gene transfer into the central genomic region genome is more likely to interrupt genes vital for virus replication (Lefkowitz *et al.*, 2006). Consistent with this, all but one of the 10 isolates we recovered integrated K3L outside a roughly defined central region of the genome containing 90 core chordopoxvirus genes (Lefkowitz *et al.*, 2006) (Fig. 2D). In the case of isolate 5, the K3L integration involved a duplication of two essential genes, G9R and L1R (Fig. 3B). In a contemporaneous study (Rahman *et. al.*, 2021), virus genomes that acquired a host gene by horizontal transfer within a duplicated gene also propagated. In both cases gene functions were preserved by duplication. Spurious gene duplications can appear frequently in poxvirus genomes (Hughes and Friedman, 2005), pointing to a flexible means of poxvirus adaptation involving horizontal transfer even into essential genes.

Previous studies found that duplication of poxvirus genes, followed by rapid gene copy amplification, can boost virus replication (Brennan *et al.*, 2014; Elde *et al.*, 2012). In our analysis, we discovered two isolates where K3L had undergone polymorphic copy number increases (isolates 4 and 5; Fig. 4A). Homologous recombination, including unequal crossovers leading to duplications, requires as little as 16bp of homology in poxviruses (Yao and Evans, 2001), suggesting that LINE-1 induced TSDs might facilitate gene amplification. Genes transferred to within the ITR are duplicated on the other end by replication-based mechanisms (McFadden and Dales, 1979), as we observed in isolates 4 and 9 (Fig. 2C). Genes integrated into the near the termini of the genome by HGT, like isolate 4, may be especially prone to duplication due to stretches of short, high copy tandem repeats found in and near the inverted terminal repeat regions (Fig. 4A and Fig. 5 – supplemental figure 1A). These observations are consistent with the idea that horizontal gene transfer occurs evenly across poxvirus genomes, but with an enrichment of transferred genes persisting near the ends, where insults to essential genes are less probable and duplication-based adaptation is favored (Fig. 5 – supplemental figure 1). Along these lines, gene families found only in specific lineages of poxviruses, indicating more recent acquisition, are more likely to be present in multiple copies than single copy host-acquired genes found among diverse poxviruses (Hughes and Friedman, 2005).

While our experimental system revealed LINE-1 driven mechanisms of HGT, the selection scheme involved viruses regaining K3L, an existing virus gene as opposed to gaining a *bona fide* host gene. To better model virus adaptation following acquisition of a host gene, we engineered viruses with the human eIF2α gene in place of K3L (Fig. 4 – supplemental figure 1; see Methods). This recombinant virus better approximates the original ancient horizontal gene transfers of the eIF2α gene that then evolved to become K3L in diverse clades of modern poxviruses (Kawagishi-kobayashi, *et al.*, 2000). Because the viruses we engineered to encode eIF2α in place of K3L also lack E3L, which encodes another inhibitor of the host PKR antiviral pathway, experiments with these viruses focus strong selective pressure on eIF2α to adapt to inhibit PKR.

We performed four serial infections of viruses expressing eIF2α or an eIF2α variant lacking the PKR phosphorylation site (S51A) in RK13 cells. While the recombinant viruses replicate weakly, we observed a nearly 10-fold increase in virus replication for both eIF2α-encoding viruses after passaging (Fig. 4B). By sequencing the eIF2α region in evolved virus isolates, we found independent gene copy increases of eIF2α in both viruses (Fig. 4C). These results support a critical role for gene copy increases following HGT in poxvirus evolution, as also observed for K3L and other genes (Brennan *et al.*, 2014; Elde *et al.*, 2012) and suggest a common pathway facilitating adaptation of host genes captured by poxviruses.

Discussion

Horizontal gene transfer is an important process facilitating virus evolution (Caprari *et al.*, 2015; Deeg et al., 2018; Hughes and Friedman, 2005; Koonin and Yutin, 2019; McLysaght *et al.*, 2003). Our discovery of LINE-1 retrotransposition-mediated HGT from host to poxvirus genomes has several notable implications. In addition to self-propagation in host genomes, retrotransposition to virus genomes could greatly extend the range of LINE-1 or related elements that subsequently mobilize to a newly infected virus host (Piskurek and Okada, 2007). The extensive host range of some poxviruses, for example between rodent and primate species, suggests the possibility of cross host lineage transfer of LINE-1s, non-autonomous SINEs, and host transcripts between mammals. A process involving the frequent appearance of DNA transposons in some insect viruses suggests an analogous mechanism for the spread of

transposable elements between insect species (Gilbert et al., 2016). In cases of retrotransposition of host transcripts, the resulting processed pseudogene products are enriched for highly expressed genes (Zhang et al., 2004). The over representation of certain transcripts might provide clues for understanding and predicting how viruses might adapt with repurposed genes mimicking host functions (Elde and Malik, 2009). Currently recognized cases of HGT involving oncogenes, regulators of translation, like eIF2α, and antiviral genes that are highly expressed in response to infection. These examples of transcripts exhibit high gene expression and encode cellular functions easily repurposed to promote virus replication.

Repetitive target site duplications generated by LINE-1 activity may also enhance poxvirus adaptation by facilitating rapid gene copy number amplification of horizontally transferred genes on arrival. Increases in gene copy number augment the probability of beneficial mutations appearing, allowing for swift diversification of newly acquired genes (Elde *et al.*, 2012). Thus, form follows function as the repetitive termini of poxvirus genomes further enable the emergence and persistence of acquired genes predominately at the ends of the genome (Fig. 5). As beneficial virus gene variants endure, they end up closer to the center of the genome as subsequently captured genes appear near the ends.

Similar mechanisms of virus adaptation by HGT may be shared by other diverse classes of nucleocytoplasmic DNA viruses (NCLDVs), in addition to poxviruses (Deeg *et al.*, 2018; Koonin and Yutin, 2019). Many giant viruses, which can exceed a megabase in genome size and infect diverse host species, encode a variety of genes acquired by HGT, including ribosomal genes. Given a primary role for LINE-1 elements in facilitating gene transfer, we speculate that other classes of transposable elements may spur NCDLV evolution by mobilizing host genes into virus genomes across diverse ecosystems.

Materials and methods

Phylogenetic analysis and sequence comparisons of vGAAP and host GAAP/TMBIM4

TMBIM sequences for various species were downloaded from NCBI (Supplemental data) and aligned using the COBALT Multiple Alignment Tool. Amino acid Phylip alignments (see

supplemental materials) were analyzed by PhyML (http://www.atgc-montpellier.fr/phyml/) for tree building with 100 bootstraps (Guindon et al., 2010).

Cell culture

BHK (ATCC) and HeLa (gift from Adam Geballe, Fred Hutchinson Cancer Research Center) cells were cultured in Dulbecco minimum essential medium (DMEM; HyClone) supplemented with 10% FBS, 2 mM L-glutamine and 100 µg/mL of pen/strep. RK13 cells were cultured in MEM-alpha supplemented with 10% fetal bovine serum (FBS; Gibco), 2 mM L-glutamine (HyClone) and 100 µg/mL of penicillin-streptomycin (pen/strep; HyClone).

RK13 cells expressing SLP-mCherry-K3L (RK13-K3L) were generated by transfecting $3x10^5$ RK13-E3K3 cells (Brennan *et al.*, 2014) with 2.5ug piggybac vector (see supplemental sequences) and 0.5 ug transposase vector (gift from Ed Grow, University of Utah) with 9uL Fugene-HD. Cells with the piggybac construct integrated were selected by the presence of 1-2 ug/mL puromycin, and populations were monitored for mCherry expression. After two weeks of selection, remaining polyclonal cells were propagated in MEM-alpha media supplemented with 10% FBS, 2 mM L-glutamine and 100 µg/mL of pen/strep. All cultures were maintained, and infections performed at 37°C in a humidified 5% CO_2 incubator.

Virus strains

VC-2 refers to the Vaccinia virus-Copenhagen strain (Goebel et al., 1990) and VCR2 is a genetically modified isolate of VC-2 deleted for E3L and K3L (Brennan *et al.*, 2014). To generate VCR2+mCherry-K3L, we cloned pBlue_165_mCherry-K3L (see supplemental sequences) into the MCS of pBluescriptIIKS (-) (Addgene). This plasmid was transfected into BHK cells, which were then infected with VCR2 virus. Recombinant viruses were plaque purified in BHK cells and checked for purity by PCR amplifying across the mCherry-K3L insertion site. The amplicon was gel purified and sequenced.

To make eIF2α viruses, sequences flanking K3L from VACV were amplified from VC-2 viral DNA: 680bp of 5' homologous sequence was amplified with the primers K2Lflank_F (5'-CTTCTTATC GATTTTTTATACCGAACATAAAAATAAGGTTAATTA) and K2Lflank_R (5'-CTTCTTCATATGG TGATTGTATTTCCTTGCAATTTAG), and 1024bp of 3'

