The impact of interspecies recombination on human herpes simplex virus evolution and host immune recognition

Amanda M. Casto[1*], Pavitra Roychoudhury[2], Hong Xie[2], Stacy Selke[1], Garrett A. Perchetti[2], Haley Wofford[1], Meei-Li Huang[2], Georges M. G. M. Verjans[3,4], Geoffrey S. Gottlieb[1,5,6], Anna Wald[1], Keith R. Jerome[2,7], David M. Koelle[1,2,5,6,7,8], Christine Johnston[1,7], Alexander L. Greninger[2*]

[1]Department of Medicine, University of Washington, Seattle, Washington, USA
[2]Department of Laboratory Medicine, University of Washington, Seattle, Washington, USA
[3]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands
[4]Research Center for Emerging Infectious and Zoonoses, University of Veterinary Medicine Hannover, Hannover, Germany
[5]Department of Global Health, University of Washington, Seattle, Washington, USA
[6]Center for Emerging and Re-Emerging Infectious Diseases, University of Washington, Seattle, Washington, USA
[7]Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
[8]Benaroya Research Institute, Seattle, Washington, USA

*To whom correspondence should be addressed:
Amanda Casto
1959 NE Pacific St.
Box 356423
Seattle, WA 98195-6423
amcasto@uw.edu

Alexander Greninger
Suite 320
1616 Eastlake Ave. E
Seattle, WA 98102
agrening@uw.edu

Running Title: Interspecies Recombination in the HSV Genomes

Abstract

Among the most ubiquitous of human pathogens, HSV-1 and HSV-2 are distinct viral species that diverged about six million years ago. At least four ancient HSV-1 x HSV-2 interspecies recombination events have affected the HSV-2 genome, with recombinants and non-recombinants at each locus circulating today. Though interspecies recombination has occurred in the past, its importance in HSV evolution remains incompletely defined. Using 255 newly-sequenced and 219 existing HSV genome sequences, we comprehensively assessed interspecies recombination in HSV. The novel recombinants we identify demonstrate that the sizes and locations of interspecies recombination events in HSV-2 are more variable than previously appreciated. One novel recombinant arose in its current host, showing for the first time that interspecies recombination occurs in contemporary HSV populations. We also demonstrate that interspecies recombination affects T-cell recognition of HSV. Our findings indicate that interspecies recombination can significantly influence genetic variation in and host immunologic response to HSV-2.

Introduction

The herpes simplex viruses (HSV-1 and HSV-2) are ubiquitous human pathogens with 3.7 billion HSV-1 and 417 million HSV-2 infected individuals worldwide (Looker et al., 2015a, 2015b). Both viruses establish lifelong infections typically characterized by mild, intermittent clinical symptoms. However, HSV can also cause significant morbidity and mortality, particularly among the immunocompromised and in neonates (Corey et al., 1983; Corey and Wald, 2009). Furthermore, genital HSV-2 infection has helped fuel the HIV epidemic by increasing the risk of HIV infection (Johnson et al., 2011; Masese et al., 2015; Zhu et al., 2009). While antivirals can reduce symptoms, they do not cure infection and do not completely prevent viral shedding and transmission, leading to the urgent need for an HSV vaccine (Gottlieb et al., 2016).

Instrumental in the development of an HSV vaccine and of new pharmaceutical therapies for HSV is a better understanding of the evolution of HSV and of the genetic variation among viral strains. HSV-1 and HSV-2 diverged from one another about 6 million years ago after which HSV-1 evolved in the human lineage and HSV-2 in the chimpanzee lineage (Wertheim et al., 2014). A human ancestor then acquired HSV-2 as a zoonotic infection from a chimpanzee ancestor 1.4 – 3 million years ago (Wertheim et al., 2014). Early studies of HSV genomes indicated that the viral species were relatively homogenous (mean pairwise distances among HSV-1 and HSV-2 strains are 0.8% and 0.2%, respectively) (Kolb et al., 2015). However, in 2015, a variant HSV-2 strain was described that was highly divergent from other HSV-2 samples at a single genomic locus (Burrel et al., 2015). This divergent region was likely affected by an ancient HSV-1 x HSV-2 interspecies recombination event with both recombinant and non-recombinant genotypes at this locus observed among HSV-2 strains today (Burrel et al., 2017; Koelle et al.,

2017). Ultimately, four loci, totaling about 1% of the HSV-2 genome, were found that carried evidence of such recombination events (Burrel et al., 2017; Koelle et al., 2017) . For three of the loci, subgenic regions within the UL29, UL30, and UL39 genes, recombinant genotypes were commonly observed among HSV-2 samples. Only one HSV-2 sample has been found that carries HSV-1 sequence at the fourth locus, which falls within UL15.

The high degree of divergence between recombinant and non-recombinant HSV-2 strains within these regions relative to the mean pairwise divergence elsewhere in the genome demonstrates that interspecies recombination could significantly contribute to variation among HSV-2 strains. However, numerous questions remain about the role of interspecies recombination in the evolution of HSV-2. In particular, it remains unknown whether interspecies recombination can affect loci other than those previously described in UL15, UL29, UL30, and UL39 and if interspecies recombination continues to affect contemporary HSV populations as all events described to date are thought to have occurred in a historical context. If HSV interspecies recombinants are still being generated, it is also unclear if the increasing number of immunocompromised hosts (Harpaz et al., 2016) or if changes in the epidemiology of HSV infection, such as the increase in genital HSV-1 (Chiam et al., 2010; Gilbert et al., 2011; Ryder et al., 2009), will alter the impact interspecies recombination has on genomic variation in HSV. We sought to answer these questions by performing a comprehensive survey of interspecies recombination on a large dataset of HSV genome sequences comprised of both previously available and newly-generated sequences.

