Cascade evolution in poxviruses: retrotransposon-mediated host gene capture, complementation and recombination

M. Julhasur Rahman[1], Sherry L. Haller[2], Jie Li[3], 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] University of Texas Medical Branch at Galveston, Department of Microbiology and Immunology, Galveston, TX 77555, USA

[3] 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 by 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 poxviruses. All identified gene capture events showed signatures of LINE-1-mediated retrotransposition. 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 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). For example, 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).

So far, the detection of HGT in large DNA viruses has relied on bioinformatic approaches, primarily phylogenetic and sequence composition analyses. 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 genus *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 frequency 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 host 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) and identified 16 unique HGT isolates. This indicates a detected transfer rate of mCherry-E3L into the VACV genome of approximately 1 in 22.7 million virions.

Horizontally acquired genes show evidence of LINE-1-mediated retrotransposition

To map the integration sites, we plaque purified each HGT candidate three times and used both Sanger sequencing of amplicons from inverse PCR (Ochman et al., 1988) and PacBio-based long-read sequencing (Eid et al., 2009). The integration sites were distributed throughout the VACV genome, and were found in intergenic spaces (three cases), as well as in non-essential and essential genes (Figure 1D, E, Table 1). All 20 isolates (HGT1-HGT20) displayed hallmarks of RNA-

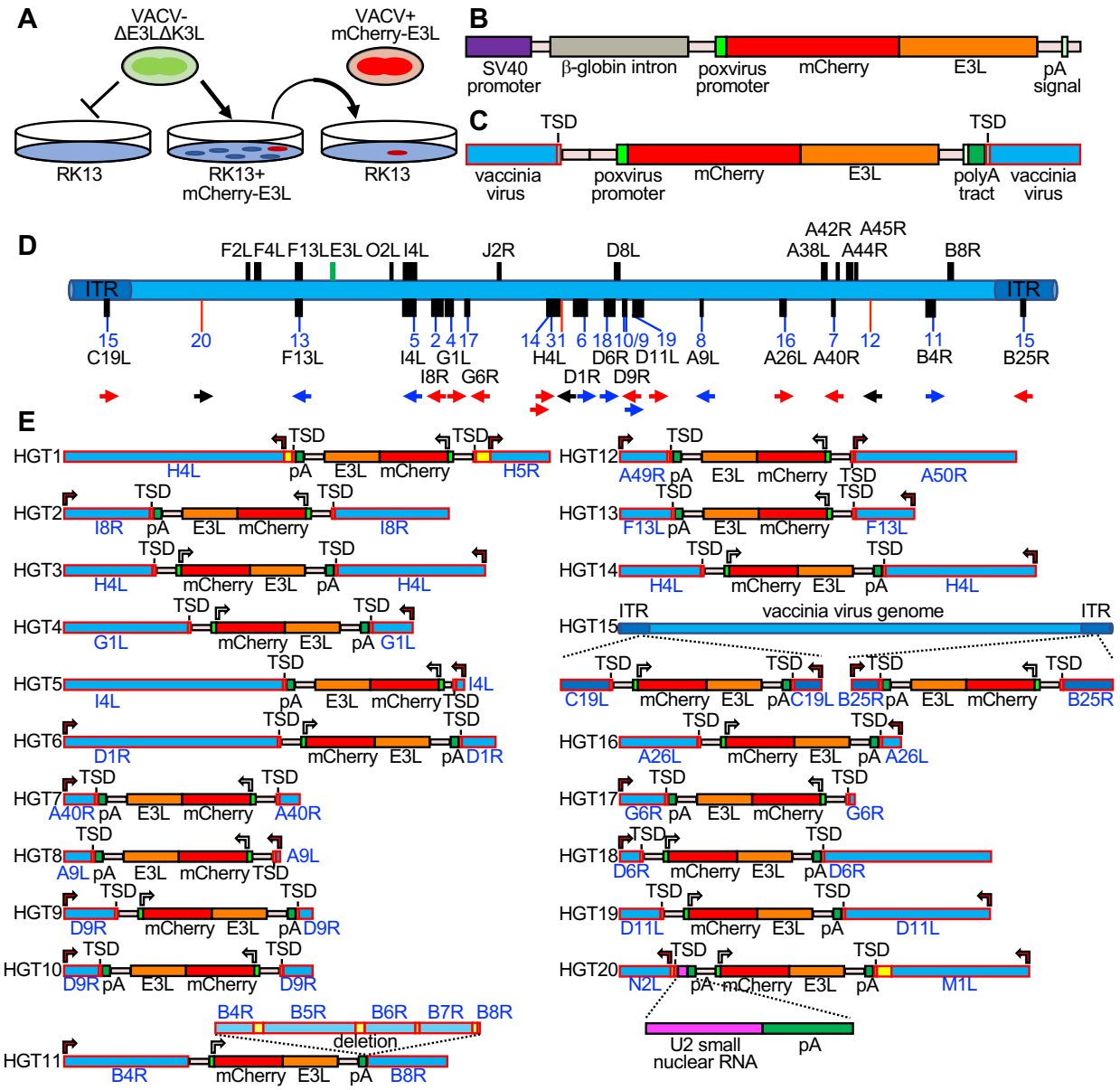


Figure 1. Detection of experimental horizontal gene transfer in vaccinia virus. (A) VACV lacking E3L and K3L cannot replicate in wild type RK13 cells, but can replicate in cells stably transfected with mCherry-E3L. Virus that acquired E3L can replicate in wild type RK13 cells (right). (**B**) Schematic 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.

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

139
140 **Table 1. Integration sites of mCherry-E3L in HGT viruses and importance of disrupted genes for**
141 **virus replication.**
142
143 dependent, retrotransposon-mediated integrations, including a spliced-out intron and long
144 untemplated stretches of poly(A) at the 3' end (Esnault et al., 2000).
145 In 19 cases, short target site duplications (TSDs) of VACV sequences surrounded the integration
146 sites (Figure 1C, E, Figure 2, Supporting Table 1). These TSDs had an average length of 15.9 base
147 pairs (bp) and a consensus (3'-AA/TTTT-5') motif at the integration site typical of the cleavage
148 consensus site of the long interspersed nuclear element-1 (LINE-1, L1) group of retrotransposons
149 (Figure 2) (Kojima, 2010, Gilbert et al., 2002). HGT11 contained a 2.7 kb deletion immediately
150 downstream of the integration site in B4R, precluding the identification of a TSD (Figure 1E). In
151 HGT20, the integrated sequence contained the intron-less mCherry-E3L cassette fused to an 88 bp
152 cellular DNA fragment encoding U2 small nuclear (sn) RNA, which contained a second poly(A)
153 tract (Figure 1E, Supporting Figure 2).

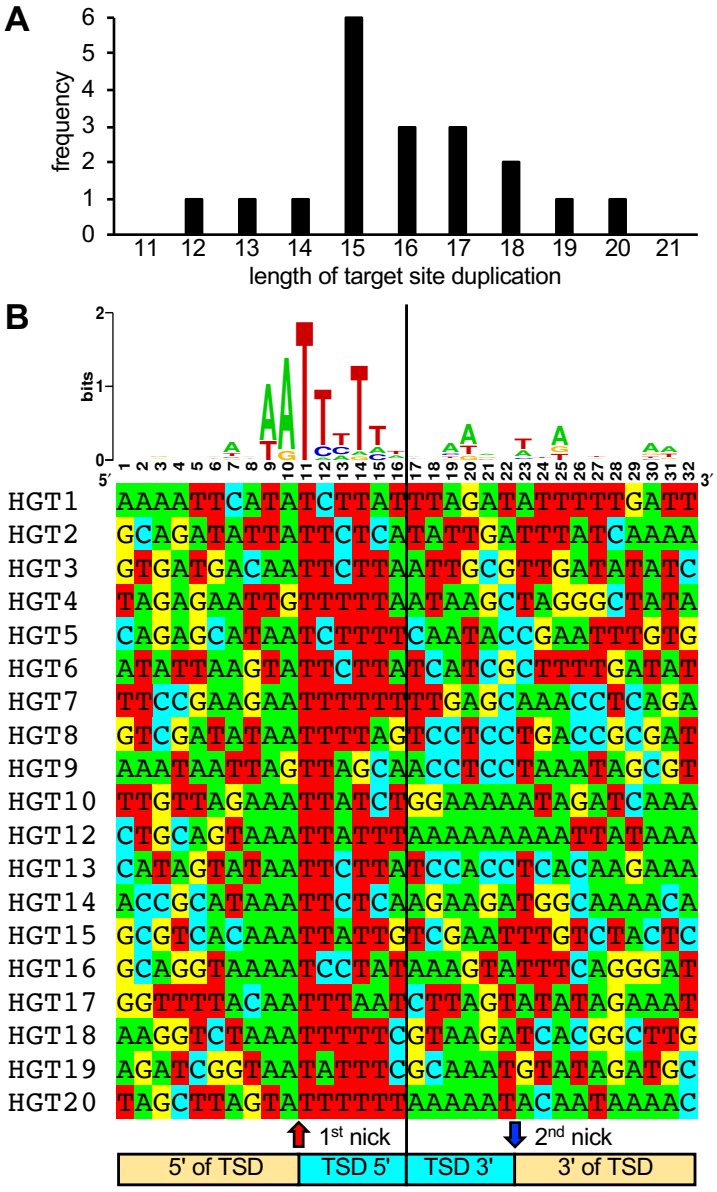


Figure 2. Length distribution and composition of target site duplications surrounding integration sites. (**A**) Length distribution of TSDs. (**B**) WebLogo of the consensus sequence and individual sequences of TSDs. Ten nucleotides 5' and 3' of the TSDs, as well as six nucleotides of the TSDs adjacent 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 substantially smaller plaques than in RK13 cells, e.g. HGT3, HGT14, and HGT19. HGT13 (insertion in F13L) formed tiny plaques in both cell lines, consistent with a previous report of

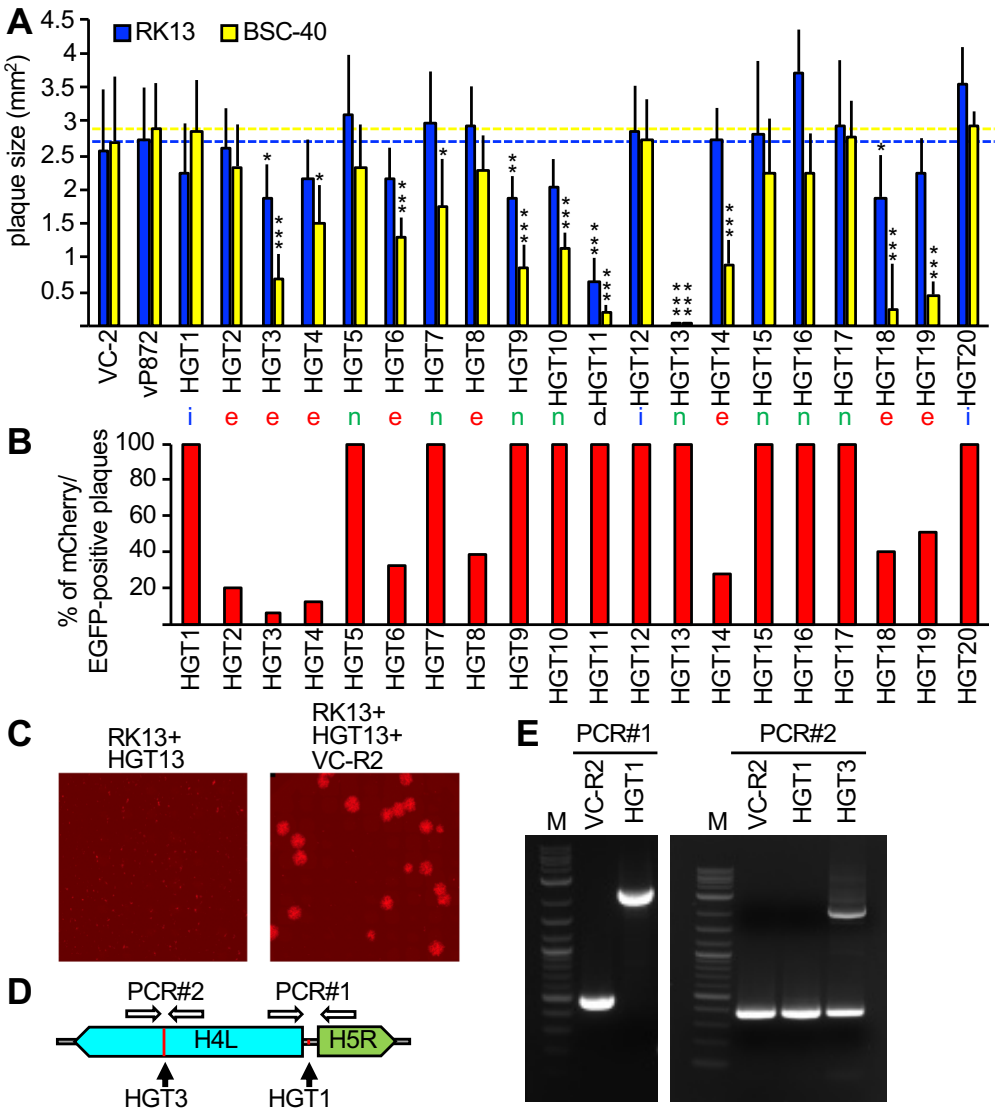


Figure 3. Complementation of viruses after mCherry-E3L integration into essential genes. (**A**) Plaque sizes formed by HGT virus isolates in RK13 and BSC-40 cells 3 days after infection. Location of integration (intergenic spaces (i), essential (e), non-essential (n), and a deletion (d)) are indicated 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).

defective viral spread in a F13L deficient VACV (Blasco and Moss, 1991). HGT11, containing a deletion stretching from B4R through B8R, also exhibited a small plaque phenotype in both cell types. These findings indicate that, at least in some cases, HGT can cause a loss-of-fitness.