306 homologous sequence (including the native K3L promoter) was amplified with the primers
307 K4Lflank_F (5'-CGTCGTGCGGCCGCCTTGTTAACGGGCTCGTAAATT) and K4Lflank_R
308 (5'-CGA GCGGAGCTCGTACGATACATAGATATTACAAATATCCTAG). A VACV
309 synthetic early/late promoter (SLP)(Chakrabarti et al., 1997) was created by annealing primers
310 SLP_F (5' TCGACAATTGGATCAGCTTTTTTTTTTTTTTTTTTGGCATATAAATA
311 AGAAGCTTCCCGGGTCTAGAC) and SLP_R (5'-AGCTCAGATCTGGGCCCTTCG
312 AAGAATAAATATACGGTTTTTTTTTTTTTTTTTTCGACTAGGTTAAC). EGFP was
313 amplified from pN1-EGFP (Addgene) using primers EGFP_F (5'-
314 GGAGGACTCGAGATGGTGAGCAAGGGCGA) and EGFP_R (5'-GGAGGTATCGA
315 TTTACTTGTACAGCTCGTCCATGC). Each PCR product was digested with restriction
316 enzymes (New England Biolabs), gel purified (Zymo Research), and sequentially cloned into
317 pB.2 as follows: 5' flank with ClaI and NdeI, 3' flank with NotI and SacI, SLP with SalI and
318 XhoI, and EGFP with XhoI and ClaI. The resulting plasmid contained EGFP following SLP,
319 between the two K3L flanking sequences (pB.2-EGFP). The K3L open reading frame was
320 amplified from VC-2 viral DNA using primers K3L_F (5'-GTTGTAG
321 GATCCATGCTTGCATTTTGTTATTCGTTGC) and K3L_R (5'-GTTCTTGTCGACTT
322 ATTGATGTCTACACATCCTTTTG). The eIF2α gene was amplified from human cDNA using
323 primers eIF2α_F (5'-GATGTAGGATCCATGCCGGGTCTAAGTTGTAGAT) and eIF2α_R
324 (5'-CTACTTGTCGACTTAATCTTCAGCTTTGGCTTCCAT). The resulting PCR products
325 were cloned into pB.2-EGFP with BamHI and SalI, placing K3L or eIF2α immediately
326 following the native K3L promoter, and upstream of SLP-EGFP to create pB.2-K3L and pB.2-
327 eIF2α, respectively. Site-directed mutagenesis was performed on pB.2-eIF2α or pB.2-eIF2α-ΔC
328 using primers eIF2α_S51A_F (5'-GATAC
329 GCCTTCTGGCTAATTCACTAAGAAGAATCATGCCTTC) and eIF2α_S51A_R (5'-
330 GAAGGCATGATTCTTCTTAGTGAATTAGCCAGAAGGCGTATC) to generate pB.2-eIF2α-
331 S51A.
332
333 Recombinant eIF2α-encoding viruses were constructed by replacing the K3L gene using
334 homologous recombination. RK13-E3K3 cells were infected with VCR2 (MOI = 1.0) and
335 transfected at 1-hour post-infection with pB.2-EGFP, pB.2-K3L, pB.2-eIF2α, or pB.2-eIF2α-
336 S51A plasmids by use of FuGENE6 (Promega) according to the manufacturer's protocol.

Infected cells were collected at 48 hours post infection, and viruses were released by one freeze-thaw cycle followed by sonication. Resulting viruses were plaque purified in RK13-E3K3 cells four times, selecting for recombinants expressing EGFP. Final virus clones were verified by PCR and sequencing of viral DNA across the K2L-K4L region of the genome.

Experimental screen for horizontal gene transfer events

Confluent 150mm dishes of RK13-K3L cells were infected with VCR2 at a MOI of 0.1. Cell-associated virus was collected after 48 hours of infection, and released from the cell by one freeze-thaw cycle followed by sonication as described (Isaacs, 2004). Confluent 150mm dishes of BHK cells were then infected with ~10^6 PFUs of this virus stock for 7 days, during which they were monitored for mCherry expression. 500 plates were screened, which accounts for an estimated 500 million viruses. For mCherry positive clones observed by fluorescence microscopy, cell-associated virus was collected and released by one freeze-thaw cycle followed by sonication. mCherry-expressing virus was then plaque purified in BHK cells as described (Isaacs, 2004).

Analysis of horizontal gene transfer events by inverse PCR (iPCR)

Virus DNA was extracted from mCherry-expressing clones as previously described from infected BHK cells (Esposito et al., 1981). Purified DNA was digested with XbaI and SpeI or BglII (New England Biolabs (NEB)), diluted, and ligated (Quick ligase; NEB) to circularize linear fragments. Primers pointing away from each other in both mCherry (F-CGTGGAACAGTACGAACGCG and R- CCATGTTATCCTCCTCGCCC) and K3L (F-GAGCATAATCCTTCTCGTATACTC and R- GAATATAGGGATAAACTGGTAGGG) were used to amplify regions flanking mCherry-K3L insertions. Resulting PCR bands were gel purified and Sanger sequenced with the iPCR primers. From this sequencing, the general location of the mCherry-K3L insert could be inferred. Primers flanking this region were then used to amplify the putative insertion, along with flanking sequences. These amplicons were then gel purified, Sanger sequenced, and compared to the parent (VCR2) genome to characterize gene integrations. Because isolates 7 and 9 incorporated mCherry-K3L within the repetitive ITR, each end was PCR amplified separately using unique flanking primers and internal primers (mCherry-R and K3L-F, above). However, only the 5' end of the insertion of isolate 9 was amplified,

despite multiple attempts. Thus, we cannot be sure whether this isolate includes a TSD or was cut at a LINE-1 endonuclease site.

Long read genome sequencing of virus isolates

Virus particles from plaques expressing mCherry-K3L were isolated from BHK cells and virus cores were purified by ultracentrifugation through a 36% sucrose cushion at 60k rcf for 1 hr. Virus DNA was extracted as previously described (Esposito *et al.*, 1981). The SQK-LSK108 library kit (Oxford Nanopore Technologies) was used to prep isolate 1 DNA, which was sequenced on a FLO-MIN106 cell. A SQK-RBK001 kit was used to prep DNA from isolates 2-5, which were multiplexed on a FLO-MIN107 cell. Isolates 6-10 were also multiplexed using the SQK-RBK004 library prep kit, and sequenced on FLO-MIN107. Reads were base-called with the Oxford Nanopore Albacore program and aligned to a reference genome that included the VCR2 genome on one contig and the cellularly-expressed mCherry-K3L construct on a separate contig using default NGMLR parameters (Sedlazeck et al., 2018). Integrative Genomics Viewer (IGV) software (Broad Institute) was used to visualize insertions (Robinson et al., 2011). Sequencing data can be found at the NCBI SRA, accession number: PRJNA614958.

Measuring titers of virus strains and isolates

We plated $5x10^6$ cells in 100mm dishes and infected them 16 hours later at an MOI of 0.1. Three plates (biological replicates) were infected with each isolate or strain. After 48 hours, media was aspirated, and cell-associated virus stock was collected by one freeze-thaw and sonication cycle. Then, 6-well plates were seeded with $5x10^5$ BHK cells per well, and, 16 hours later, infected with 10-fold serial dilutions of virus stock (2 wells per dilution) in 200uL media. After 2 hours at 37ºC, 2mL of media was added to each well. Cells were fixed and stained with 20% Methanol + 0.2% crystal violet 48 hours post infection. Wells with 10-100 plaques were counted and averaged to calculate virus titer. Each of the biological replicates is shown in figures 2 and 4, along with an average of the three and standard deviation.

Alu assay of RK13 cells

$5x10^5$ HA-HeLa or RK13 cells were plated per 100mm dish (3x each cell type). 16 hours later, each plate was transfected with 5ug Alu-Neo (Dewannieux et al., 2003). After 2 days, all cells

were treated with 2ug/mL G418. Media was replaced daily for 7 days, after which time it was removed, and cells were fixed and stained with 20% Methanol + 0.2% crystal violet.

Detection of duplication events

Single Nanopore reads with multiple tandem copies of the mCherry-K3L fusion gene were analyzed using IGV (Robinson *et al.*, 2011). Duplications were confirmed by PCR, using the inverse PCR primers in mCherry (F- CGTGGAACAGTACGAACGCG and R-CCATGTTATCCTCCTCGCCC). PCR amplicons were then gel purified and Sanger sequenced to determine break points. However, due to the repetitive nature of the sequence flanking the isolate 7 insertion, the exact breakpoint was not able to be determined.

Serial passage of VACV-eIF2α virus strains

For each passage, 150-mm dishes were seeded with an aliquot from the same stock RK13- cells (5×10^6 cells/dish). For P1, dishes were infected with eIF2α-GFP or eIF2α^{S51A}-GFP virus (MOI = 1.0) for 2 hours in 5mL and then supplemented with 15mL medium. After 48 hours, cells were washed, pelleted, and resuspended in 1mL of medium. Virus was released by one freeze-thaw cycle followed by sonication. 900µl of virus was then used to infect a new dish of cells for P2, and the process was repeated for subsequent passages. Viral titers were determined using the remaining 100µl of reserved virus stocks from each passage by 48-hour plaque assay in RK13-E3K3 cells.

Analysis of viral-encoded eIF2α genes

RK13-E3K3 cells were infected with P4 eIF2α-GFP or eIF2α^{S51A}-GFP viruses (MOI = 0.1) for 24 hours. Virus-infected cells were collected, and total viral DNA extracted as previously described (Esposito *et al.*, 1981). The region between K2L and K4L containing the different viral-encoded eIF2α genes was amplified by PCR with primers K2L_seq_F and K4L_seq_R (5'-GGCATTGGTAAATCCTTGCAGA and 5'-CACCTTTTAGTAGGACTAGTATCGTACAA, respectively). SNV detection was performed by sequencing across eIF2α using primers eIF2α_F and eIF2α_R for eIF2α-GFP and eIF2α^{S51A}-GFP PCR products, CNV analysis was performed by PCR using primers eIF2α_rep_F (5'-CCTCCTATGGAAGCCAAAGCTGAAGATGAA) and eIF2α_rep_R (5'-CCTCCTATCTACAACTTAGACC CGGCAT) for eIF2α-GFP and

eIF2α^{S51A}-GFP viral DNA. Any CNV PCR products formed were sequenced using the same primers for breakpoint detection.

References

Bahar, M.W., Graham, S.C., Chen, R.A., Cooray, S., Smith, G.L., Stuart, D.I., and Grimes, J.M. (2011). How vaccinia virus has evolved to subvert the host immune response. J Struct Biol *175*, 127-134. 10.1016/j.jsb.2011.03.010.

Beattie, E., Tartaglia, J., and Paoletti, E. (1991). Vaccinia virus-encoded eIF-2 alpha homolog abrogates the antiviral effect of interferon. Virology *183*, 419-422. 10.1016/0042-6822(91)90158-8.

Bellehumeur, C., Nielsen, O., Measures, L., Harwood, L., Goldstein, T., Boyle, B., and Gagnon, C.A. (2016). Herpesviruses Including Novel Gammaherpesviruses Are Widespread among Phocid Seal Species in Canada. J Wildl Dis *52*, 70-81. 10.7589/2015-01-020.

Bratke, K.A., McLysaght, A., and Rothenburg, S. (2013). A survey of host range genes in poxvirus genomes. Infect Genet Evol *14*, 406-425. 10.1016/j.meegid.2012.12.002.

