LINE-1 retrotransposons facilitate horizontal gene transfer into poxviruses

M. Julhasur Rahman[1, 2], Sherry L. Haller[3], Ana M. M. Stoian[1], Jie Li[4], Greg Brennan[1] and Stefan Rothenburg[1*]

[1] School of Medicine, University of California Davis, Department of Medial Microbiology and Immunology, Davis, CA 95616, USA

[2] Current address: Genentech Inc., 1 DNA Way, South San Francisco, CA 94080

[3] University of Texas Medical Branch at Galveston, Department of Microbiology and Immunology, Galveston, TX 77555, USA

[4] Genome Center, University of California Davis, Davis, CA 95616, USA

[*]correspondence: rothenburg@ucdavis.edu. ORCID iD:0000-0002-2525-8230

Keywords: horizontal gene transfer; LINE-1; retrotransposons; poxvirus; evolution, PKR; E3L

Abstract

There is ample phylogenetic evidence that many critical virus functions, like immune evasion, evolved by the acquisition of genes from their hosts through horizontal gene transfer (HGT). However, the lack of an experimental system has prevented a mechanistic understanding of this process. We developed a model to elucidate the mechanisms of HGT into vaccinia virus, the prototypic poxvirus. All identified gene capture events showed signatures of long interspersed nuclear element-1 (LINE-1)-mediated retrotransposition, including spliced-out introns, polyadenylated tails and target site duplications. In one case, the acquired gene integrated together with a polyadenylated host U2 small nuclear RNA. Integrations occurred across the genome, in some cases knocking out essential viral genes. These essential gene knockouts were rescued through a process of complementation by the parent virus followed by non-homologous recombination during serial passaging to generate a single competent virus. This work links multiple evolutionary mechanisms into one adaptive cascade and identifies host retrotransposons as major drivers for virus evolution.

Introduction

Horizontal gene transfer (HGT) is the transmission of genetic material between different organisms. This process shapes the evolution and functional diversity of all major life forms, including viruses (Keeling, 2009, Gilbert and Cordaux, 2017). The genomes of large DNA viruses, including poxviruses and herpesviruses, contain numerous genes that likely have been acquired from their hosts by HGT. Many of these captured genes are involved in immune evasion or protection from environmental damage (Hughes and Friedman, 2005, Bratke and McLysaght, 2008, Bratke et al., 2013, Wilson and Brooks, 2011, Iyer et al., 2006). So far, the detection of HGT in large DNA viruses has relied on bioinformatic approaches, primarily phylogenetic and sequence composition analyses. For example, Odom, et al. performed a family wide comparison of poxvirus coding genes to different taxonomic subsets, such as "all eukaryotic genes" or "all other viral genes". They found that poxvirus ORFs are on average more similar to eukaryotic genes than other viruses, suggesting substantial HGT into poxviruses over evolutionary time (Odom et al., 2009). In an orthogonal approach, Monier, et al, used a Bayesian methodology to identify large DNA virus genes with anomalous nucleotide composition relative to the viral genome as a whole. Using this approach, they determined that a large number of the compositionally anomalous genes in *Poxviridae* are associated with host immune control (Monier et al., 2007). From these and other studies, it has become apparent that viral homologs of host interleukin 10 (IL-10) genes are among the best supported examples of these HGT events. Viral IL-10 genes, which presumably provide a selective advantage through inhibition and modulation of the antiviral response, appear to have been independently acquired by several herpesviruses and poxviruses (Hughes and Friedman, 2005, Schonrich et al., 2017). While these methods can detect putative HGT they can do little to elucidate the frequency or mechanisms of HGT.

The two most likely mechanisms for HGT into DNA viruses are either a DNA-mediated process such as integration via recombination or the help of DNA transposons, or an RNA-mediated integration, such as gene transfer via retrotransposons or retroviruses (Schonrich et al., 2017, Sekiguchi and Shuman, 1997). While it is unclear how most viral orthologs of host genes were acquired, the vIL-10 gene in ovine herpesvirus 2 has retained the intron structure of the host gene, supporting a mechanism of DNA-mediated HGT in this instance (Jayawardane et al., 2008). However, because most herpesvirus genes and all known poxvirus genes lack introns, it is unclear how common DNA-mediated HGT events may be. In fact, the taterapox virus genome contains a host-derived short interspersed nuclear element (SINE), which is flanked by a 16 bp target-site duplication (TSD), indicating that the long interspersed nuclear element-1 (LINE1, L1) group of retrotransposons can also facilitate the transfer of host genetic material into poxviruses through an RNA-mediated process (Piskurek and Okada, 2007).

The family *Poxviridae* comprise many viruses of medical and veterinary importance, including variola virus, the causative agent of human smallpox, vaccinia virus (VACV) the best-studied poxvirus, which was used as the vaccine against smallpox, and monkeypox virus, an emerging human pathogen. Poxviruses can enter a large number of mammalian cell types in a species-independent manner; therefore, productive infection largely depends on the successful subversion of the host immune response (Haller et al., 2014). More than 40% of poxvirus gene families show evidence that they were host-derived, including many genes involved in immune evasion (Hughes and Friedman, 2005, Bratke and McLysaght, 2008). This high proportion of host-acquired genes suggests that poxviruses may represent ideal candidates to study the process of HGT.

In order to establish an experimental model of HGT, we exploited the dependence of VACV replication on effective inhibition of PKR, which is an antiviral protein activated by double-stranded (ds) RNA formed during infection with both DNA and RNA viruses. Activated PKR phosphorylates the alpha subunit of eukaryotic translation initiation factor 2 (eIF2α), which inhibits cap-dependent protein synthesis, thereby preventing virus replication (Wu and Kaufman, 1997). Because PKR exerts a strong antiviral effect, many viruses, including poxviruses, have evolved inhibitors that can block PKR activation (Langland et al., 2006). Most poxviruses that infect mammals contain two PKR inhibitors (Bratke et al., 2013), which are called E3 (encoded by E3L) and K3 (encoded by K3L) in VACV. E3 is a dsRNA-binding protein, which prevents PKR homodimerization and thereby inhibits PKR activation (Davies et al., 1993, Chang et al., 1992). K3L, which may itself be a horizontally transferred gene, possesses homology to the PKR-binding S1 domain of eIF2α, and acts as a pseudosubstrate inhibitor of PKR (Beattie et al., 1991, Dar and Sicheri, 2002).

In this work we have established a model system to study HGT from the host genome into VACV. In all of the HGT events we detected, the captured genes showed multiple signatures that host LINE-1 retrotransposons were involved in the gene capture. In some cases, integration of the horizontally acquired gene occurred in essential virus genes. In those cases, complementation of both the parental virus and the virus that acquired the gene was necessary for virus replication. After only a few rounds of continued passaging, these complementing viruses recombined to generate single replication-competent viruses. Taken together, retrotransposon-mediated HGT followed by complementation and recombination describes an unexpected cascade of evolutionary processes to rapidly integrate host genes into the viral genome.

Results

An experimental system to detect host gene capture by vaccinia virus

In order to detect HGT into VACV, we designed an experimental system that uses the selective pressure imposed by PKR on VACV replication. In this model, we use a VACV strain (VC-R2) that lacks both PKR antagonists and can therefore only replicate in PKR-deficient cells or in complementing cells that express viral PKR inhibitors, such as VACV E3 or K3 (Hand et al., 2015). We stably transfected PKR-competent RK13 cells with E3L-tagged mCherry (mCherry-E3L) (Figure 1A), selected an individual clone that showed robust mCherry-E3 expression by fluorescence microscopy, and verified that it contained the entire expression cassette (Supporting Figure 1). The PKR sensitive virus VC-R2 can infect these cells as efficiently as wild type VACV, but not wild type RK13 cells. However, if during the course of infection VC-R2 acquired mCherry-E3L from the host cell, the virus would be able to replicate in PKR-competent cells and fluoresce red (Figure 1A). The expression cassette also contains the rabbit β-globin intron (Figure 1B) enabling us to distinguish between direct DNA-mediated integration or indirect RNA-mediated integration, the two most likely mechanisms for HGT into DNA viruses. We infected the RK13-mCherry-E3L cell line with VC-R2 and titered the resulting viruses on permissive RK13+E3+K3 cells (Rahman et al., 2013). Then, we infected wild type RK13 cells, which are non-permissive for VC-R2, with these viruses and isolated plaques that expressed both mCherry and EGFP (Figure 1A). Overall, we identified 20 recombinant viruses with unique integration sites (Figure 1D, E). In total, we screened approximately 7.5 x 10^8 virions while optimizing this system. After optimization, we screened 3.63 x 10^8 virions derived from a single 48-hour infection of RK13 cells (multiplicity of infection (MOI) = 0.01). We screened this population in RK13 cells (MOI = 0.2)

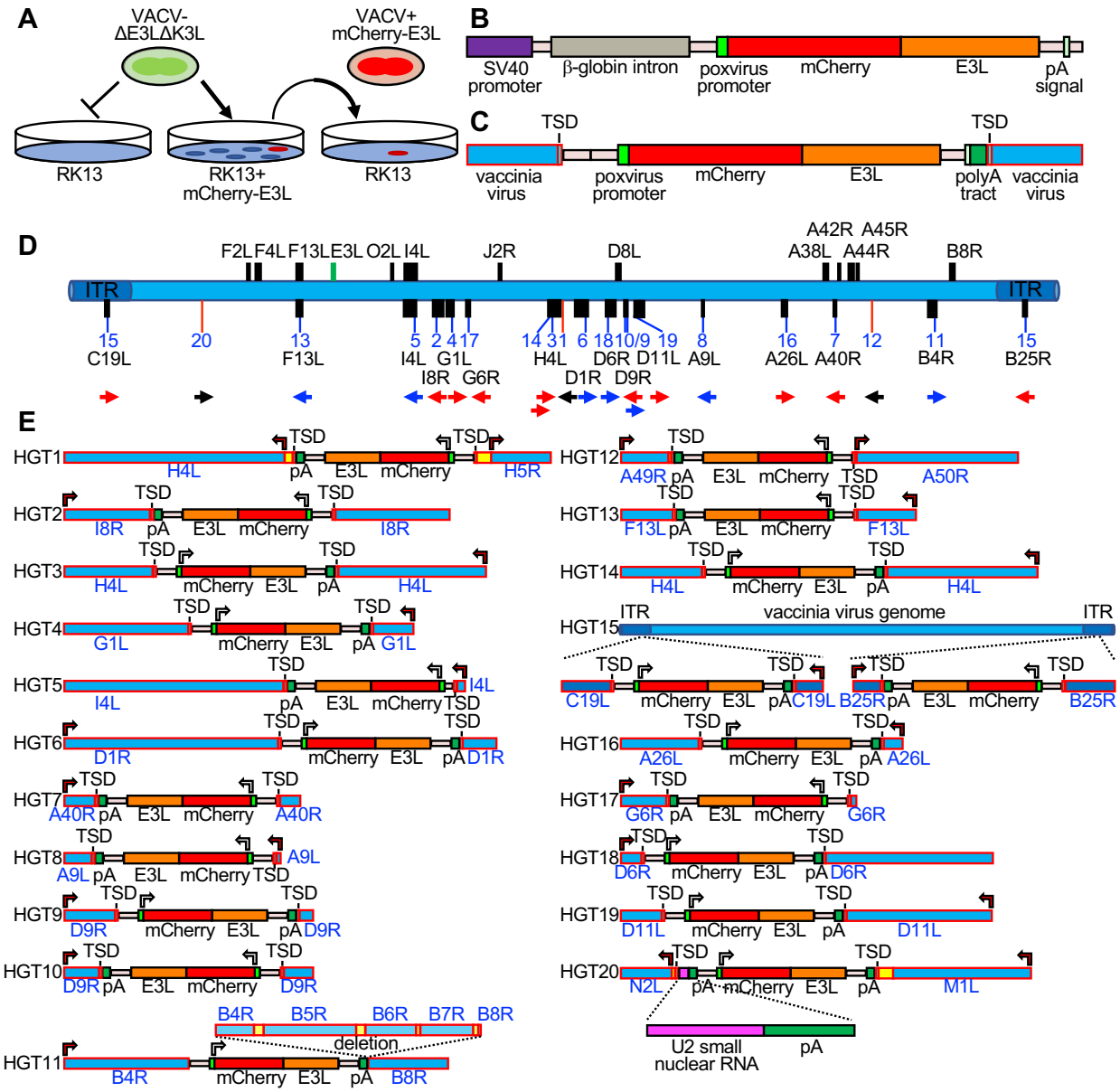


Figure 1. Detection of experimental horizontal gene transfer in vaccinia virus. (A) VACV lacking E3L and K3L cannot replicate in wild type RK13 cells, but can replicate in cells stably transfected with mCherry-E3L. Virus that acquired E3L can replicate in wild type RK13 cells (right). (**B**) Schematic of the mCherry-E3L vector that was stably transfected into RK13 cells. (**C**) Schematic of the general genetic architecture of horizontally transferred genes identified in VACV isolates. (**D**) HGT integration sites in the VACV genome. The VACV genome is represented in blue. Genes highlighted above the genome were described to have likely originated from HGT (Hughes and Friedman, 2005, Bratke and McLysaght, 2008). Features shown below the genome are: integration sites into genes (black boxes, blue lines), or into intergenic regions (red lines). The orientation of the transferred genes is indicated by the arrow, colors of arrows indicate the orientation of mCherry-E3L relative to target genes (blue: same direction; red opposite direction; black arrow: intergenic). (**E**) Maps of mCherry-E3L integration sites in HGT1 to HGT20. Arrows indicate the direction of transcription for VACV and mCherry-E3L. Intergenic regions are depicted in yellow. The position of an integrated U2 small nuclear RNA and the associated poly(A) tract is shown for HGT20 by dashed lines.

142 and identified 16 unique HGT isolates. This indicates a detected transfer rate of mCherry-E3L into
143 the VACV genome of approximately 1 in 22.7 million virions.
144
145 **Horizontally acquired genes show evidence of LINE-1-mediated retrotransposition**
146 To map the integration sites, we plaque purified each HGT candidate three times and used both
147 Sanger sequencing of amplicons from inverse PCR (Ochman et al., 1988) and PacBio-based long-
148 read sequencing (Eid et al., 2009). The integration sites were distributed throughout the VACV
149 genome, and were found in intergenic spaces (three cases), as well as in non-essential and essential
150 genes (Figure 1D, E, Table 1). All 20 isolates (HGT1-HGT20) displayed hallmarks of RNA-
151

Isolate #	Integration site	Essential vs. non-essential
HGT1	Intergenic between H4L and H5R	Presumably non-essential
HGT2	I8R	Essential (Gross and Shuman, 1996)
HGT3	H4L	Essential (Kane and Shuman, 1992)
HGT4	G1L	Essential (Hedengren-Olcott et al., 2004))
HGT5	I4L	Non-essential (Child et al., 1990, Gammon et al., 2010)
HGT6	D1R	Essential (Hassett et al., 1997)
HGT7	A40R	Non-essential (Wilcock et al., 1999)
HGT8	A9L	Essential (Yeh et al., 2000)
HGT9 and HGT10	D9R	Non-essential (Parrish and Moss, 2006)
HGT11	B4R	Non-essential (Burles et al., 2014)
HGT12	Intergenic between A49R and A50R	Presumably non-essential
HGT13	F13L	Non-essential (Blasco and Moss, 1991)
HGT14	H4L	Essential (Kane and Shuman, 1992)
HGT15	C19L/B25R	Non-essential (Perkus et al., 1991)
HGT16	A26L	Non-essential (Howard et al., 2008, Chang et al., 2019)
HGT17	G6R	Non-essential (Senkevich et al., 2008)
HGT18	D6R	Essential (Broyles and Fesler, 1990)
HGT19	D11L	Essential (Seto et al., 1987)
HGT20	Intergenic between N2L and M1L	Presumably non-essential

152
153 **Table 1. Integration sites of mCherry-E3L in HGT viruses and importance of disrupted genes for**
154 **virus replication.**
155

156 dependent, retrotransposon-mediated integrations, including a spliced-out intron and long
157 untemplated stretches of poly(A) at the 3' end (Esnault et al., 2000).
158 In 19 cases, short target site duplications (TSDs) of VACV sequences surrounded the integration
159 sites (Figure 1C, E, Figure 2, Supporting Table 1). These TSDs had an average length of 15.9 base
160 pairs (bp) and a consensus (3'-AA/TTTT-5') motif at the integration site typical of the cleavage
161 consensus site of the long interspersed nuclear element-1 (LINE-1, L1) group of retrotransposons
162 (Figure 2) (Kojima, 2010, Gilbert et al., 2002). HGT11 contained a 2.7 kb deletion immediately
163 downstream of the integration site in B4R, precluding the identification of a TSD (Figure 1E). In
164 HGT20, the integrated sequence contained the intron-less mCherry-E3L cassette fused to an 88 bp
165 cellular DNA fragment encoding U2 small nuclear (sn) RNA, which contained a second poly(A)
166 tract (Figure 1E, Supporting Figure 2).
167

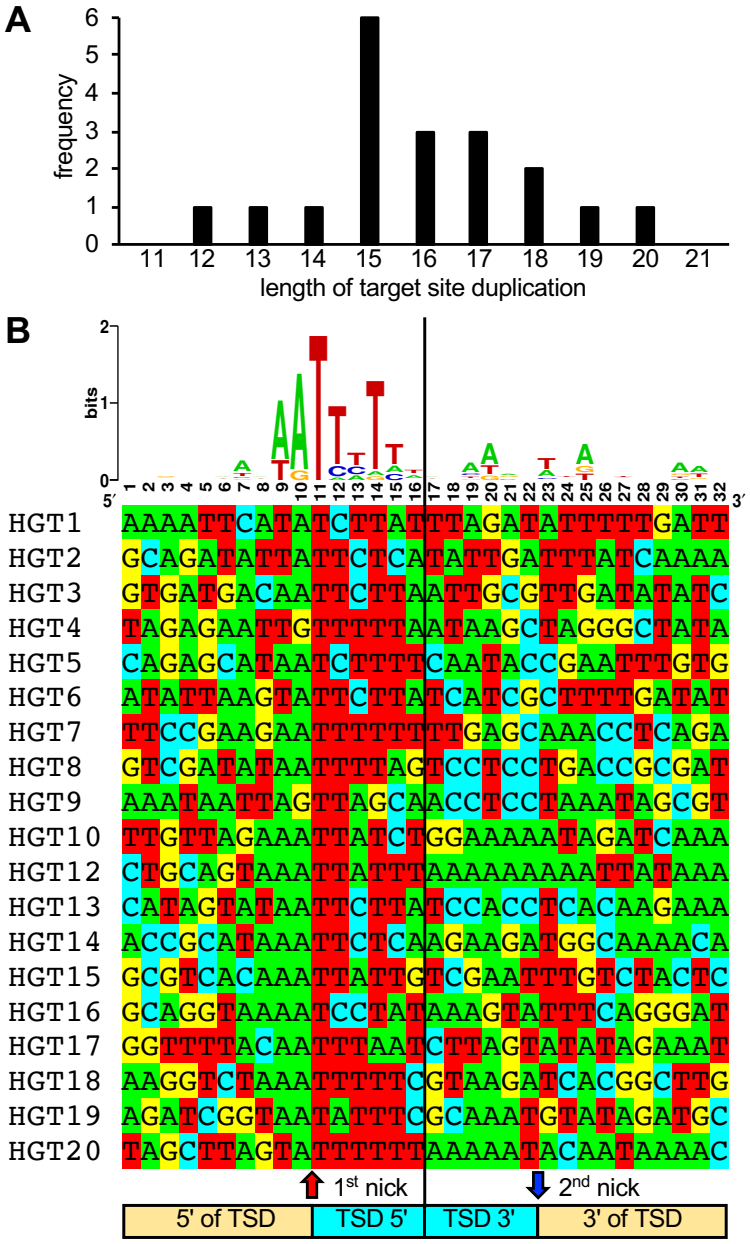


Figure 2. Length distribution and composition of target site duplications surrounding integration sites. (**A**) Length distribution of TSDs. (**B**) WebLogo of the consensus sequence and individual sequences of TSDs. Ten nucleotides 5' and 3' of the TSDs, as well as six nucleotides of the TSDs adjacent to the putative first and second LINE-1 endonuclease cleavage sites, respectively, are shown.