Results

New sequencing more than doubles pool of available HSV genomes

We generated genome sequences for a total of 59 HSV-1 clinical samples, nearly doubling the number of HSV-1 genomes available for analysis (Supplementary Table 1). Fifty-eight of these samples were collected in Seattle, WA, USA, and one was from Uganda (Supplementary Table 2). Nine out of these newly-sequenced samples were from 6 persons who are genitally co-infected with HSV-1 and HSV-2 (Supplementary Table 3). For HSV-2, we generated 196 new genome sequences, increasing the total number of available HSV-2 genomes by 137% (Supplementary Table 4). Most (161 out of 196, or 82%) of these newly-sequenced samples were collected in Seattle, though samples collected in Cameroon (20), Peru (3), Senegal (10), and Uganda (2) were also sequenced. Eleven of these samples were from 6 persons genitally co-infected with HSV-1 and HSV-2. The countries of origin as well as the gender and HIV status of the carriers of viruses are reported for all newly-generated HSV genomes and for all the HSV genomes that were available in GenBank as of April 2018 in Supplementary Tables 1, 2, and 4.

No evidence of interspecies recombination found within HSV-1 genomes

We examined the newly-sequenced HSV-1 genomes for HSV-1 x HSV-2 recombination events using both the Recombination Detection Program (RDP) software and manual alignment review (see Methods). All HSV-1 genomes were aligned with an HSV-2 reference (SD90e, KF781518) and a chimpanzee herpesvirus, or ChHV, reference (NC_023677) separately to avoid detection of intraspecies recombination events. As a positive control, we first reviewed the

output of RDP for the regions of UL29 and UL30 where the HSV-2 reference strain SD90e carries recombinant HSV-1 sequence. RDP detected the UL29 and UL30 events in all 59 of the HSV-1-SD90e-ChHV alignments. RDP did return other putative events for all HSV-1 alignment trios. However, these events generated higher p-values than the UL29 and UL30 events for the same metric (Supplementary Note 1, Supplementary Figure 1) and we did not see evidence for recombination at the indicated loci on manual review. Overall, in line with previous results (Burrel et al., 2017; Koelle et al., 2017), we found no evidence of HSV-1 x HSV-2 recombination in the 59 newly-sequenced HSV-1 genomes.

Interspecies recombination event spans multiple ORFs and is stable within host

Next, we analyzed 230 HSV-2 (196 newly generated and 34 previously unanalyzed) genome sequences for interspecies recombination again using both RDP and manual review. Five previously undescribed interspecies recombination events were observed. A sample collected from an HIV negative woman in Seattle (2015-14086, MF510363) contained 7 kilobases (kb) of HSV-1 sequence spanning half of UL29 and all of UL30 and UL31 (nucleotides 60,761 - 67,777 in SD90e reference) (Figure 1A, Supplementary Figure 2). The longest interspecies recombination event described prior to this measured 538 basepairs (bp) (Burrel et al., 2017; Koelle et al., 2017). Sanger sequencing confirmed the breakpoints of this event. We sequenced two additional samples from the same person (2013-34209, MH790566; 2014-13047, MH790604; Figure 1B). The 3 sequenced samples were collected over a period of 659 days. The 2 additional samples also contained the 7 kb UL29 – UL31 recombination event. Across the recombinant region, the 3 sequences were identical (only coding sequence was

considered in sequence comparisons). The three HSV-1 sequences with the most similarity to the recombinant within the recombinant region were all collected in Seattle, including two that differed by 5 SNPs and one that differed by 6 (Figure 1C). Outside the recombinant region, there was only one polymorphic site among the three sequences; they formed a single clade in a phylogenetic tree of all available HSV-2 sequences (Figure 1D).

Because it spans the entire UL30 ORF, this event encompasses the region affected by the previously-defined recombination event in UL30. It does not encompass the part of the UL29 ORF where the UL29 recombination event is observed (Burrel et al., 2017; Koelle et al., 2017). We hypothesized that the UL30 recombinant genotype could have descended from the UL29 – UL31 recombinant through backcrosses. However, in a phylogenetic tree of the UL30 recombinant region, the UL29 – UL31 recombinant does not cluster with samples carrying the UL30 recombination event, suggesting that the UL29 – UL31 recombinant and the UL30 recombinants were derived from different HSV-1 parental strains (Figure 1E).

Interspecies recombinant generated in HSV-1/HSV-2 genitally co-infected host

A second novel interspecies recombination event was observed in a sample from a different HIV negative woman from Seattle (1996-26333, MH790638). This person has had both HSV-1 and HSV-2 detected in genital swabs. This event is more than 6 kb in length, spanning half of UL47 and all of UL48, UL49, and UL49A, and most of UL50 (nucleotides 102,443 to 108,479 in the SD90e reference) (Figure 2A, Supplementary Figure 3). Sanger sequencing also confirmed the breakpoints of this event. Within the recombinant region, the recombinant HSV-2 sample differed from an HSV-1 sample from the same person (1995-63175, MG999862) by

just one SNP (Figure 2B). We then compared the recombinant region to the rest of our HSV-1 dataset. The next most similar sequence had 6 nucleotide differences and had been collected in Germany, followed by a sequence with 7 differences that was collected in Seattle. These data suggest that the HSV-1 strain collected from the person with the UL47-UL50 recombinant is most likely the HSV-1 parent of the recombinant strain (Figure 2C).