Brennan, G., Kitzman, J.O., Rothenburg, S., Shendure, J., and Geballe, A.P. (2014). Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog *10*, e1004002. 10.1371/journal.ppat.1004002.

Caprari, S., Metzler, S., Lengauer, T., and Kalinina, O.V. (2015). Sequence and Structure Analysis of Distantly-Related Viruses Reveals Extensive Gene Transfer between Viruses and Hosts and among Viruses. Viruses *7*, 5388-5409. 10.3390/v7102882.

Chakrabarti, S., Sisler, J., and Moss, B. (1997). Compact, synthetic, vaccinia virus early/late promoter for protein expression. Biotechniques *23*, 1094-1097.

de Groof, A., Guelen, L., Deijs, M., van der Wal, Y., Miyata, M., Ng, K.S., van Grinsven, L., Simmelink, B., Biermann, Y., Grisez, L., et al. (2015). A Novel Virus Causes Scale Drop Disease in Lates calcarifer. PLoS Pathog *11*, e1005074. 10.1371/journal.ppat.1005074.

Deeg, C.M., Chow, C.T., and Suttle, C.A. (2018). The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. Elife *7*. 10.7554/eLife.33014.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41-48. 10.1038/ng1223.

Dolja, V.V. and Koonin, E.V (2018). Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Res 244, 36-52. 10.1016/j.virusres.2017.10.020

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry.Nat Rev Microbiol *7*, 787-97. 10.1038/nrmicro2222.

Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., and Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150*, 831-841. 10.1016/j.cell.2012.05.049.

Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature *457*, 485-489. 10.1038/nature07529.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol *7*, 787-797. 10.1038/nrmicro2222.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat Genet *24*, 363-367. 10.1038/74184.

Esposito, J., Condit, R., and Obijeski, J. (1981). The preparation of orthopoxvirus DNA. J Virol Methods *2*, 175-179. 10.1016/0166-0934(81)90036-7.

Gilbert, C., Peccoud, J., Chateigner, A., Moumen, B., Cordaux, R., and Herniou, E.A. (2016). Continuous Influx of Genetic Material from Host to Virus Populations. PLoS Genet *12*, e1005838. 10.1371/journal.pgen.1005838.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247-266, 517-263. 10.1016/0042-6822(90)90294-2.

Gubser, C., Bergamaschi, D., Hollinshead, M., Lu, X., van Kuppeveld, F.J., and Smith, G.L. (2007). A new inhibitor of apoptosis from vaccinia virus and eukaryotes. PLoS Pathog *3*, e17. 10.1371/journal.ppat.0030017.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology *59*, 307-321.

Haller, S.L., Peng, C., McFadden, G., and Rothenburg, S. (2014). Poxviruses and the evolution of host range and virulence. Infect Genet Evol *21*, 15-40. 10.1016/j.meegid.2013.10.014.

Hughes, A.L., and Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. Mol Phylogenet Evol *35*, 186-195. 10.1016/j.ympev.2004.12.008.

Isaacs, S.N., ed. (2004). Vaccinia Virus and Poxvirology (Humana Press).

Kawagishi-Kobayashi, M., Silverman, J.B., Ung, T.L., and Dever, T.E. (1997). Regulation of the protein kinase PKR by the vaccinia virus pseudosubstrate inhibitor K3L is dependent on residues

519 conserved between the K3L protein and the PKR substrate eIF2alpha. Mol Cell Biol *17*, 4146-
520 58. 10.1128/MCB.17.7.4146
521
522 Kawagishi-Kobayashi, M., Cao, C., Lu, J., Ozato, K., and Dever, T.E. (2000). Pseudosubstrate
523 inhibition of protein kinase PKR by swine pox virus C8L gene product. Virology *276*, 424-34.
524 10.1006/viro.2000.0561.
525
526 Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat Rev
527 Genet *9*, 605-618. 10.1038/nrg2386.
528
529 Koonin, E.V., and Yutin, N. (2019). Evolution of the Large Nucleocytoplasmic DNA Viruses of
530 Eukaryotes and Convergent Origins of Viral Gigantism. Adv Virus Res *103*, 167-202.
531 10.1016/bs.aivir.2018.09.002.
532
533 Langland, J.O. and Jacobs, B.L. (2002). The role of the PKR-inhibitory genes, K3L and E3L, in
534 determining vaccinia virus host range. Virology *299*, 133-41. 10.1006/viro.2002.1479.
535
536 Lefkowitz, E.J., Wang, C., and Upton, C. (2006). Poxviruses: past, present and future. Virus Res
537 *117*, 105-118. 10.1016/j.virusres.2006.01.016.
538
539 McFadden, G., and Dales, S. (1979). Biogenesis of poxviruses: mirror-image deletions in
540 vaccinia virus DNA. Cell *18*, 101-108. 10.1016/0092-8674(79)90358-1.
541
542 McLysaght, A., Baldi, P.F., and Gaut, B.S. (2003). Extensive gene gain associated with adaptive
543 evolution of poxviruses. Proc Natl Acad Sci U S A *100*, 15655-15660.
544 10.1073/pnas.2136653100.
545
546 Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer,
547 M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1
548 retrotransposition. Nat Genet *31*, 159-165. 10.1038/ng898.
549
550 Park, C., Peng, C., Brennan, G., and Rothenburg, S. (2019). Species-specific inhibition of
551 antiviral protein kinase R by capripoxviruses and vaccinia virus. Ann N Y Acad Sci. *1438*, 18-
552 29. 10.1111/nyas.14000.
553
554 Park, C., Peng, C., Rahman, M.J., Haller, S.L., Tazi, L., Brennan, G., Rothenburg, S. (2021).
555 Orthopox K3 orthologs show virus- and host- specific ihibition of the antiviral protein PKR.
556 PLoS Pathog. *17,* e1009183. 10.1371/journal.ppat.1009183
557
558 Perez-Carmona, N., Farre, D., Martinez-Vicente, P., Terhorst, C., Engel, P., and Angulo, A.
559 (2015). Signaling Lymphocytic Activation Molecule Family Receptor Homologs in New World
560 Monkey Cytomegaloviruses. J Virol *89*, 11323-11336. 10.1128/JVI.01296-15.
561
562 Piskurek, O., and Okada, N. (2007). Poxviruses as possible vectors for horizontal transfer of
563 retroposons from reptiles to mammals. Proc Natl Acad Sci U S A *104*, 12046-12051.
564 10.1073/pnas.0700531104.

Rahman, M.J., Haller, S.L., Stoian, A.M.M., Li, J., Brennan, G., and Rothenburg, S. (2021) LINE-1 retrotransposons facilitate horizontal gene transfer into poxviruses. BioRxiv, https://www.biorxiv.org/content/10.1101/2020.10.26.355610v2.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24-26.

Saraiva, N., Prole, D.L., Carrara, G., Maluquer de Motes, C., Johnson, B.F., Byrne, B., Taylor, C.W., and Smith, G.L. (2013). Human and viral Golgi anti-apoptotic proteins (GAAPs) oligomerize via different mechanisms and monomeric GAAP inhibits apoptosis and modulates calcium. J Biol Chem *288*, 13057-13067. 10.1074/jbc.M112.414367.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods *15*, 461-468. 10.1038/s41592-018-0001-7.

Sklenovska, N., and Van Ranst, M. (2018). Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. Front Public Health *6*, 241. 10.3389/fpubh.2018.00241.

Swanstrom, R., Parker, R.C., Varmus, H.E., and Bishop, J.M. (1983). Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc Natl Acad Sci U S A *80*, 2519-2523. 10.1073/pnas.80.9.2519.

Wijayawardena, B.K., Minchella, D.J., and DeWoody, J.A. (2013). Hosts, parasites, and horizontal gene transfer. Trends Parasitol *29*, 329-338. 10.1016/j.pt.2013.05.001.

Yao, X.D., and Evans, D.H. (2001). Effects of DNA structure and homology length on vaccinia virus recombination. J Virol *75*, 6923-6932. 10.1128/JVI.75.15.6923-6932.2001.

Zhang, Z., Carriero, N., and Gerstein, M. (2004). Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends Genet *20*, 62-67. 10.1016/j.tig.2003.12.005.

Acknowledgments We thank D. Hancks, E. Choung, D. Downhour, and E. Grow for reagents and advice. This work was supported by NIH grants R35GM134936 (N.C.E.), T32GM007464 (S.M.F. and T.A.S), T32AI055434 (K.R.C.) and R01AI146915 (S.R.). N.C.E. was supported by a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Disease award and a H.A. and Edna Benning Presidential Endowed Chair.

Competing interests The authors declare no competing interests.

Figure Legends

Figure 1. Retrotransposon-mediated transfer of GAAP into poxvirus genomes. (**A**) The poxvirus gene vGAAP (orange) is encoded near the inverted terminal repeat (ITR; gray) of cowpox genomes. Signatures of retrotransposition include TSDs (blue), a LINE-1-like endonuclease site (EN; underlined), and a partially degraded poly(A) tail (pink). Some cowpox genomes (CPXV_5'; top line) contain a pseudo-empty site, with a single copy of the TSD sequence. (**B**) Phylogeny of TMBIM proteins, including TMBIM4, or GAAP, and virus encoded TMBIM4s (red). Bootstrap values >50 are indicated.

Figure 1 – figure supplement 1. (A) v-GAAP (orange bar) is encoded by many orthopox species, including most camelpox strains, many cowpox strains, and at least one vaccinia strain. In camelpox viruses, v-GAAP is encoded on the left end of the genome; in. all other strains, v-GAAP is on the right end. In all strains, v-GAAP is flanked by CrmE and vaccinia-B22R. In vaccinia strains that do not encode v-GAAP, CrmE is also missing, suggesting recombination of the region between B22R and B25R, which is truncated in the Lister strain. (B) Alignment of target site duplications (blue) and flanking sequence in 21 v-GAAP-encoding poxvirus strains. MPXV, monkeypox; CMLV, camelpox; CPXV, cowpox; VACV, vaccinia. Asterisks denote gene truncations.