HGT acquisition in essential genes is facilitated by mutual complementation

To analyze whether HGT imposed any fitness costs, we compared the plaque sizes of HGT viruses in RK13 and BSC-40 cells with plaques formed by replication-competent viruses VC-2 and vP872 (Beattie et al., 1991). For most HGT viruses, plaque sizes were comparable to the replication-competent viruses in RK13 cells (Figure 3A). However, in BSC-40 cells some viruses formed

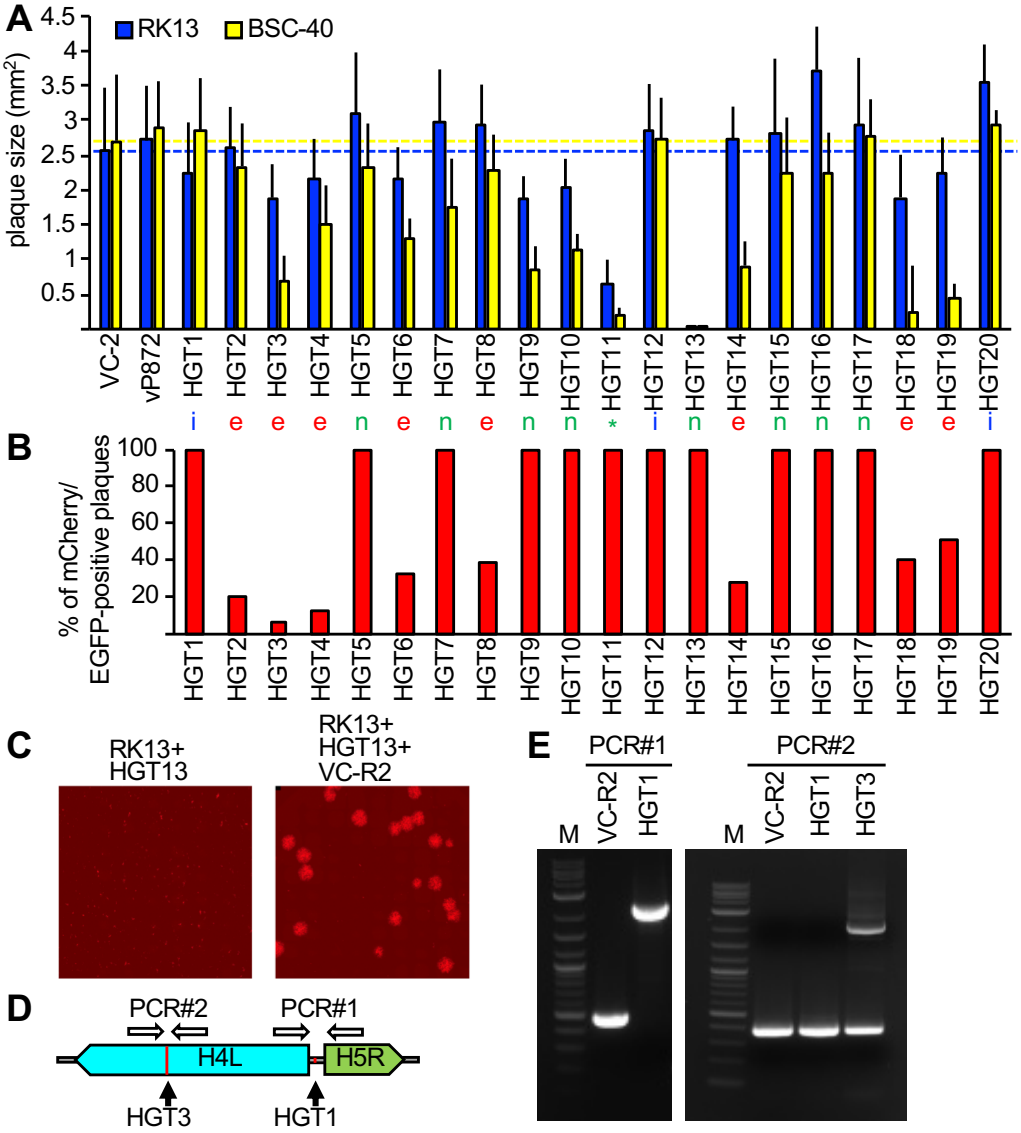


Figure 3. Complementation of viruses after mCherry-E3L integration into essential genes. (**A**) Plaque sizes formed by HGT virus isolates in RK13 and BSC-40 cells 3 days after infection. Location of integration (intergenic spaces (i), essential (e), non-essential (n), and a deletion (d)) are indicated below each bar. Dotted lines indicate plaque sizes caused by vP872. Asterisks denote significant differences between vP872 and HGT isolates (* $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$). (**B**) Percentage of foci that were mCherry-positive 3 days after infecting RK13+E3+K3 cells with the indicated HGT viruses. (**C**) Complementation of HGT13 by VC-R2. RK13 cells were infected with either HGT13 alone (left) or co-infected with HGT13 and VC-R2 (right). Foci were visualized 3 days after infection. (**D**) Location of PCR primers for amplifying integration sites in HGT1 and HGT3. (**E**) PCR amplification for integration site identified for HGT1 (PCR#1, left) and HGT3 (PCR#1, right).

217 substantially smaller plaques than in RK13 cells, e.g. HGT3, HGT14, and HGT19. HGT13
218 (insertion in F13L) formed tiny plaques in both cell lines, consistent with a previous report of
219 defective viral spread in a F13L deficient VACV (Blasco and Moss, 1991). HGT11, containing a
220 deletion stretching from B4R through B8R, also exhibited a small plaque phenotype in both cell
221 types. These findings indicate that, unsurprisingly, HGT can cause a loss-of-fitness.
222 Surprisingly, eight of the VACV genes that were disrupted by HGT were reported to be essential
223 for virus replication (Table 1). Because these insertions should presumably inactivate these
224 essential genes, it was unclear how this HGT resulted in viable viruses. One possibility could be
225 that although these viruses were plaque purified three times, the parent virus may have persisted
226 to complement the otherwise non-viable HGT virus. To test this hypothesis, we infected
227 RK13+E3+K3 cells, which are also permissive for the parental virus, with each plaque-purified
228 HGT isolate. In all these infections, we identified the expected EGFP/mCherry double-positive
229 foci. Additionally, foci that were EGFP-positive but mCherry-negative were observed in isolates
230 in which mCherry-E3L insertions occurred in essential genes, indicating the continued presence
231 of parental virus. The percentage of mCherry positive to mCherry negative foci in these viruses
232 ranged between 7% - 51% (Figure 3B, Supporting Figure 3). We hypothesized that the two
233 fluorescent phenotypes indicated mutual complementation, wherein the HGT virus supplied
234 mCherry-E3 and VC-R2 supplied the disrupted essential gene product, thus permitting replication
235 of both viruses in co-infected cells. To test this hypothesis, we co-infected RK13 cells with VC-
236 R2 and the F13L-disrupted HGT13 virus, which did not contain detectable VC-R2 but formed only
237 minute foci in either RK13 or BSC-40 cells. As predicted, in the co-infection we observed larger
238 foci in addition to the very small foci produced by HGT13 alone (Figure 3C). In an additional,
239 PCR-based test of this hypothesis, we amplified DNA spanning the integration sites in HGT1,
240 which acquired mCherry-E3L in an intergenic region, or HGT3, in which the essential gene H4L
241 was disrupted. As expected, HGT1 yielded a single 2.3 kB amplicon of the expected size for
242 mCherry-E3L integration. However, we amplified two products from HGT3: a 396 bp amplicon,
243 consistent with wild type H4L, and a 2.2 kb amplicon consistent with mCherry-E3L integration
244 into H4L (Figure 3D, E). Taken together, these assays support our hypothesis of mutual
245 complementation of these viruses.
246
247 **Recombination of complementing viruses generates replication-competent individual viruses**
248 If both viruses complemented each other efficiently, the expected ratio of HGT virus to parent to
249 virus should be close to 1:1. However, for most HGT isolates this ratio was lower, indicating that
250 virus complementation did not necessarily result in optimal replication of both viruses. In these
251 cases, there may be strong selective pressure to evolve a single replication-competent virus that is
252 not reliant on a complementing virus. To test this hypothesis, we serially passaged the
253 HGT3/parental virus population in PKR-competent RK13 cells. In three independent replicates,
254 we observed a rapid increase in double-positive fluorescent foci (Figure 4A, B), consistent with
255 an increase in HGT3 fitness. In order to directly test for a fitness increase, we infected RK13 and
256 BSC-40 cells with the passaged virus populations for 48 hours and then titered the viruses on
257 RK13+E3+K3 cells. Whereas in RK13 cells the passaged viruses showed only a modest increase
258 in replication (about 5-fold) in passages 9 and 17, an approximately 100-fold higher replication
259 was observed for BSC-40 cells infected with passages 9 and 17 as compared to passage 0 (Figure
260 4C, D).
261 To investigate possible structural changes around the disrupted H4L locus, we designed outward-
262

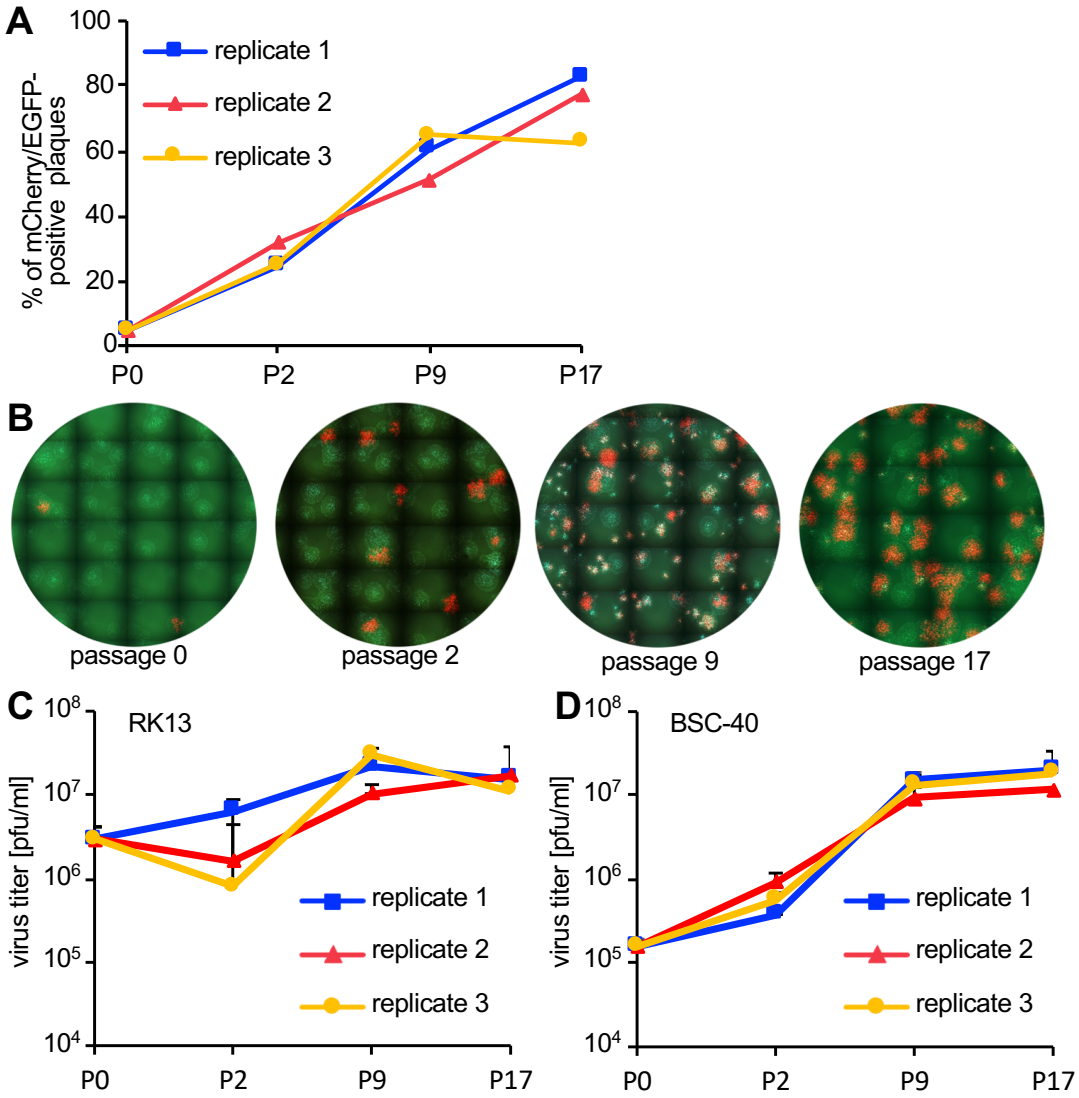


Figure 4. Fitness increase of HGT3 after serial passaging. (**A**) The percentage of mCherry-positive foci formed on RK13+E3+K3 cells by three serially passaged replicates of HGT3. Percentages were determined for the indicated passages (P). (**B**) Increase in double-positive mCherry and EGFP expressing foci of replicate 3 during serial passaging. (**C**) RK13 and (**D**) BSC-40 cells were infected with serially passaged HGT3 (MOI = 0.01) for 48 hours and viruses were titered on RK13+E3+K3 cells.

facing primers in H4L similar to an approach to detect gene amplification in poxviruses (Elde et al., 2012, Brennan et al., 2014). This PCR strategy only yields products if two or more copies of H4L are adjacent to each other (Figure 5A). We amplified products from passaged virus DNA, but not from DNA isolated from VC-R2 or the original plaque-purified virus (P0) (Figure 5B). Because of this evidence for genomic structural changes, we used PacBio long-read sequencing to more completely define these changes in the different populations. This approach demonstrated that in each HGT3 virus population, there were genomes in which an intact copy of H4L existed in close proximity to the mCherry-E3L disrupted H4L locus, although we found several different recombinant viruses with different genomic architectures (Figure 5C). Only 3 shared nucleotides are present at the the breakpoints, indicating non-homologous recombination between the parent

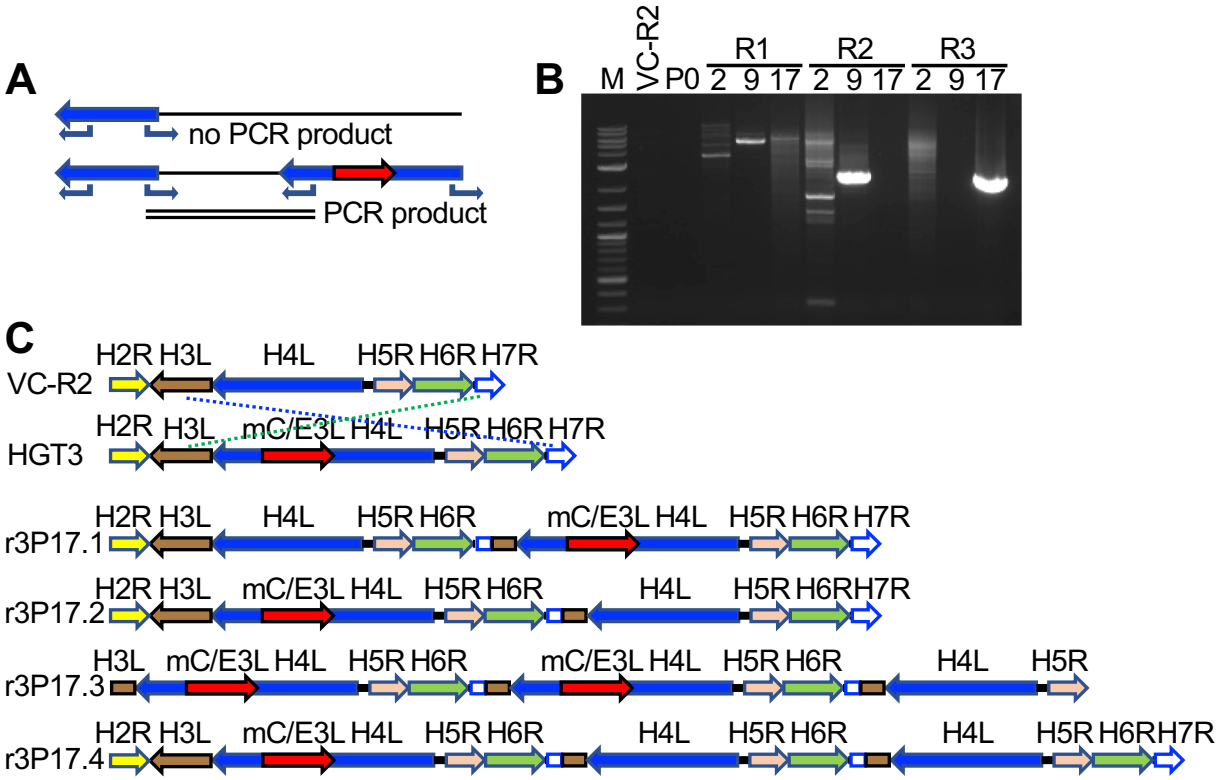


Figure 5. Recombination of complementing viruses after serial passaging. (**A**) PCR strategy to detect gene arrays with outward-facing primers. (**B**) Representative PCR products amplified with outward-facing primers for VC-R2 and three independent HGT3 passages. (**C**) Recombination of complementing viruses resulted in tandem arrays of the H4L locus as determined by long-read sequencing.

and HGT viruses (Supporting Figure 4). Taken together, these observations suggest that recombination occurred early in the population during serial passage, supporting the hypothesis that there may be selective pressure to generate a single viable virus. Alternatively, mCherry-E3L may have integrated into a pre-existing H4L duplication; however, this scenario is unlikely, because virus complementation would be unnecessary.