Surprisingly, eight of the VACV genes that were disrupted by HGT were reported to be essential for virus replication (Table 1). Because these insertions should presumably inactivate these essential genes, it is unclear how this HGT resulted in viable viruses. One possibility could be that

186 although these viruses were plaque purified three times, the parent virus may have persisted to
187 complement the otherwise non-viable HGT virus. To test this hypothesis, we infected
188 RK13+E3+K3 cells, which are also permissive for the parental virus, with each plaque-purified
189 HGT isolate. In all these infections, we identified the expected EGFP/mCherry double-positive
190 foci. Additionally, foci that were EGFP-positive but mCherry-negative were observed in isolates
191 in which mCherry-E3L insertions occurred in essential genes, indicating the continued presence
192 of parental virus. The percentage of mCherry positive to mCherry negative foci in these viruses
193 ranged between 7% - 51% (Figure 3B, Supporting Figure 3). We hypothesized that the two
194 fluorescent phenotypes indicated mutual complementation, wherein the HGT virus supplied
195 mCherry-E3 and VC-R2 supplied the disrupted essential gene product, thus permitting replication
196 of both viruses in co-infected cells. To test this hypothesis, we co-infected RK13 cells with VC-
197 R2 and the F13L-disrupted HGT13 virus, which did not contain detectable VC-R2 but formed only
198 minute foci in either RK13 or BSC-40 cells. As predicted, in the co-infection we observed larger
199 foci in addition to the very small foci produced by HGT13 alone (Figure 3C). In an additional,
200 PCR-based test of this hypothesis, we amplified DNA spanning the integration sites in HGT1,
201 which acquired mCherry-E3L in an intergenic region, or HGT3, in which the essential gene H4L
202 was disrupted. As expected, HGT1 yielded a single 2.3 kB amplicon of the expected size for
203 mCherry-E3L integration. However, we amplified two products from HGT3: a 396 bp amplicon,
204 consistent with wild type H4L, and a 2.2 kb amplicon consistent with mCherry-E3L integration
205 into H4L (Figure 3D, E). Taken together, these assays support our hypothesis of mutual
206 complementation of these viruses.
207
208 **Recombination of complementing viruses generates replication-competent individual viruses**
209 If both viruses complemented each other efficiently, the expected ratio of HGT virus to parent to
210 virus should be close to 1:1. However, for most HGT isolates this ratio was lower, indicating that
211 virus complementation did not necessarily result in optimal replication of both viruses. In these
212 cases, there may be strong selective pressure to evolve a single replication-competent virus that is
213 not reliant on a complementing virus. To test this hypothesis, we serially passaged the
214 HGT3/parental virus population in PKR-competent RK13 cells. In three independent replicates,
215 we observed a rapid increase in double-positive fluorescent foci (Figure 4A, B), consistent with
216 an increase in HGT3 fitness. In order to directly test for a fitness increase, we infected RK13 and
217 BSC-40 cells with the passaged virus populations for 48 hours and then titered the viruses on
218 RK13+E3+K3 cells. Whereas in RK13 cells the passaged viruses showed only a modest increase
219 in replication (about 5-fold) in passages 9 and 17, an approximately 100-fold higher replication
220 was observed for BSC-40 cells infected with passages 9 and 17 as compared to passage 0 (Figure
221 4C, D).
222 To investigate possible structural changes around the disrupted H4L locus, we designed outward-
223 facing primers in H4L similar to an approach to detect gene amplification in poxviruses (Elde et
224 al., 2012, Brennan et al., 2014). This PCR strategy only yields products if two or more copies of
225 H4L are adjacent to each other (Figure 5A). We amplified products from passaged virus DNA, but
226 not from DNA isolated from VC-R2 or the original plaque-purified virus (P0) (Figure 5B).
227 Because of this evidence for genomic structural changes, we used PacBio long-read sequencing to
228 more completely define these changes in the different populations. This approach demonstrated
229 that in each HGT3 virus population, there were genomes in which an intact copy of H4L existed
230 in close proximity to the mCherry-E3L disrupted H4L locus, although we found several different
231 recombinant viruses with different genomic architectures (Figure 5C). No apparent sequence

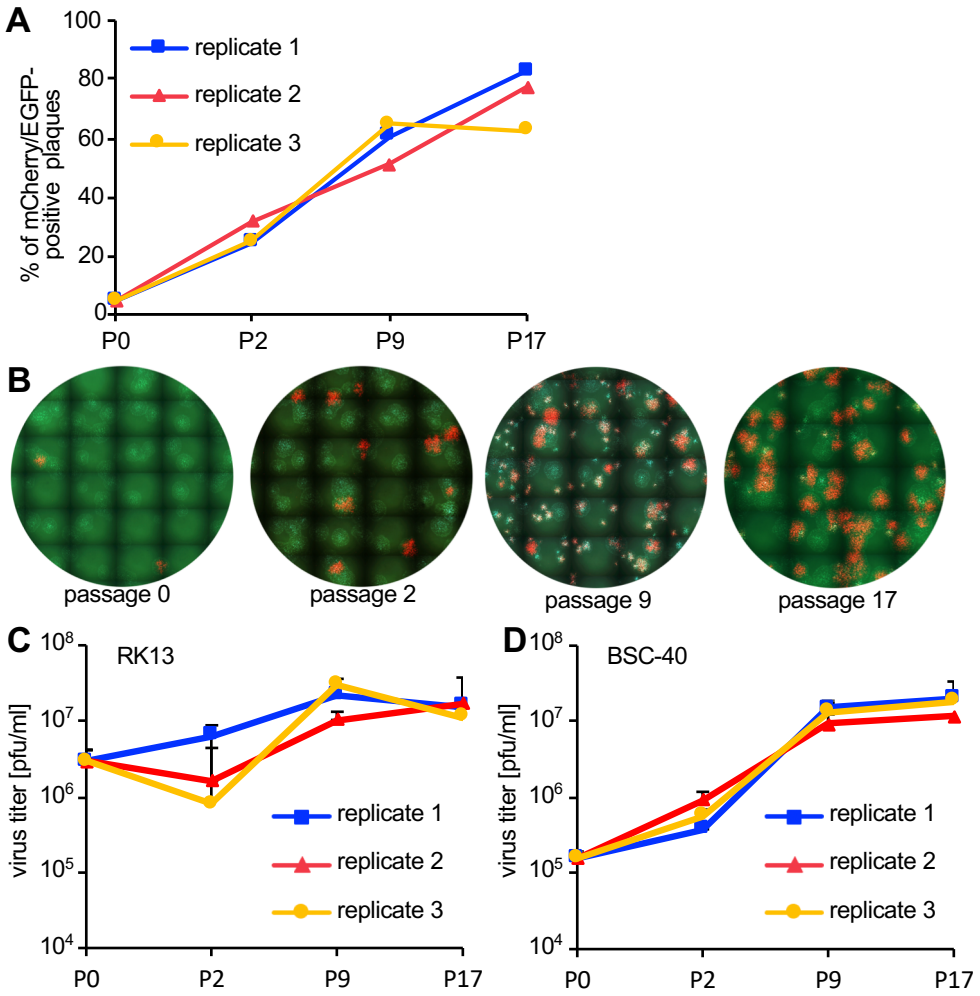


Figure 4. Fitness increase of HGT3 after serial passaging. (A) The percentage of mCherry-positive foci formed on RK13+E3+K3 cells by three serially passaged replicates of HGT3. Percentages were determined for the indicated passages (P). (**B**) Increase in double-positive mCherry and EGFP expressing 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.

similarities were detected at the breakpoints, indicating non-homologous recombination between the parent 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

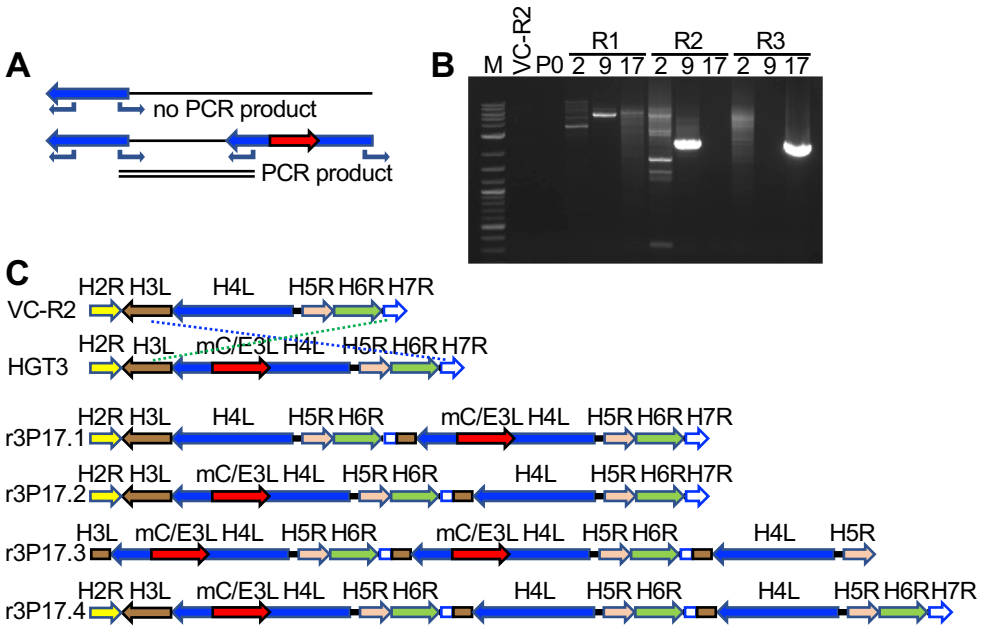


Figure 5. Recombination of complementing viruses after serial passaging. (**A**) PCR strategy to detect gene arrays with outward-facing primers. (**B**) Representative PCR products amplified with outward-facing primers for VC-R2 and three independent HGT3 passages. (**C**) Recombination of complementing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

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).