Figure 2. Experimental capture of K3L by horizontal gene transfer. (**A**) Viruses lacking K3L (VCR2) replicated in RK13 cells expressing mCherry-K3L and were transferred to BHK cells lacking K3L to select for capture of mCherry-K3L. (**B**) Replication of VCR2 in RK13-K3L cells and BHK cells compared to VCR2+mCherry-K3L (see Methods). (**C**) Replication of recovered isolates in BHK cells. Three biological replicates of each strain/isolate are shown in (B) and (C). (**D**) Vaccinia genome illustrating K3L integrations in recovered isolates indicated by colored triangles. Endogenous K3L location is shown. The central region of the genome is highlighted in yellow. Triangles above and below the genome denote positive and negative sense orientation respectively. Virus genes interrupted by K3L integrations are indicated. The asterisk denotes a genomic rearrangement in isolate 10 (see Fig 3A).

Figure 2 – figure supplement 1. Protein Kinase R (PKR) is activated by binding dsRNA which is made in the cytoplasm during a viral infection. When activated, PKR binds and phosphorylates the eukaryotic initiation factor, eIF2α, leading to a block in translation and preventing viral replication. K3L acts as a pseudo-substrate of PKR, preventing activated PKR from binding eIF2α, which is then free to initiate translation. In the complementing RK13-K3L cells, an mCherry-K3L fusion gene was integrated into cellular chromosomes. The cellular mCherry-K3L expression construct is shown below. The fusion gene is driven by a strong EF1a promoter, but the construct also includes a poxvirus-specific promoter, the SLP, so that mCherry-K3L could be immediately expressed if transferred.

Figure 3. Each virus isolate exhibits signatures of LINE-1 mediated retrotransposition. (**A**) A detailed view of the region K3L inserted in each experimentally captured virus isolate. Arrows above cartoons indicate reading frame orientation. Flanking and/or interrupted (*) viral genes (blue/purple boxes), 3'/5' untranslated regions (UTRs; gray), and poly(A) tails (A_n; white) are shown. Genomic rearrangements included in isolate 10 are also shown. (**B**) Schematic of construct integrated in RK13-K3L cells (top) compared to K3L integrations in recovered viruses

(bottom). Shared features include spliced introns, 5' and 3' untranslated regions (UTRs), and poly(A) tails (A_n). Seven isolates have target site duplications (TSDs, blue) with LINE-1 endonuclease cut sites (EN, cyan). Six isolates encode a guanine (G cap) adjacent to the 5'UTR, indicative of 7-methylguanylate mRNA capping.

Figure 3 – supplemental figure 1. (A) Genetic features of each isolate recovered in the screen, including endonuclease (EN) cut site, target site duplication (TSD) length, presence or absence of transcript features: 5'-guanosine triphosphate (G) cap, 5' untranslated region (UTR), splicing, 3-UTR, and poly adenosine (A) tail, and whether there is evidence of a dsDNA break. ND (not determined) (**B**) A sequence logo plot of nucleotide preference from EN sites detected among all virus isolates.

Figure 3 – supplemental figure 2. An Alu-Neo reporter (cartoon above) was used to measure endogenous LINE-1 activity of the RK13-K3L cells in comparison to HeLa-HA cells. Representative plate wells from several experiments show visible Neo-resistant colonies. 4x magnification of plate wells shown to the right.

Figure 4. Homology driven gene duplication of K3L and eIF2*α*. (**A**) Schematics of mCherry-K3L duplications in isolates 4 and 5. (**B**) Virus titers after serial infections of eIF2α viruses in HeLa cells (see Methods). (**C**) Schematics of eIF2α duplications following virus passaging. Asterisks denote gene truncations.

Figure 4 – supplemental figure 1. Diagrams of the genomic region encompassing K3L in wildtype (WT) vaccinia virus, VCR2 strain, and viruses engineered to express eIF2α and eIF2α S51A. Arrows indicate reading frame orientation.

Figure 5. Model for horizontal transfer of host genes to poxvirus genomes. The schematic highlights how host genes (purple) are retrotransposed by LINE-1 reverse transcriptase (RT, blue), into virus genomes (yellow box). Following horizontal transfer, gene duplication is facilitated by unequal crossover recombination between TSDs (blue) or flanking repeat sequence (gray).

Figure 5 – supplemental figure 1. Models of evolution of newly acquired virus genes. (**A**) Genes that land in the repetitive region of the ITR are likely to be duplicated by both replication mechanisms (to the other ITR) and recombination mechanisms (tandem duplications). Duplications provide more chances for advantageous mutations to arise, increasing the likelihood of fixation. (**B**) Genes acquired outside the ITR may still be duplicated via TSD-mediated recombination. (**C**) Acquired genes that interrupt essential genes are unlikely to be maintained.

Data and materials availability Bioproject PRJNA614958.