Signatures of naturally occurring LINE-1 mediated HGT in the GAAP encoding gene

To determine whether this mechanism is relevant in nature, we asked whether we could detect evidence of RNA-mediated HGT in any phylogenetically predicted HGT candidate genes. We selected and analyzed the phylogenetically predicted HGT gene encoding Golgi anti-apoptotic protein (GAAP) in the cowpox virus Finland_2000_MAN strain, which is also present in some other orthopoxviruses. GAAP shows approximately 76% protein identity with its mammalian homologs and therefore was likely acquired relatively recently (Gubser et al., 2007). Analyzing this gene for signatures of HGT, we identified putative remnants of a poly(A) tail (16 out of 21 bp) 3' of the gene, and a perfect 18 bp TSD flanking the gene, which contains AA/TTGT at the presumed site for the first nick during integration similar to the AA/TTTT motif identified in our screen. Moreover, adjacent to each TSD are two 59 bp homology regions, which share 83% nucleotide identity, and could be remnants of a larger duplication due to recombination, as seen in this study (Fig. 6).

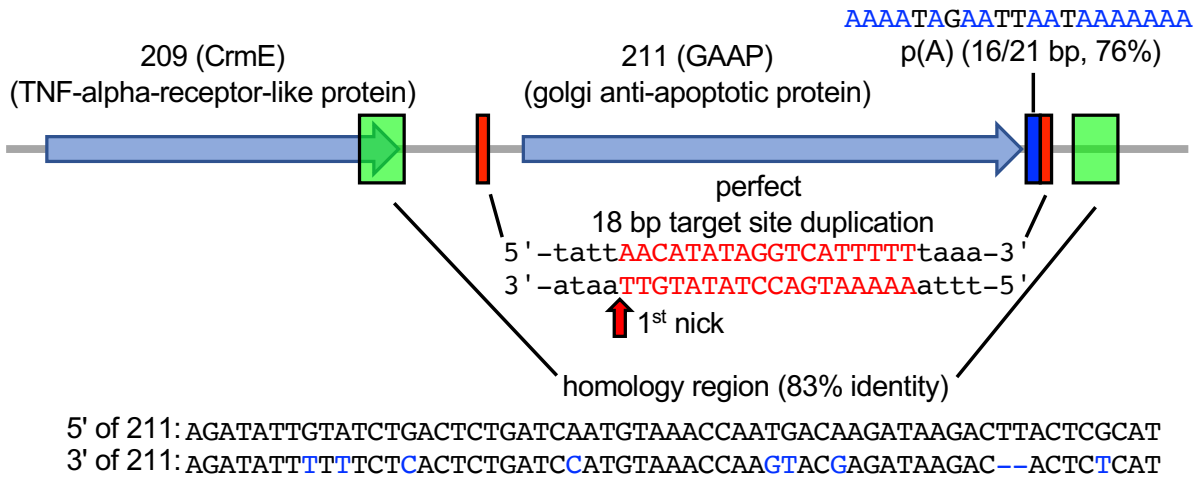


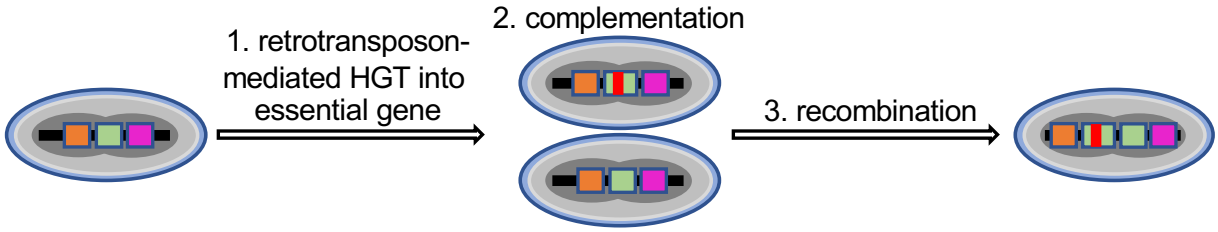
Figure 6. Signatures of retrotransposon-mediated HGT adjacent to cowpox virus gene 211. The 211 gene encodes Golgi anti-apoptotic protein (GAAP). Red boxes mark a perfect 18 bp putative target site duplication. p(A) indicates a polyadenylate stretch (16/21 bp) adjacent to the TSD. Green boxes show a duplicated 59 bp stretch with 83% sequence identity (homology region).

Discussion

In this study, we show that HGT from the host cell into poxviruses can be recapitulated in an experimental system. All HGT events showed evidence for an RNA-mediated mechanism, likely mediated by LINE-1 retrotransposons, as they contain TSD of approximately 16 bp surrounding the integration sites, as well as spliced-out introns and untemplated poly(A) tracts 3' of the integrated genes. LINE-1s are ancient and ubiquitous retrotransposons, which are found in both plants and animals. The number of active LINE-1s varies across species (Ivancevic et al., 2016). Active LINE-1s encode two proteins called ORF1p and ORF2p, which possess either RNA-binding properties or endonuclease and reverse transcriptase (RT) activity, respectively (Dombroski et al., 1991, Khazina et al., 2011, Doucet et al., 2010, Feng et al., 1996, Mathias et al., 1991). LINE-1-mediated reverse transcription is generally thought to occur in the nucleus. Because poxviruses replicate exclusively in the cytoplasm, the detected signatures of LINE-1-mediated retrotransposition after HGT reported here suggests that LINE-1 RT and integrase activities are not limited to the nucleus. The notion that LINE-1s can facilitate the transfer of host genetic material into poxviruses through an RNA-mediated process during the natural course of viral infection is further supported by the presence of a naturally-occurring SINE element in taterapox virus, which is surrounded by a 16 bp target site duplication (Piskurek and Okada, 2007). This observation shows that poxvirus evolution can be impacted by host transposable elements providing them with selective advantages and accelerating their evolution. It also defines a new function for retrotransposons, in addition to their roles in genomic and epigenetic diversification of the host (Beck et al., 2011).

Even though phylogenetically predicted HGT candidates in VACV are distributed throughout the genome (Figure 1D) (Hughes and Friedman, 2005, Bratke and McLysaght, 2008), a common notion in the field is that uptake of new genes into poxviruses would be biased towards the ends of poxvirus genomes, because the central genome region contains more essential genes and shorter intergenic spaces and thus insertions there would be less likely to result in viable progeny (Yao

339 and Evans, 2001, Lefkowitz et al., 2006). Also, poxvirus lineage-specific genes tend to be more
340 often found near the termini of poxvirus genomes (McLysaght et al., 2003). However, the
341 integration sites that we discovered in this study were distributed throughout the VACV genome,
342 which indicates that integration *per se* is not disfavored in the central genomic region. The
343 concentration of insertions around the H4L locus might indicate that some regions of the genome
344 are indeed hotspots for integration. However, the number of integration events detected in this
345 study precludes a definitive answer, and will require further analysis of more HGT events. It is
346 also possible that the insertion site is influenced by the cell type. Because complementation of
347 replication-deficient viruses was more efficient in RK13 as opposed to BSC-40 cells, it appears
348 that propagation of viruses with insertions in essential genes might be favored in the former cells.
349 In a companion paper, Fixsen et al. used a similar strategy to detect HGT by expressing the other
350 VACV PKR inhibitor K3 in RK13 cells, which were also used in our study, and selecting for virus
351 replication in Syrian hamster BHK cells. They found a lower proportion of HGT integrations in
352 the central genome region (Fixsen et al., 2020). The reasons for the different integration patterns
353 could be due to the different cells used for selection and might reflect differences in the relative
354 importance of viral genes in these cells or different, cell line-dependent complementation
355 efficiencies.
356 Our results show that HGT into essential poxvirus genes can result in viable progeny virus in a
357 process that we call cascade evolution. This process sequentially combines HGT, virus
358 complementation and recombination, three classic mechanisms of virus evolution (Figure 7). Our
359 experimental system relies on the complete failure for viruses to replicate in the absence of a PKR
360 inhibitor, creating a strong selective pressure. In eight instances in this study we also identified
361 integrations into essential genes, creating a similarly strong selective pressure against virus
362 replication. In these cases, complementation followed by recombination was sufficient to
363 overcome this selective pressure *in vitro*. The homology regions surrounding the cowpox virus
364 GAAP gene (Fig. 6) suggest a similar process can occur during natural infection as well. In
365 instances in which the selective pressure to maintain the horizontally transferred gene is weaker or
366 the inactivated gene is non-essential but still contributes to fitness, we hypothesize that similar
367 molecular events might occur. This balance between the negative effects of genome disruption and
368 improved fitness might also display some degree of cell- and species-specificity.



369
370 **Figure 7. Cascade virus evolution after horizontal gene transfer into an essential gene.** After
371 retrotransposon-mediated horizontal gene integration into an essential locus (1) the virus initially
372 depends on a helper virus to complement the essential gene (2), which is followed by
373 recombination and results in a single virus with increased fitness.
374
375 Our PacBio analyses identified tandem copies of the integrated gene and intact gene, in both
376 orientations relative to each other. However, at the population level the genomic structure
377 surrounding this locus was complex. For example, we identified loci including both intact H4L

378 and the horizontally acquired mCherry-E3L in which multiple copies of each gene were found in
379 *cis*. Furthermore, in some genomes we identified three or more copies of some genes adjacent to
380 the integration site. In addition to the immediate impact of generating a replication-competent virus
381 through recombination, these variant products of recombination might provide the raw material
382 for subsequent sub- or neofunctionalization of the additional gene copies (Bayer et al., 2018).
383 The majority of the HGT viruses that required complementation with the parental virus formed
384 larger plaques in RK13 cells as compared to BSC-40 cells. This difference may indicate that the
385 complementation is more efficient in RK13 cells. In line with this explanation, we observed that
386 the parent HGT3 population replicated poorly in BSC-40 cells, and the passaged population
387 replicated about 100-fold more efficiently, whereas in RK13 cells, the parent HGT3 population
388 replicated better and only gained a modest increase in replication efficiency during passaging. It
389 is striking that for most observed cases of virus complementation, the ratio of parent to HGT virus
390 was skewed to the parent virus, in some cases making up more than 90% of the population (Figure
391 3B). This observation suggests that there may be competition for cellular resources or viral gene
392 products between the co-infecting viruses. While mCherry-E3 localizes to the cytoplasm and likely
393 antagonizes PKR equally well for each co-infecting virus, the transcription and translation of many
394 other viral gene products are tightly associated with viral factories derived from a single genome
395 (Katsafanas and Moss, 2007). Conceptually, this competition may occur when the complementing
396 molecule is required to diffuse or to be transported into another viral factory to exert its effect, or
397 if the viral gene product is not delivered until virus factories fuse late in the replication cycle
398 (Paszkowski et al., 2016).
399 An interesting observation was that in one of the HGT isolates, mCherry-E3L was fused to a
400 polyadenylated fragment of the cellular U2 snRNA. This genetic structure may have been
401 generated by the fusion of mCherry-E3L RNA with U2 snRNA prior to integration. This
402 observation suggests that other cellular genes could piggyback during HGT, which suggests in rare
403 instances multiple genes may be integrated into a single genome. Of note, snRNAs encoding DNA
404 including U1, U2, U5 and U6 can be found in host genomes fused to LINE-1 as a result of RNA
405 ligation (Buzdin et al., 2003, Garcia-Perez et al., 2007, Doucet et al., 2015, Moldovan et al., 2019).
406 However, it is unclear what, if any, impact the potential to integrate multiple genes would have on
407 viral evolution.
408 The calculated approximate transfer rate of E3L-mCherry into VACV in this experimental setting
409 of 1 in 23 million viable virions, is probably an underestimation, as integrations in essential loci
410 have detrimental effects on virus replication unless complementing viruses are present. It should
411 be noted that we used a system with a strong synthetic poxvirus promoter and a strong VACV PKR
412 inhibitor in order to detect HGT. Retention of a transferred host gene that has not been optimized
413 is expected to happen with a far lower frequency. However, we selectively looked for the uptake
414 of only one of several thousand genes that are expressed by the host cells. Thus, HGT of host
415 genes, including transposable elements such as SINEs and LINEs, into poxviruses can be expected
416 to occur at a high frequency, but in most cases the inserted gene will not be maintained because of
417 negative or neutral effects on virus replication.
418 This work, and the companion study by Fixsen et al. (Fixsen et al., 2020), provide the first
419 experimental evidence of horizontal gene transfer into poxviruses. The experimental system
420 described here can be extended to study the evolution of transgenes after HGT. In addition, it can
421 be adapted to elucidate the mechanisms of HGT in other groups of viruses for which HGT is
422 believed to play an important evolutionary role. While we only detected RNA-mediated HGT
423 facilitated by LINE-1, it is possible that in other viruses or cell types additional RNA or DNA

transposons, or recombination might be important. Similarly, other facilitators, such as endogenous or exogenous retroviruses may also promote HGT and increase virus fitness. Taken together, our work demonstrates that cascade evolution is a powerful and rapid mechanism to integrate new genetic material in poxviral genomes. It also demonstrates that HGT in poxviruses is traceable in experimental settings, and that this mechanism may be relevant to other viruses, which appropriate host genes for their own ends in the ongoing host-virus arms race.

Materials and Methods

Cell lines

RK13 cells (ATCC CCL-37) and BSC-40 (ATCC CRL-2761) cell lines were kindly provided by Dr. Bernard Moss and Dr. Adam Geballe, respectively. RK13 cells expressing E3 and K3 (designated RK13+E3+K3) were previously described (Rahman et al., 2013). All cell lines were maintained in Dulbecco's Modified Essential Medium (DMEM, Life Technologies) supplemented with 5% fetal bovine serum (FBS, Fisher) and 25µg/ml gentamycin (Quality Biologicals) at 37˚C and 5%CO_2. RK13+E3+K3 cells were supplemented with 500µg/ml geneticin and 300µg/ml zeocin (Life Technologies). Cell lines were seeded in 6-well (at 5 X 10^5 cells/well) or 12-well (at 2.5 X 10^5 cells/well) plates 24 hours before infections. Cell lines used in this study were negative for mycoplasma contamination as determined with Lookout Mycoplasma PCR Detection Kit (Millipore Sigma). During the course of this study, the RK13 cells we used were confirmed to be of European rabbit (*Oryctolagus cuniculus*) origin by PacBio sequencing (ArrayExpress accession: E-MTAB-9682). PKR expressed in BSC-40 was amplified from cDNA and sequenced, which confirmed that the cells are of African green monkey (*Chlorocebus aethiops*) origin.
RK13 cells expressing mCherry-E3 (RK13-mCherry-E3L) were generated by stable transfection of *ApaL*I-linearized pmCherry-E3L using GenJet (SignaGen) according to the manufacturer's protocol. The pmCherry-E3L vector contains, in 5' to 3' order, the SV40 promoter, rabbit β-globin intron, a synthetic poxvirus early/late promoter, a mCherry-E3L fusion gene, cloned into the backbone of pEGFP-N1 (Clontech). During this process, the CMV promoter and EGFP were removed from pEGFP-N1. Transfected cells were selected with 500 µg/ml geneticin (G418, Invitrogen) for two weeks. Individual clones were isolated by seeding cells at a low density (0.3-1 cell/ well) in 96-well plates. Wells containing a single cell, verified by fluorescent microscopy, were selected to establish clonal cell lines. Genomic DNA was prepared to verify the structure of the integrated cassette by PCR with primer JR253-intron-4F (5'-TTC CAG AAG TAG TGA AGA GGC TT-3') and JR254-E3L-5R (5'-AGC AAG TAA AAC CTC TAC AAA TG-3'). All subsequent experiments were carried out with one RK13+mCherry-E3L cell clone that showed robust mCherry-E3 expression (Supplementary figure 1).