We next sequenced three additional HSV-2 samples collected before the original sample (1996-26333) from the same person (1995-56864, MH790588; 1995-57888, MH790664; 2018-5934, MH790582). The samples (4 HSV-2 and one HSV-1) sequenced for this person were collected over a 151-day period (see Figure 2B). One of the 3 additional HSV-2 samples (2018-5934), collected one day before the original sample (1996-26333), carried the UL47-UL50 recombination event. However, two other samples (1995-56864, 1995-57888) both collected earlier than the recombinant HSV-2 samples did not contain the recombination event. Outside the recombinant region all four of the HSV-2 samples phylogenetically clustered together (Figure 2D), suggesting that that the non-recombinant HSV-2 strain from this person is the most likely HSV-2 parent for the UL47 – UL50 recombinant.

Novel events vary widely in size and location

In addition to the two large recombination events, we observed three other small novel events. The first of these was a 40 bp event in UL17 (Supplementary Figure 4), which resulted in 9 nucleotide and 3 amino acid changes. This was observed in a single sample from a man with HIV infection from Seattle (2005-42278, MH790643). The second small event was a 52 bp event in UL28 that resulted in 4 nucleotide and no amino acid changes. It was observed in 4 samples

194 from 4 different individuals, all of which were collected in East Africa (three in Uganda, one in
195 Kenya) (2012-18385, MF510324; 2009-3495, MF510280; 2012-18420, MF510337; 2009-4463,
196 MF621258). This small event was detected only by manual review. Finally, we observed a 259
197 bp event in UL32 which resulted in 25 nucleotide and 4 amino acid changes (Supplementary
198 Figure 5). This event was observed in 3 samples collected over a 12-year period from the same
199 HIV negative man in Seattle (2002-14972, MH790585; 2014-14807, MF510306; 2014-14811,
200 MF510292).
201
202 *Four rare genotypes found at UL29 recombinant locus*
203 We next sought to characterize variation at the 3 loci in HSV-2 (UL29, UL30, and UL39)
204 where interspecies recombinant genotypes are common (Burrel et al., 2017; Koelle et al., 2017)
205 by assigning a genotype to each HSV-2 sequence for each of these 3 loci (Supplementary Note
206 2). We report first on the results for UL29, which encodes the HSV single stranded binding
207 protein, ICP8 (Mapelli et al., 2005) (Figure 3A). The interspecies recombination event in UL29
208 affects a 121 amino acid (363 bp) region in the C-terminal half of the protein. The full length
209 recombinant genotype, which we called the common recombinant or CR genotype, was present
210 in 98.1% of HSV-2 sequences (209 out of 213) (see methods for a description of the sequences
211 included in this count). Only one sequence out of 213 (0.5%) carried the non-recombinant (NR)
212 genotype (Figure 3B). This HSV-2 sample was collected from an HIV positive woman in
213 Cameroon. We additionally noted a total of 4 rare recombinant genotypes at this locus, 3 of
214 which are previously undescribed (the fourth rare genotype is described in Koelle et al., 2017).
215 Each of these rare genotypes was seen in a single sample from Cameroon (R1), Seattle (R2),

DRC (R3), and Kenya (R4). The CR genotype at UL29 is characterized by two stretches of HSV-1 sequence separated by a non-recombinant region and as such has 4 different breakpoints (UL29 nucleotide positions 2,076 and 2,136 mark of the ends of the HSV-1 sequence for the first block; 2,196 and 2,439 mark the ends of the second block). All of the rare genotypes share at least 2 breakpoints with the CR genotype except for R1, which has just one stretch of HSV-1 sequence with 2 unique breakpoints (UL29 nucleotides 1,815 and 2,196). Intriguingly, the recombinant portion of UL29 maps to functionally critical portions of the protein which allow it to resist proteolysis and interact with single-stranded DNA (Mapelli et al., 2005) (Figure 3C).

UL30 has two recombinant genotypes

UL30 encodes the catalytic subunit of the DNA polymerase (Liu et al., 2006). The interspecies recombination event in UL30 affects a 180 amino acid (538 bp) region in the C-terminal half of the protein (Figure 4A). The full length or common recombinant (CR) genotype was observed in 95.7% (202 out of 211 with 2 excluded for missing data) of HSV-2 sequences (Figure 4B). The non-recombinant genotype (NR) was observed in 8 sequences (3.8%). These sequences were from Uganda (4), Cameroon (2), and Seattle, WA, USA (2). Two (10.5%) out of 19 sequences from Central Africa and four (15.3%) out of 26 sequences from East Africa had the NR genotype. There was only one rare genotype (R1) observed for UL30. This genotype has not been described previously and has a shorter HSV-1 block than the common recombinant, with a unique 5' breakpoint (UL30 nucleotide 2,926) but the same 3' breakpoint (UL30 nucleotide 3,409) as the CR genotype. The R1 genotype has no amino acid differences relative to the CR

237 genotype. The sample carrying R1 at UL30 was collected in Seattle. Five other samples from the
238 same person also carried this rare genotype.

239 The recombinant region of UL30 encodes a portion of the thumb domain of the DNA
240 polymerase (Liu et al., 2006) (Figure 4C). This critical domain interacts with double-stranded
241 DNA as it leaves the catalytic center of the enzyme. Though none of the 23 residues affected by
242 the recombination event have been associated with drug resistance to date, 21.5% of drug-
243 resistance (acyclovir, penciclovir, foscarnet, cidofovir) mutations identified in the HSV-1 DNA
244 polymerase are in the thumb domain (Sauerbrei et al., 2016; Topalis et al., 2016).

245

246 *UL39 contains a complex recombination locus*

247 UL39 encodes the large subunit of the HSV ribonucleotide reductase (Conner et al.,
248 1994). The interspecies recombination event affects a 152 amino acid (456 bp) region of this
249 gene. This recombinant locus exhibits high sequence diversity with several common and
250 numerous rare genotypes (Koelle et al., 2017). We defined genotypes at this locus first at the
251 nucleotide level (see Supplementary Note 2). However, given the large number of nucleotide
252 genotypes observed (21), we ultimately elected to group genotypes together if they had the
253 same amino acid sequence. The most common nucleotide sequence corresponding to each
254 amino acid sequence is illustrated in Figure 5A.