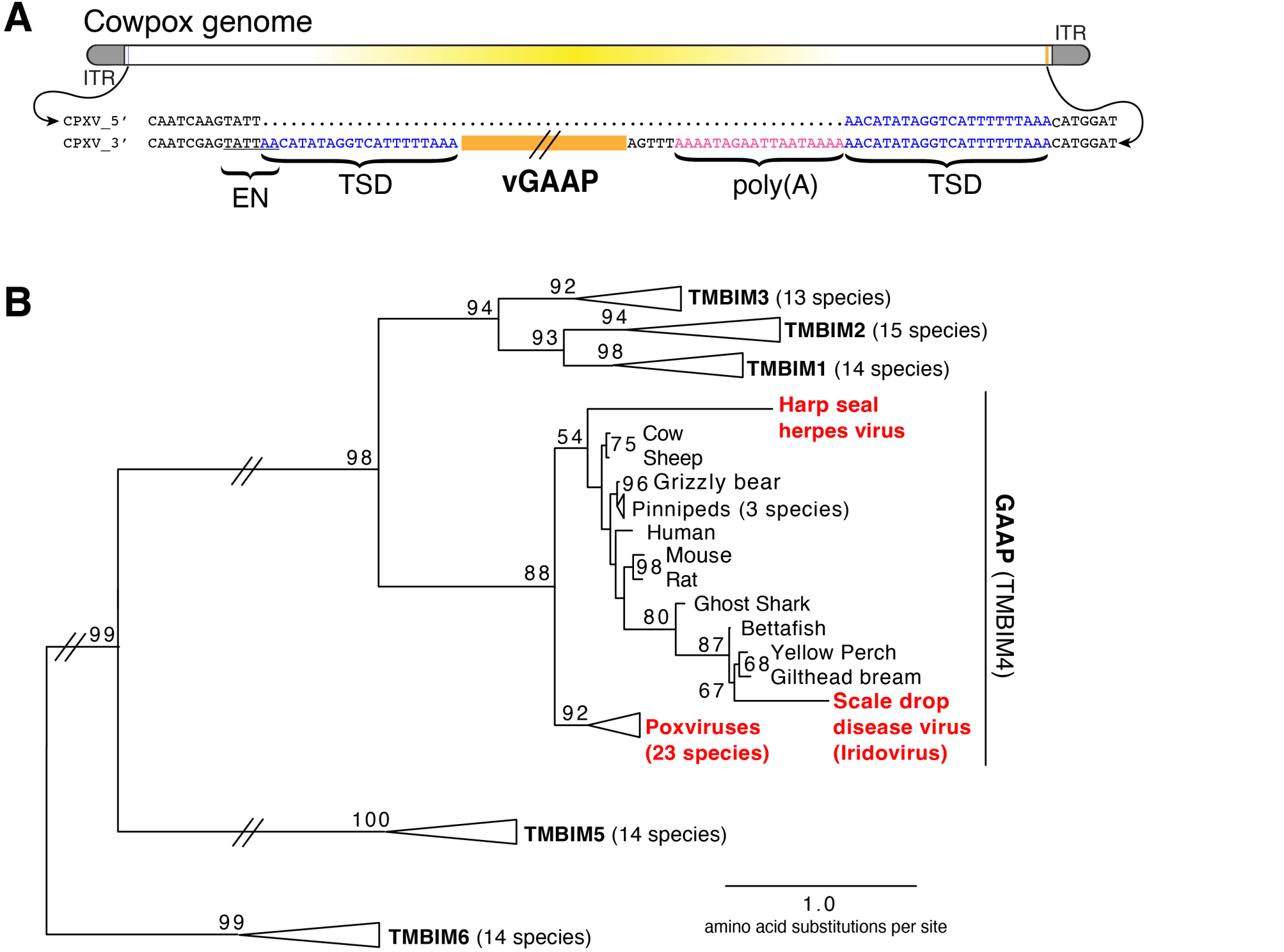
List of Supplementary Materials

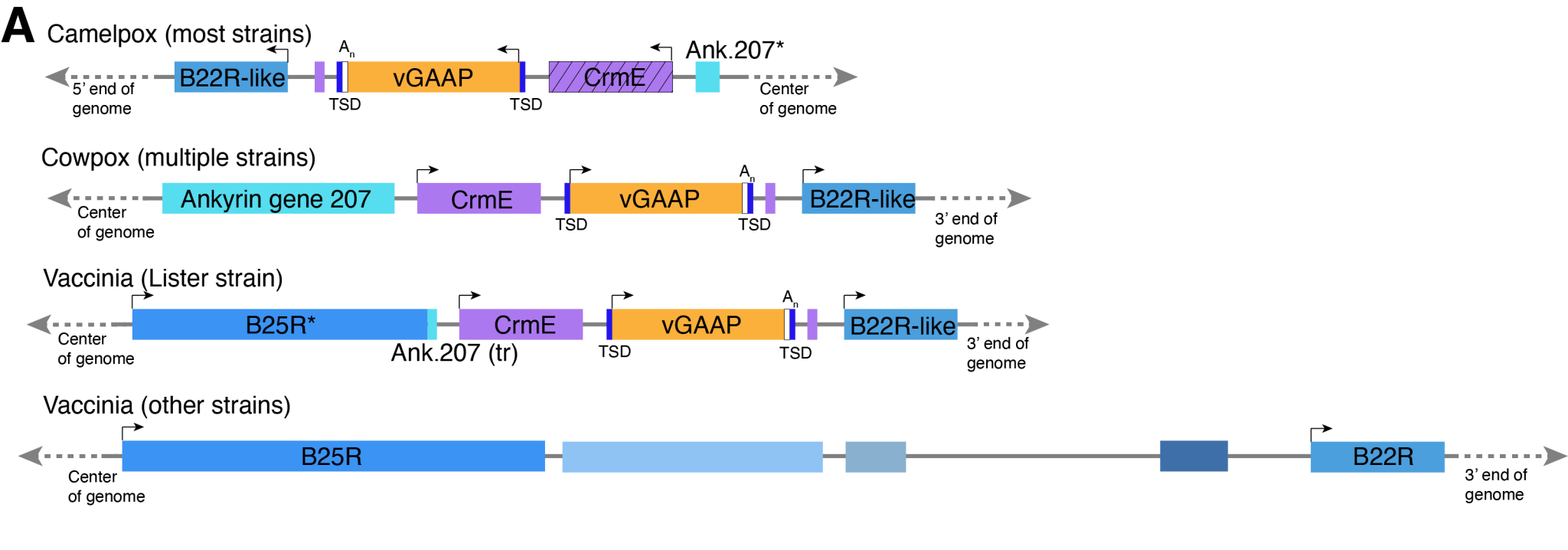
Supplementary File 1

Accession numbers for sequences used in phylogenetic analysis shown in Figure 1B

Amino acid alignment used in phylogenetic analysis shown in Figure 1B

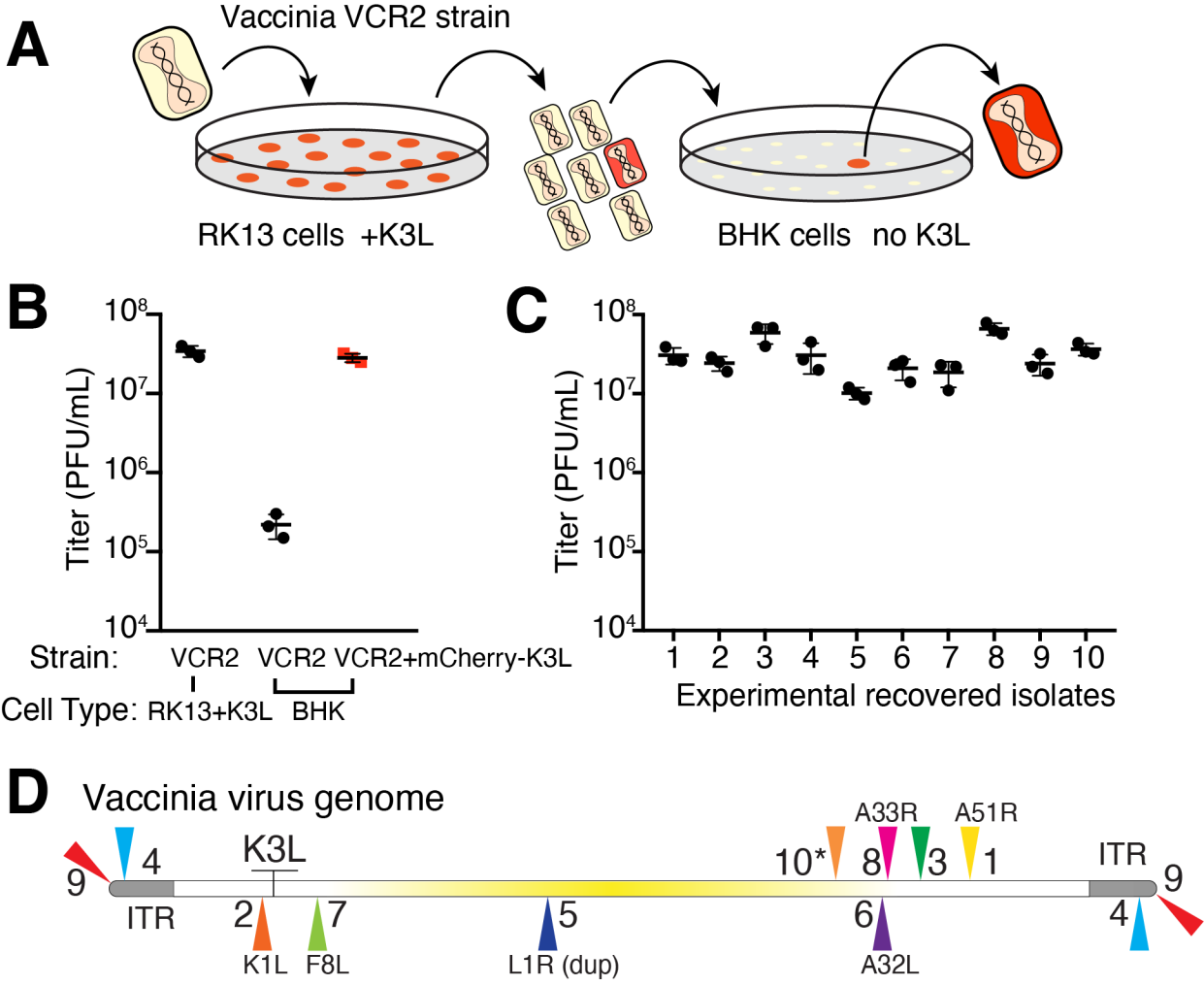
Supplementary sequences

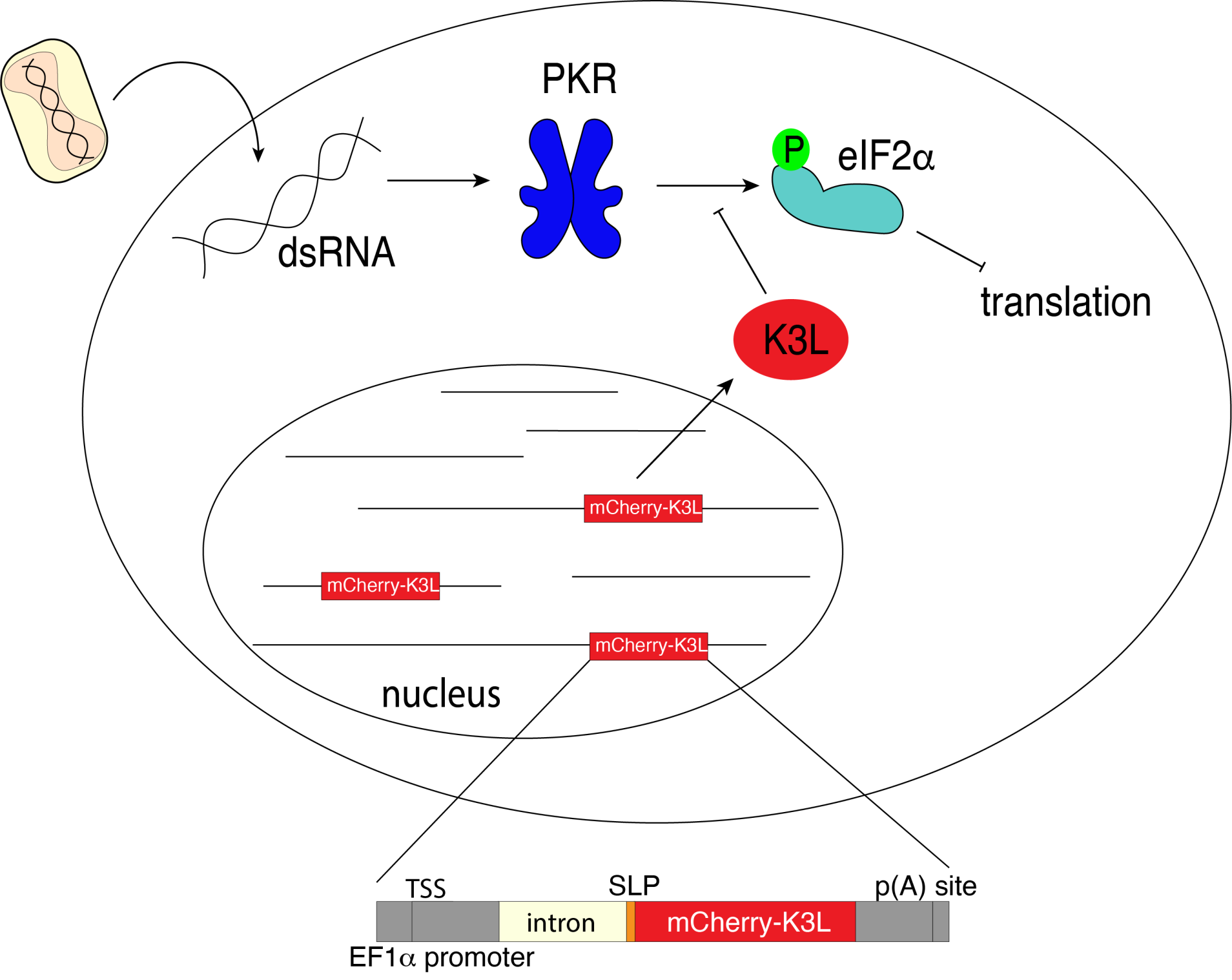

A
Cowpox genome
ITR
ITR
CPXV_5′ CAATCAAGTATT.....AACATATAGGTCATTTTTTAAACATGGAT
CPXV_3′ CAATCGAGTATTAACATATAGGTCATTTTTAAA AGTTTAAAATAGAATTAATAAAAAACATATAGGTCATTTTTTAAACATGGAT
EN
TSD
vGAAP
poly(A)
TSD
B
TMBIM3 (13 species)
TMBIM2 (15 species)
TMBIM1 (14 species)
Harp seal herpes virus
Cow
Sheep
Grizzly bear
Pinnipeds (3 species)
Human
Mouse
Rat
Ghost Shark
Bettafish
Yellow Perch
Gilthead bream
Scale drop disease virus (Iridovirus)
Poxviruses (23 species)
GAAP (TMBIM4)
TMBIM5 (14 species)
TMBIM6 (14 species)
1.0
amino acid substitutions per site