Viruses

Vaccinia virus VC-2 (Copenhagen strain) and vP872 (ΔK3L) (Beattie et al., 1991) were kindly provided by Dr. Bertram Jacobs and propagated in RK13 cells. VC-R2, which lacks both E3L and K3L (Brennan et al., 2014), was propagated in RK13+E3+K3 cells. VC-2, vP872, and VC-R2 were purified by zonal sucrose gradient centrifugation as previously described (Cotter et al., 2017). Virus clones HGT1 to HGT20 were plaque purified three times in RK13 cells with carboxymethyl cellulose (CMC, 1% final concentration in DMEM+5%FBS) overlay. Virus clones HGT1 to HGT7 were purified by zonal sucrose gradient centrifugation. For HGT8 to HGT20, crude cell

470 lysates were used as stocks. All viruses were titered in RK13+E3+K3 cells on 12-well plates. Virus
471 samples were 10-fold serially diluted and cells were infected in duplicate. One hour after infection,
472 the infecting medium was replaced with CMC and incubated for 3 days at 37˚C in 5% CO_2. The
473 CMC overlays were then aspirated, and plaques were identified by staining the monolayers with
474 0.1% crystal violet in 20% methanol.
475 Serial passaging of HGT3 was performed as described (Elde et al., 2012, Brennan et al., 2014).
476 Briefly, confluent RK13 cells in 6-well plates were infected with plaque-purified HGT3 virus
477 (MOI = 0.1) in triplicate. 48 hours after infection (hpi), cell lysates were collected, freeze-thawed
478 three times, sonicated, and titered on RK13+E3+K3 cells before serially infecting new RK13 cells.
479 Genomic DNA from passaged viruses was isolated from infected RK13 cells (MOI = 1.0) as
480 previously described (Esposito et al., 1981). RK13 and BSC-40 cells were infected with passaged
481 virus populations in duplicate at MOI = 0.01 for 48 hours and viruses were titered on
482 RK13+E3+K3 cells in duplicate.
483
484 **Screening for viruses that captured mCherry-E3L**
485 RK13-mCherry-E3L cells were infected with VC-R2 (MOI = 0.01) for three days. Viruses were
486 titered on RK13+E3+K3 cells. Subsequently, viruses were screened on confluent monolayers of
487 RK13 cells in 6-well plates. Cells were infected at MOI = 0.2 to screen for fluorescence to avoid
488 cell death observed at higher MOIs. For most isolates, plaques were identified after 3 days but
489 some took as long as 5 days. Initially, plates were screened for double-positive mCherry and EGFP
490 expressing plaques using an inverted microscope (Leica DMi8 automated, total magnification
491 X6.25) using the FITC filter (460–500 nm) to visualize EGFP and the Texas Red filter (540 – 580
492 nm) to visualize mCherry. In order to increase throughput, screening for double-positive plaques
493 was then performed with an automated microscope (EVOS FL Auto 2, Thermo Fisher Scientific)
494 using comparable filters. Double-positive virus plaques were picked using a sterile pipette tip.
495 Harvested plaques were subjected to three rounds of freeze-thaw followed by sonication (2 times
496 30s). These viruses were then serially diluted to infect RK13 cells for a total of three rounds of
497 plaque purification. After the third round of plaque purification, viruses were amplified in RK13
498 cells and titered on RK13+E3+K3 cells. Viral genomic DNA was isolated from infected RK13
499 cells according to a published protocol (Seto et al., 1987).
500
501 **Detection of mCherry-E3L integration sites in the virus genome**
502 The mCherry-E3L integration sites in the purified HGT virus genomes were located by inverse
503 PCR amplification and by PacBio sequencing. For inverse PCR, we used a *Xba*I restriction site in
504 between mCherry-E3L. VC-2 contains 123 *Xba*I sites. Viral DNA from HGT isolates was digested
505 with *Xba*I, and then ligated by T4 DNA ligase (NEB) to make circular DNA. After ligation, E3L-
506 specific, outward-facing primers: HGT-E3L-1F (5'-TGG CAG TAG ATA AAC TTC TTG GTT
507 ACG-3') and HGT-E3L-1R (5'-AGC CTC ACA CAC AAT CTC TGC G-3') were used to amplify
508 DNA circles containing mCherry-E3L and neighboring genomic viral DNA. These amplicons
509 were gel purified and Sanger sequenced to define the approximate mCherry-E3L integration sites,
510 because the long poly(A) tails precluded identification of the virus sequences 3' of the poly(A)
511 tails. To define integration sites and target site duplications, primers (Supplementary table 2) from
512 the VACV genome that are adjacent to the integration sites were designed to amplify whole
513 integrated captured sequences, which were then Sanger sequenced. Because for HGT5 and
514 HGT19, PCR amplification of the whole insert did not work, two separate PCR amplification with
515 primers located in E3L and the adjacent vaccinia virus genes were performed and sequenced.

516 Insertion start sites and polyadenylation sites of transgenes in HGT viruses are shown in supporting
517 Figure 5.
518 To determine if mCherry-E3L integrated into both sites of the ITR in the C19L/B25R locus in
519 HGT15, PCR with DNA of HGT15 and VC-R2, as control, was performed with primers that
520 combined primer JR179-C19L-2F (5'-CGA GGA CTA TGT TTG GTA TAC TG-3') located in
521 the ITR, towards the termini of the integration site and primers JR180-C13L-2R(5'-CTG GAC
522 TAT CCA CAC CTG-3) for the "left" region of the VACV genome, and JR181-B19R-1F (5'-
523 CTA GTA GAG GCG GTA TCA C-3'), for the "right" region of the VACV genome outside of
524 the ITR. Both primer combinations showed larger bands when DNA HGT15 DNA was used as
525 template, in comparison to VC-R2 DNA, which is indicative of mCherry-E3L in both ITR copies
526 (Supporting Figure 6).
527

528 **Measurement of plaque sizes**

529 Confluent 6-well plates of either RK13 or BSC-40 cells were infected with 100 PFU of either each
530 HGT isolate, VC-R2, or vP872 in triplicate. After one hour, cells were overlaid with CMC and
531 incubated at 37℃ in 5% CO2 for 3 days followed by crystal violet staining. The stained plates and
532 an adjacent ruler were imaged with the EVOS FL 2 microscope (Thermo Fisher Scientific) using
533 a bright-field filter. Plaque sizes were calculated using Adobe Photoshop CC 2019. To calculate
534 this, a global-scale was set by measuring the average number pixels in a 1 mm-distance from the
535 imaged rulers (133 pixels = 1mm, SD = 0.98 pixels) using the Adobe Photoshop CC 2019 ruler
536 tool. The Photoshop quick selection tool was used to mark the perimeter of individual plaques
537 followed by area measurement from the analysis tab. Between 30 and 387 plaques were measured
538 for each virus. Kruskal-Wallis one-way ANOVA followed by Dunn's test of multiple comparisons
539 were used to compare the mean plaque area of HGT samples with vP872 (alpha =.05) by GraphPad
540 Prism (version 8.3.0). Graphs are presented as mean ± SD and adjusted p-values are shown.
541

542 **Identifying virus complementation**

543 Confluent 6-well plates of RK13+E3+K3 cells were infected with 100 PFU of HGT1 through
544 HGT20, and the passaged HGT3 from passages 2, 9, and 17. The percentage of mCherry and EGFP
545 double-positive plaques were determined three days after infection using the automated EVOS FL
546 Auto 2 microscope (Thermo Fisher Scientific). Images were taken at 4 X 0.13 NA magnification
547 by using filter channels Texas red (540–580 nm) for mCherry and EGFP (465–495 nm) to visualize
548 EGFP. Single-channel autofocus was used with phase-contrast using the first field to autofocus all
549 other fields. 100% of well areas were photographed, stitched and tiled by using the EVOS FL Auto
550 2 software and analyzed to quantify the number of single-positive EGFP plaques and double-
551 positive EGFP and mCherry expressing plaques. For each virus sample more than 86 plaques were
552 analyzed.
553 For the VC-R2 and HGT13 complementation assay, confluent 6-well plates of RK13 cells were
554 co-infected in duplicate with each virus (MOI = 0.05 for each virus), or with HGT13 alone (MOI
555 = 0.1). Images for mCherry fluorescence were taken 3 days after infection with EVOS FL Auto 2.
556

557 **Detection of tandem duplications of the H4L locus using PCR**

558 Genomic DNA from HGT3 passages 0, 2, 9 and 17 as well as VC-R2 , were used as templates for
559 PCR using outward-facing primers: JR81-H4L-2F (5'-GTC TAG TAG ATA TGC TTT TAT TTT
560 TG-3') and JR80-H4L-2R (5'- CGA AAA TAT AAC TCG TAT TAA AGA G-3') or JR42-H4L-
561 3F (5'- CAC GGA GAT GGC GTA TTT AAG AG -3') and JR88-H4L-3R (5'- GAG CTA ACG

562 TGT GAC GAA G-3'). PCR amplified products were cloned into the pCR2.1-TOPO-TA
563 (Invitrogen) vector for amplification and Sanger sequencing using M13F (-21) and M13R primers
564 (UC Davis sequencing facility).
565
566 **Library preparation, PacBio CCS sequencing and data analysis**
567 HiFi SMRTbell library construction and sequencing were performed using Sequel II System 2.0
568 with P2/C2 (polymerase 2.0 and chemistry 2.0) chemistry at UC Davis DNA Technologies Core
569 according to the manufacturer's instructions (Pacific Biosciences). Before library preparation, the
570 quality of genomic DNA was evaluated by pulsed-field gel electrophoresis (Sage Science Pippin
571 Pulse) and genomic DNA concentration was quantified using a Qubit 3.0 Fluorometer (Life
572 Technologies Q33216). Briefly, 10 µg of each viral genomic DNA was used for HiFi SMRT-bell
573 library preparation by using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences 100-
574 938-900). The genomic DNA was sheared to ~15-20 kb size by Megaruptor (Diagenode
575 B06010001). The sheared DNA samples were then concentrated and purified by AMPure PB
576 Beads (Pacific Biosciences 100-265-900) by adding 0.45 X volume of AMPure PB magnetic beads
577 to each sheared DNA samples and subsequent 80% ethanol precipitation and elution with 100 µl
578 elution buffer. Following genomic DNA shearing and purification, each genomic DNA sample
579 was subjected to single-strand DNA (ssDNA) overhang removal and DNA damage repair. In short,
580 ssDNA overhangs were removed using ssDNA overhang removal reaction mix (DNA prep buffer,
581 NAD, DNA prep additive and DNA prep enzyme) and incubated at 37˚C for 15 mins. The DNA
582 fragments of each sample were then repaired by using a DNA damage repair mix and incubated at
583 37˚C for 30 mins before proceeding to End repair/A-tailing by mixing with 1 X End prep mix and
584 incubated at 20˚C for 10 mins followed by 30 mins incubation at 65˚C. After that, different
585 barcoded adapters were ligated to each fragmented viral DNA sample by mixing with adapter
586 ligation reaction mix (Overhang adapter v3, ligation mix, ligation additive, and ligation enhancer)
587 and incubated at 20˚C for an hour followed by incubation at 65˚C for 10 mins to inactivate the
588 ligase. After ligation reaction, each adapter-ligated SMRTbell library was treated with
589 exonucleases to remove damaged or un-ligated DNA fragments for one hour at 37˚C followed by
590 a second purification step with 0.45 X AMPure® PB Beads. After purification, the libraries were
591 multiplexed into 2 library pools for size selection: 6-plex pool 1 and 10-plex pool 2. These pooled
592 SMRTbell libraries were then purified by mixing with 0.45 X volume of AMPure PB beads
593 followed by 80 % ethanol wash and elution with 31 µl EB for size selection. Size selection was
594 done using the BluePippin Size-Selection System (Sage Science BLU0001) to remove <6kb
595 SMRTbell templates and eluted the size selected libraries into two size fractions: 9-13 kb and >15
596 kb per pooled library. The only fraction with >15 kb SMRTbell libraries was purified with 0.50X
597 AMPure® PB Beads and used for primer annealing and polymerase binding. Sequencing Primer
598 v2 (PN 101-847-900) was annealed to each size selected SMRTbell library fraction of >15 kb
599 length at primer: template ratio of 20:1 by denaturation step at 80 °C for 150 mins followed by
600 slow cooling at 0.1 °C/s to 25 °C. Prior to sequencing, the polymerase-template complex was
601 bound to the P2 enzyme with 10:1 polymerase to the SMRTbell template ratio for 4 hours. The
602 polymerase bound SMRTbell libraries were then sequenced by placing libraries in an 8M SMRT
603 cell at a sequencing concentration of 63 pM and 30-hour movie run time in a Sequel II System 2.0
604 machine.
605 For data analysis, PacBio CCS reads were first demultiplexed using lima from SMRT Link
606 8.0 command-line tools. For all samples, the location of gene H4L and mCherry-E3L sequences
607 were identified in each CCS reads by using BLAST (v2.9+) (Camacho et al., 2009). The search

608 results from BLAST were processed using custom scripts to identify the different modes of
609 mCherry-E3L insertions, as well as the alterations of gene structures of H4L. Representative
610 sequences of CCS reads were extracted to showcase selected scenarios. Those sequences were
611 further analyzed using Seqbuilder, Seqman Pro, Editseq of DNASTAR software package
612 (DNASTAR, Madison, WI). A logo plot from 19 target site duplications (first 6 nucleotides 5' and
613 3' of the TSD) and 10 nucleotides of adjacent sequences was created with WebLogo (Crooks et
614 al., 2004). PacBio CCS reads were submitted to ArrayExpress (accession: E-MTAB-9682).
615
616
617 **Acknowledgments**
618
619 This work was supported by grant AI146915 (to S.R.) from the National Institute of Allergy and
620 Infectious Diseases, National Institutes of Health. We thank Dr. Adam Geballe, Dr. Bertram
621 Jacobs and Dr. Bernard Moss for cell lines and viruses, and current and former Rothenburg lab
622 members for helpful discussions.
623
624

625 # References
626

627 BAYER, A., BRENNAN, G. & GEBALLE, A. P. 2018. Adaptation by copy number variation in monopartite
628 viruses. *Curr Opin Virol,* 33**,** 7-12.
629 BEATTIE, E., TARTAGLIA, J. & PAOLETTI, E. 1991. Vaccinia virus-encoded eIF-2 alpha homolog abrogates
630 the antiviral effect of interferon. *Virology,* 183**,** 419-22.
631 BECK, C. R., GARCIA-PEREZ, J. L., BADGE, R. M. & MORAN, J. V. 2011. LINE-1 elements in structural variation
632 and disease. *Annu Rev Genomics Hum Genet,* 12**,** 187-215.
633 BLASCO, R. & MOSS, B. 1991. Extracellular vaccinia virus formation and cell-to-cell virus transmission are
634 prevented by deletion of the gene encoding the 37,000-Dalton outer envelope protein. *J Virol,* 65**,**
635 5910-20.
636 BRATKE, K. A. & MCLYSAGHT, A. 2008. Identification of multiple independent horizontal gene transfers
637 into poxviruses using a comparative genomics approach. *BMC Evol Biol,* 8**,** 67.
638 BRATKE, K. A., MCLYSAGHT, A. & ROTHENBURG, S. 2013. A survey of host range genes in poxvirus
639 genomes. *Infect Genet Evol,* 14**,** 406-25.
640 BRENNAN, G., KITZMAN, J. O., ROTHENBURG, S., SHENDURE, J. & GEBALLE, A. P. 2014. Adaptive gene
641 amplification as an intermediate step in the expansion of virus host range. *PLoS Pathog,* 10**,**
642 e1004002.
643 BROYLES, S. S. & FESLER, B. S. 1990. Vaccinia virus gene encoding a component of the viral early
644 transcription factor. *J Virol,* 64**,** 1523-9.
645 BURLES, K., IRWIN, C. R., BURTON, R. L., SCHRIEWER, J., EVANS, D. H., BULLER, R. M. & BARRY, M. 2014.
646 Initial characterization of vaccinia virus B4 suggests a role in virus spread. *Virology,* 456-457**,** 108-
647 20.
648 BUZDIN, A., GOGVADZE, E., KOVALSKAYA, E., VOLCHKOV, P., USTYUGOVA, S., ILLARIONOVA, A., FUSHAN,
649 A., VINOGRADOVA, T. & SVERDLOV, E. 2003. The human genome contains many types of chimeric
650 retrogenes generated through in vivo RNA recombination. *Nucleic Acids Res,* 31**,** 4385-90.
651 CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L.
652 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.
653 CHANG, H. W., WATSON, J. C. & JACOBS, B. L. 1992. The E3L gene of vaccinia virus encodes an inhibitor of
654 the interferon-induced, double-stranded RNA-dependent protein kinase. *Proc Natl Acad Sci U S*
655 *A,* 89**,** 4825-9.
656 CHANG, H. W., YANG, C. H., LUO, Y. C., SU, B. G., CHENG, H. Y., TUNG, S. Y., CARILLO, K. J. D., LIAO, Y. T.,
657 TZOU, D. M., WANG, H. C. & CHANG, W. 2019. Vaccinia viral A26 protein is a fusion suppressor of
658 mature virus and triggers membrane fusion through conformational change at low pH. *PLoS*
659 *Pathog,* 15**,** e1007826.
660 CHILD, S. J., PALUMBO, G. J., BULLER, R. M. & HRUBY, D. E. 1990. Insertional inactivation of the large
661 subunit of ribonucleotide reductase encoded by vaccinia virus is associated with reduced
662 virulence in vivo. *Virology,* 174**,** 625-9.
663 COTTER, C. A., EARL, P. L., WYATT, L. S. & MOSS, B. 2017. Preparation of Cell Cultures and Vaccinia Virus
664 Stocks. *Curr Protoc Mol Biol,* 117**,** 16 16 1-16 16 18.
665 CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator.
666 *Genome Res,* 14**,** 1188-90.
667 DAR, A. C. & SICHERI, F. 2002. X-ray crystal structure and functional analysis of vaccinia virus K3L reveals
668 molecular determinants for PKR subversion and substrate recognition. *Mol Cell,* 10**,** 295-305.
669 DAVIES, M. V., CHANG, H. W., JACOBS, B. L. & KAUFMAN, R. J. 1993. The E3L and K3L vaccinia virus gene
670 products stimulate translation through inhibition of the double-stranded RNA-dependent protein
671 kinase by different mechanisms. *J Virol,* 67**,** 1688-92.