255 Most HSV-2 samples (135 out of 211, or 64.0%) had the non-recombinant (NR) genotype
256 at UL39 (Figure 5B). There are 2 common recombinant genotypes at the UL39 locus. Common
257 recombinant 1 (CR1) has 456 bp of HSV-1 sequence (breakpoints at UL39 nucleotides 2,373 and
258 2,829) and was seen in 54 (25.6%) of 211 (two with missing data excluded) samples. Common

recombinant 2 (CR2) has 234 bp of HSV-1 sequence (breakpoints at UL39 nucleotides 2,373 and 2,607) and was seen in 16 (7.6%) of 211 samples. The remaining seven samples have 7 distinct, rare recombinant genotypes, 5 of which are previously undescribed (R3 and R7 described in Koelle et al., 2017). Samples with rare genotypes at UL39 were from the US (3), Cameroon (1), the DRC (1), Finland (1), and Kenya (1). Interestingly, the samples from Cameroon and from Kenya both also had rare genotypes for the UL29 locus.

Given the lack of a crystal structure for the HSV ribonuclease reductase, the protein structure of yeast ribonucleotide reductase (see Methods) is shown in Figure 5C (Xu et al., 2006). Three of the main murine subdominant epitopes recognized by HSV-1 specific CD8 cells have been mapped to the UL39 gene product (St Leger et al., 2011). One of these epitopes, which includes HSV-1 UL39 amino acids 822 – 829 (corresponding to HSV-2 UL39 amino acids 826 – 833), is within the region affected by interspecies recombination (St Leger et al., 2011). The NR version of this epitope contains two amino acid changes relative to the CR1 and CR2 genotypes. Functional experiments have demonstrated that the HSV-1 specific CD8+ T-cells that recognize the CR1/CR2 version of this epitope do not recognize the NR version (Salvucci et al., 1995).

Recombinant genotypes at UL29, UL30, and UL39 are stable over time

To determine whether genotypes at the UL29, UL30, and UL39 loci are stable over time within hosts, we identified people who had more than one sequenced HSV-2 sample. Out of 65 such persons, the samples for 57 (87.7%) had the same genotypes at the UL29, UL30, and UL39

loci. The remaining 8 persons had all previously been identified as being infected with at least 2 distinct HSV-2 strains (Johnston et al., 2017).

One of these 8 persons, a man with HIV infection from Seattle, had a total of 4 sequenced samples. The two sequences that were analyzed in the Johnston et al., 2017 study (2003-18061, MF510298; 2007-22031, MF510341) had the same genotypes at the UL29, UL30, and UL39 loci with the NR genotype at the UL39 locus. However, we noted that a third sequence (2003-16029, KX574903) from this person had the CR1 genotype at UL39. We confirmed UL39 genotype assignments for all three of these samples using Sanger sequencing. We then compared these 3 sequences across the entire length of the genome. The two previously studied sequences differed from one another at 251 SNPs while the third sequence differed from the first two at 105 and 120 SNPs, respectively. In the Johnston et al. study, the strains in persons infected with two strains differed from one another by 87 to 305 SNPs (range based on 13 sequence pairs). Taken together, this suggests that this person is infected with at least 3 different HSV-2 strains.

One HSV-2 sequence from Cameroon had no observed HSV-1 recombination events

We noted that one sample from Cameroon had the non-recombinant genotype at the UL29, UL30, and UL39 loci and did not contain any other interspecies recombination events. This sequence (2006-16150, MH790600) would be a useful alternative reference sequence to SD90e (KF781518) and HG52 (JX112656) for studies of HSV evolution as both of these sequences contain recombinant HSV-1 sequence. This sample has been sequenced to high quality with 50x or higher coverage for 95.8% of the full genome.

Interspecies recombination impacts recognition of HSV by T-cells

In order to understand the functional importance of HSV interspecies recombination *in vivo*, we investigated whether interspecies recombination alters HSV recognition by CD4 and CD8 T-cells. We found that a polyclonal CD4 T-cell line (from an HSV-1-infected donor, GU14669) (Supplementary Table 5) strongly recognized full-length UL30 from both an HSV-1 strain (lab strain E115) and HSV-2 strains with the UL30 CR genotype (strains HG52 and 186, JX112656 and JN561323, respectively). However, these T-cells failed to recognize the NR version of UL30 (2008_15116, MF621257) (Figure 6A). We next examined T-cell recognition of the UL47-UL50 recombinant (1996_26333, MH790638) relative to HSV-1 and HSV-2 samples without the recombination event. A CD8 T-cell clone (1874.1191.22) (Supplementary Table 5) recognizing an epitope in HSV-2 UL47 laying outside the UL47 – UL50 recombination event (amino acids 551-559) but not recognizing HSV-1 recognized the UL47 – UL50 recombinant (Figure 6B). However, a CD8 T-cell clone (5101.1999.23) recognizing an epitope in HSV-2 UL47 at amino acids 289 – 298, which lies within the HSV-1 portion of UL47 in the recombinant, failed to recognize the recombinant. Conversely, two CD8 cell lines (TG3 UL48 N, TG3 UL48 C) that recognize HSV-1 UL48 but do not recognize HSV-2 UL48 gained novel recognition of the HSV-2 UL47-UL50 recombinant (Figure 6C, 6D, 6E). As a negative control, we also tested a CD4 cell clone (9447.28) that recognizes a peptide in UL49 identical in HSV-1 and HSV-2. As expected, this clone recognized the recombinant (Figure 6F). These results illustrate that interspecies recombination can serve as a mechanism of T-cell immune evasion.