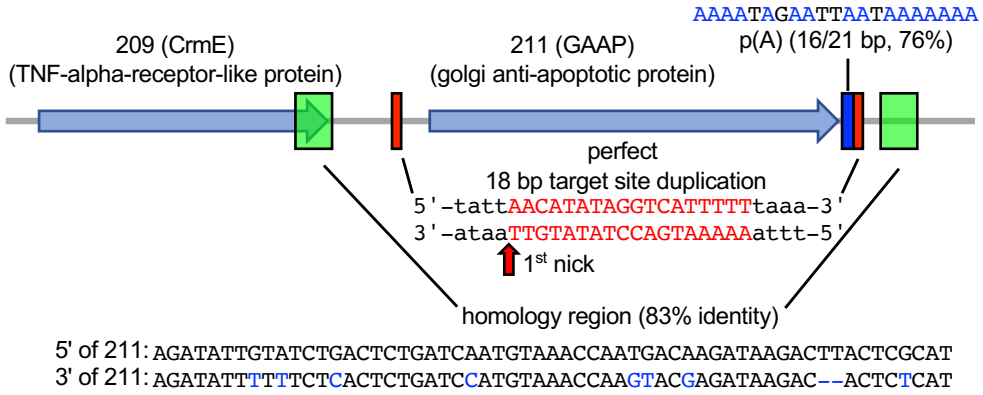
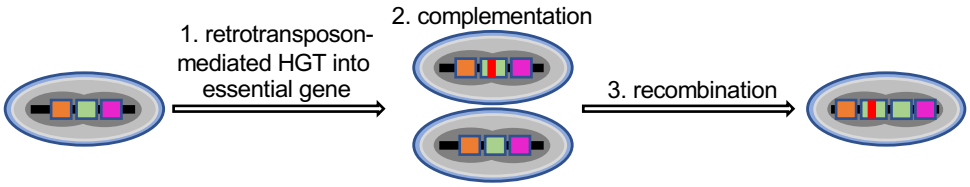


Figure 6. Signatures of retrotransposon-mediated HGT adjacent to cowpox virus gene 211. The 211 gene encodes Golgi anti-apoptotic protein (GAAP). Red boxes mark a perfect 18 bp putative target site duplication. p(A) indicates a polyadenylate stretch (16/21 bp) adjacent to the TSD. Green boxes show a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

271 **Discussion**
272
273 In this study, we show that HGT from the host cell into poxviruses can be recapitulated in an
274 experimental system. All HGT events showed evidence for an RNA-mediated mechanism, likely
275 mediated by LINE-1 retrotransposons, as they contain TSD of approximately 16 bp surrounding
276 the integration sites, as well as spliced-out introns and untemplated poly(A) tracts 3' of the
277 integrated genes. This observation shows that poxvirus evolution can be impacted by host
278 transposable elements providing them with selective advantages and accelerating their evolution.
279 It also defines a new function for retrotransposons, in addition to their role in genomic and
280 epigenetic diversification of the host (Beck et al., 2011).
281 Even though phylogenetically predicted HGT candidates in VACV are distributed throughout the
282 genome (Figure 1D) (Hughes and Friedman, 2005, Bratke and McLysaght, 2008), a common
283 notion in the field is that uptake of new genes into poxviruses would be biased towards the ends
284 of poxvirus genomes, because the central genome region contains more essential genes and shorter
285 intergenic spaces and thus insertions there would be less likely to result in viable progeny (Yao
286 and Evans, 2001, Lefkowitz et al., 2006). Also, poxvirus lineage-specific genes tend to be more
287 often found near the termini of poxvirus genomes (McLysaght et al., 2003). However, the
288 integration sites that we discovered in this study were distributed throughout the VACV genome,
289 which indicates that integration *per se* is not disfavored in the central genomic region. In fact, our
290 results show that HGT into essential poxvirus genes can result in viable progeny virus in a process
291 that we call cascade evolution. This process sequentially combines HGT, virus complementation
292 and recombination, three mechanisms of virus evolution (Figure 7). Our PacBio analyses identified
293 tandem copies of the integrated gene and intact gene, in both orientations relative to each other.
294 However, at the population level the genomic structure surrounding this locus was complex. For
295 example, we identified loci including both intact H4L and the horizontally acquired mCherry-E3L
296 in which multiple copies of each gene were found in *cis*. Furthermore, in some genomes we
297 identified three or more copies of some genes adjacent to the integration site. In addition to the
298 immediate impact of generating a replication-competent virus through recombination, these
299 variant products of recombination might provide the raw material for subsequent sub- or
300 neofunctionalization of the additional gene copies (Bayer et al., 2018).
301



302
303 **Figure 7. Cascade virus evolution after horizontal gene transfer into an essential gene.** After
304 retrotransposon-mediated horizontal gene integration into an essential locus (1) the virus initially
305 depends on a helper virus to complement the essential gene (2), which is followed by
306 recombination and results in a single virus with increased fitness.
307
308 The majority of the HGT viruses that required complementation with the parental virus formed
309 larger plaques in RK13 cells as compared to BSC-40 cells. This difference may indicate that the
310 complementation is more efficient in RK13 cells. In line with this explanation, we observed that

311 the parent HGT3 population replicated poorly in BSC-40 cells, and the passaged population
312 replicated about 100-fold more efficiently, whereas in RK13 cells, the parent HGT3 population
313 replicated better and only gained a modest increase in replication efficiency during passaging. It
314 is striking that for most observed cases of virus complementation, the ratio of parent to HGT virus
315 was skewed to the parent virus, in some cases making up more than 90% of the population (Figure
316 3B). This observation suggests that there may be competition for cellular resources or viral gene
317 products between the co-infecting viruses. While mCherry-E3 localizes to the cytoplasm and likely
318 antagonizes PKR equally well for each co-infecting virus, the transcription and translation of many
319 other viral gene products are tightly associated with viral factories derived from a single genome
320 (Katsafanas and Moss, 2007). Conceptually, this competition may occur when the complementing
321 molecule is required to diffuse or to be transported into another viral factory to exert its effect, or
322 if the viral gene product is not delivered until virus factories fuse late in the replication cycle
323 (Paszkowski et al., 2016).
324 An interesting observation was that in one of the HGT isolates, mCherry-E3L was fused to a
325 polyadenylated fragment of the cellular U2 snRNA. This genetic structure may have been
326 generated by the fusion of mCherry-E3L RNA with U2 snRNA prior to integration. This
327 observation suggests that other cellular genes could piggyback during HGT, which suggests in rare
328 instances multiple genes may be integrated into a single genome. Of note, snRNAs encoding DNA
329 including U1, U2, U5 and U6 can be found in host genomes fused to LINE-1 as a result of RNA
330 ligation (Buzdin et al., 2003, Garcia-Perez et al., 2007, Doucet et al., 2015, Moldovan et al., 2019).
331 However, it is unclear what, if any, impact the potential to integrate multiple genes would have on
332 viral evolution.
333 This work is the first experimental evidence of horizontal gene transfer into poxviruses. The
334 experimental system described here can be extended to study the evolution of transgenes after
335 HGT. In addition, it can be adapted to elucidate the mechanisms of HGT in other groups of viruses
336 for which HGT is believed to play an important evolutionary role. While we only detected RNA-
337 mediated HGT facilitated by LINE-1, it is possible that in other viruses or cell types additional
338 RNA or DNA transposons, or recombination might be important. Similarly, other facilitators, such
339 as endogenous or exogenous retroviruses may also promote HGT and increase virus fitness. Taken
340 together, our work demonstrates that cascade evolution is a powerful and rapid mechanism to
341 integrate new genetic material in poxviral genomes. It also demonstrates that HGT in poxviruses
342 is traceable in experimental settings, and that this mechanism may be relevant to other viruses,
343 which appropriate host genes for their own ends in the ongoing host-virus arms race.
344
345
346 **Materials and Methods**
347
348 **Cell lines**
349 RK13 cells (ATCC CCL-37) and BSC-40 (ATCC CRL-2761) cell lines were kindly provided by
350 Dr. Bernard Moss and Dr. Adam Geballe, respectively. RK13 cells expressing E3 and K3
351 (designated RK13+E3+K3) were previously described (Rahman et al., 2013). All cell lines were
352 maintained in Dulbecco's Modified Essential Medium (DMEM, Life Technologies) supplemented
353 with 5% fetal bovine serum (FBS, Fisher) and 25µg/ml gentamycin (Quality Biologicals) at 37˚C
354 and 5%CO_2. RK13+E3+K3 cells were supplemented with 500µg/ml geneticin and 300µg/ml
355 zeocin (Life Technologies). Cell lines were seeded in 6-well (at 5 X 10^5 cells/well) or 12-well (at
356 2.5 X 10^5 cells/well) plates 24 hours before infections.

357 RK13 cells expressing mCherry-E3 (RK13-mCherry-E3L) were generated by stable transfection
358 of *ApaL*I-linearized pmCherry-E3L using GenJet (SignaGen) according to the manufacturer's
359 protocol. The pmCherry-E3L vector contains, in 5' to 3' order, the SV40 promoter, rabbit β-globin
360 intron, a synthetic poxvirus early/late promoter, a mCherry-E3L fusion gene, cloned into the
361 backbone of pEGFP-N1 (Clontech). During this process, the CMV promoter and EGFP were
362 removed from pEGFP-N1. Transfected cells were selected with 500 µg/ml geneticin (G418,
363 Invitrogen) for two weeks. Individual clones were isolated by seeding cells at a low density (0.3-
364 1 cell/ well) in 96-well plates. Wells containing a single cell, verified by fluorescent microscopy,
365 were selected to establish clonal cell lines. Genomic DNA was prepared to verify the structure of
366 the integrated cassette by PCR with primer JR253-intron-4F (5'-TTC CAG AAG TAG TGA AGA
367 GGC TT-3') and JR254-E3L-5R (5'-AGC AAG TAA AAC CTC TAC AAA TG-3'). All
368 subsequent experiments were carried out with one RK13+mCherry-E3L cell clone that showed
369 robust mCherry-E3 expression (Supplementary figure 1).
370
371 **Viruses**
372 Vaccinia virus VC-2 (Copenhagen strain) and vP872 (ΔK3L) (Beattie et al., 1991) were kindly
373 provided by Dr. Bertram Jacobs and propagated in RK13 cells. VC-R2, which lacks both E3L and
374 K3L (Brennan et al., 2014), was propagated in RK13+E3+K3 cells. VC-2, vP872, and VC-R2
375 were purified by zonal sucrose gradient centrifugation as previously described (Cotter et al., 2017).
376 Virus clones HGT1 to HGT20 were plaque purified three times in RK13 cells with carboxymethyl
377 cellulose (CMC, 1% final concentration in DMEM+5%FBS) overlay. Virus clones HGT1 to
378 HGT7 were purified by zonal sucrose gradient centrifugation. For HGT8 to HGT20, crude cell
379 lysates were used as stocks. All viruses were titered in RK13+E3+K3 cells on 12-well plates. Virus
380 samples were 10-fold serially diluted and cells were infected in duplicate. One hour after infection,
381 the infecting medium was replaced with CMC and incubated for 3 days at 37˚C in 5% CO_2. The
382 CMC overlays were then aspirated, and plaques were identified by staining the monolayers with
383 0.1% crystal violet in 20% methanol.
384 Serial passaging of HGT3 was performed as described (Elde et al., 2012, Brennan et al., 2014).
385 Briefly, confluent RK13 cells in 6-well plates were infected with plaque-purified HGT3 virus
386 (MOI = 0.1) in triplicate. 48 hours after infection (hpi), cell lysates were collected, freeze-thawed
387 three times, sonicated, and titered on RK13+E3+K3 cells before serially infecting new RK13 cells.
388 Genomic DNA from passaged viruses was isolated from infected RK13 cells (MOI = 1.0) as
389 previously described (Esposito et al., 1981). RK13 and BSC-40 cells were infected with passaged
390 virus populations in duplicate at MOI = 0.01 for 48 hours and viruses were titered on
391 RK13+E3+K3 cells in duplicate.
392
393 **Screening for viruses that captured mCherry-E3L**
394 RK13-mCherry-E3L cells were infected with VC-R2 (MOI = 0.01) for three days. Viruses were
395 titered on RK13+E3+K3 cells. Subsequently, viruses were screened on confluent monolayers of
396 RK13 cells in 6-well plates. Cells were infected at MOI = 0.2 to screen for fluorescence to avoid
397 cell death observed at higher MOIs. For most isolates, plaques were identified after 3 days but
398 some took as long as 5 days. Initially, plates were screened for double-positive mCherry and EGFP
399 expressing plaques using an inverted microscope (Leica DMi8 automated, total magnification
400 X6.25) using the FITC filter (460–500 nm) to visualize EGFP and the Texas Red filter (540 – 580
401 nm) to visualize mCherry. In order to increase throughput, screening for double-positive plaques
402 was then performed with an automated microscope (EVOS FL Auto 2, Thermo Fisher Scientific)

403 using comparable filters. Double-positive virus plaques were picked using a sterile pipette tip.
404 Harvested plaques were subjected to three rounds of freeze-thaw followed by sonication (2 times
405 30s). These viruses were then serially diluted to infect RK13 cells for a total of three rounds of
406 plaque purification. After the third round of plaque purification, viruses were amplified in RK13
407 cells and titered on RK13+E3+K3 cells. Viral genomic DNA was isolated from infected RK13
408 cells according to a published protocol (Seto et al., 1987).
409