672 DOMBROSKI, B. A., MATHIAS, S. L., NANTHAKUMAR, E., SCOTT, A. F. & KAZAZIAN, H. H., JR. 1991. Isolation
673 of an active human transposable element. *Science,* 254**,** 1805-8.
674 DOUCET, A. J., DROC, G., SIOL, O., AUDOUX, J. & GILBERT, N. 2015. U6 snRNA Pseudogenes: Markers of
675 Retrotransposition Dynamics in Mammals. *Mol Biol Evol,* 32**,** 1815-32.
676 DOUCET, A. J., HULME, A. E., SAHINOVIC, E., KULPA, D. A., MOLDOVAN, J. B., KOPERA, H. C., ATHANIKAR,
677 J. N., HASNAOUI, M., BUCHETON, A., MORAN, J. V. & GILBERT, N. 2010. Characterization of LINE-
678 1 ribonucleoprotein particles. *PLoS Genet,* 6.
679 EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P., BETTMAN,
680 B., BIBILLO, A., BJORNSON, K., CHAUDHURI, B., CHRISTIANS, F., CICERO, R., CLARK, S., DALAL, R.,
681 DEWINTER, A., DIXON, J., FOQUET, M., GAERTNER, A., HARDENBOL, P., HEINER, C., HESTER, K.,
682 HOLDEN, D., KEARNS, G., KONG, X., KUSE, R., LACROIX, Y., LIN, S., LUNDQUIST, P., MA, C., MARKS,
683 P., MAXHAM, M., MURPHY, D., PARK, I., PHAM, T., PHILLIPS, M., ROY, J., SEBRA, R., SHEN, G.,
684 SORENSON, J., TOMANEY, A., TRAVERS, K., TRULSON, M., VIECELI, J., WEGENER, J., WU, D., YANG,
685 A., ZACCARIN, D., ZHAO, P., ZHONG, F., KORLACH, J. & TURNER, S. 2009. Real-time DNA
686 sequencing from single polymerase molecules. *Science,* 323**,** 133-8.
687 ELDE, N. C., CHILD, S. J., EICKBUSH, M. T., KITZMAN, J. O., ROGERS, K. S., SHENDURE, J., GEBALLE, A. P. &
688 MALIK, H. S. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral
689 defenses. *Cell,* 150**,** 831-41.
690 ESNAULT, C., MAESTRE, J. & HEIDMANN, T. 2000. Human LINE retrotransposons generate processed
691 pseudogenes. *Nat Genet,* 24**,** 363-7.
692 ESPOSITO, J., CONDIT, R. & OBIJESKI, J. 1981. The preparation of orthopoxvirus DNA. *J Virol Methods,* 2**,**
693 175-9.
694 FENG, Q., MORAN, J. V., KAZAZIAN, H. H., JR. & BOEKE, J. D. 1996. Human L1 retrotransposon encodes a
695 conserved endonuclease required for retrotransposition. *Cell,* 87**,** 905-16.
696 FIXSEN, S. M., CONE, K. R., GOLDSTEIN, S. A., SASANI, T. A., QUINLAN, A. R., ROTHENBURG, S. & ELDE, N.
697 C. 2020. Poxviruses capture host genes by LINE-1 retrotransposition. *bioRxiv,* 2020.10.26.356402.
698 GAMMON, D. B., GOWRISHANKAR, B., DURAFFOUR, S., ANDREI, G., UPTON, C. & EVANS, D. H. 2010.
699 Vaccinia virus-encoded ribonucleotide reductase subunits are differentially required for
700 replication and pathogenesis. *PLoS Pathog,* 6**,** e1000984.
701 GARCIA-PEREZ, J. L., DOUCET, A. J., BUCHETON, A., MORAN, J. V. & GILBERT, N. 2007. Distinct mechanisms
702 for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res,*
703 17**,** 602-11.
704 GILBERT, C. & CORDAUX, R. 2017. Viruses as vectors of horizontal transfer of genetic material in
705 eukaryotes. *Curr Opin Virol,* 25**,** 16-22.
706 GILBERT, N., LUTZ-PRIGGE, S. & MORAN, J. V. 2002. Genomic deletions created upon LINE-1
707 retrotransposition. *Cell,* 110**,** 315-25.
708 GROSS, C. H. & SHUMAN, S. 1996. Vaccinia virions lacking the RNA helicase nucleoside triphosphate
709 phosphohydrolase II are defective in early transcription. *J Virol,* 70**,** 8549-57.
710 GUBSER, C., BERGAMASCHI, D., HOLLINSHEAD, M., LU, X., VAN KUPPEVELD, F. J. & SMITH, G. L. 2007. A
711 new inhibitor of apoptosis from vaccinia virus and eukaryotes. *PLoS Pathog,* 3**,** e17.
712 HALLER, S. L., PENG, C., MCFADDEN, G. & ROTHENBURG, S. 2014. Poxviruses and the evolution of host
713 range and virulence. *Infect Genet Evol,* 21**,** 15-40.
714 HAND, E. S., HALLER, S. L., PENG, C., ROTHENBURG, S. & HERSPERGER, A. R. 2015. Ectopic expression of
715 vaccinia virus E3 and K3 cannot rescue ectromelia virus replication in rabbit RK13 cells. *PLoS One,*
716 10**,** e0119189.
717 HASSETT, D. E., LEWIS, J. I., XING, X., DELANGE, L. & CONDIT, R. C. 1997. Analysis of a temperature-
718 sensitive vaccinia virus mutant in the viral mRNA capping enzyme isolated by clustered charge-to-
719 alanine mutagenesis and transient dominant selection. *Virology,* 238**,** 391-409.

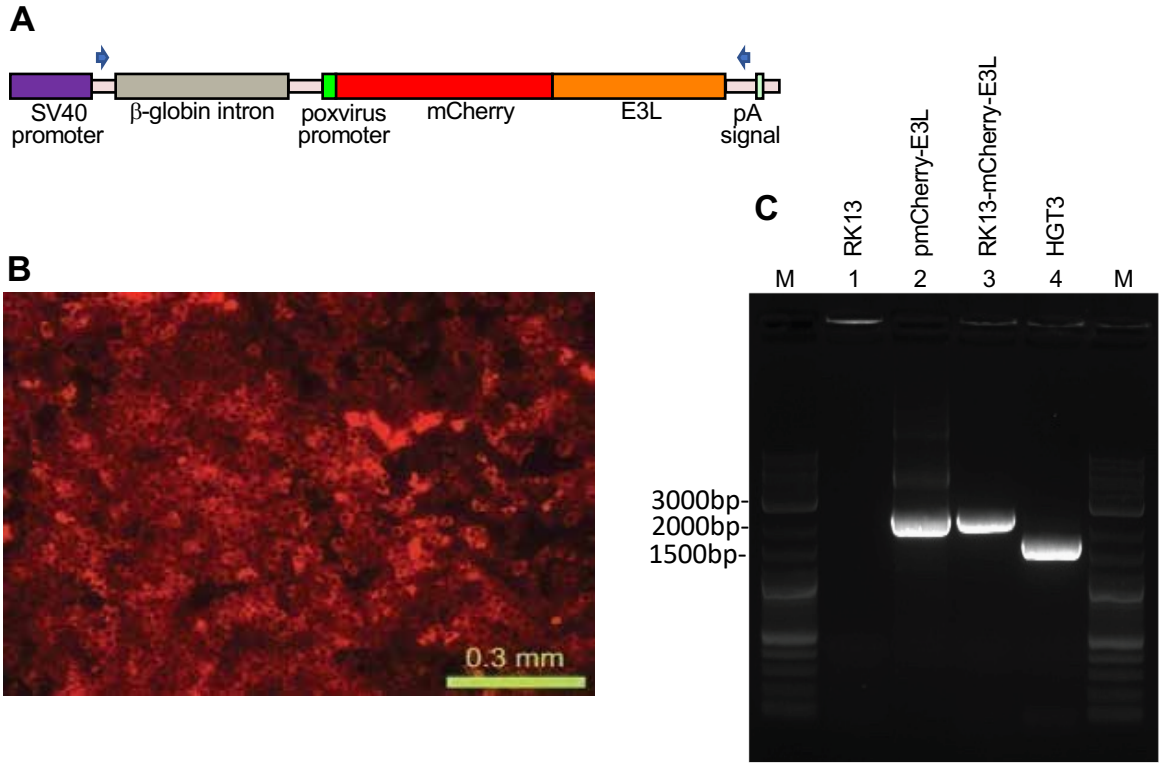
720 HEDENGREN-OLCOTT, M., BYRD, C. M., WATSON, J. & HRUBY, D. E. 2004. The vaccinia virus G1L putative
721 metalloproteinase is essential for viral replication in vivo. *J Virol,* 78**,** 9947-53.
722 HOWARD, A. R., SENKEVICH, T. G. & MOSS, B. 2008. Vaccinia virus A26 and A27 proteins form a stable
723 complex tethered to mature virions by association with the A17 transmembrane protein. *J Virol,*
724 82**,** 12384-91.
725 HUGHES, A. L. & FRIEDMAN, R. 2005. Poxvirus genome evolution by gene gain and loss. *Mol Phylogenet*
726 *Evol,* 35**,** 186-95.
727 IVANCEVIC, A. M., KORTSCHAK, R. D., BERTOZZI, T. & ADELSON, D. L. 2016. LINEs between Species:
728 Evolutionary Dynamics of LINE-1 Retrotransposons across the Eukaryotic Tree of Life. *Genome Biol*
729 *Evol,* 8**,** 3301-3322.
730 IYER, L. M., BALAJI, S., KOONIN, E. V. & ARAVIND, L. 2006. Evolutionary genomics of nucleo-cytoplasmic
731 large DNA viruses. *Virus Res,* 117**,** 156-84.
732 JAYAWARDANE, G., RUSSELL, G. C., THOMSON, J., DEANE, D., COX, H., GATHERER, D., ACKERMANN, M.,
733 HAIG, D. M. & STEWART, J. P. 2008. A captured viral interleukin 10 gene with cellular exon
734 structure. *J Gen Virol,* 89**,** 2447-55.
735 KANE, E. M. & SHUMAN, S. 1992. Temperature-sensitive mutations in the vaccinia virus H4 gene encoding
736 a component of the virion RNA polymerase. *J Virol,* 66**,** 5752-62.
737 KATSAFANAS, G. C. & MOSS, B. 2007. Colocalization of transcription and translation within cytoplasmic
738 poxvirus factories coordinates viral expression and subjugates host functions. *Cell Host Microbe,*
739 2**,** 221-8.
740 KEELING, P. J. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin*
741 *Genet Dev,* 19**,** 613-9.
742 KHAZINA, E., TRUFFAULT, V., BUTTNER, R., SCHMIDT, S., COLES, M. & WEICHENRIEDER, O. 2011. Trimeric
743 structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol,*
744 18**,** 1006-14.
745 KOJIMA, K. K. 2010. Different integration site structures between L1 protein-mediated retrotransposition
746 in cis and retrotransposition in trans. *Mob DNA,* 1**,** 17.
747 LANGLAND, J. O., CAMERON, J. M., HECK, M. C., JANCOVICH, J. K. & JACOBS, B. L. 2006. Inhibition of PKR
748 by RNA and DNA viruses. *Virus Res,* 119**,** 100-10.
749 LEFKOWITZ, E. J., WANG, C. & UPTON, C. 2006. Poxviruses: past, present and future. *Virus Res,* 117**,** 105-
750 18.
751 MATHIAS, S. L., SCOTT, A. F., KAZAZIAN, H. H., JR., BOEKE, J. D. & GABRIEL, A. 1991. Reverse transcriptase
752 encoded by a human transposable element. *Science,* 254**,** 1808-10.
753 MCLYSAGHT, A., BALDI, P. F. & GAUT, B. S. 2003. Extensive gene gain associated with adaptive evolution
754 of poxviruses. *Proc Natl Acad Sci U S A,* 100**,** 15655-60.
755 MOLDOVAN, J. B., WANG, Y., SHUMAN, S., MILLS, R. E. & MORAN, J. V. 2019. RNA ligation precedes the
756 retrotransposition of U6/LINE-1 chimeric RNA. *Proc Natl Acad Sci U S A,* 116**,** 20612-20622.
757 MONIER, A., CLAVERIE, J. M. & OGATA, H. 2007. Horizontal gene transfer and nucleotide compositional
758 anomaly in large DNA viruses. *BMC Genomics,* 8**,** 456.
759 OCHMAN, H., GERBER, A. S. & HARTL, D. L. 1988. Genetic applications of an inverse polymerase chain
760 reaction. *Genetics,* 120**,** 621-3.
761 ODOM, M. R., HENDRICKSON, R. C. & LEFKOWITZ, E. J. 2009. Poxvirus protein evolution: family wide
762 assessment of possible horizontal gene transfer events. *Virus Res,* 144**,** 233-49.
763 PARRISH, S. & MOSS, B. 2006. Characterization of a vaccinia virus mutant with a deletion of the D10R gene
764 encoding a putative negative regulator of gene expression. *J Virol,* 80**,** 553-61.
765 PASZKOWSKI, P., NOYCE, R. S. & EVANS, D. H. 2016. Live-Cell Imaging of Vaccinia Virus Recombination.
766 *PLoS Pathog,* 12**,** e1005824.

767 PERKUS, M. E., GOEBEL, S. J., DAVIS, S. W., JOHNSON, G. P., NORTON, E. K. & PAOLETTI, E. 1991. Deletion
768 of 55 open reading frames from the termini of vaccinia virus. *Virology,* 180**,** 406-10.
769 PISKUREK, O. & OKADA, N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons
770 from reptiles to mammals. *Proc Natl Acad Sci U S A,* 104**,** 12046-51.
771 RAHMAN, M. M., LIU, J., CHAN, W. M., ROTHENBURG, S. & MCFADDEN, G. 2013. Myxoma virus protein
772 M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to
773 promote expanded host tropism and viral replication. *PLoS Pathog,* 9**,** e1003465.
774 SCHONRICH, G., ABDELAZIZ, M. O. & RAFTERY, M. J. 2017. Herpesviral capture of immunomodulatory host
775 genes. *Virus Genes,* 53**,** 762-773.
776 SEKIGUCHI, J. & SHUMAN, S. 1997. Ligation of RNA-containing duplexes by vaccinia DNA ligase.
777 *Biochemistry,* 36**,** 9073-9.
778 SENKEVICH, T. G., WYATT, L. S., WEISBERG, A. S., KOONIN, E. V. & MOSS, B. 2008. A conserved poxvirus
779 NlpC/P60 superfamily protein contributes to vaccinia virus virulence in mice but not to replication
780 in cell culture. *Virology,* 374**,** 506-14.
781 SETO, J., CELENZA, L. M., CONDIT, R. C. & NILES, E. G. 1987. Genetic map of the vaccinia virus HindIII D
782 Fragment. *Virology,* 160**,** 110-9.
783 WILCOCK, D., DUNCAN, S. A., TRAKTMAN, P., ZHANG, W. H. & SMITH, G. L. 1999. The vaccinia virus A4OR
784 gene product is a nonstructural, type II membrane glycoprotein that is expressed at the cell
785 surface. *J Gen Virol,* 80 (Pt 8)**,** 2137-2148.
786 WILSON, E. B. & BROOKS, D. G. 2011. The role of IL-10 in regulating immunity to persistent viral infections.
787 *Curr Top Microbiol Immunol,* 350**,** 39-65.
788 WU, S. & KAUFMAN, R. J. 1997. A model for the double-stranded RNA (dsRNA)-dependent dimerization
789 and activation of the dsRNA-activated protein kinase PKR. *J Biol Chem,* 272**,** 1291-6.
790 YAO, X. D. & EVANS, D. H. 2001. Effects of DNA structure and homology length on vaccinia virus
791 recombination. *J Virol,* 75**,** 6923-32.
792 YEH, W. W., MOSS, B. & WOLFFE, E. J. 2000. The vaccinia virus A9L gene encodes a membrane protein
793 required for an early step in virion morphogenesis. *J Virol,* 74**,** 9701-11.
794
795