Discussion

It has long been known that HSV-1 and HSV-2 could recombine *in vitro* (Halliburton, 1980; Morse et al., 1977), but only recently appreciated that interspecies recombination has impacted the genome of nearly all HSV-2 clinical samples sequenced to date. Because HSV-2 has the potential to evolve more rapidly via recombination than through point mutation, a greater understanding of when, where, and why *in vivo* HSV interspecies recombination occurs and what its impacts are on human immunological response to HSV has important clinical and public health implications. Here we show that interspecies recombination events are more pervasive in the HSV-2 genome than previous appreciated and that they continue to be generated in circulating HSV-2 strains. We also demonstrate that such recombination can alter the recognition of HSV by CD4 and CD8 T-cells.

Consistent with earlier findings, we found no evidence of HSV-2 sequence in any of the HSV-1 genomes that we analyzed. While interspecies recombination events in HSV-1 may still exist, a profound asymmetry between the receptivity of the HSV-1 and HSV-2 genomes to sequence from the other HSV species is evident, given that almost all HSV-2 genomes contain some HSV-1 sequence. Since HSV-1 interspecies recombinants have been observed *in vitro*, the barrier to the presence of HSV-2 DNA in HSV-1 genomes is presumably downstream from the recombination event itself. It has been purposed that this barrier is at the level of transmission (Burrel et al., 2017). However, the failure to detect even one instance of recombinant HSV-1 in 135 HSV-1 genomes, including 9 from genitally co-infected persons (where the viruses are not relegated to different anatomic sites) is suggestive that the barrier may, in fact, be earlier in the HSV-1 life cycle. Overall, the one-way exchange of DNA between HSV-1 and HSV-2 *in vivo* is a

346 curious feature of the evolution of these two viruses which does not yet have a clear
347 explanation.

348 In HSV-2, recombinant genotypes at the UL29, UL30, and UL39 are observed at high
349 frequencies, suggesting they have a selective advantage. Given the scarcity of the UL29 and
350 UL30 NR genotypes, the recombinant regions in these two essential genes may prove useful as
351 drug or vaccine targets given the sequence identity between HSV-1 and most HSV-2 strains at
352 these loci. Our analysis of the UL29, UL30, and UL39 recombinant regions also established that
353 there are both common and rare recombinant genotypes at all three loci. This finding suggests
354 that these recombinant genotypes are not static but continue to evolve both through inter- and
355 intraspecies recombination. Indeed, this process may be accelerating as rare genotypes, such as
356 the NR genotypes at UL29 and UL30, are shuffled around the world by globalization. The
357 potential for rapid creation of new genotypes through recombination at these loci is perhaps
358 most concerning at the UL30 locus, as the region of the gene were HSV-1 sequence is
359 commonly observed is also where about 20% of all described drug resistance mutations in UL30
360 are located (Sauerbrei et al., 2016; Topalis et al., 2016).

361 We also present significant additional data that the effects of interspecies
362 recombination on HSV-2 are not restricted to the previously defined events in UL15, UL29,
363 UL30, and UL39. Among the novel events we describe are two that are several kilobases in
364 length and span multiple genes; both of these events are more than 10 times larger than the
365 largest event described in Burrel et al., 2017 and Koelle et al., 2017. Furthermore, we provide
366 evidence that these recombinant viruses are stable and can be persistently shed for years. Even
367 if recombinant HSV-2 with such large recombination events are rare, any advantageous HSV-1

368 alleles in the region could theoretically disperse throughout the HSV-2 population via
369 subsequent intraspecies recombination.

370 Our findings also challenge the supposition that new interspecies recombinant
371 genotypes are no longer being generated in modern HSV-2 populations. In particular, the large
372 sizes of the UL29-UL31 and UL47-UL50 events are by themselves suggestive that they were
373 generated recently as intraspecies recombination has not had time to reduce their size through
374 backcrosses with non-recombinant genomes. Furthermore, we present evidence that one of
375 these two events, UL47-UL50, arose in the person from whom it was collected. This particular
376 person is one of six genitally co-infected people from whom we have a sequenced HSV-2
377 sample. The high incidence of recombinant HSV-2 with large HSV-1 blocks in genitally co-
378 infected persons is striking relative to the incidence of such recombinants among all individuals
379 with a sequenced HSV-2 sample (1 in 6 compared to 2 in 231). This finding, coupled with our
380 documentation of the generation of the UL47 – UL50 recombinant in a genitally co-infected
381 person, suggests that genital co-infection fosters interspecies recombination. This is
382 noteworthy as the incidence of genital HSV-1 has been shown in multiple studies to be
383 increasing (Chiam et al., 2010; Gilbert et al., 2011; Ryder et al., 2009). The evidence we present
384 here indicates that this epidemiologic shift could lead to an increase in the frequency of
385 recombinant HSV-2.

386 The existence of interspecies HSV-2 recombinants has numerous potential impacts on
387 clinical phenotype and on the development of new therapeutics, including an HSV vaccine. We
388 demonstrate here that interspecies recombination can profoundly alter T-cell recognition of
389 HSV. In particular, we showed that HSV-2 strains can completely gain or lose recognition by a

particular T-cell clone through a single interspecies recombination event suggesting a mechanism by which such recombination could result in immune escape. Given the ability of interspecies recombination to significantly alter the T-cell repertoire that recognizes an HSV-2 strain, a better understanding of the natural history of HSV recombination and of the worldwide distribution of interspecies recombinants is absolutely crucial to the development of an effective HSV-2 vaccine. Other curative therapeutic approaches for HSV such as the disruption of latent HSV by targeted endonucleases could take advantage of the high prevalence of interspecies recombinant genotypes for UL29 and UL30 in HSV-2. These two essential genes have already been successfully targeted by CRISPR/Cas9 (van Diemen et al., 2016).