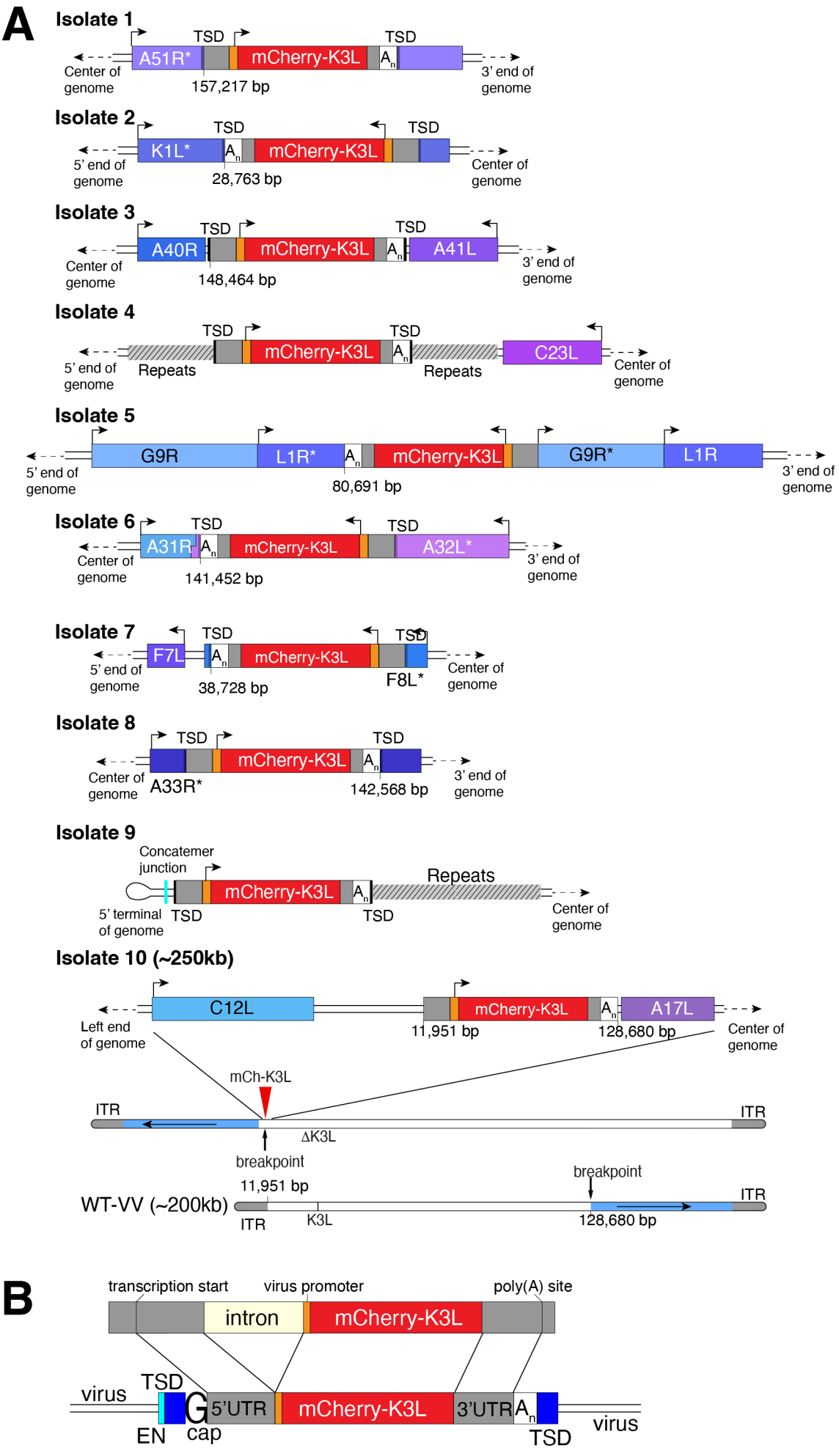


B

```
MXPV_Zaire                CAATCGAG---TATTAACATATAAGTTATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
MXPV_Cote_d'Ivoire        CAATCGAG---TATTAACATATAAGTTATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CMLV_Negev                CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CMLV_Kazakhstan           CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CMLV_CMS                  CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Germany_91-3         CAATCGAGTATTATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_FM2292               CAATCGAGTATTATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Ger/2015/Cat4        CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Germany_1998_2       CAATCGAGTATTATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_HumMag07/1           CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Germany_1971_EP1     CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Germany_1980_EP4     CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Germany_2002_MKY     CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Ger/2010/Rat         CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Ratpox09             CAATCGAT---TATTAACATATAGG---TTTTTAAA.........AACATATAGGTCATTTTTTAA--ATGGAT
CPXV_HumAac09/1           CAATCGAT---TATTAACATATAGG---TTTTTAAA.........AACATATAGGTCATTTTTTAA--ATGGAT
CPXV_Austria_1999         CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Kostroma_2015        CAATCGAG---TATTAACATATAAGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_GRI-90               CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
CPXV_Finland_2000_MAN     CAATCGAG---TATTAACATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
VACV_Lister               CAATCGAG---TATTAATATATAGGTCATTTTTAAA.........AACATATAGGTCATTTTTTAAACATGGAT
```


A
Vaccinia VCR2 strain
RK13 cells +K3L
BHK cells no K3L
B
Titer (PFU/mL)
10^8
10^7
10^6
10^5
10^4
Strain:
VCR2
VCR2
VCR2+mCherry-K3L
Cell Type:
RK13+K3L
BHK
C
Titer (PFU/mL)
1 2 3 4 5 6 7 8 9 10
Experimental recovered isolates
D
Vaccinia virus genome
ITR
K3L
K1L
F8L
L1R (dup)
A33R
A51R
A32L
9
4
2
7
5
10*
8
3
1
6

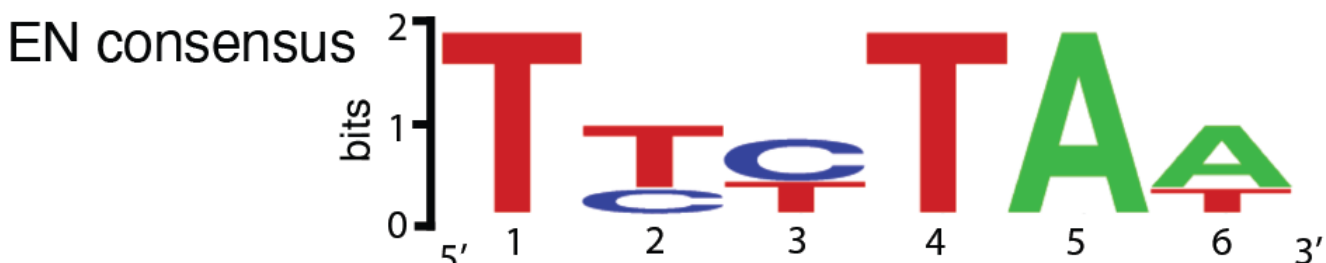

PKR
dsRNA
P
eIF2α
translation
K3L
mCherry-K3L
mCherry-K3L
mCherry-K3L
nucleus
TSS
SLP
p(A) site
intron
mCherry-K3L
EF1α promoter

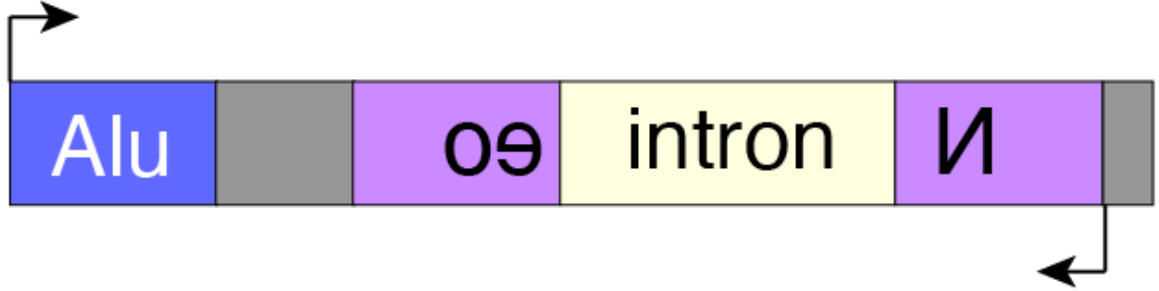

A
Isolate 1
TSD
TSD
A51R*
mCherry-K3L
An
Center of genome
157,217 bp
3' end of genome
Isolate 2
TSD
TSD
K1L*
An
mCherry-K3L
5' end of genome
28,763 bp
Center of genome
Isolate 3
TSD
TSD
A40R
mCherry-K3L
An
A41L
Center of genome
148,464 bp
3' end of genome
Isolate 4
TSD
TSD
mCherry-K3L
An
C23L
5' end of genome
Repeats
Repeats
Center of genome
Isolate 5
G9R
L1R*
An
mCherry-K3L
G9R*
L1R
5' end of genome
80,691 bp
3' end of genome
Isolate 6
TSD
TSD
A31R
An
mCherry-K3L
A32L*
Center of genome
141,452 bp
3' end of genome
Isolate 7
TSD
TSD
F7L
An
mCherry-K3L
5' end of genome
38,728 bp
F8L*
Center of genome
Isolate 8
TSD
TSD
mCherry-K3L
An
Center of genome
A33R*
142,568 bp
3' end of genome
Isolate 9
Concatemer junction
Repeats
mCherry-K3L
An
5' terminal of genome
TSD
TSD
Center of genome
Isolate 10 (~250kb)
C12L
mCherry-K3L
An
A17L
Left end of genome
11,951 bp
128,680 bp
Center of genome
mCh-K3L
ITR
ITR
ΔK3L
breakpoint
11,951 bp
breakpoint
WT-VV (~200kb)
ITR
ITR
K3L
128,680 bp
B
transcription start
virus promoter
poly(A) site
intron
mCherry-K3L
virus
TSD
G
5'UTR
mCherry-K3L
3'UTR
An
virus
EN
cap
TSD