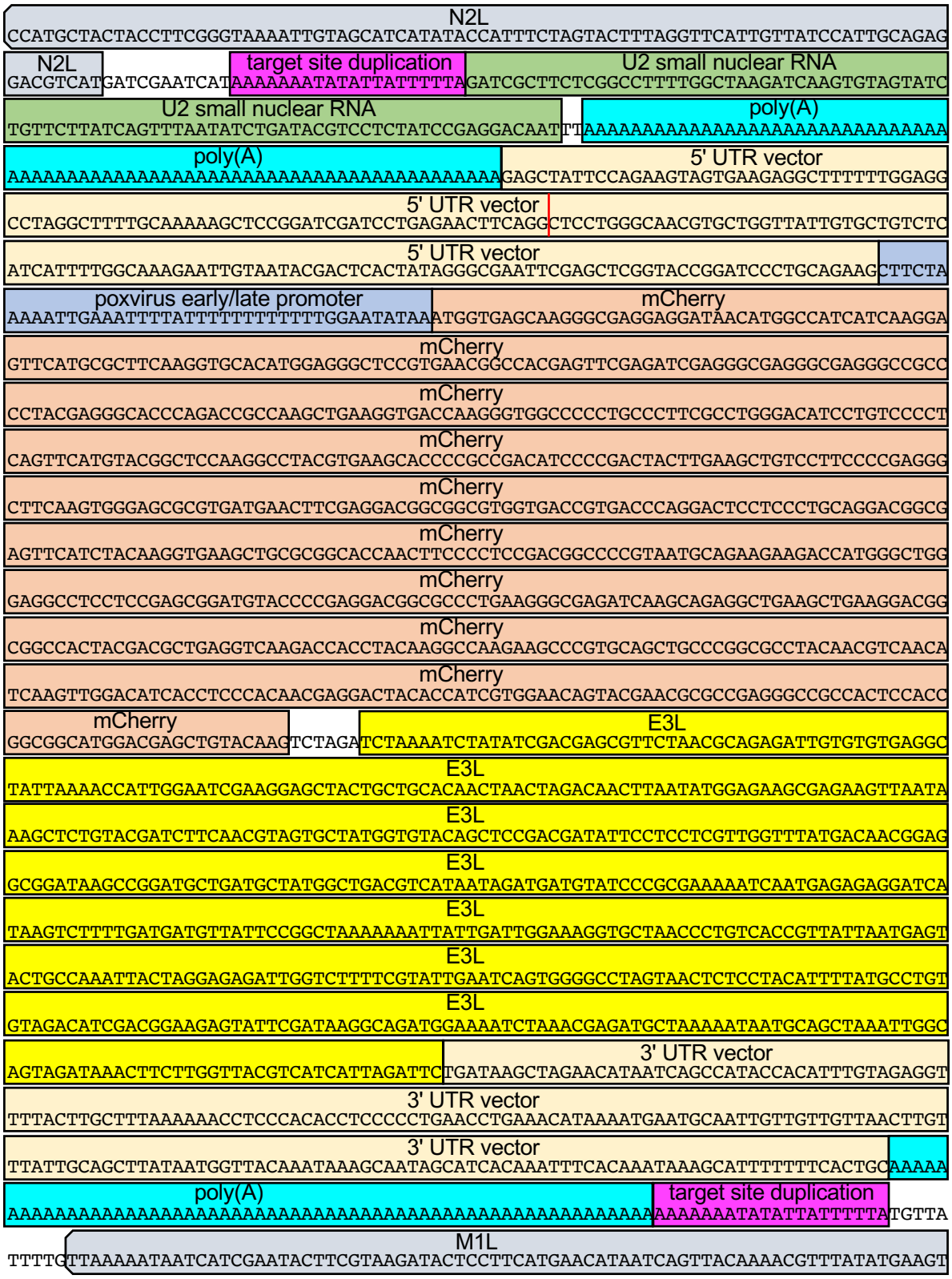
796 # Figure legends:
797
798 **Figure 1. Detection of experimental horizontal gene transfer in vaccinia virus. (A)** VACV lacking E3L
799 and K3L cannot replicate in wild type RK13 cells, but can replicate in cells stably transfected with mCherry-
800 E3L. Virus that acquired E3L can replicate in wild type RK13 cells (right). (**B**) Schematic of the mCherry-
801 E3L vector that was stably transfected into RK13 cells. (**C**) Schematic of the general genetic architecture
802 of horizontally transferred genes identified in VACV isolates. (**D**) HGT integration sites in the VACV
803 genome. The VACV genome is represented in blue. Genes highlighted above the genome were described
804 to have likely originated from HGT (Hughes and Friedman, 2005, Bratke and McLysaght, 2008). Features
805 shown below the genome are: integration sites into genes (black boxes, blue lines), or into intergenic regions
806 (red lines). The orientation of the transferred genes is indicated by the arrow, colors of arrows indicate the
807 orientation of mCherry-E3L relative to target genes (blue: same direction; red opposite direction; black
808 arrow: intergenic). (**E**) Maps of mCherry-E3L integration sites in HGT1 to HGT20. Arrows indicate the
809 direction of transcription for VACV and mCherry-E3L. Intergenic regions are depicted in yellow. The
810 position of an integrated U2 small nuclear RNA and the associated poly(A) tract is shown for HGT20 by
811 dashed lines.
812
813 **Figure 2. Length distribution and composition of target site duplications surrounding integration**
814 **sites.** (**A**) Length distribution of TSDs. (**B**) WebLogo of the consensus sequence and individual sequences
815 of TSDs. Ten nucleotides 5' and 3' of the TSDs, as well as six nucleotides of the TSDs adjacent to the
816 putative first and second LINE-1 endonuclease cleavage sites, respectively, are shown.
817
818 **Figure 3. Complementation of viruses after mCherry-E3L integration into essential genes.** (**A**) Plaque
819 sizes formed by HGT virus isolates in RK13 and BSC-40 cells 3 days after infection. Location of integration
820 (intergenic spaces (i), essential (e), non-essential (n), and a deletion (d)) are indicated below each bar.
821 Dotted lines indicate plaque sizes caused by vP872. Asterisks denote significant differences between vP872
822 and HGT isolates (* $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$). (**B**) Percentage of foci that were mCherry-
823 positive 3 days after infecting RK13+E3+K3 cells with the indicated HGT viruses. (**C**) Complementation
824 of HGT13 by VC-R2. RK13 cells were infected with either HGT13 alone (left) or co-infected with HGT13
825 and VC-R2 (right). Foci were visualized 3 days after infection. (**D**) Location of PCR primers for amplifying
826 integration sites in HGT1 and HGT3. (**E**) PCR amplification for integration site identified for HGT1
827 (PCR#1, left) and HGT3 (PCR#1, right).
828
829 **Figure 4. Fitness increase of HGT3 after serial passaging. (A)** The percentage of mCherry-positive foci
830 formed on RK13+E3+K3 cells by three serially passaged replicates of HGT3. Percentages were determined
831 for the indicated passages (P). (**B**) Increase in double-positive mCherry and EGFP expressing foci of
832 replicate 3 during serial passaging. (**C**) RK13 and (**D**) BSC-40 cells were infected with serially passaged
833 HGT3 (MOI = 0.01) for 48 hours and viruses were titered on RK13+E3+K3 cells.
834
835 **Figure 5. Recombination of complementing viruses after serial passaging. (A)** PCR strategy to detect
836 gene arrays with outward-facing primers. (**B**) Representative PCR products amplified with outward-facing
837 primers for VC-R2 and three independent HGT3 passages. (**C**) Recombination of complementing viruses
838 resulted in tandem arrays of the H4L locus as determined by long-read sequencing.
839
840 **Figure 6. Signatures of retrotransposon-mediated HGT adjacent to cowpox virus gene 211.** The 211
841 gene encodes Golgi anti-apoptotic protein (GAAP). Red boxes mark a perfect 18 bp putative target site
842 duplication. p(A) indicates a polyadenylate stretch (16/21 bp) adjacent to the TSD. Green boxes show a
843 duplicated 59 bp stretch with 83% sequence identity (homology region).
844

845 **Figure 7. Cascade virus evolution after horizontal gene transfer into an essential gene.** After
846 retrotransposon-mediated horizontal gene integration into an essential locus (1) the virus initially
847 depends on a helper virus to complement the essential gene (2), which is followed by
848 recombination and results in a single virus with increased fitness.
849
850

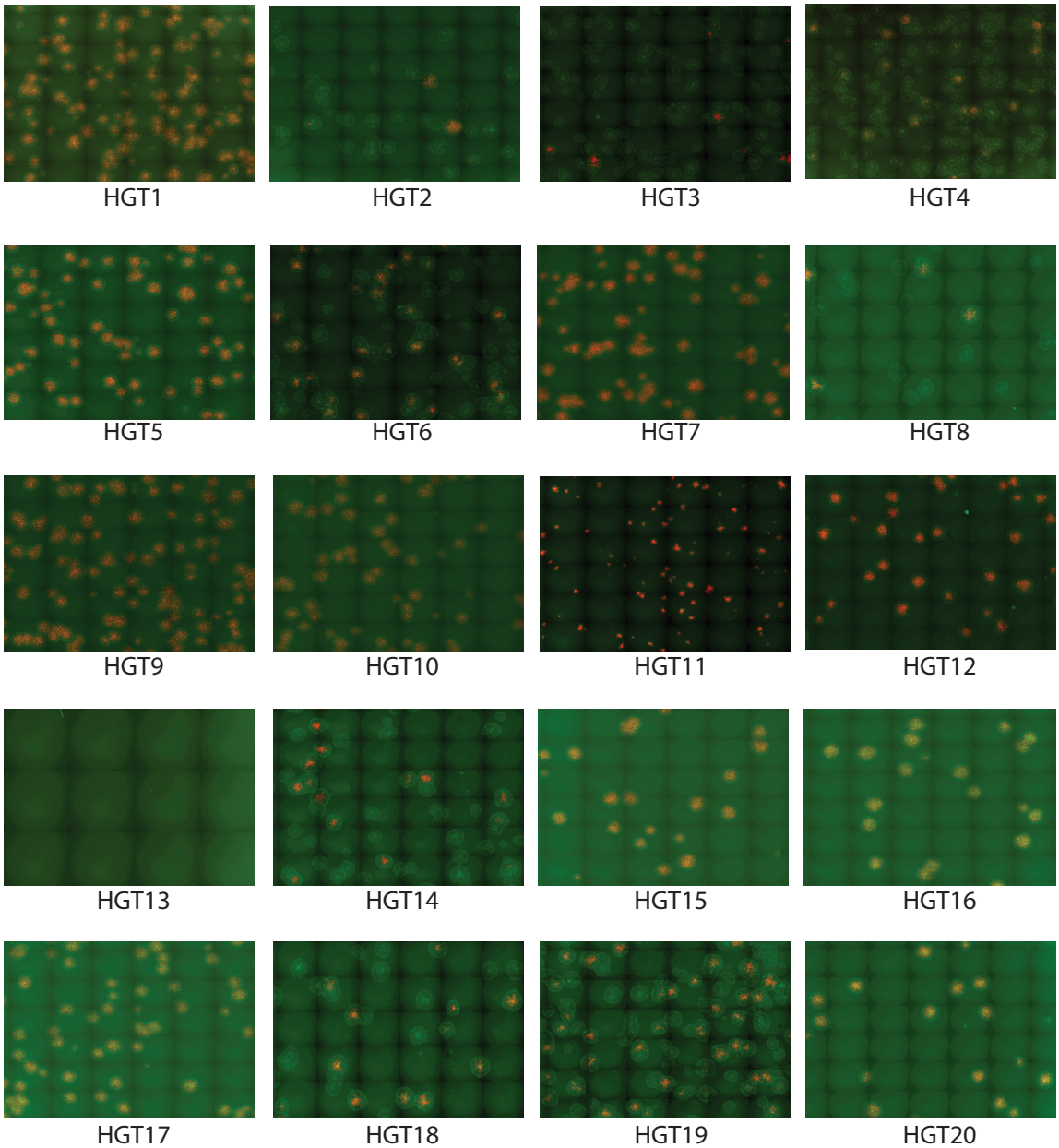
851
852 **Supporting information**
853



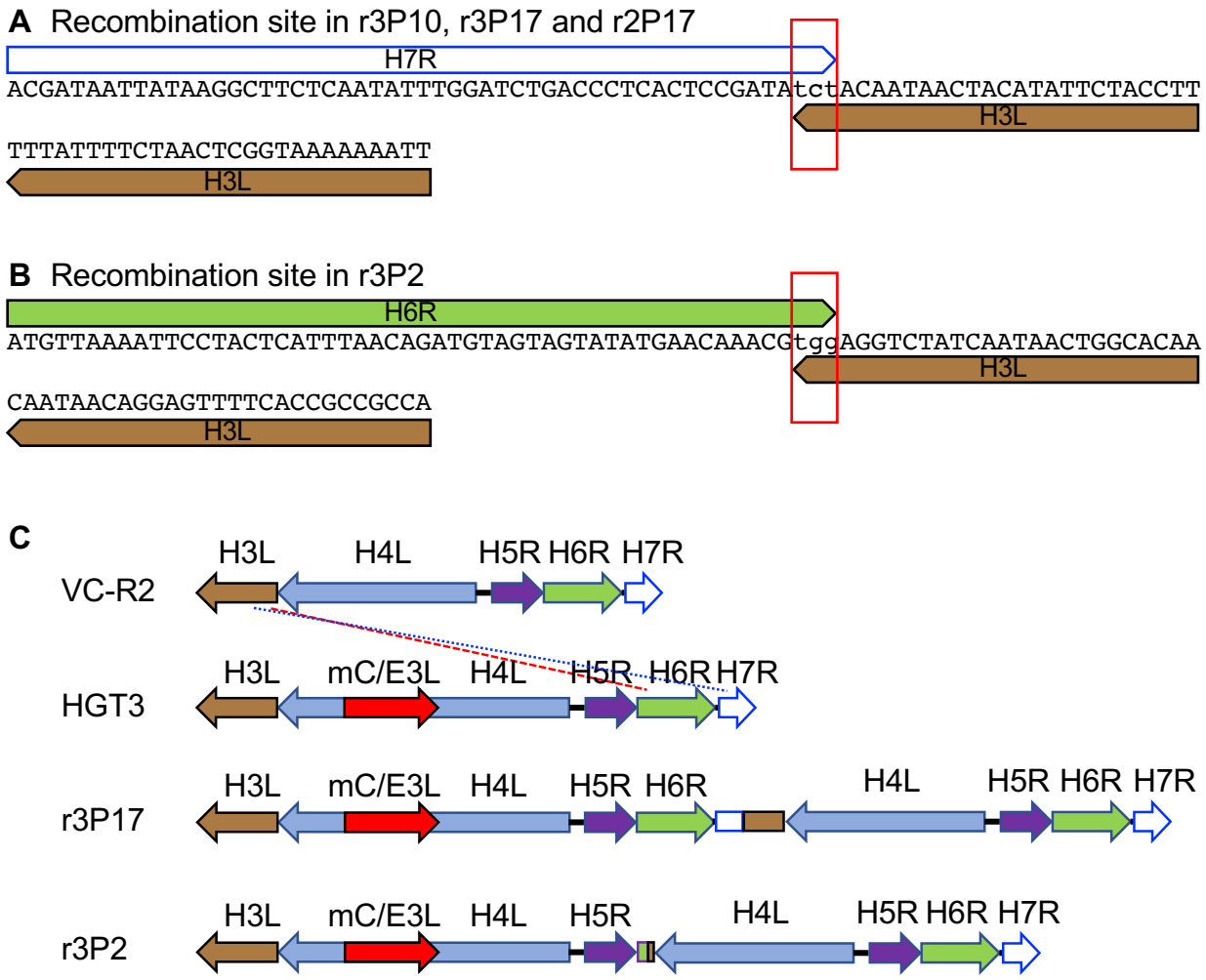
854
855
856 **Supporting Figure 1. Stable expression of mCherry-E3L in RK13 cells.** (**A**) Map of the
857 expression cassette in the pmCherry-E3L plasmid used to stably transfect RK13 cells. The
858 positions of oligonucleotides used to amplify the parts of the expression cassette are indicated by
859 arrows. (**B**) Fluorescent micrograph of stably transfected RK13-mCherry-E3L cells. (**C**) PCR
860 specific for mCherry-E3L did not amplify a product from RK13 cells (lane 1), but amplified
861 comparably sized products from the plasmid pmCherry-E3L (lane 2) and genomic DNA of the
862 RK13-mCherry-E3L cell line (lane 3). A smaller band was amplified from HGT3 DNA due to the
863 spliced-out intron (lane 4). M = molecular weight marker.
864



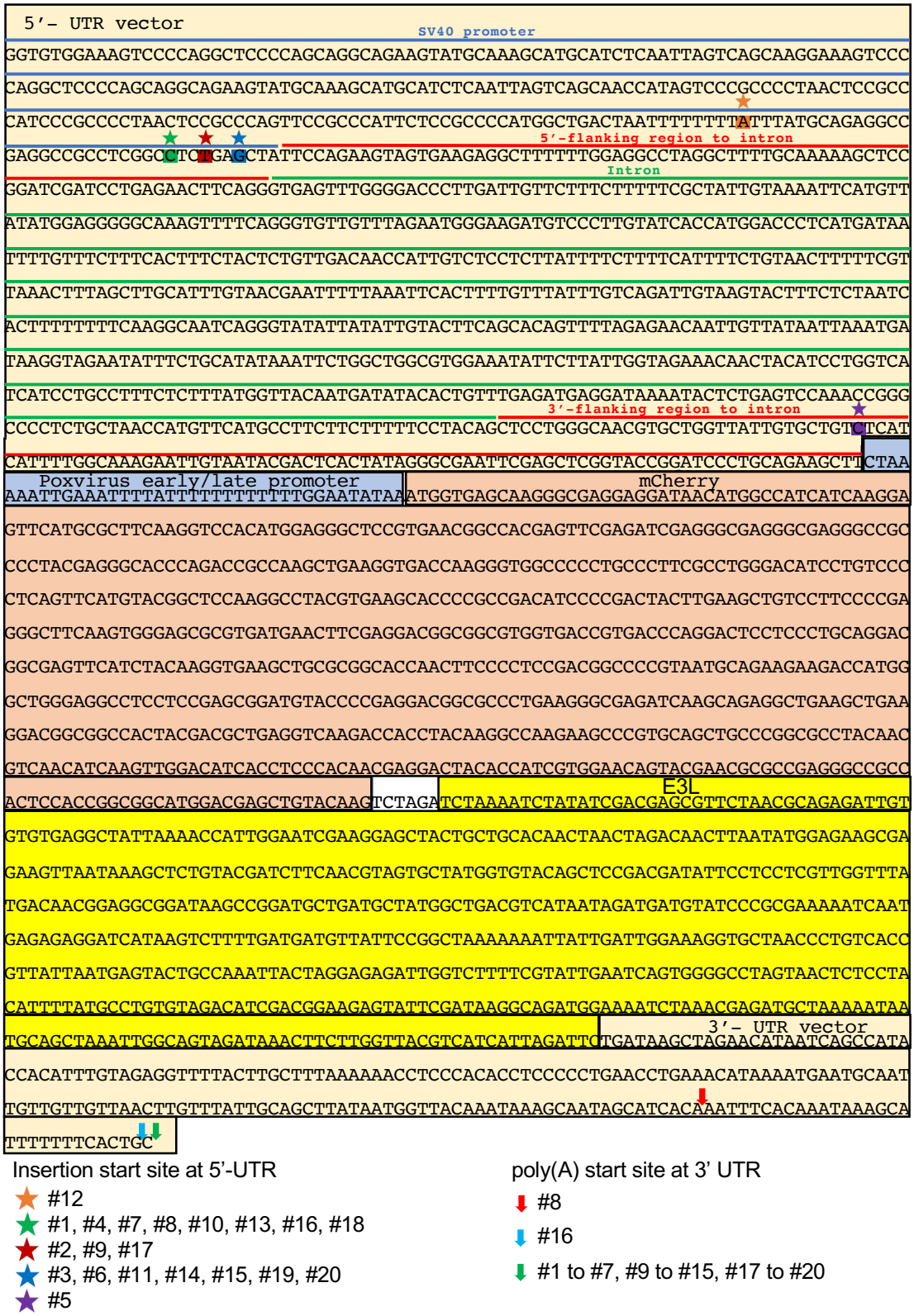
Supporting Figure 2. Integration site and composition of the U2-mCherry/E3L fusion in HGT20.