We recognize several limitations of our work. First of all, our analysis of recombinant genotypes at the UL29, UL30, and UL39 loci is limited by our sample set, which is not ideal for approximating the true prevalences of these genotypes in the populations covered. Our study constitutes a secondary analysis of HSV samples collected for various clinical studies and not at random from the represented populations. Sample sizes for geographic locations are also often limited. Because of this, the samples from any given location may represent a sub-population of the HSV-2 circulating at that location rather than being reflective of the population at large. Our results serve to put a prior probability on interspecies recombination prevalence in HSV and suggest that more rigorous sequencing of HSV from diverse human populations is required to more accurately measure this prevalence.

We also recognize that we may have underestimated the degree of HSV-1 x HSV-2 recombination, as even though the capture technique that we used for HSV sequencing greatly

412 improved our ability to obtain high-quality, genome sequences, there remain portions of the
413 HSV genome that are difficult to accurately sequence due to particularly high GC content and
414 the presence of repeats. In general, these occur outside genic loci though some coding regions
415 (such as UL36) contain difficult to sequence repeats. Because of these sequencing limitations, it
416 remains challenging to confidently assess whether HSV-1 x HSV-2 recombination has an impact
417 on these regions in either HSV genome.

418 In summary, our work demonstrates that the effects of interspecies recombination on
419 the HSV-2 genome are less constrained than previously thought. We show that in addition to
420 the 3 loci where recombinant genotypes are common (UL29, UL30, and UL39), there are a
421 number of other loci that have been affected by interspecies recombination. Our results also
422 demonstrate that recombination events can vary widely in size from about 50bp to multiple
423 kilobases in length. We also show that interspecies recombination events continue to be
424 generated in circulating HSV-2 clinical strains and that they are stable within hosts. Finally, we
425 demonstrate that interspecies recombination can alter immune response to HSV by adding or
426 removing T-cell epitopes. Together these findings suggest that interspecies recombination has
427 significant impacts on HSV-2 genetic diversity and evolution with major implications for HSV
428 vaccine and drug design.

429

430 **Methods**

431 *Samples*

432 All sequenced samples were collected from 1994 to 2016 as part of clinical research
433 studies on HSV infection at the University of Washington Virology Research Clinic (for Seattle

samples) or at international study sites in Cameroon, Peru, Senegal, and Uganda. Oral and genital swabs were either self-collected by participants or collected by clinicians. Swabs were collected directly from genital lesions or from mixed anogenital swabs as previously described (Tronstein et al., 2011). Written informed consent to collect swabs and demographic information was obtained from all participants. All studies were approved by the University of Washington (UW) Human Subjects Division and the local institutional review boards/ethics committees for the international sites.

All sequenced samples had HSV viral loads greater than 1,000 copies/mL due to limitations in recovering high confidence genomic sequences from samples with lower viral loads (Greninger et al., 2018). We selected samples in our dataset with respect to geographic origin and collection date. We also specifically included samples from individuals who had had both HSV-1 and HSV-2 detected in genital tract swabs (genital co-infection) as defined using our UL27 (gB)-based genotyping platform (Corey et al., 2005) and from individuals superinfected with more than one HSV-2 strain (as identified in Johnston et al., 2017). See Supplementary Note 3 for more details on sample selection.

Sequencing and Consensus Generation

Samples selected for sequencing underwent the next-generation, direct-capture sequencing method for HSV previously described in Greninger et al., 2018. DNA was extracted directly from sample swabs. Pooled DNA from multiple samples was then sequenced following enrichment using HSV-1 and HSV-2 specific oligonucleotide capture panels.

Consensus sequences and .bam files for newly-generated sequences were created from raw sequencing reads using the computational pipeline as previously described in Greninger et al., 2018, which is publicly available at https://github.com/proychou/HSV/. For HSV-1 sequences, strain 17 (NC_001806) was used as the reference (Davison, 2011). For HSV-2 sequences, SD90e (KF781518) was used as the reference (Colgrove et al., 2014). Of the 70 HSV-1 and 238 HSV-2 samples we attempted to sequence, we were successfully able to generate 59 HSV-1 and 196 HSV-2 high confidence consensus sequences. Sequencing data and GenBank accession numbers for all newly-generated sequences are listed in Supplementary Tables 1 and 4. Samples we were unable to sequence are marked as "failed".

For HSV-2, in addition to the 196 newly-generated sequences, we downloaded an additional 34 consensus sequences that had not previously been analyzed for recombination directly from GenBank. These included KY922720 – KY922726 (direct submission) and 28 sequences from Johnston et al., 2017 (see Supplementary Table 4).

Detection of Novel Recombination Events

Each HSV-1 sequence in our dataset of 59 newly-generated sequences was aligned with HSV-2 SD90e (KF781518) and a chimpanzee herpesvirus (ChHV, NC_023677) (Severini et al., 2013) reference sequence using MAFFT (Katoh and Standley, 2013). Each of these sequence trios were then run through RDP4, version Beta 4.95 (Martin et al., 2015). The RDP program was run from the command line with the default settings. This program uses the RDP, GENECONV, Chimaera, and MaxChi algorithms to both detect recombination events and verify events identified by other algorithms (Martin and Rybicki, 2000; Padidam et al., 1999; Posada

and Crandall, 2001; Smith, 1992). The algorithms BootScan, SiScan, and 3Seq are computationally intensive when used to detect new events and so are only used to verify other events when using the default settings (Boni et al., 2007; Gibbs et al., 2000; Martin et al., 2005). Putative events detected by RDP were subjected to manual review with apparently false positive events (those due to sequence misalignment or in regions of poor sequence quality) excluded from further analyses.