410 **Detection of mCherry-E3L integration sites in the virus genome**

411 The mCherry-E3L integration sites in the purified HGT virus genomes were located by inverse
412 PCR amplification and by PacBio sequencing in the case of HGT14 and HGT19. For inverse PCR,
413 we used a *Xba*I restriction site in between mCherry-E3L. VC-2 contains 123 *Xba*I sites. Viral DNA
414 from HGT isolates was digested with *Xba*I, and then ligated by T4 DNA ligase (NEB) to make
415 circular DNA. After ligation, E3L-specific, outward-facing primers: HGT-E3L-1F (5'-TGG CAG
416 TAG ATA AAC TTC TTG GTT ACG-3') and HGT-E3L-1R (5'-AGC CTC ACA CAC AAT CTC
417 TGC G-3') were used to amplify DNA circles containing mCherry-E3L and neighboring genomic
418 viral DNA. These amplicons were gel purified and Sanger sequenced to define the approximate
419 mCherry-E3L integration sites, because the long poly(A) tails precluded identification of the virus
420 sequences 3' of the poly(A) tails. To define integration sites and target site duplications, primers
421 Supplementary table 2) from the VACV genome that are adjacent to the integration sites were
422 designed to amplify whole integrated captured sequences, which were then Sanger sequenced.
423 Because for HGT5 and HGT19, PCR amplification of the whole insert did not work, two separate
424 PCR amplification with primers located in E3L and the adjacent vaccinia virus genes were
425 performed and sequenced. Insertion start sites and polyadenylation sites of transgenes in HGT
426 viruses are shown in supporting Figure 5.
427 To determine if mCherry-E3L integrated into both sites of the ITR in the C19L/B25R locus in
428 HGT15, PCR with DNA of HGT15 and VC-R2, as control, was performed with primers that
429 combined primer JR179-C19L-2F (5'-CGA GGA CTA TGT TTG GTA TAC TG-3') located in
430 the ITR, towards the termini of the integration site and primers JR180-C13L-2R(5'-CTG GAC
431 TAT CCA CAC CTG-3) for the "left" region of the VACV genome, and JR181-B19R-1F (5'-
432 CTA GTA GAG GCG GTA TCA C-3'), for the "right" region of the VACV genome outside of
433 the ITR. Both primer combinations showed larger bands when DNA HGT15 DNA was used as
434 template, in comparison to VC-R2 DNA, which is indicative of mCherry-E3L in both ITR copies
435 (Supporting Figure 6).
436

437 **Measurement of plaque sizes**

438 Confluent 6-well plates of either RK13 or BSC-40 cells were infected with 100 PFU of either each
439 HGT isolate, VC-R2, or vP872 in triplicate. After one hour, cells were overlaid with CMC and
440 incubated at 37℃ in 5% CO2 for 3 days followed by crystal violet staining. The stained plates and
441 an adjacent ruler were imaged with the EVOS FL 2 microscope (Thermo Fisher Scientific) using
442 a bright-field filter. Plaque sizes were calculated using Adobe Photoshop CC 2019. To calculate
443 this, a global-scale was set by measuring the average number pixels in a 1 mm-distance from the
444 imaged rulers (133 pixels = 1mm, SD = 0.98 pixels) using the Adobe Photoshop CC 2019 ruler
445 tool. The Photoshop quick selection tool was used to mark the perimeter of individual plaques
446 followed by area measurement from the analysis tab. Between 30 and 387 plaques were measured
447 for each virus. Kruskal-Wallis one-way ANOVA followed by Dunn's test of multiple comparisons

448 were used to compare the mean plaque area of HGT samples with vP872 (alpha =.05) by GraphPad
449 Prism (version 8.3.0). Graphs are presented as mean ± SD and adjusted p-values are shown.
450

451 **Identifying virus complementation**

452 Confluent 6-well plates of RK13+E3+K3 cells were infected with 100 PFU of HGT1 through
453 HGT20, and the passaged HGT3 from passages 2, 9, and 17. The percentage of mCherry and EGFP
454 double-positive plaques were determined three days after infection using the automated EVOS FL
455 Auto 2 microscope (Thermo Fisher Scientific). Images were taken at 4 X 0.13 NA magnification
456 by using filter channels Texas red (540–580 nm) for mCherry and EGFP (465–495 nm) to visualize
457 EGFP. Single-channel autofocus was used with phase-contrast using the first field to autofocus all
458 other fields. 100% of well areas were photographed, stitched and tiled by using the EVOS FL Auto
459 2 software and analyzed to quantify the number of single-positive EGFP plaques and double-
460 positive EGFP and mCherry expressing plaques. For each virus sample more than 86 plaques were
461 analyzed.
462 For the VC-R2 and HGT13 complementation assay, confluent 6-well plates of RK13 cells were
463 co-infected in duplicate with each virus (MOI = 0.05 for each virus), or with HGT13 alone (MOI
464 = 0.1). Images for mCherry fluorescence were taken 3 days after infection with EVOS FL Auto 2.
465

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

467 Genomic DNA from HGT3 passages 0, 2, 9 and 17 as well as VC-R2 , were used as templates for
468 PCR using outward-facing primers: JR81-H4L-2F (5'-GTC TAG TAG ATA TGC TTT TAT TTT
469 TG-3') and JR80-H4L-2R (5'- CGA AAA TAT AAC TCG TAT TAA AGA G-3') or JR42-H4L-
470 3F (5'- CAC GGA GAT GGC GTA TTT AAG AG -3') and JR88-H4L-3R (5'- GAG CTA ACG
471 TGT GAC GAA G-3'). PCR amplified products were cloned into the pCR2.1-TOPO-TA
472 (Invitrogen) vector for amplification and Sanger sequencing using M13F (-21) and M13R primers
473 (UC Davis sequencing facility).
474

475 **Library preparation, PacBio CCS sequencing and data analysis**

476 HiFi SMRTbell library construction and sequencing were performed using Sequel II System 2.0
477 with P2/C2 (polymerase 2.0 and chemistry 2.0) chemistry at UC Davis DNA Technologies Core
478 according to the manufacturer's instructions (Pacific Biosciences). Before library preparation, the
479 quality of genomic DNA was evaluated by pulsed-field gel electrophoresis (Sage Science Pippin
480 Pulse) and genomic DNA concentration was quantified using a Qubit 3.0 Fluorometer (Life
481 Technologies Q33216). Briefly, 10 µg of each viral genomic DNA was used for HiFi SMRT-bell
482 library preparation by using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences 100-
483 938-900). The genomic DNA was sheared to ~15-20 kb size by Megaruptor (Diagenode
484 B06010001). The sheared DNA samples were then concentrated and purified by AMPure PB
485 Beads (Pacific Biosciences 100-265-900) by adding 0.45 X volume of AMPure PB magnetic beads
486 to each sheared DNA samples and subsequent 80% ethanol precipitation and elution with 100 µl
487 elution buffer. Following genomic DNA shearing and purification, each genomic DNA sample
488 was subjected to single-strand DNA (ssDNA) overhang removal and DNA damage repair. In short,
489 ssDNA overhangs were removed using ssDNA overhang removal reaction mix (DNA prep buffer,
490 NAD, DNA prep additive and DNA prep enzyme) and incubated at 37˚C for 15 mins. The DNA
491 fragments of each sample were then repaired by using a DNA damage repair mix and incubated at
492 37˚C for 30 mins before proceeding to End repair/A-tailing by mixing with 1 X End prep mix and
493 incubated at 20˚C for 10 mins followed by 30 mins incubation at 65˚C. After that, different

494 barcoded adapters were ligated to each fragmented viral DNA sample by mixing with adapter
495 ligation reaction mix (Overhang adapter v3, ligation mix, ligation additive, and ligation enhancer)
496 and incubated at 20˚C for an hour followed by incubation at 65˚C for 10 mins to inactivate the
497 ligase. After ligation reaction, each adapter-ligated SMRTbell library was treated with
498 exonucleases to remove damaged or un-ligated DNA fragments for one hour at 37˚C followed by
499 a second purification step with 0.45 X AMPure® PB Beads. After purification, the libraries were
500 multiplexed into 2 library pools for size selection: 6-plex pool 1 and 10-plex pool 2. These pooled
501 SMRTbell libraries were then purified by mixing with 0.45 X volume of AMPure PB beads
502 followed by 80 % ethanol wash and elution with 31 µl EB for size selection. Size selection was
503 done using the BluePippin Size-Selection System (Sage Science BLU0001) to remove <6kb
504 SMRTbell templates and eluted the size selected libraries into two size fractions: 9-13 kb and >15
505 kb per pooled library. The only fraction with >15 kb SMRTbell libraries was purified with 0.50X
506 AMPure® PB Beads and used for primer annealing and polymerase binding. Sequencing Primer
507 v2 (PN 101-847-900) was annealed to each size selected SMRTbell library fraction of >15 kb
508 length at primer: template ratio of 20:1 by denaturation step at 80 °C for 150 mins followed by
509 slow cooling at 0.1 °C/s to 25 °C. Prior to sequencing, the polymerase-template complex was
510 bound to the P2 enzyme with 10:1 polymerase to the SMRTbell template ratio for 4 hours. The
511 polymerase bound SMRTbell libraries were then sequenced by placing libraries in an 8M SMRT
512 cell at a sequencing concentration of 63 pM and 30-hour movie run time in a Sequel II System 2.0
513 machine.
514 For data analysis, PacBio CCS reads were first demultiplexed using lima from SMRT Link
515 8.0 command-line tools. For all samples, the location of gene H4L and mCherry-E3L sequences
516 were identified in each CCS reads by using BLAST (v2.9+) (Camacho et al., 2009). The search
517 results from BLAST were processed using custom scripts to identify the different modes of
518 mCherry-E3L insertions, as well as the alterations of gene structures of H4L. Representative
519 sequences of CCS reads were extracted to showcase selected scenarios. Those sequences were
520 further analyzed using Seqbuilder, Seqman Pro, Editseq of DNASTAR software package
521 (DNASTAR, Madison, WI). A logo plot from 19 target site duplications (first 6 nucleotides 5' and
522 3' of the TSD) and 10 nucleotides of adjacent sequences was created with WebLogo (Crooks et
523 al., 2004). PacBio CCS reads were submitted to ArrayExpress (accession: E-MTAB-9682).
524
525
526 **Acknowledgments**
527
528 This work was supported by grant AI146915 (to S.R.) from the National Institute of Allergy and
529 Infectious Diseases, National Institutes of Health. We thank Dr. Adam Geballe, Dr. Bertram
530 Jacobs and Dr. Bernard Moss for cell lines and viruses, and all current and former Rothenburg lab
531 members for helpful discussions.
532
533