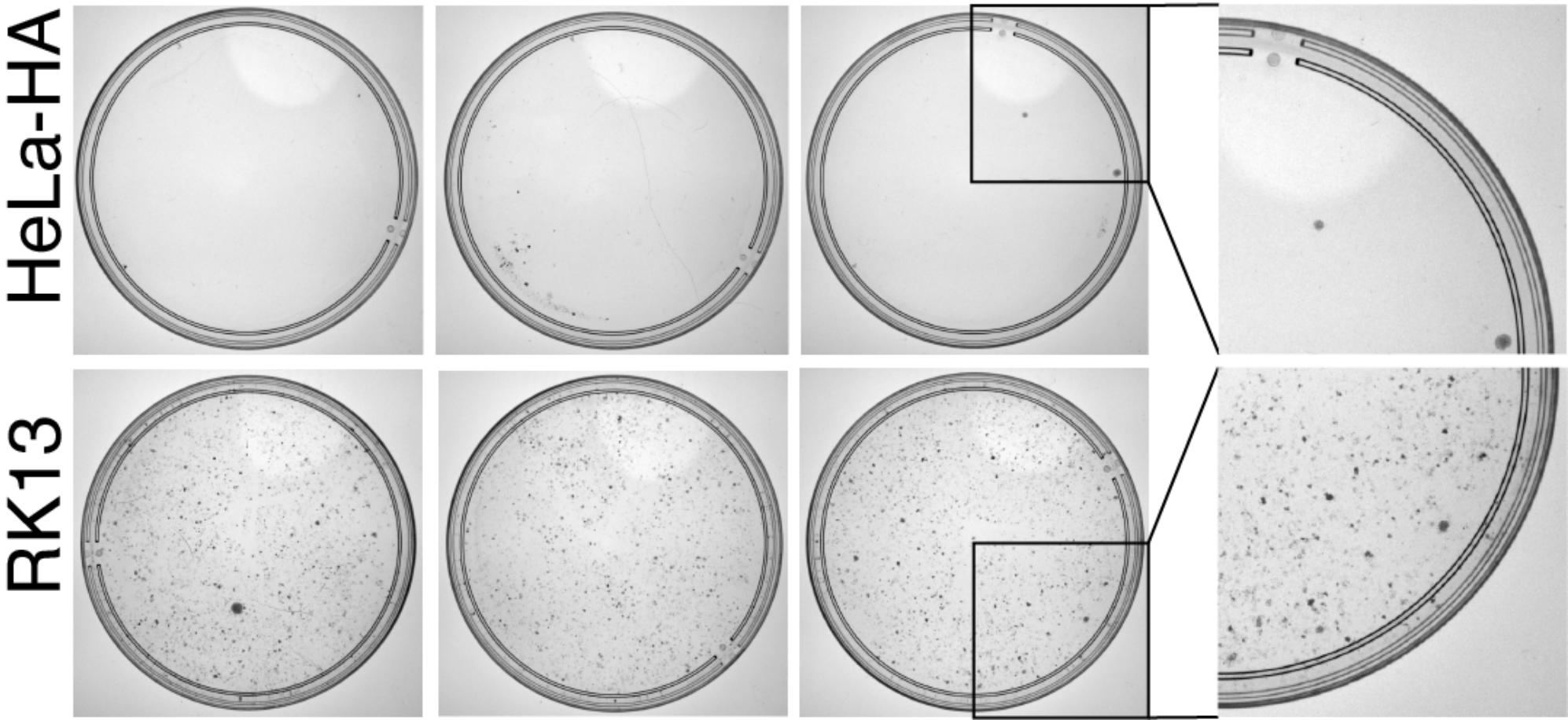
A

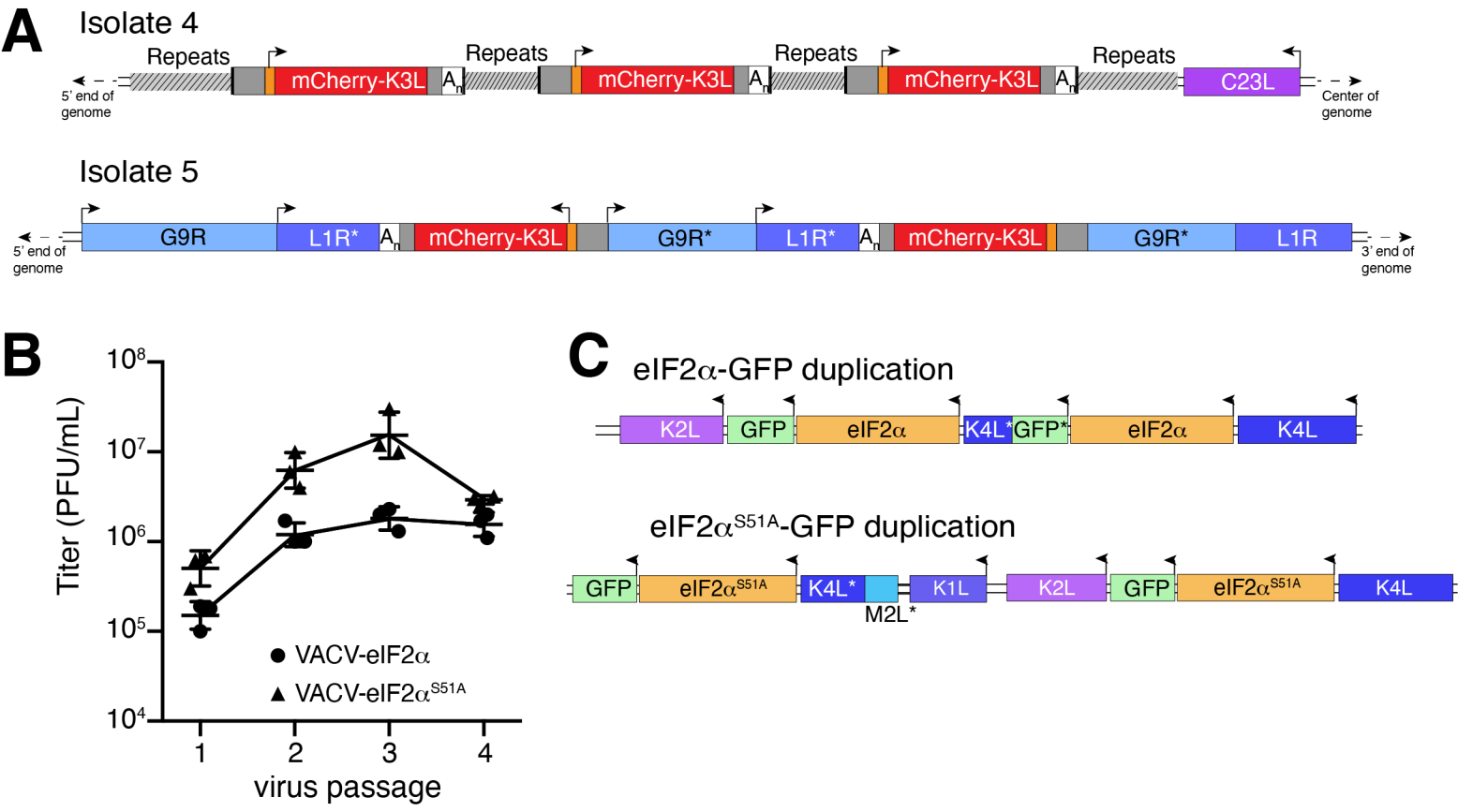
Isolate	EN site	TSD length	G cap	5'UTR	spliced	3'UTR	poly(A)	dsDNA break
1	TCTT/AA	17bp	Y	Y	Y	Y	Y	None
2	TTCT/AA	16bp	Y	Y	Y	Y	Y	None
3	TTCT/AA	19bp	Y	Y	Y	Y	Y	None
4	TTTT/AT	14bp	N	Y	Y	Y	Y	None
5	None	None	Y	Y	Y	Y	Y	tandem dup
6	TTCT/AA	15bp	N	Y	Y	Y	Y	None
7	TCTT/AA	18bp	Y	Y	Y	Y	Y	None
8	TTCT/AT	16bp	N	Y	Y	Y	Y	None
9	ND	ND	N	Y	Y	Y	Y	concatemer
10	None	None	Y	Y	Y	Y	Y	translocation