This process was then performed for 230 HSV-2 sequences that had not previously been analyzed for recombination with each sequence aligned to HSV-1 KOS (JQ673480) and the ChHV reference (NC_023677).

Novel Recombination Event Verification

We confirmed novel recombination events observed in just one sample by performing Sanger sequencing across the entire event or across each of the two breakpoints, depending on the size of the event, directly from the original sample swab. Additional samples from persons with novel recombination events were sequenced when such samples were available to further confirm the recombination event and to examine the stability of the novel recombinant *in vivo*.

Phylogenetic Trees

All phylogenetic trees were constructed using concatenated coding sequences from the genomic region of interest. Non-coding sequence was excluded due to the increased frequency of ambiguous base calls in these regions. The phylogenetic trees in Figures 1C, 1E, and 2C were created using MrBayes (Huelsenbeck and Ronquist, 2001) with a HKY85 substitution model,

gamma distribution for mutation rate variation across sites with 4 rate categories, chain length of 1,100,000, and burn-in length of 100,000. PHYML with a HKY85 substitution model was used to generate the trees in Figures 1D and 2D (Guindon et al., 2010). PHYML was used in place of MrBayes for the trees in 1D and 2D given the prolonged run time required by MrBayes for alignments of long sequences.

To better illustrate the full range of genomic variation in HSV-1 and HSV-2, we used all HSV genomes available in GenBank as of April 2018 in the generation of phylogenetic trees in addition to the newly-generated genome sequences. The accessions for the GenBank sequences are listed in Supplementary Tables 1 and 4. References for these sequences are as follows: HSV-1 Colgrove et al., 2016; Davison, 2011; Greninger et al., 2018; Kolb et al., 2011; Macdonald et al., 2012a, 2012b; Parsons et al., 2015; Szpara et al., 2014; Watson et al., 2012; HSV-2 Burrel et al., 2017; Colgrove et al., 2014; Johnston et al., 2017; Koelle et al., 2017; Kolb et al., 2015; Newman et al., 2015. Within the genomic regions of interest for each tree, only one sequence per unique genotype was used. For trees constructed from alignments of relatively short sequences (1C, 1E, and 2C), sequences with any ambiguous bases were excluded. For trees constructed from alignments of longer sequences (1D and 2D), sequences with greater than 10% ambiguous base calls were excluded.

Assignment of Genotypes for Recombinant Regions in UL29, UL30, UL39

We assigned genotypes for the UL29, UL30, and UL39 recombinant regions (as described in Burrel et al., 2017; Koelle et al., 2017) to all newly-generated HSV-2 sequences plus all those available in GenBank as of April 2018. For UL29 and UL30, we assigned genotypes based on

each sample's nucleotide sequence in the recombinant region. For UL39, genotypes were assigned based on amino acid sequence given the large number of different nucleotide genotypes observed in the region. See Supplementary Note 2 for details on how genotypes were assigned.

For the purpose of reporting the frequencies of the genotypes at the UL29, UL30, and UL39 loci, we first excluded all HSV-2 sequences with missing data at all three loci of interest. Next, we separated out samples from Burrel et al., 2017 as these samples were selected for sequencing from a larger sample set because of their unusual (non-recombinant) UL30 sequence (and therefore, the frequency of recombinant genotypes particularly at the UL30 locus in this set cannot be considered to approximate their true frequency in the represented geographic areas). We scanned these sequences for rare recombinant genotypes but did not include them in our calculations of the frequencies of common recombinant and non-recombinant genotypes. Finally, from the remaining sequences, we selected a subset which included just one sample per person. This left us with a total of 213 sequences for which we report genotype frequencies. Note that for individuals with HSV-2 superinfection (as noted in Supplementary Table 4) this means that only one of their HSV-2 strains was considered in the reported genotype frequencies.

Structural Analyses

Depictions of the protein products of HSV-1 UL29 (ssDNA binding protein, ICP8) and HSV-1 UL30 (DNA polymerase) as determined by x-ray crystallography (Liu et al., 2006; Mapelli

et al., 2005) were uploaded into Chimera (Pettersen et al., 2004), which was used to highlight the structures corresponding to genic regions affected by interspecies recombination.

We used HHPred to search for the protein with known structure that had the greatest predicted structural homology to the UL39 protein product (ribonuclease reductase) in the Research Collaboratory for Structural Bioinformatics protein database (RCSB PDB) (Berman et al., 2000). The best hit, regardless of which HSV protein sequence was used for UL39, was the large subunit of the yeast ribonuclease reductase (Xu et al., 2006). The region of this protein corresponding to the recombinant genic region was highlighted using Chimera as above.

T-Cell Analysis

Informed written consent was obtained from subjects contributing T-cells and viral isolates. The consent process for autopsy-derived T-cells has been documented (van Velzen et al., 2013). Lab HSV-2 strains 186 (JX112656) (Anderson et al., 1980) and HG52 (JN561323), (Harland and Brown, 1985), lab HSV-1 strain E115 (Spruance and Chow, 1980), and HSV-2 sample 1996_26333 (MH790638) were grown and titered on Vero cells. Both strain 186 and HG52 have the common recombinant or CR genotype at UL30. Full-length UL30 of HSV-1 strain 17 and HSV-2 strain 186 were cloned into pDEST103 and expressed by transient transfection of Cos-7 cells as described (Jing et al., 2012; Johnston et al., 2014). For HSV-2 UL30 NR, we similarly cloned full-length UL30 from HSV-2 sample 2008_15116 (MF621257) (Johnston et al., 2017) with the same PCR primers used for HSV-2 strain 186 (Johnston et al., 2014). Accurate amplification was confirmed by sequencing and UL30 was subcloned into pDEST103 after initial cloning into pENTR221 (Jing et al., 2012). UL30 of HSV-1 17, HSV-2 186, and HSV-2 2008_15116