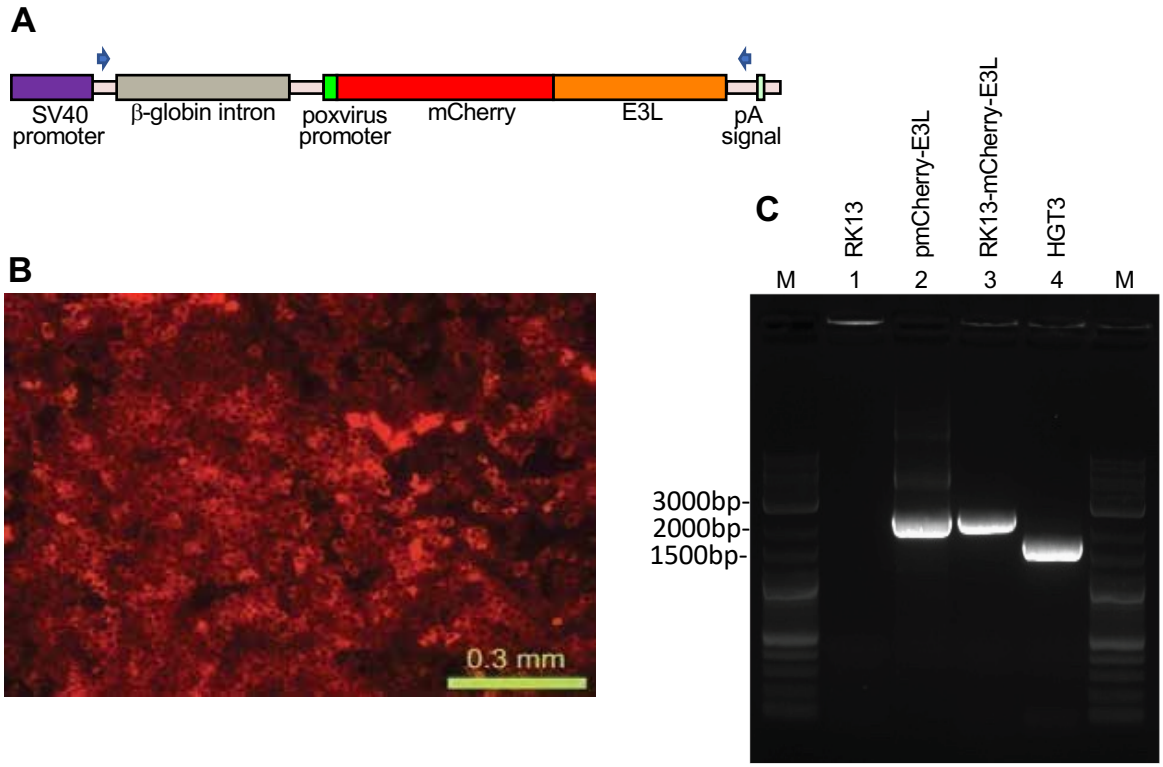
534 # References
535
536 BAYER, A., BRENNAN, G. & GEBALLE, A. P. 2018. Adaptation by copy number variation in monopartite
537 viruses. *Curr Opin Virol,* 33**,** 7-12.
538 BEATTIE, E., TARTAGLIA, J. & PAOLETTI, E. 1991. Vaccinia virus-encoded eIF-2 alpha homolog abrogates
539 the antiviral effect of interferon. *Virology,* 183**,** 419-22.
540 BECK, C. R., GARCIA-PEREZ, J. L., BADGE, R. M. & MORAN, J. V. 2011. LINE-1 elements in structural variation
541 and disease. *Annu Rev Genomics Hum Genet,* 12**,** 187-215.
542 BLASCO, R. & MOSS, B. 1991. Extracellular vaccinia virus formation and cell-to-cell virus transmission are
543 prevented by deletion of the gene encoding the 37,000-Dalton outer envelope protein. *J Virol,* 65**,**
544 5910-20.
545 BRATKE, K. A. & MCLYSAGHT, A. 2008. Identification of multiple independent horizontal gene transfers
546 into poxviruses using a comparative genomics approach. *BMC Evol Biol,* 8**,** 67.
547 BRATKE, K. A., MCLYSAGHT, A. & ROTHENBURG, S. 2013. A survey of host range genes in poxvirus
548 genomes. *Infect Genet Evol,* 14**,** 406-25.
549 BRENNAN, G., KITZMAN, J. O., ROTHENBURG, S., SHENDURE, J. & GEBALLE, A. P. 2014. Adaptive gene
550 amplification as an intermediate step in the expansion of virus host range. *PLoS Pathog,* 10**,**
551 e1004002.
552 BROYLES, S. S. & FESLER, B. S. 1990. Vaccinia virus gene encoding a component of the viral early
553 transcription factor. *J Virol,* 64**,** 1523-9.
554 BURLES, K., IRWIN, C. R., BURTON, R. L., SCHRIEWER, J., EVANS, D. H., BULLER, R. M. & BARRY, M. 2014.
555 Initial characterization of vaccinia virus B4 suggests a role in virus spread. *Virology,* 456-457**,** 108-
556 20.
557 BUZDIN, A., GOGVADZE, E., KOVALSKAYA, E., VOLCHKOV, P., USTYUGOVA, S., ILLARIONOVA, A., FUSHAN,
558 A., VINOGRADOVA, T. & SVERDLOV, E. 2003. The human genome contains many types of chimeric
559 retrogenes generated through in vivo RNA recombination. *Nucleic Acids Res,* 31**,** 4385-90.
560 CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L.
561 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.
562 CHANG, H. W., WATSON, J. C. & JACOBS, B. L. 1992. The E3L gene of vaccinia virus encodes an inhibitor of
563 the interferon-induced, double-stranded RNA-dependent protein kinase. *Proc Natl Acad Sci U S*
564 *A,* 89**,** 4825-9.
565 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.,
566 TZOU, D. M., WANG, H. C. & CHANG, W. 2019. Vaccinia viral A26 protein is a fusion suppressor of
567 mature virus and triggers membrane fusion through conformational change at low pH. *PLoS*
568 *Pathog,* 15**,** e1007826.
569 CHILD, S. J., PALUMBO, G. J., BULLER, R. M. & HRUBY, D. E. 1990. Insertional inactivation of the large
570 subunit of ribonucleotide reductase encoded by vaccinia virus is associated with reduced
571 virulence in vivo. *Virology,* 174**,** 625-9.
572 COTTER, C. A., EARL, P. L., WYATT, L. S. & MOSS, B. 2017. Preparation of Cell Cultures and Vaccinia Virus
573 Stocks. *Curr Protoc Mol Biol,* 117**,** 16 16 1-16 16 18.
574 CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator.
575 *Genome Res,* 14**,** 1188-90.
576 DAR, A. C. & SICHERI, F. 2002. X-ray crystal structure and functional analysis of vaccinia virus K3L reveals
577 molecular determinants for PKR subversion and substrate recognition. *Mol Cell,* 10**,** 295-305.
578 DAVIES, M. V., CHANG, H. W., JACOBS, B. L. & KAUFMAN, R. J. 1993. The E3L and K3L vaccinia virus gene
579 products stimulate translation through inhibition of the double-stranded RNA-dependent protein
580 kinase by different mechanisms. *J Virol,* 67**,** 1688-92.

581 DOUCET, A. J., DROC, G., SIOL, O., AUDOUX, J. & GILBERT, N. 2015. U6 snRNA Pseudogenes: Markers of
582 Retrotransposition Dynamics in Mammals. *Mol Biol Evol,* 32**,** 1815-32.
583 EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P., BETTMAN,
584 B., BIBILLO, A., BJORNSON, K., CHAUDHURI, B., CHRISTIANS, F., CICERO, R., CLARK, S., DALAL, R.,
585 DEWINTER, A., DIXON, J., FOQUET, M., GAERTNER, A., HARDENBOL, P., HEINER, C., HESTER, K.,
586 HOLDEN, D., KEARNS, G., KONG, X., KUSE, R., LACROIX, Y., LIN, S., LUNDQUIST, P., MA, C., MARKS,
587 P., MAXHAM, M., MURPHY, D., PARK, I., PHAM, T., PHILLIPS, M., ROY, J., SEBRA, R., SHEN, G.,
588 SORENSON, J., TOMANEY, A., TRAVERS, K., TRULSON, M., VIECELI, J., WEGENER, J., WU, D., YANG,
589 A., ZACCARIN, D., ZHAO, P., ZHONG, F., KORLACH, J. & TURNER, S. 2009. Real-time DNA
590 sequencing from single polymerase molecules. *Science,* 323**,** 133-8.
591 ELDE, N. C., CHILD, S. J., EICKBUSH, M. T., KITZMAN, J. O., ROGERS, K. S., SHENDURE, J., GEBALLE, A. P. &
592 MALIK, H. S. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral
593 defenses. *Cell,* 150**,** 831-41.
594 ESNAULT, C., MAESTRE, J. & HEIDMANN, T. 2000. Human LINE retrotransposons generate processed
595 pseudogenes. *Nat Genet,* 24**,** 363-7.
596 ESPOSITO, J., CONDIT, R. & OBIJESKI, J. 1981. The preparation of orthopoxvirus DNA. *J Virol Methods,* 2**,**
597 175-9.
598 GAMMON, D. B., GOWRISHANKAR, B., DURAFFOUR, S., ANDREI, G., UPTON, C. & EVANS, D. H. 2010.
599 Vaccinia virus-encoded ribonucleotide reductase subunits are differentially required for
600 replication and pathogenesis. *PLoS Pathog,* 6**,** e1000984.
601 GARCIA-PEREZ, J. L., DOUCET, A. J., BUCHETON, A., MORAN, J. V. & GILBERT, N. 2007. Distinct mechanisms
602 for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res,*
603 17**,** 602-11.
604 GILBERT, C. & CORDAUX, R. 2017. Viruses as vectors of horizontal transfer of genetic material in
605 eukaryotes. *Curr Opin Virol,* 25**,** 16-22.
606 GILBERT, N., LUTZ-PRIGGE, S. & MORAN, J. V. 2002. Genomic deletions created upon LINE-1
607 retrotransposition. *Cell,* 110**,** 315-25.
608 GROSS, C. H. & SHUMAN, S. 1996. Vaccinia virions lacking the RNA helicase nucleoside triphosphate
609 phosphohydrolase II are defective in early transcription. *J Virol,* 70**,** 8549-57.
610 GUBSER, C., BERGAMASCHI, D., HOLLINSHEAD, M., LU, X., VAN KUPPEVELD, F. J. & SMITH, G. L. 2007. A
611 new inhibitor of apoptosis from vaccinia virus and eukaryotes. *PLoS Pathog,* 3**,** e17.
612 HALLER, S. L., PENG, C., MCFADDEN, G. & ROTHENBURG, S. 2014. Poxviruses and the evolution of host
613 range and virulence. *Infect Genet Evol,* 21**,** 15-40.
614 HAND, E. S., HALLER, S. L., PENG, C., ROTHENBURG, S. & HERSPERGER, A. R. 2015. Ectopic expression of
615 vaccinia virus E3 and K3 cannot rescue ectromelia virus replication in rabbit RK13 cells. *PLoS One,*
616 10**,** e0119189.
617 HASSETT, D. E., LEWIS, J. I., XING, X., DELANGE, L. & CONDIT, R. C. 1997. Analysis of a temperature-
618 sensitive vaccinia virus mutant in the viral mRNA capping enzyme isolated by clustered charge-to-
619 alanine mutagenesis and transient dominant selection. *Virology,* 238**,** 391-409.
620 HEDENGREN-OLCOTT, M., BYRD, C. M., WATSON, J. & HRUBY, D. E. 2004. The vaccinia virus G1L putative
621 metalloproteinase is essential for viral replication in vivo. *J Virol,* 78**,** 9947-53.
622 HOWARD, A. R., SENKEVICH, T. G. & MOSS, B. 2008. Vaccinia virus A26 and A27 proteins form a stable
623 complex tethered to mature virions by association with the A17 transmembrane protein. *J Virol,*
624 82**,** 12384-91.
625 HUGHES, A. L. & FRIEDMAN, R. 2005. Poxvirus genome evolution by gene gain and loss. *Mol Phylogenet*
626 *Evol,* 35**,** 186-95.
627 IYER, L. M., BALAJI, S., KOONIN, E. V. & ARAVIND, L. 2006. Evolutionary genomics of nucleo-cytoplasmic
628 large DNA viruses. *Virus Res,* 117**,** 156-84.