Supporting Figure 3. mCherry and/or EGFP positive foci formed by HGT1 – HGT20 on RK13+E3+K3 cells. VC-R2 permissive RK13+E3+K3 cells were infected with plaque-purified HGT virus isolates at 100 pfu/well. 48 hours after infection, red and green fluorescent images were taken and overlayed in FiJi.

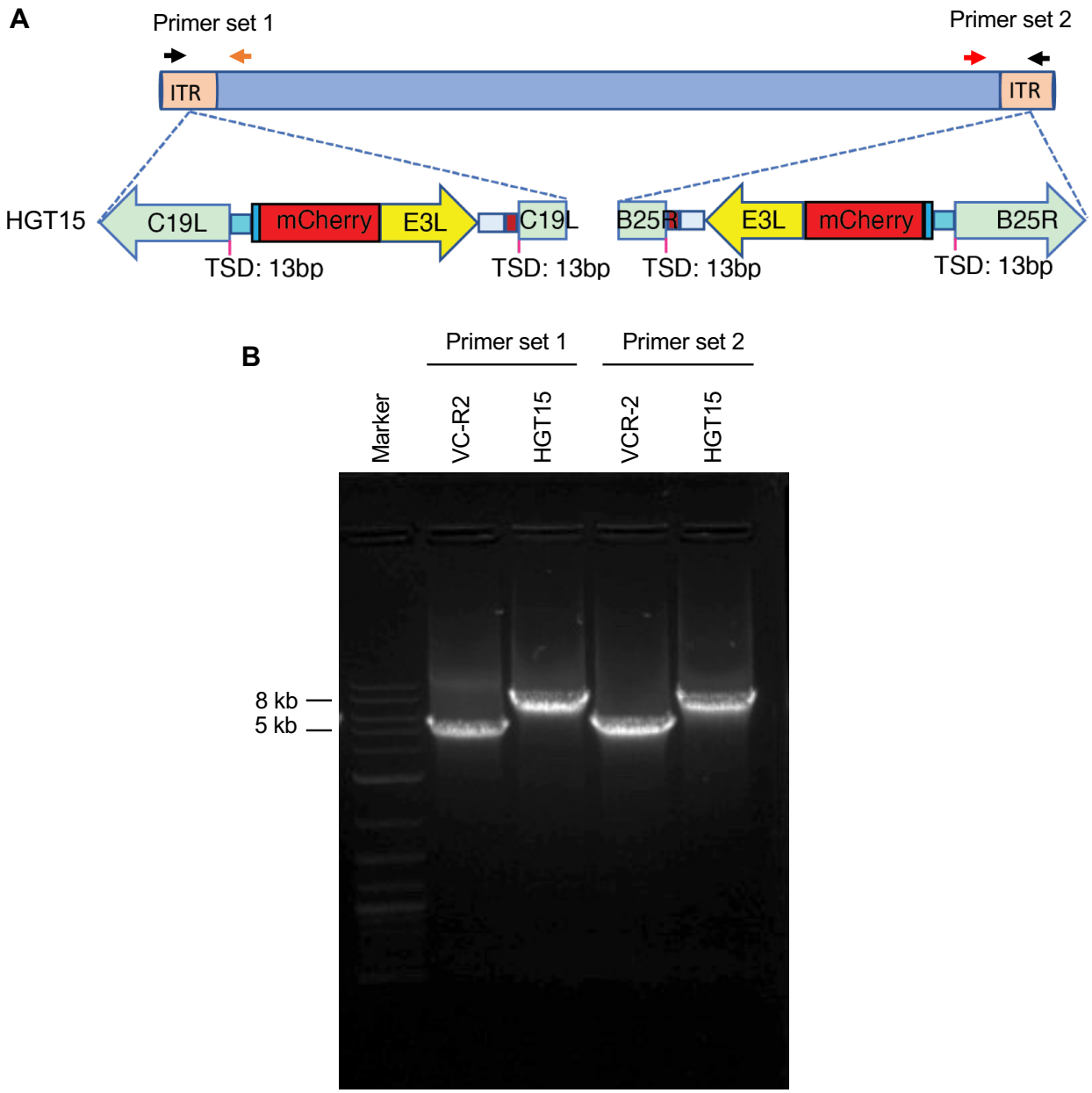


Supporting Figure 4. Recombination sites in passaged HGT3. (**A**) Recombination site in HGT3 between H7R and H3L identified in passage 10 of replicate 3 and passage 17 of replicate 2 by PCR and Sanger sequencing, as well as in passage 17 of replicate 3 by PacBio sequencing. The red rectangle shows 3 three shared nucleotides between the genes at the breakpoint. (**B**) Recombination site in HGT3 between H6R and H3L identified in passage 2 of replicate 3 by PCR and Sanger sequencing. The red rectangle shows 3 three shared nucleotides between the genes at the breakpoint. (**C**) Schematic representation of recombined loci in parental VC-R2 and HGT3, and recombined loci in r3P17 (blue dotted line) and r3P2 (red dashed line).



Supporting Figure 5. Map of pmCherry-E3L. Insertion start sites and polyadenylation sites of transgenes in HGT viruses #1 to #20 are indicated.

891



892
893 **Supporting Figure 6. HGT15 contains one mCherry-E3L copy in each inverted terminal**
894 **repeat (ITR).** (**A**)To determine whether mCherry-E3L was present in both ITRs or only in one,
895 primers were designed to span the integration site and the unique region outside of the ITR. (**B**)
896 Both primer sets yielded larger PCR products for HGT15 than for the parental VC-R2, indicating
897 that mCherry-E3L was present in both ITRs.
898

Clones	Target Site Duplication (TSD)
HGT1 15 bp TSD	5′-AGAATAAAGAATCTA-3′ 3′-TCTTATTTCTTAGAT-5′
HGT2-15bp 15 bp TSD	5′-AAGAGTGCCATAACT-3′ 3′-TTCTCACGGTATTGA-5′
HGT3 18 bp TSD	5′-AAGAATGATTACTAACGC-3′ 3′-TTCTTACTAATGATTGCG-5′
HGT4 17 bp TSD	5′-AAAAATTACGATATTCG-3′ 3′-TTTTTAATGCTATAAGC-5′
HGT5 16 bp TSD	5′-AGAAAATTATGTTATG-3′ 3′-TCTTTTAATACAATAC-5′
HGT6 18 bp TSD	5′-AAGAATTTACCTAGTAGC-3′ 3′-TTCTTAAATGGATCATCG-5′
HGT7 15 bp TSD	5′-AAAAAACTTAACTCG-3′ 3′-TTTTTTGAATTGAGC-5′
HGT8 15 bp TSD	5′-AAAATCTGTAGGAGG-3′ 3′-TTTTAGACATCCTCC-5′
HGT9 12 bp TSD	5′-AATCGTTGGAGG-3′ 3′-TTAGCAACCTCC-5′
HGT10 14 bp TSD	5′-AATAGATTCCTTTT-3′ 3′-TTATCTAAGGAAAA-5′
HGT12 19 bp TSD	5′-AATAAAAAAAGTATTTTTT-3′ 3′-TTATTTTTTTCATAAAAAA-5′
HGT13 17 bp TSD	5′-AAGAATCCTATAGGTGG-3′ 3′-TTCTTAGGATATCCACC-5′
HGT14 16 bp TSD	5′-AAGAGTGGATTCTTCT-3′ 3′-TTCTCACCTAAGAAGA-5′
HGT15 13 bp TSD	5′-AATAACTAGCTTA-3′ 3′-TTATTGATCGAAT-5′
HGT16 15 bp TSD	5′-AGGATATGCTTTCAT-3′ 3′-TCCTATACGAAAGTA-5′
HGT17 15 bp TSD	5′-AAATTATTAGAATCA-3′ 3′-TTTAATAATCTTAGT-5′
HGT18 17 bp TSD	5′-AAAAAGGTTTACATTCT-3′ 3′-TTTTTCCAAATGTAAGA-5′
HGT19 16 bp TSD	5′-ATAAAGTGCACGTTTA-3′ 3′-TATTTCACGTGCAAAT-5′
HGT20 20 bp TSD	5′-AAAAAAATATATTATTTTTA-3′ 3′-TTTTTTTATATAATAAAAAT-5′

Supporting Table 1. Target site duplications identified in HGT viruses.

	Forward primer (5'-3')	Reverse primer (5'-3')
HGT1	JR70 vvH4L-1F (H4L): CTGTGATAGTCGATACGTTATAAAGG	JR73 vvH5R-2F (H5R) GTTTGGTTTTGGGCTTAGTAGATGG
HGT2	JR75 VV-I8R-1F (I8R) GGGATTTAAGGTACTAGATGGATCTCC	JR74 VV-I8R-1R (I8R) GCCGTTCCCTGTCATCCTCTAACG
HGT3	JR35-HGT-H4L-2F (H4L) CTCGCAAATTCTAGTCTTAACC	JR34-HGT-H4L-2R (H4L) GAGAGGTCTGTTCTTTCTTCG
HGT4	JR86-G1L-1F (G1L) GGAATAATCTCCGGACATGCTGG	JR87-G1L-1R (G1L) CGGTGGGAGAATAGACATGATAG
HGT5	JR117-I4L-4F (I4L) GATGTTGTGTAGTACAAGTGGC	HGT-E3L-1F (as reverse primer from E3L) TGGCAGTAGATAAACTTCTTGGTTACG
	HGT-E3L-1R (as forward primer from E3L) AGCCTCACACACAATCTCTGCG	JR114-I5L-2R (I5L) CCGTAACGATTTTCAAATATATAGGACTC
HGT6	JR118-D1R-1F (D1R) GGAGGCAAGGTATTAATCACTACC	JR119-D1R-1R (D1R) CGTTAAACACTCTGACTATATCGTTC
HGT7	JR165-HGT7-A40-1F (A40R) GGGAGGAAGGACGTAATAC	JR166-HGT7-A40-1R (A40R) CTAACCGAAGTAGTGGTATG
HGT8	JR130-A9L-2F (A9L) GGACTGGAGTTAGAATTTATAGAC	JR129-HGT-A10L-1R (A10L) GACTAACGAAATCACAGATATG
HGT9	JR131-D9R-1F (D9R) CGTTCCAGGCGATATTATCTC	JR132-D10R-1R (D10R) GACTTAGCTAGTCGTCTATTATAC
HGT10	JR131-D9R-1F (D9R) CGTTCCAGGCGATATTATCTC	JR132-D10R-1R (D10R) GACTTAGCTAGTCGTCTATTATAC
HGT11	JR190-B4R-1F (B4R) GAGACGATTACTACTAGTGGATG	JR134-B8R-1R (B8R) CGACGCATAGTCCCGTAC
HGT12	JR135-A48R-1F (A48R) CACACGGTTACTGGACC	JR136-A50R-1R (A50R) GGAAGCAATAGCTTAATGATC
HGT13	JR137-F13L-1F (F13L) CTATACTAGTCTTAGCTGACC	JR138-F13L-1R (F13L) CTGGAGGATCTATACATACG
HGT14	JR124-H3L-2F (H3L) CCGATGATAGACCTCCAG	JR85-H5R-2R (H5R) GGTGACTAGATCAGATAGTGTTG
HGT15	JR161-HGT15-C19L-1F (C19L) GTCATCAATGGTGTCCGTTAG	JR162-HGT15-C19L-1R (intergenic space between C19L and C18L) CATTTACCGGCATCATAAACACG
HGT16	JR186-A26L-1F (A26L) GATGTTCTGGTTCGACATCC	JR170-A28L-1R (A28L) CAGGCAGAAGTTGGACC
HGT17	JR171-G5R-1F (G5R) GCGTTAGAACCACGCAAG	JR187-G7L-1R (G7L) GATGGAGCTGTGACTAGTC
HGT18	JR173-D6R-1F (D6R) CGCTTTGTTGTTCGCCTTG	JR174-D6R-1R (D6R) GAGTTATTGTAGCGAGATAATCC
HGT19	JR242-D11L-1F (D11L) GGTCTTCGTCTACAGTAGG	HGT-CHR-1R ATGGCCATGTTATCCTCCTCGC
	HGT-E3L-1F TGGCAGTAGATAAACTTCTTGGTTACG	JR236-D11L-1R (D11L) GAATCAAGGCGGTGGC
HGT20	JR175-C2L-1F (C2L) GATGGTGATCCTGAATGCG	JR178-M1L-1R (M1L) CTCTGTAGTTAACGGTCATACATG

Supporting Table 2. Oligonucleotides used to amplify the integration sites of mCherry-E3L in HGT viruses.

Supporting sequences

The 5' and 3' UTR of the mCherry-E3L gene are shown in light green. Dark red indicates the rabbit β-globin intron. A synthetic early/late poxvirus promoter sequence is marked in grey. mCherry is shown in red. E3L is highlighted in yellow. poly(A) tracts are shown in turquoise. VACV genomic sequences surrounding the integration sites in HGT1 through 20 are shown in lower case letters. Blue shades highlight untemplated nucleotides 5' of the 5'UTR. Target site duplications at the mCherry-E3L integration sites are indicated in pink. For HGT20, dark green indicates a fragment of a U2 small nuclear RNA encoding sequence. For HGT1-20, sequences from Sanger sequencing of PCR products are shown. Note that the poly(A) tracts were not completely sequenced and might contain sequencing errors due to problems with sequencing of long mononucleotide stretches and possible heterogeneity in the population.

<u>pmCherry-E3L:</u>

```
CTGATTCTGTGGATAACCGTATTAGCGGCCGCGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAAGGAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTAT
TTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTT
GGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGGTGAGTTTGGGGA
CCCTTGATTGTTCTTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGAGGGGGCAAAGTTTT
CAGGGTGTTGTTTAGAATGGGAAGATGTCCCTTGTATCACCATGGACCCTCATGATAATTTTGT
TTCTTTCACTTTCTACTCTGTTGACAACCATTGTCTCCTCTTATTTTCTTTTCATTTTCTGTAA
CTTTTTCGTTAAACTTTAGCTTGCATTTGTAACGAATTTTTAAATTCACTTTTGTTTATTTGTC
AGATTGTAAGTACTTTCTCTAATCACTTTTTTTTCAAGGCAATCAGGGTATATTATATTGTACT
TCAGCACAGTTTTAGAGAACAATTGTTATAATTAAATGATAAGGTAGAATATTTCTGCATATAA
ATTCTGGCTGGCGTGGAAATATTCTTATTGGTAGAAACAACTACATCCTGGTCATCATCCTGCC
TTTCTCTTTATGGTTACAATGATATACACTGTTTGAGATGAGGATAAAATACTCTGAGTCCAAA
CCGGGCCCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGC
TGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATT
CGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGA
ATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCA
AGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCG
CCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCC
TGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA
GGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAG
GTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCT
GGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAG
GCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAG
CCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGG
ACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGA
GCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAG
GCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGG
AGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGA
CGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCT
GACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATG
```

954 ATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGA
955 GTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCT
956 CCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTA
957 AACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCAT
958 TAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAA
959 AAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTT
960 GTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
961 TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAATT
962 GTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACC
963 AATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGT
964 TGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAA
965 ACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGA
966 GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAA
967 GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCA
968 AGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCG
969 CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACA
970 TTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
971 AAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCC
972 AGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGA
973 AAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA
974 TAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCC
975 CCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
976 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGA
977 TGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGG
978 AGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
979 GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
980 CTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGC
981 TCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
982 CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTG
983 CATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAC
984 GTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGC
985 GCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACC
986 CATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACT
987 GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGA
988 AGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCG
989 CAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAAT
990 GACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAA
991 AGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
992 TGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGA
993 AGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTT
994 CATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGG
995 GCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGG
996 GCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAG
997 ATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCA
998 TGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA
999 AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
1000 CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT

TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT
GGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT
CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGT
ATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGC
TGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC

HGT1:

cgctttatttgtaattgaccacgccattacgatacaaacttaacggatatcgcgataatgaaat
aatttatgattatttctcgctttcaatttaacacaaccctcaagaacctttgtatttattttca
ttttttaagtatagaataaagaatctaCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTT
GGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACG
TGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGA
ATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTT
GGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTC
GCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCG
ACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT
CGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGG
GCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACG
AGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGA
CGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGT
GAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATA
TGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTC
CGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATG
GCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTG
ATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAA
TGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAAC
TCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAAT
CTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCAT
CATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAA
CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA
GCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAGGAAAAAAAAAACCAAAAAAAAAAAAAAAAagaataaagaatctataaaaactaa
aaaaattatacatcataaaccaatttcctagttgtttgtaactttaaatggactctaaagagac
tattctaattgagatcattcc

HGT2:

gccgttccctgtcatcctctaacgtggcagtcattaaaaacatagaatctattttcgtatgatg
ctttctcgctactgctataataatatctcctatttgatcatgctcatgaacttcgtctataata
agagtgccataactTGTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTT
TTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTG
CTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTA
CCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGT
GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATG
GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGG
GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTAC
TTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG
TGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG
CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC
TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA
AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCT
GCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATC
GTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGT
CTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAAC
CATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAA
GTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTC
CTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAAT
AGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCG
GCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAA
TTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTA
TGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCT
AAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGAT
AAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
GCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaga
gtgccataactaaatagttttgttagagataacttatgggtagaaaatacaattccatattttt
ttggttgtttgtttattaattcttccggtatagatccgtaccgtaaagaaataggagatccatc
tagtaccttaaatccc