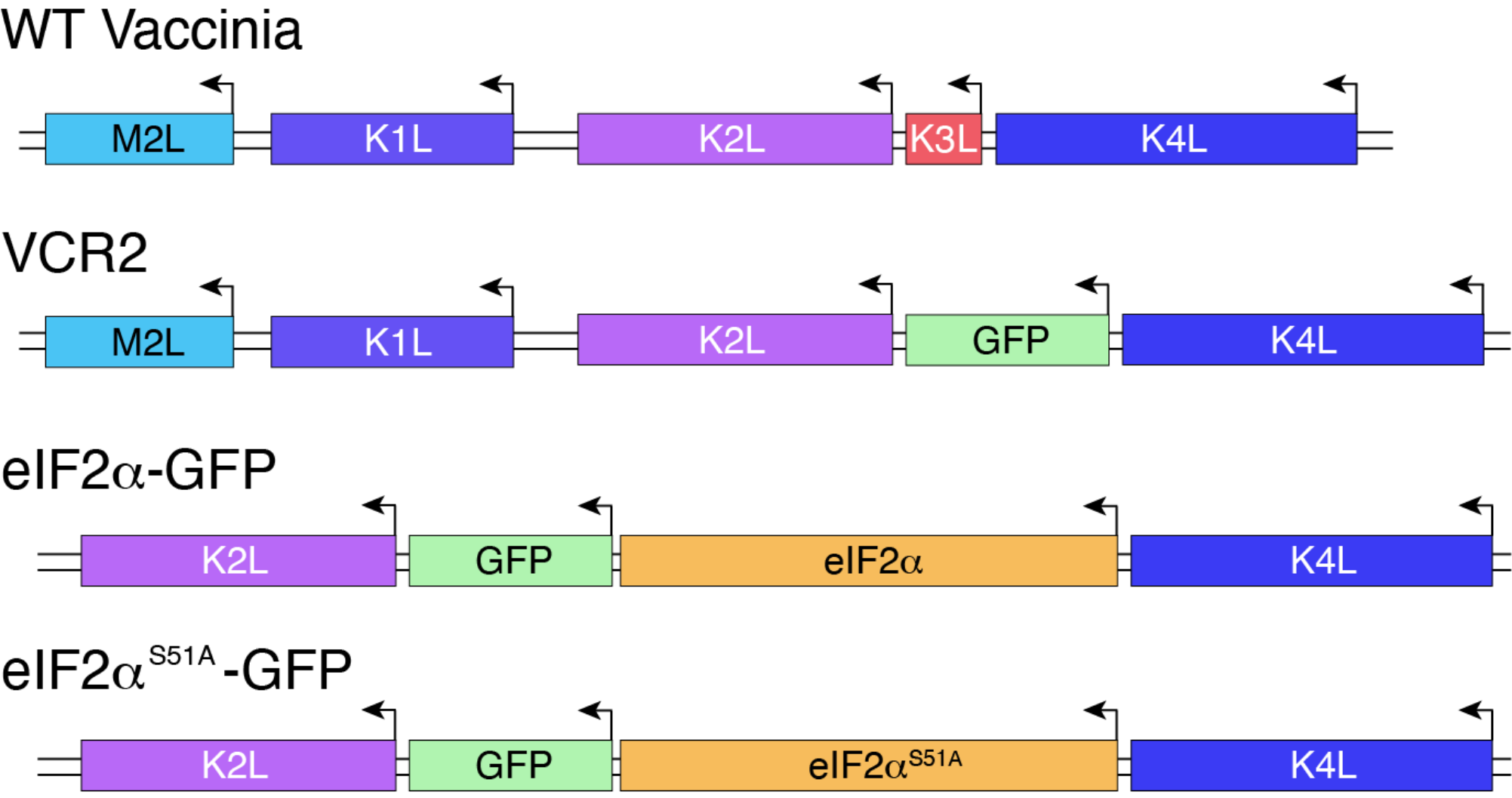
B

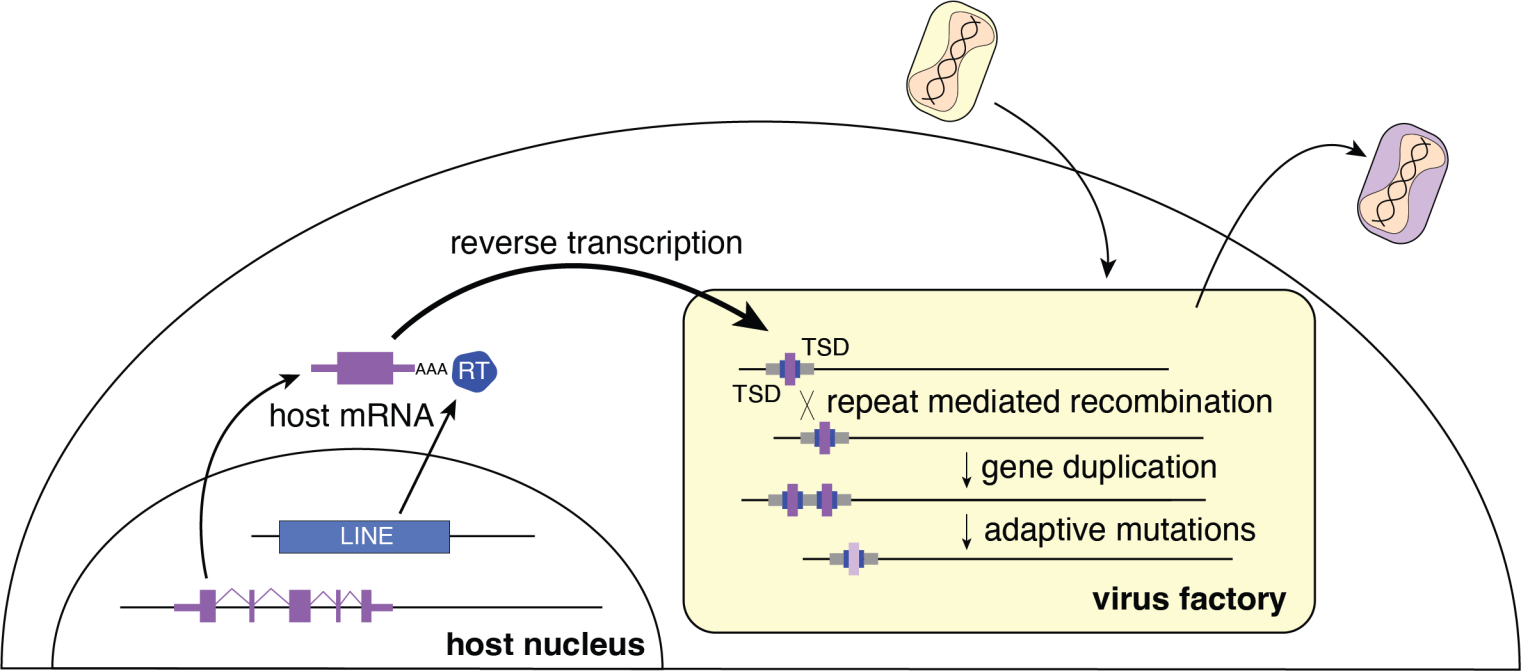


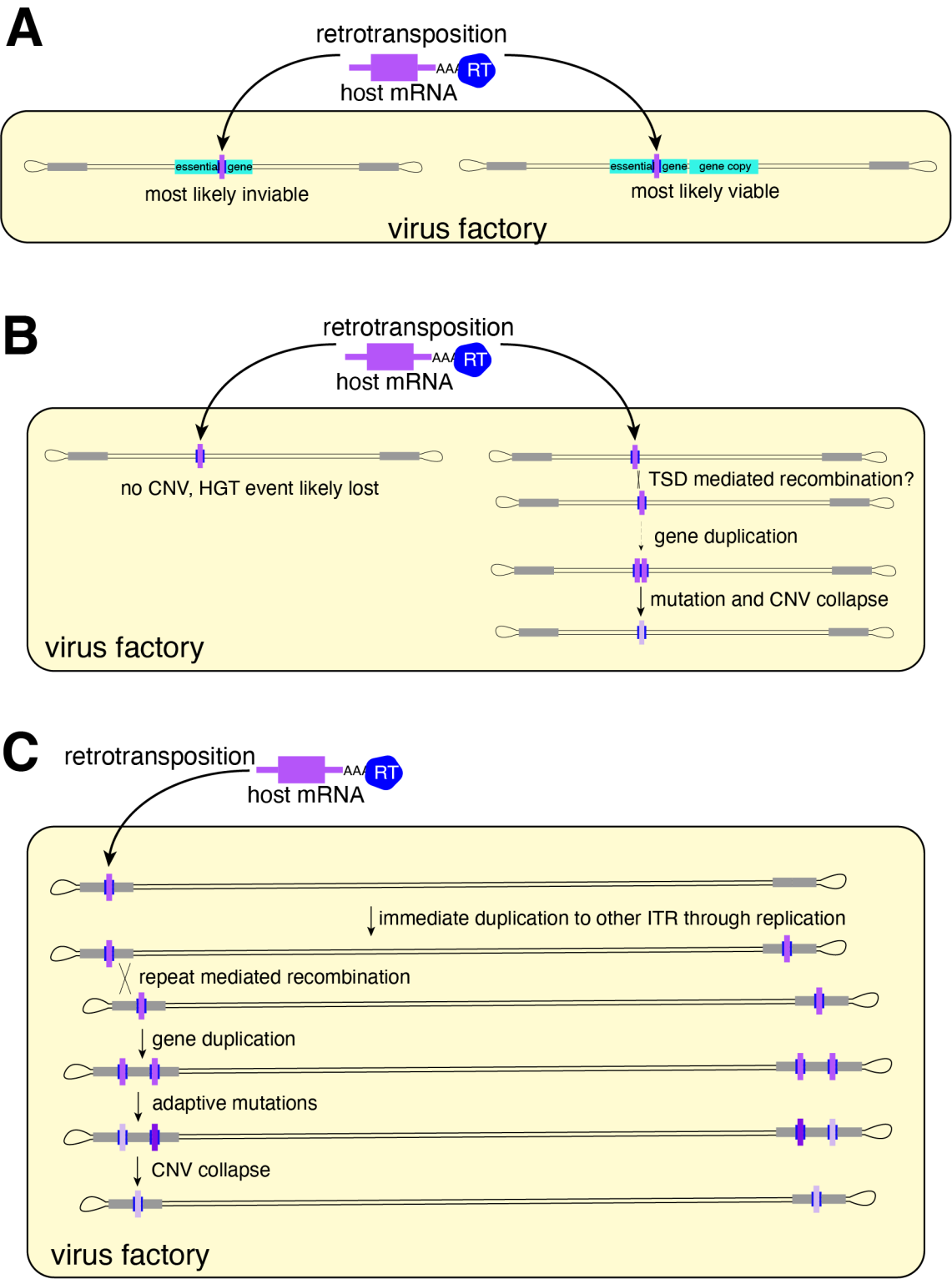

Alu
оэ
intron
И


HeLa-HA
RK13


A
Isolate 4
Repeats
mCherry-K3L
An
C23L
5' end of genome
Center of genome
Isolate 5
G9R
L1R*
G9R*
3' end of genome
B
Titer (PFU/mL)
virus passage
VACV-eIF2α
VACV-eIF2αS51A
C
eIF2α-GFP duplication
K2L
GFP
eIF2α
K4L*
GFP*
K4L
eIF2αS51A-GFP duplication
K1L
M2L*


WT Vaccinia
M2L
K1L
K2L
K3L
K4L
VCR2
M2L
K1L
K2L
GFP
K4L
eIF2α-GFP
K2L
GFP
eIF2α
K4L
eIF2α S51A -GFP
K2L
GFP
eIF2α S51A
K4L


reverse transcription
AAA
RT
host mRNA
LINE
host nucleus
TSD
TSD
repeat mediated recombination
gene duplication
adaptive mutations
virus factory


A
retrotransposition
AAA
RT
host mRNA
essential gene
most likely inviable
essential gene
gene copy
most likely viable
virus factory
B
retrotransposition
AAA
RT
host mRNA
no CNV, HGT event likely lost
TSD mediated recombination?
gene duplication
mutation and CNV collapse
virus factory
C
retrotransposition
AAA
RT
host mRNA
immediate duplication to other ITR through replication
repeat mediated recombination
gene duplication
adaptive mutations
CNV collapse
virus factory