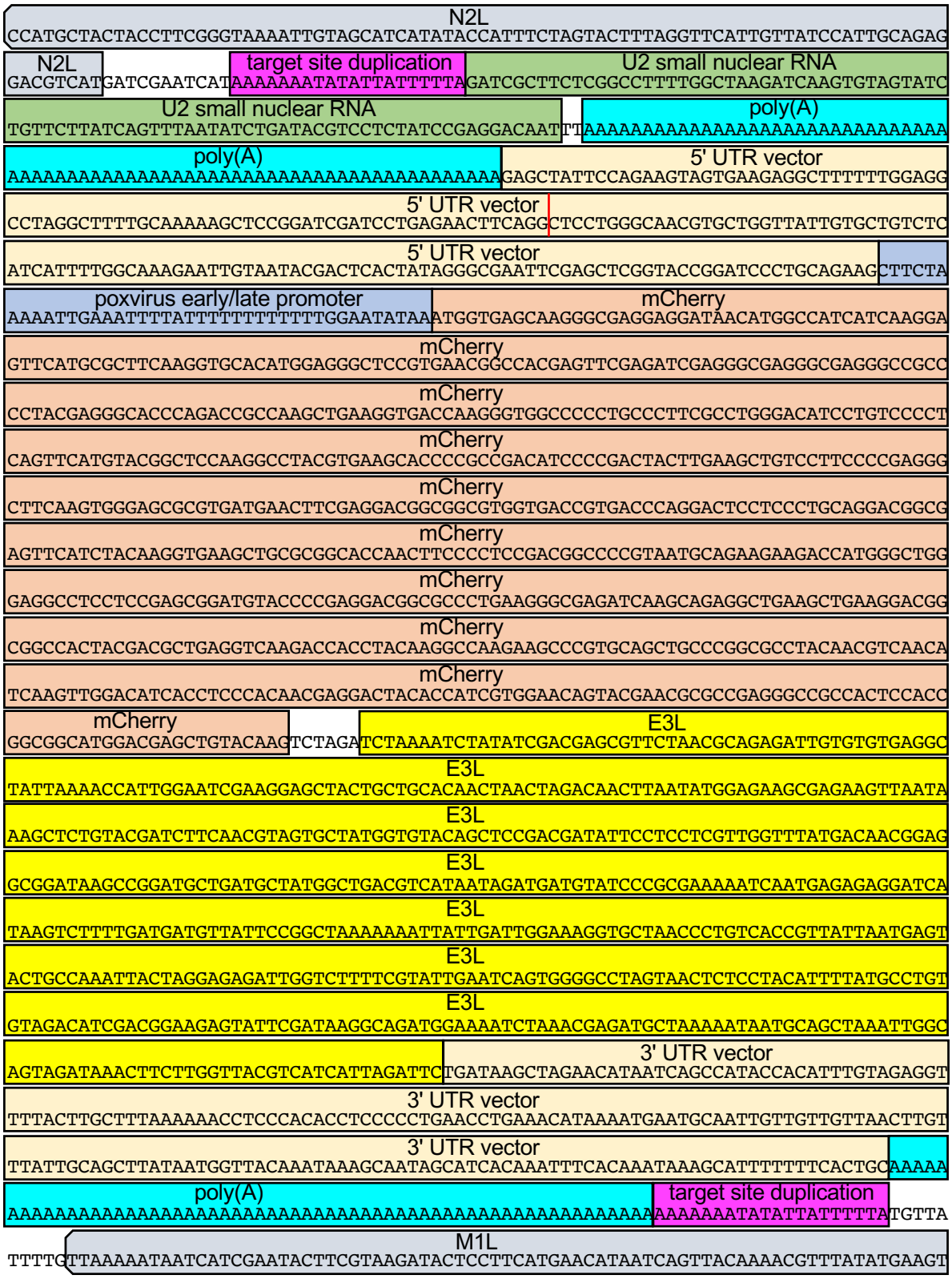
629 JAYAWARDANE, G., RUSSELL, G. C., THOMSON, J., DEANE, D., COX, H., GATHERER, D., ACKERMANN, M.,
630 HAIG, D. M. & STEWART, J. P. 2008. A captured viral interleukin 10 gene with cellular exon
631 structure. *J Gen Virol,* 89**,** 2447-55.
632 KANE, E. M. & SHUMAN, S. 1992. Temperature-sensitive mutations in the vaccinia virus H4 gene encoding
633 a component of the virion RNA polymerase. *J Virol,* 66**,** 5752-62.
634 KATSAFANAS, G. C. & MOSS, B. 2007. Colocalization of transcription and translation within cytoplasmic
635 poxvirus factories coordinates viral expression and subjugates host functions. *Cell Host Microbe,*
636 2**,** 221-8.
637 KEELING, P. J. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin*
638 *Genet Dev,* 19**,** 613-9.
639 KOJIMA, K. K. 2010. Different integration site structures between L1 protein-mediated retrotransposition
640 in cis and retrotransposition in trans. *Mob DNA,* 1**,** 17.
641 LANGLAND, J. O., CAMERON, J. M., HECK, M. C., JANCOVICH, J. K. & JACOBS, B. L. 2006. Inhibition of PKR
642 by RNA and DNA viruses. *Virus Res,* 119**,** 100-10.
643 LEFKOWITZ, E. J., WANG, C. & UPTON, C. 2006. Poxviruses: past, present and future. *Virus Res,* 117**,** 105-
644 18.
645 MCLYSAGHT, A., BALDI, P. F. & GAUT, B. S. 2003. Extensive gene gain associated with adaptive evolution
646 of poxviruses. *Proc Natl Acad Sci U S A,* 100**,** 15655-60.
647 MOLDOVAN, J. B., WANG, Y., SHUMAN, S., MILLS, R. E. & MORAN, J. V. 2019. RNA ligation precedes the
648 retrotransposition of U6/LINE-1 chimeric RNA. *Proc Natl Acad Sci U S A,* 116**,** 20612-20622.
649 OCHMAN, H., GERBER, A. S. & HARTL, D. L. 1988. Genetic applications of an inverse polymerase chain
650 reaction. *Genetics,* 120**,** 621-3.
651 PARRISH, S. & MOSS, B. 2006. Characterization of a vaccinia virus mutant with a deletion of the D10R gene
652 encoding a putative negative regulator of gene expression. *J Virol,* 80**,** 553-61.
653 PASZKOWSKI, P., NOYCE, R. S. & EVANS, D. H. 2016. Live-Cell Imaging of Vaccinia Virus Recombination.
654 *PLoS Pathog,* 12**,** e1005824.
655 PERKUS, M. E., GOEBEL, S. J., DAVIS, S. W., JOHNSON, G. P., NORTON, E. K. & PAOLETTI, E. 1991. Deletion
656 of 55 open reading frames from the termini of vaccinia virus. *Virology,* 180**,** 406-10.
657 PISKUREK, O. & OKADA, N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons
658 from reptiles to mammals. *Proc Natl Acad Sci U S A,* 104**,** 12046-51.
659 RAHMAN, M. M., LIU, J., CHAN, W. M., ROTHENBURG, S. & MCFADDEN, G. 2013. Myxoma virus protein
660 M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to
661 promote expanded host tropism and viral replication. *PLoS Pathog,* 9**,** e1003465.
662 SCHONRICH, G., ABDELAZIZ, M. O. & RAFTERY, M. J. 2017. Herpesviral capture of immunomodulatory host
663 genes. *Virus Genes,* 53**,** 762-773.
664 SEKIGUCHI, J. & SHUMAN, S. 1997. Ligation of RNA-containing duplexes by vaccinia DNA ligase.
665 *Biochemistry,* 36**,** 9073-9.
666 SENKEVICH, T. G., WYATT, L. S., WEISBERG, A. S., KOONIN, E. V. & MOSS, B. 2008. A conserved poxvirus
667 NlpC/P60 superfamily protein contributes to vaccinia virus virulence in mice but not to replication
668 in cell culture. *Virology,* 374**,** 506-14.
669 SETO, J., CELENZA, L. M., CONDIT, R. C. & NILES, E. G. 1987. Genetic map of the vaccinia virus HindIII D
670 Fragment. *Virology,* 160**,** 110-9.
671 WILCOCK, D., DUNCAN, S. A., TRAKTMAN, P., ZHANG, W. H. & SMITH, G. L. 1999. The vaccinia virus A4OR
672 gene product is a nonstructural, type II membrane glycoprotein that is expressed at the cell
673 surface. *J Gen Virol,* 80 (Pt 8)**,** 2137-2148.
674 WILSON, E. B. & BROOKS, D. G. 2011. The role of IL-10 in regulating immunity to persistent viral infections.
675 *Curr Top Microbiol Immunol,* 350**,** 39-65.

676 WU, S. & KAUFMAN, R. J. 1997. A model for the double-stranded RNA (dsRNA)-dependent dimerization
677 and activation of the dsRNA-activated protein kinase PKR. *J Biol Chem,* 272**,** 1291-6.
678 YAO, X. D. & EVANS, D. H. 2001. Effects of DNA structure and homology length on vaccinia virus
679 recombination. *J Virol,* 75**,** 6923-32.
680 YEH, W. W., MOSS, B. & WOLFFE, E. J. 2000. The vaccinia virus A9L gene encodes a membrane protein
681 required for an early step in virion morphogenesis. *J Virol,* 74**,** 9701-11.
682
683