<u>HGT3:</u>

ctcgcaaattctagtcttaaccaaaaaatcgtatataaccacggagatggcgtatttaagagtg
gattcttctaccgttttgttcttggatgtcatataggaaactataaagtccgcactactgttaa
gaatgattactaacgcGGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTT
TTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTG
CTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTA
CCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGT
GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATG
GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGG
GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTAC
TTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG
TGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG

1094 CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC
1095 TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA
1096 AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCT
1097 GCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATC
1098 GTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGT
1099 CTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAAC
1100 CATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAA
1101 GTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTC
1102 CTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAAT
1103 AGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCG
1104 GCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAA
1105 TTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTA
1106 TGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCT
1107 AAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGAT
1108 AAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
1109 CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG
1110 CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
1111 GCAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAACAAAAAAAAAAAaagaatgattactaa
1112 cgcaactatatagttcaaattaagcattttggaaacataaaataactctgtagacgatacttga
1113 ctttcgaataagtttgcagacaaacgaagaaagaacagacctctcttaatttcagaagaaaact
1114 ttttttcgtattcctgacgtctagagtttatatcaataagaaagttaagaattagtcggttaat
1115 gttgtatttcattacccaagtttgagatttcataatattatcaaaagacatgataatattaaag
1116 ataaagcgctgactatgaacgaaatagctatatggttcgctcaaaaatatagtcttgttaaac
1117

1118 HGT4:

1119 ctcgccacgaagaaagctctcgtattcagtttcatcgataaaggataccgttaaatataactgg
1120 ttgccgatagtctcatagtctattaagtggtaagtttcgtacaaatacagaatccctaaaatat
1121 tatctaatgttggattaatctttaccataactgtataaaatggagacggagtcataactatttt
1122 accgtttgtacttactggaatagatgaaggaataatctccggacatgctggtaaagacccaaat
1123 gtctgtttgaagaaatccaatgttccaggtcctaatctcttaacaaaaattacgatattcgCTC
1124 TGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGG
1125 ATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTG
1126 GCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAA
1127 GCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAG
1128 GATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACG
1129 GCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAA
1130 GCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATG
1131 TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCC
1132 CCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCA
1133 GGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCC
1134 TCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACC
1135 CCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTA
1136 CGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
1137 GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAAC
1138 GCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTA
1139 TATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGA
1140 GCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGT

ACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGAC
AACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGC
GAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTG
ATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTG
GTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATC
GACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTA
AATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATC
AGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACC
TGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAA
ATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaaaattacg
atattcgatcccgatatcctttgcattctatttaccagcatatcacgaactatattaagattat
ctatcatgtctattctcccaccgttatataaatcgcctccgctaagaaacgttagtatatccat
acaatggaatacttcatttctaaaatagtattcgttttctaattctttaatgtgaaatcgtata
ctagaaagggaaaaattatctttgagttttccgttagaaaagaaccacgaaactaatgttctga
ttgcgtccgattccgttgctgaattaatggatttacaccaaaaactcatataacttctagatgt
agaagcattcgctaaaaaattagtagaatcaaaggatataagtagatgttccaacaagtgagca
attcccaag

<u>HGT5:</u>

atgtttgtcattaaacgaaatggatacaaggaaaatgtcatgtttgataaaatcacgtctcgta
ttagaaaattatgttatgCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGA
ATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTT
GGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTC
GCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCG
ACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT
CGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGG
GCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACG
AGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGA
CGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGT
GAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATA
TGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTC
CGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATG
GCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTG
ATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAA
TGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAAC
TCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAAT
CTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCAT
CATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAA
CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA
GCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AAAAAAAAAagaaaattatgttatggcttaaacacggatcatatagatcctattaaaatagcta
tgaaggttattcaaggaatatataatggagtaacaacggtagaattggacactctggcagccga
aatagcagccacttgtactacacaacatc

HGT6:

ggaggcaaggtattaatcactaccatggacggagacaaattatcaaaattaacagataaaaaga
cttttataattcataagaatttacctagtagcGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTT
TGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAAC
GTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCG
AATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTT
TGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGC
TTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGG
GCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCC
GACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACT
TCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTA
CAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATG
GGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGC
AGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAA
GAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAAC
GAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGG
ACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTG
TGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAAT
ATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCT
CCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTAT
GGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTT
GATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTA
ATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAA
CTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAA
TCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCA
TCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCT
TTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTA
ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAACAAAAAAAAAaa
gaatttacctagtagcgaaaactatatgtctgtagaaaaaatagctgatgatagaatagtggta
tataatccatcaacaatgtctactccaatgactgaatacattatcaaaaagaacgatatagtca
gagtgtttaacg

HGT7:

ctaaccgaagtagtggtatgataaaaatgtagtgtaattgttatatagtgtaacacgaatgcag
tttgggagtattcgttgtgctacaaatatatttccattcttcgtggcatttttagctataaaat
tatatggggttgtctggtttaatatttcggattctcctacccaatagcctcgtctaaggcttct
taaaaaacttaactcgCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGG
CTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATT
GTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCG
GTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAAT
GGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCAC

ATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACG
AGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACAT
CCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGAC
TACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCG
GCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCT
GCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCC
TCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGC
TGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCA
GCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACC
ATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACA
AGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAA
AACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGA
GAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTC
CTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCAT
AATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATT
CCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCC
AAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATT
TTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGAT
GCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCT
GATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTC
CCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTG
CAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTC
ACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaaaaactaactcgtttggagtctctat
cttaattagatccgaatttggatttagagctttgcatgtattacgtccttcctccc

HGT8:

gactaacgaaatcacagatatgataaacgcctcattgaagaatacaatttctaaagataataat
atgctagtcagccaagcgttgaactctgtagctaatcgttctaaacaaacgattggagacttga
ggcaatcatcgtgtaaaatggcattgttgtttaaaaatcttgctacatccatctacacaataga
acgtattttcaatgctaaagtaggcgatgatgttaaggcatcgatgttggagaagtataaagta
ttcacagatatttccatgtcattgtataaagacttgatagctatggagaatctcaaagcgatgc
tatacattattcgacgaagcggatgcagaatagacgatgcacaaattactactgacgatctagt
caagtcttactcattgatccgtcctaaaattctaagtatgataaactattataatgaaatgagt
agaggatactttgaacacatgaaaaaaaatctaaatatgacagatggtgactctgtctcttttg
atgatgaataaatgtcatgttatacagctatattaaaatctgtaggaggCTCTGAGCTATTCCA
GAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAG
AACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGT
AATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAAT
TGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCC
ATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCG
AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGAC
CAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAG
GCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCA
AGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCT
GCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCC
GTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCG
CCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGT

1282 CAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAG
1283 TTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCC
1284 GCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCG
1285 TTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCA
1286 CAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAAC
1287 GTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGA
1288 TAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATG
1289 AGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTG
1290 CTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTAT
1291 TGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTA
1292 TTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAG
1293 ATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACA
1294 TTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAA
1295 TGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAG
1296 CATCACATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaaatctgtaggag
1297 gactggcgctatttcaagtagccaatggcgccatagatttatgtagacatttctttatgtattt
1298 ttgtgaacaaaagctacgaccaaattcattttggttcgttgttgttagagccattgcgagcatg
1299 ataatgtatttagtattaggtatagcattgctgtatatttctgaacaagatgacaagaagaata
1300 ctaataatgatggtagtaataatgataaacgaaatgagtcgtctataaattctaactccagtcc
1301

1302 HGT9:

1303 cgttccaggcgatattatctcaacaaaattcagattctatctttagagtatccactaaactatt
1304 acggtttatgtactacaatgaactaagagaaatctttagacggttgagaaaaggttctatcaac
1305 aatatcgatcctcacttcgaagagttaatattattgggtggtaaactagataaaaaggaatcta
1306 ttaaagattgtttaagaagagaattaaaagaggaaagtgatgaacgtataacagtaaaagaatt
1307 tggaaatgtaattctaaaacttacaacacgggataaattatttaataaagtatatataagttat
1308 tgcatggcgtgttttattaatcaatcgttggaggTGAGCTATTCCAGAAGTAGTGAAGAGGCTT
1309 TTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGC
1310 AACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGG
1311 GCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTT
1312 TTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATG
1313 CGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCG
1314 AGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCC
1315 CTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCC
1316 GCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGA
1317 ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCAT
1318 CTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC
1319 ATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCA
1320 AGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGC
1321 CAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCAC
1322 AACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCA
1323 TGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGT
1324 GTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTT
1325 AATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACA
1326 GCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGC
1327 TATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCT
1328 TTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTA

TTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAG
TAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGA
AAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACG
TCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTT
GCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTG
TTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAA
TAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAACAAAAAAAA
AAAaatcgttggaggatttatcgcatactagtatttacaatgtagaaattagaaagattaaatc
attaaatgattgtattaacgacgataaatacgaatatctgtcttatatttataatatgctagtt
aatagtaaatgaacttttacagatctagtataattagtcagattattaagtataatagacgact
agctaagtc

HGT10:

tcctccaacgattgattaataaaacacgccatgcaataacttatatatactttattaaataatt
tatcccgtgttgtaagttttagaattacatttccaaattcttttactgttatacgttcatcact
ttcctcttttaattctcttcttaaacaatctttaatagattccttttCTCTGAGCTATTCCAGA
AGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAA
CTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAA
TACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTG
AAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCAT
CATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAG
ATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCA
AGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGC
CTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAG
TGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC
AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGT
AATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCC
CTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCA
AGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTT
GGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGC
CACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTT
CTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACA
ACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGT
AGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATA
AGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAG
AGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCT
AACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTG
AATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATT
CGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGAT
AAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATT
TGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATG
AATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCA
TCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAMAAAAAAAatagattccttttatctagtttaccaccca
ataatattaactcttcgaagtgaggatcgatattgttgatagaacctttctcaaccgtctaaa

gatttctcttagttcattgtagtacataaaccgtaatagtttagtggatactctaaagatagaa
tctgaattttgttgagataatatcgcctggaacg

HGT11:

gagacgattactactagtggatgtacatgtatttcggaagcagtcgcaaacaacaacaaaataa
taatggaagtactattgtctaaacgaccatctttgaaaattatgatacagtctatgatagcaat
tactaaaaataaacaacataatgcagatttattgaaaatgtgtataaaatatactgGGAGCTAT
TCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCC
TGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAA
TTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAA
AAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACAT
GGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAG
TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGG
TGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTC
CAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGC
TTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCT
CCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGAC
GGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTG
AGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACAT
CAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAG
GGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACG
AGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGC
TGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTT
CAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGG
CGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATC
AATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAA
GGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTC
GTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAG
AGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCA
GTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATAC
CACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACAT
AAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCA
ATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAGAA
AAAAAAAAAAAAACAAAAAAAAAAAtaattctcgcagttttgttcattaatagtatacacgcta
aaataactagttataagtttgaatccgtcaattttgattccaaaattgaatggactggggatgg
tctatacaatatatcccttaaaaattatggcatcaagacgtggcaaacaatgtatacaaatgta
ccagaaggaacatacgac

HGT12:

ggaagcaatagcttaatgatcaaatatttatcatccctatctgttataaagtctctaattaatt
tagatttttctttatatcctgatgcgtgatatatatcacagcataatttctaaattcgcgaag
cgacgtcatttaataaaaaaagtatttttttCATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCT
GAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGA
TCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGG
CAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAG
CTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGG

ATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGG
CCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAG
CTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGT
ACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCC
CGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAG
GACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCT
CCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCC
CGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTAC
GACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACG
TCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACG
CGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTAT
ATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAG
CTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTA
CGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACA
ACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCG
AAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGA
TTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGG
TCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCG
ACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAA
ATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCA
GCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCT
GAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAA
TAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAataaaaaaagtatttttttttaat
atttttcacaaatatcgttcgcggatatcattagacaattgtagtatattcttacgtatcactt
gtttaatatctatatcggctattctggaaccgag

HGT13:

ctggaggatctatacatacgattaaaacgttaggtgtatattctgattatcccccgctggccac
agatcttcgtagaagatttgatacttttaaagcctttaatagcgcaaaaaattcatggttgaat
ttatgctctgcggcttgttgtctgccagttagcactgcgtatcatattaagaatcctataggtg
gCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCT
CCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCAT
TTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGC
AGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGA
GGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTG
AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCG
CCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTT
CATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCC
TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGA
CCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTT
CCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATG
TACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCC
ACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTA
CAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTAC
GAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAA
TCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGA

AGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCT
CTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTA
TGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATC
CCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATT
ATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAG
ATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGA
CATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCA
GCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACAT
AATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTG
AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTT
ACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaagaatcc
tataggtggagtgttctttactgattctccggaacacctattgggatattctagagatctagat
accgatgtagttattgataaactcaggtcagctaagactagtatag

HGT14:

ctcgcaaattctagtcttaaccaaaaaatcgtatataaccacggagatggcgtatttaagagtg
gattcttctGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAA
AGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCA
TCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCC
CTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGG
GCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAG
ACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTC
AGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCT
GTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC
GTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCA
ACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCG
GATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGC
GGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCG
CCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACA
GTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCT
AAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAA
TCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAA
AGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGG
TTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATG
TATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAA
AATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGG
AGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTG
TAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAA
TGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGA
ACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCC
CCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAAT
GGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAA
AAAAAAAAAAAAAGAAAAAAAAAAAAAAAACAAAAAAAAAAAaagagtggattcttctaccgttt
tgttcttggatgtcatataggaaactataaagtccgcactactgttaagaatgattactaacgc
aactatatagttcaaattaagcattttggaaacataaaataactctgtagacgatacttgac

HGT15:

gtcatcaatggtgtccgttagtatatcaaagatcttgttatcgattgatagtgaatgaatcaga
tagtggtgtccttttcatccttgctatcaaagttacgcatgccgtggtgtaacaatatcttta
atacagatggattaaatcgtgtattcatcgtatagcaatgtaatggagagttacctcgtttatt
cagatcgcagtgtttaataactagcttaGGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGG
AGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTG
CTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAAT
TCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGG
AATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTC
AAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCC
GCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGC
CTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGAC
ATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCG
AGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC
TGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGA
GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAA
GCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAG
GACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACG
AGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGA
GGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATG
GAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCG
ACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGC
TGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGAT
GATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATG
AGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTC
TCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCT
AAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCA
TTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTA
AAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACT
TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGC
ATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaat
aactagcttaaacagatgagacgatgtatccacatcaaagaacgtaaaatacatatgacaaaca
ttgt

HGT16:

ggagtaccatcgtgattcttaccagatattatacaaaatactatatataaaatatattgaccaa
cgttagtaatcatataaatgtttaacgttttaaattttgtattcaatgatccattatcatacgc
tagcatggtcttatgatattcattctttaaaatataatattgtgttagccattgcattggagct
cctaatggagattttctattctcgtccatttaggatatgctttcatCTCTGAGCTATTCCAGA
AGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAA
CTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAA
TACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTG
AAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCAT
CATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAG

ATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCA
AGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGC
CTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAG
TGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC
AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGT
AATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCC
CTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCA
AGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTT
GGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGC
CACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTT
CTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACA
ACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGT
AGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATA
AGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAG
AGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCT
AACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTG
AATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATT
CGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGAT
AAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATT
TGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATG
AATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCA
TCACAAATTTCACAAATAAAGCATTTTTTTCACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAaggatatgctttcataaagtccctaataacttcgt
gaataatgtttctatgttttctactgatgcatgtatttgcttcgatttttttatcccatgtttc
atctatcatagatttaaacgcagtaatgctcgcaacattaacatcttgaaccgttggtacaatt
ccgttccataaatttataatgttcgccatttatataactcatttttgaatatacttttaatta
acaaaagagttaagttactcatatggacgccgtccagtctgaac

HGT17:

gatatgcaaaagtttaccattttagaatatctttacattatgagagtgatggcaaacaacgtta
agaaaaagaacgtgggaaaaaacaacggaggagtagttatgcatattaactcacccttaaggt
aatcaatttgccaaaatgttaaattattagaatcaGTTGAGCTATTCCAGAAGTAGTGAAGAGG
CTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTG
GGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTAT
AGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTT
TTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTC
ATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCT
GCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCAC
CCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGA
TGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTT
CATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAG
ACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGA
TCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA
GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCC
CACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCG

1611 GCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGAT
1612 TGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAA
1613 CTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGT
1614 ACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGA
1615 TGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAG
1616 TCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCG
1617 TTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCC
1618 TAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGAT
1619 GGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTT
1620 ACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTA
1621 CTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTG
1622 TTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC
1623 AAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
1624 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaattattagaat
1625 catatatctttatccatctattatcgtctgtgaagttttggctcagaaaaatatctttacctaa
1626 tattcgtttagacgatgtttcaacaccggcatttttataacagtcagctaccaatttaaaacaa
1627 tacattctattatgaccaaatccatacggaatacccaataaagttaatgacgtatcagctgc
1628

1629 <u>HGT18:</u>
1630 attgttccatattatgggatcaggtaaaacaataatcgctttgttgttcgccttggtagcttcc
1631 agatttaaaaaggtttacattctCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAG
1632 GCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCT
1633 GGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTC
1634 GAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAA
1635 TATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAA
1636 GGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGC
1637 CCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCT
1638 GGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACAT
1639 CCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG
1640 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGG
1641 TGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
1642 GGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGG
1643 CTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGC
1644 CCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGA
1645 CTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAG
1646 CTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGG
1647 CTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGA
1648 GAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGAC
1649 GATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTG
1650 ACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGA
1651 TGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAG
1652 TACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTC
1653 CTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAA
1654 ACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATT
1655 AGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
1656 AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTG
1657 TTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCAT

TTTTTTCACTGCAAA

1705 GGACAATTTAAA