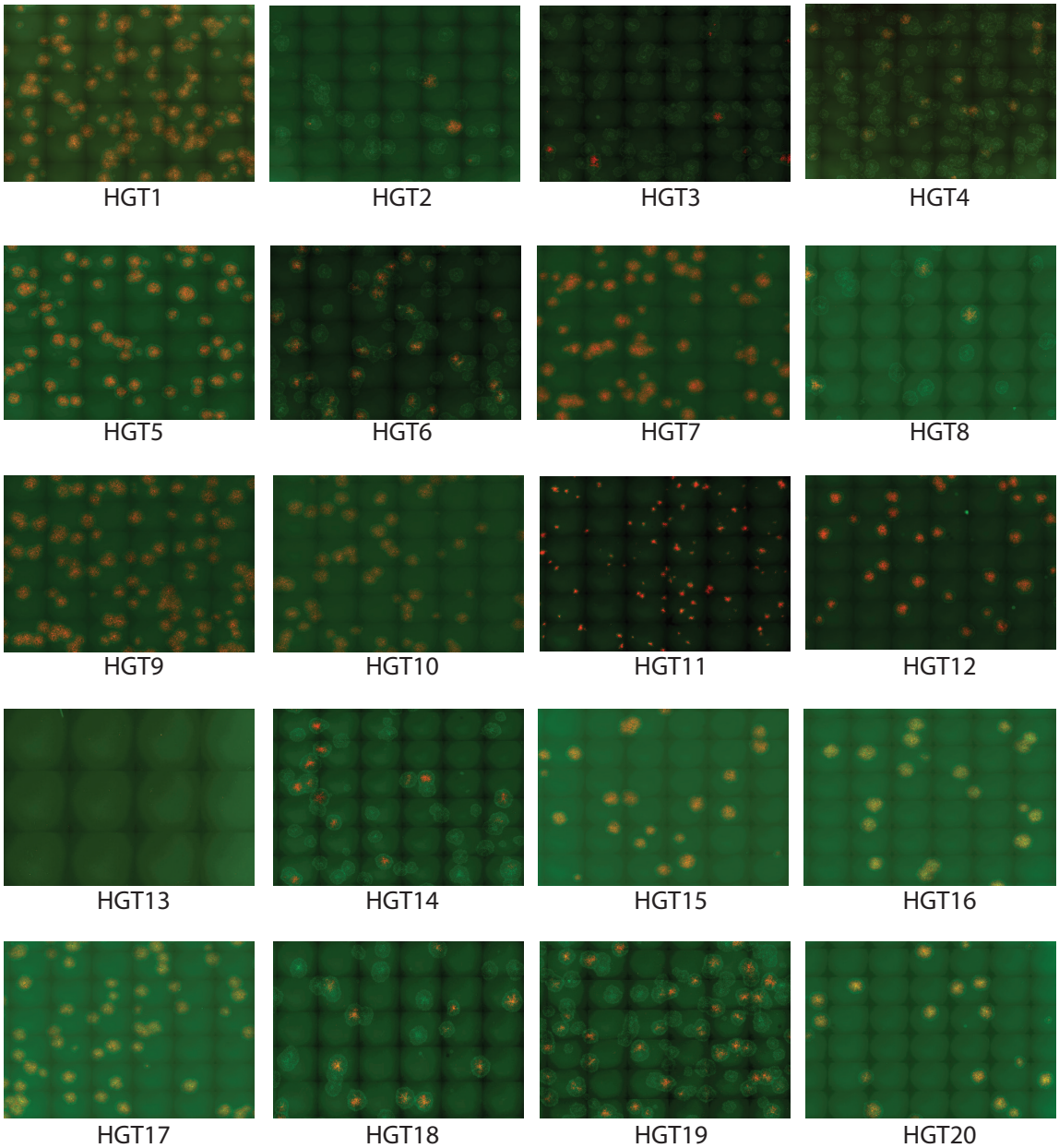
684 **Supporting information**
685



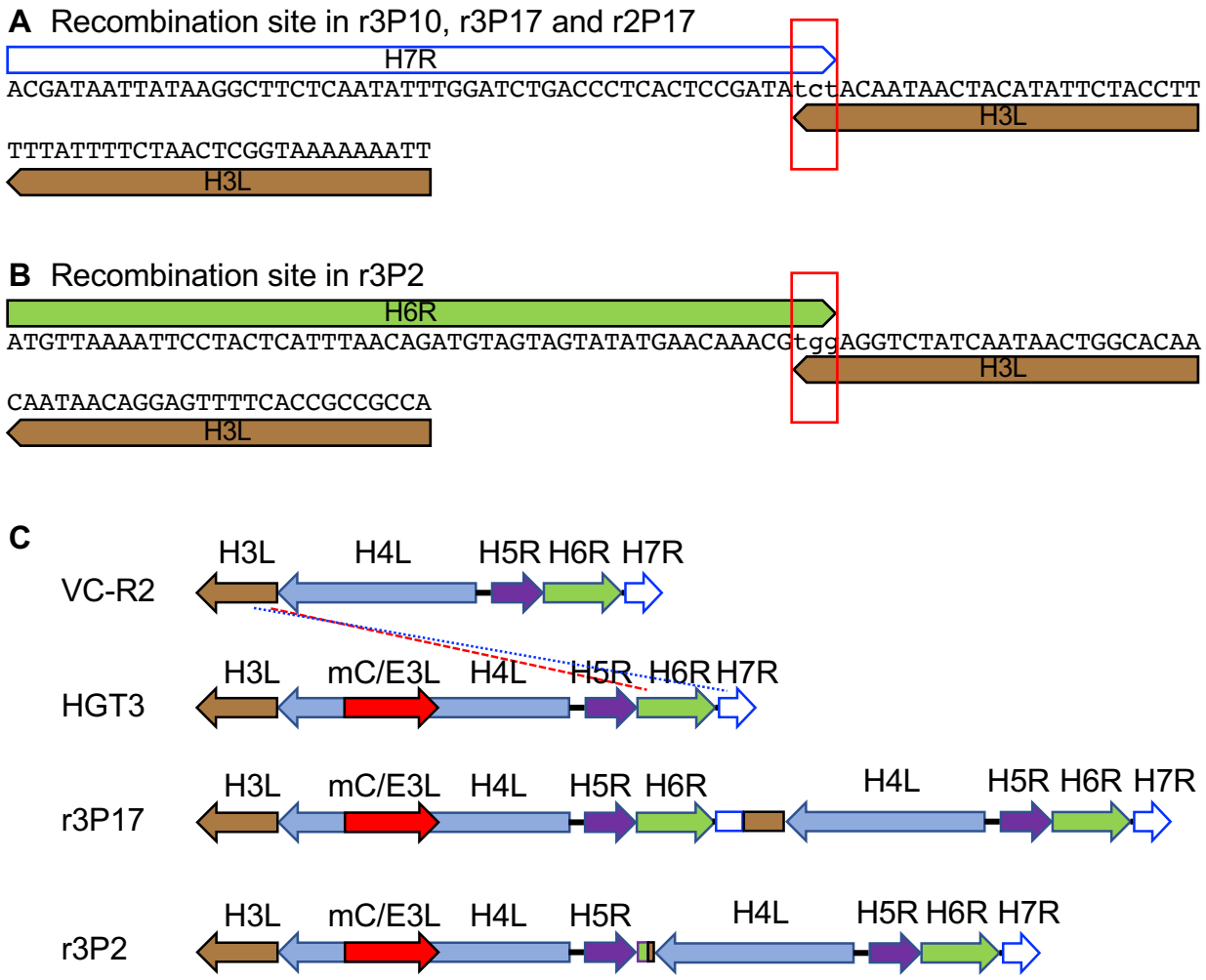
686
687
688 **Supporting Figure 1. Stable expression of mCherry-E3L in RK13 cells.** (**A**) Map of the
689 expression cassette in the pmCherry-E3L plasmid used to stably transfect RK13 cells. The
690 positions of oligonucleotides used to amplify the parts of the expression cassette are indicated by
691 arrows. (**B**) Fluorescent micrograph of stably transfected RK13-mCherry-E3L cells. (**C**) PCR
692 specific for mCherry-E3L did not amplify a product from RK13 cells (lane 1), but amplified
693 comparably sized products from the plasmid pmCherry-E3L (lane 2) and genomic DNA of the
694 RK13-mCherry-E3L cell line (lane 3). A smaller band was amplified from HGT3 DNA due to the
695 spliced-out intron (lane 4). M = molecular weight marker.
696



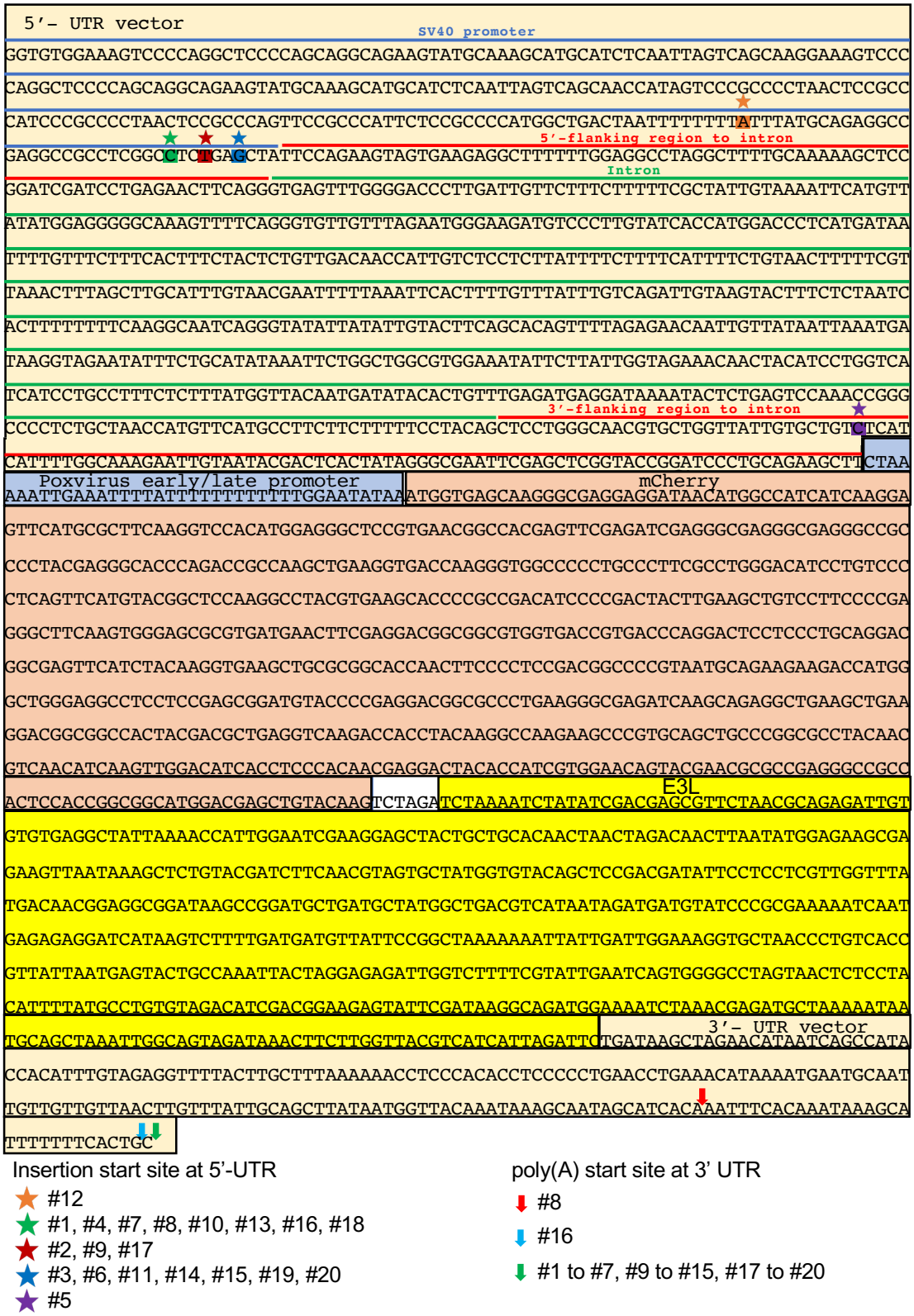
Supporting Figure 2. Integration site and composition of the U2-mCherry/E3L fusion in HGT20.



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.

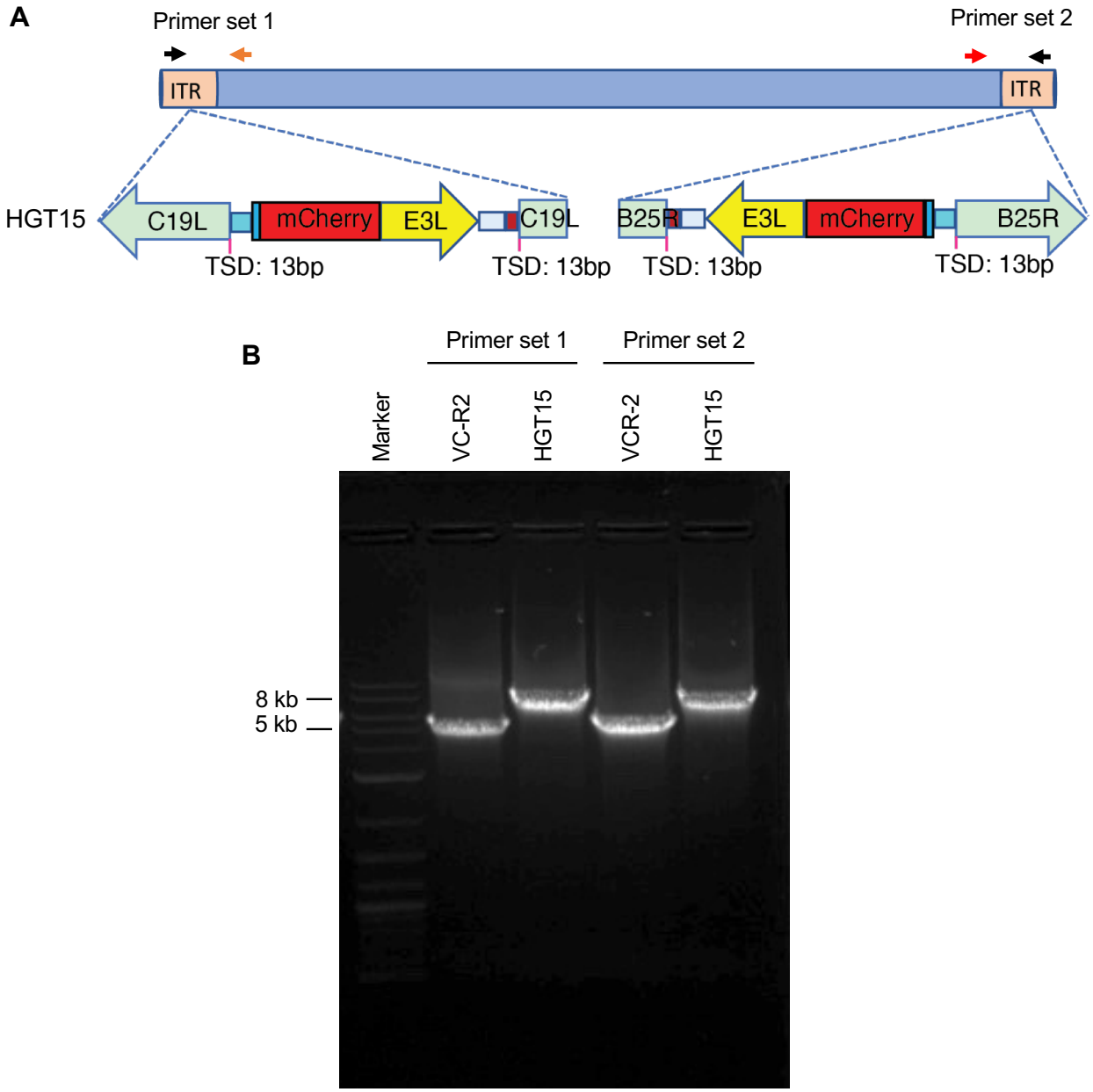


Supporting Figure 4. Recombination sites in passaged HGT3. (**A**) Recombination site in HGT3 between H7R and H3L identified in passage 10 of replicate 3 and passage 17 of replicate 2 by PCR and Sanger sequencing, as well as in passage 17 of replicate 3 by PacBio sequencing. A red rectangle shows a 3 nucleotide overlap between both 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. A red rectangle shows a 3 nucleotide overlap between both genes at the breakpoint. (**C**) Schematic presentation 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.

723



724
725 **Supporting Figure 6. HGT15 contains one mCherry-E3L copy in each inverted terminal**
726 **repeats (ITR).** (**A**)To determine whether mCherry-E3L was present in both ITRs or only in one,
727 primers were designed to span the integration site and the unique region outside of the ITR. (**B**)
728 Both primer sets yielded larger PCR products for HGT15 than for the parental VC-R2, indicating
729 that mCherry-E3L was present in both ITRs.
730

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

735

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

738

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
```

786 ATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGA
787 GTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCT
788 CCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTA
789 AACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCAT
790 TAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAA
791 AAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTT
792 GTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
793 TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAATT
794 GTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACC
795 AATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGT
796 TGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAA
797 ACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGA
798 GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAA
799 GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCA
800 AGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCG
801 CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACA
802 TTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
803 AAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCC
804 AGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGA
805 AAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA
806 TAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCC
807 CCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
808 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGA
809 TGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGG
810 AGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
811 GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
812 CTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGC
813 TCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
814 CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTG
815 CATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAC
816 GTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGC
817 GCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACC
818 CATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACT
819 GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGA
820 AGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCG
821 CAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAAT
822 GACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAA
823 AGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
824 TGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGA
825 AGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTT
826 CATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGG
827 GCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGG
828 GCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAG
829 ATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCA
830 TGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA
831 AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
832 CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT

833 TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
834 GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT
835 GGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT
836 CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
837 ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
838 CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGT
839 ATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
840 AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGC
841 TGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC
842
843 HGT1:
844 cgctttatttgtaattgaccacgccattacgatacaaacttaacggatatcgcgataatgaaat
845 aatttatgattatttctcgctttcaatttaacacaaccctcaagaacctttgtatttattttca
846 ttttttaagtatagaataaagaatctaCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTT
847 GGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACG
848 TGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGA
849 ATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTT
850 GGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
851 TCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
852 CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTC
853 GCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCG
854 ACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT
855 CGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
856 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGG
857 GCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
858 GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
859 AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACG
860 AGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGA
861 CGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGT
862 GAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATA
863 TGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTC
864 CGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATG
865 GCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTG
866 ATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAA
867 TGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAAC
868 TCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAAT
869 CTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCAT
870 CATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
871 TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAA
872 CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA
873 GCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
874 AAAAAAAAAGGAAAAAAAAAACCAAAAAAAAAAAAAAAAagaataaagaatctataaaaactaa
875 aaaaattatacatcataaaccaatttcctagttgtttgtaactttaaatggactctaaagagac
876 tattctaattgagatcattcc
877
878 HGT2:

879 gccgttccctgtcatcctctaacgtggcagtcattaaaaacatagaatctattttcgtatgatg
880 ctttctcgctactgctataataatatctcctatttgatcatgctcatgaacttcgtctataata
881 agagtgccataactTGTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTT
882 TTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTG
883 CTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTA
884 CCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGT
885 GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATG
886 GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGG
887 GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
888 GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTAC
889 TTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG
890 TGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG
891 CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC
892 TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA
893 AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCT
894 GCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATC
895 GTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGT
896 CTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAAC
897 CATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAA
898 GTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTC
899 CTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAAT
900 AGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCG
901 GCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAA
902 TTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTA
903 TGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCT
904 AAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGAT
905 AAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
906 CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG
907 CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
908 GCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAga
909 gtgccataactaaatagttttgttagagataacttatgggtagaaaatacaattccatattttt
910 ttggttgtttgtttattaattcttccggtatagatccgtaccgtaaagaaataggagatccatc
911 tagtaccttaaatccc
912
913 <u>HGT3:</u>
914 ctcgcaaattctagtcttaaccaaaaaatcgtatataaccacggagatggcgtatttaagagtg
915 gattcttctaccgttttgttcttggatgtcatataggaaactataaagtccgcactactgttaa
916 gaatgattactaacgcGGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTT
917 TTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTG
918 CTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTA
919 CCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGT
920 GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATG
921 GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGG
922 GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
923 GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTAC
924 TTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG
925 TGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG

CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC
TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA
AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCT
GCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATC
GTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGT
CTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAAC
CATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAA
GTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTC
CTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAAT
AGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCG
GCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAA
TTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTA
TGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCT
AAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGAT
AAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
GCAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAACAAAAAAAAAAAaagaatgattactaa
cgcaactatatagttcaaattaagcattttggaaacataaaataactctgtagacgatacttga
ctttcgaataagtttgcagacaaacgaagaaagaacagacctctcttaatttcagaagaaaact
ttttttcgtattcctgacgtctagagtttatatcaataagaaagttaagaattagtcggttaat
gttgtatttcattacccaagtttgagatttcataatattatcaaaagacatgataatattaaag
ataaagcgctgactatgaacgaaatagctatatggttcgctcaaaaatatagtcttgttaaac

HGT4:

ctcgccacgaagaaagctctcgtattcagtttcatcgataaaggataccgttaaatataactgg
ttgccgatagtctcatagtctattaagtggtaagtttcgtacaaatacagaatccctaaaatat
tatctaatgttggattaatctttaccataactgtataaaatggagacggagtcataactatttt
accgtttgtacttactggaatagatgaaggaataatctccggacatgctggtaaagacccaaat
gtctgtttgaagaaatccaatgttccaggtcctaatctcttaacaaaaattacgatattcgCTC
TGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGG
ATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTG
GCAAAGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCGGATCCCTGCAGAA
GCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAATATAAATGGTGAGCAAGGGCGAGGAG
GATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTCCACATGGAGGGCTCCGTGAACG
GCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAA
GCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATG
TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCC
CCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCA
GGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCC
TCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACC
CCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTA
CGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAAC
GCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTA
TATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGA
GCTACTGCTGCACAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGT

973 ACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGAC
974 AACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGC
975 GAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTG
976 ATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTG
977 GTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATC
978 GACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTA
979 AATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATC
980 AGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACC
981 TGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAA
982 ATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAA
983 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaaaattacg
984 atattcgatcccgatatcctttgcattctatttaccagcatatcacgaactatattaagattat
985 ctatcatgtctattctcccaccgttatataaatcgcctccgctaagaaacgttagtatatccat
986 acaatggaatacttcatttctaaaatagtattcgttttctaattctttaatgtgaaatcgtata
987 ctagaaagggaaaaattatctttgagttttccgttagaaaagaaccacgaaactaatgttctga
988 ttgcgtccgattccgttgctgaattaatggatttacaccaaaaactcatataacttctagatgt
989 agaagcattcgctaaaaaattagtagaatcaaaggatataagtagatgttccaacaagtgagca
990 attcccaag
991
992 <u>HGT5:</u>
993 atgtttgtcattaaacgaaatggatacaaggaaaatgtcatgtttgataaaatcacgtctcgta
994 ttagaaaattatgttatgCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGA
995 ATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTT
996 GGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
997 TCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
998 CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTC
999 GCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCG
1000 ACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT
1001 CGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
1002 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGG
1003 GCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
1004 GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
1005 AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACG
1006 AGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGA
1007 CGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGT
1008 GAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATA
1009 TGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTC
1010 CGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATG
1011 GCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTG
1012 ATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAA
1013 TGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAAC
1014 TCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAAT
1015 CTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCAT
1016 CATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
1017 TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAA
1018 CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA
1019 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

1114 CAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAG
1115 TTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCC
1116 GCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCG
1117 TTCTAACGCAGAGATTGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCA
1118 CAACTAACTAGACAACTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAAC
1119 GTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGA
1120 TAAGCCGGATGCTGATGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATG
1121 AGAGAGGATCATAAGTCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTG
1122 CTAACCCTGTCACCGTTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTAT
1123 TGAATCAGTGGGGCCTAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTA
1124 TTCGATAAGGCAGATGGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAG
1125 ATAAACTTCTTGGTTACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACA
1126 TTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAA
1127 TGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAG
1128 CATCACATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaaatctgtaggag
1129 gactggcgctatttcaagtagccaatggcgccatagatttatgtagacatttctttatgtattt
1130 ttgtgaacaaaagctacgaccaaattcattttggttcgttgttgttagagccattgcgagcatg
1131 ataatgtatttagtattaggtatagcattgctgtatatttctgaacaagatgacaagaagaata
1132 ctaataatgatggtagtaataatgataaacgaaatgagtcgtctataaattctaactccagtcc
1133

1134 HGT9:

1135 cgttccaggcgatattatctcaacaaaattcagattctatctttagagtatccactaaactatt
1136 acggtttatgtactacaatgaactaagagaaatctttagacggttgagaaaaggttctatcaac
1137 aatatcgatcctcacttcgaagagttaatattattgggtggtaaactagataaaaaggaatcta
1138 ttaaagattgtttaagaagagaattaaaagaggaaagtgatgaacgtataacagtaaaagaatt
1139 tggaaatgtaattctaaaacttacaacacgggataaattatttaataaagtatatataagttat
1140 tgcatggcgtgttttattaatcaatcgttggaggTGAGCTATTCCAGAAGTAGTGAAGAGGCTT
1141 TTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGC
1142 AACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGG
1143 GCGAATTCGAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTT
1144 TTTTGGAATATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATG
1145 CGCTTCAAGGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCG
1146 AGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCC
1147 CTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCC
1148 GCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGA
1149 ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCAT
1150 CTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC
1151 ATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCA
1152 AGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGC
1153 CAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCAC
1154 AACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCA
1155 TGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGT
1156 GTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTT
1157 AATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACA
1158 GCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGC
1159 TATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCT
1160 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
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAataaaaaaagtatttttttttaat
atttttcacaaatatcgttcgcggatatcattagacaattgtagtatattcttacgtatcactt
gtttaatatctatatcggctattctggaaccgag

<u>HGT13:</u>

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

GCATGGACGAGCTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGAT
TGTGTGTGAGGCTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAA
CTTAATATGGAGAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGT
ACAGCTCCGACGATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGA
TGCTATGGCTGACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAG
TCTTTTGATGATGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCG
TTATTAATGAGTACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCC
TAGTAACTCTCCTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGAT
GGAAAATCTAAACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTT
ACGTCATCATTAGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTA
CTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTG
TTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC
AAATAAAGCATTTTTTTCACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAaaattattagaat
catatatctttatccatctattatcgtctgtgaagttttggctcagaaaaatatctttacctaa
tattcgtttagacgatgtttcaacaccggcatttttataacagtcagctaccaatttaaaacaa
tacattctattatgaccaaatccatacggaatacccaataaagttaatgacgtatcagctgc

HGT18:

attgttccatattatgggatcaggtaaaacaataatcgctttgttgttcgccttggtagcttcc
agatttaaaaaggtttacattctCTCTGAGCTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAG
GCCTAGGCTTTTGCAAAAAGCTCCGGATCGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCT
GGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTC
GAGCTCGGTACCGGATCCCTGCAGAAGCTTCTAAAAATTGAAATTTTATTTTTTTTTTTTGGAA
TATAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAA
GGTCCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGC
CCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCT
GGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACAT
CCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG
GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGG
TGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
GGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGG
CTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGC
CCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGA
CTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAG
CTGTACAAGTCTAGATCTAAAATCTATATCGACGAGCGTTCTAACGCAGAGATTGTGTGTGAGG
CTATTAAAACCATTGGAATCGAAGGAGCTACTGCTGCACAACTAACTAGACAACTTAATATGGA
GAAGCGAGAAGTTAATAAAGCTCTGTACGATCTTCAACGTAGTGCTATGGTGTACAGCTCCGAC
GATATTCCTCCTCGTTGGTTTATGACAACGGAGGCGGATAAGCCGGATGCTGATGCTATGGCTG
ACGTCATAATAGATGATGTATCCCGCGAAAAATCAATGAGAGAGGATCATAAGTCTTTTGATGA
TGTTATTCCGGCTAAAAAAATTATTGATTGGAAAGGTGCTAACCCTGTCACCGTTATTAATGAG
TACTGCCAAATTACTAGGAGAGATTGGTCTTTTCGTATTGAATCAGTGGGGCCTAGTAACTCTC
CTACATTTTATGCCTGTGTAGACATCGACGGAAGAGTATTCGATAAGGCAGATGGAAAATCTAA
ACGAGATGCTAAAAATAATGCAGCTAAATTGGCAGTAGATAAACTTCTTGGTTACGTCATCATT
AGATTCTGATAAGCTAGAACATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTG
TTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCAT

TTTTTTCACTGCAAA

1537 GGACAATTTAAA