were also cloned into pDEST203 and expressed by in vitro transcription/translation (IVTT) as described (Jing et al., 2012). HSV-specific T-cell clones 9447.28, 1874.1991.22, and 5101.1999.23 (Supplementary Table 5) from Seattle donors have been described (Dong et al., 2010; Koelle et al., 2001, 2000). Polyclonal, mono-specific, HSV-specific CD8 T-cells were enriched from T-cells expanded from autopsy HSV-1-infected trigeminal ganglia TG3 using tetramers of HLA A*0101 and either HSV-1 UL48 AA 90-99 or HSV-1 UL48 AA 479-488, as published (van Velzen et al., 2013). Live cells positive for CD3, CD8, and tetramer were sorted (FacsAria II, Becton Dickinson) and expanded polyclonally (Koelle et al., 2001). To obtain bulk HSV-specific CD4 T-cells, HSV-1-reactive CD4 T-cells were enriched from PBMC from a Seattle HSV-1 seropositive, HSV-2 seronegative donor with no clinical history of herpes infection, as published (Jing et al., 2012). Epstein-Barr virus-transformed lymphocyte continuous lines (LCL) from persons of defined HLA type were cultured as described (Tigges et al., 1992) for use as antigen-presenting cells (APC). For functional readouts of CD8 T-cells and cloned CD4 T-cells, LCL were infected overnight with HSV and then co-cultured for an additional 24 hours with T-cells (50,000 LCL and 50,000 T-cells/well in 200 µL in triplicate). Supernatants were then assayed for IFNγ by ELISA (Koelle et al., 2001). For polyclonal CD4 T-cells, a triplicate ^{3}H thymidine incorporation proliferation assay used autologous, irradiated PBMC as APC and responder T-cells (100,000 each in 200 µl/well final volumes) as published (Johnston et al., 2014). Antigens were sonicated, transfected Cos-7 cells or sonicated, UV-killed HSV-infected Vero cell preparations at final dilutions of 1:100, or IVTT protein preparations at 1:1000.

Acknowledgements

We would like to thank the clinical study participants from whom HSV samples were collected. We would also like to thank local study teams at the international sites in Cameroon, Peru, Senegal, and Uganda. At the Senegal site, we thank Marie-Pierre Sy for swab collection. At UW, we would like to thank Chris McClurkan and Marliis Ott for assistance with T-cell experiments.

Figures

Figure 1: UL29 – UL31 Recombinant. A) Schematic showing the size and position of the recombination event relative to the HSV-2 genome and within its genic neighborhood. Green represents coding regions and purple represents non-coding regions. The recombinant region is represented by a blue bar. B) List of all sequenced samples collected from the person carrying the UL29 – UL31 recombinant with the day of sample collection relative to the collection date of the first sample. C) Phylogenetic tree of the recombinant region extracted from HSV-1 sequences and from the UL29 – UL31 recombinant. The bar below the tree shows the extracted region (colored in bright blue) relative to the rest of the genome (faded out). The branch representing the UL29-UL31 recombinants is marked with blue circle. D) Phylogenetic tree of all coding sequences (concatenated) excluding UL29, UL30, and UL31 extracted from HSV-2 sequences. The bar below the tree shows the extracted regions (colored in bright green) relative to the rest of the genome (faded out). All 3 HSV-2 sequences from the person carrying the UL29 – UL31 recombinant clustered on the branch marked with the blue circle. E) Phylogenetic tree of the UL30 recombinant region extracted from HSV-1 sequences, HSV-2

608 sequences with the UL30 CR genotype, and HSV-2 sequences with the UL29 – UL31
609 recombination event. The bar below the tree shows the extracted region (colored in bright
610 blue) relative to the rest of the genome (faded out). The UL29 – UL31 recombinant branch is
611 marked with a blue circle. HSV-1 branches are marked in orange and HSV-2 branches are
612 marked in green.

613 **Figure 2: UL47 – UL50 Recombinant.** A) Schematic showing the size and position of the
614 recombination event relative to the HSV-2 genome and within its genic neighborhood. Green
615 represents coding regions and purple represents non-coding regions. The recombinant region is
616 represented by a blue bar. B) List of all sequenced samples collected from the person carrying
617 the UL47 – UL50 recombinant with the day of sample collection relative to the collection date
618 of the first sample. C) Phylogenetic tree of the recombinant region extracted from all HSV-1 and
619 HSV-2 sequences. The bar below the tree shows the extracted region (colored in bright blue)
620 relative to the rest of the genome (faded out). The branches marked with colored circles
621 represent the position in the tree of the various samples collected from the person carrying the
622 UL47 – UL50 recombinant. These colors correspond to those in the table in B. D) Phylogenetic
623 tree of all coding sequences (concatenated) excluding UL47, UL48, UL49, UL49a, and UL50
624 extracted from all HSV sequences. The bar below the tree shows the extracted regions (colored
625 in bright green) relative to the rest of the genome (faded out). The branches marked with
626 colored circles represent the position in the tree of the various samples collected from the
627 person carrying the UL47 – UL50 recombinant. These colors correspond to those in the table in
628 B.

629 **Figure 3: UL29 Recombinant Region.** A) Schematic showing the HSV-1 sequence blocks for each
630 UL29 recombinant genotype. HSV-1 sequence is colored in blue. ChHV/non-recombinant HSV-2
631 coding sequence is colored in green. Intergenic regions are purple. Letters denote amino acids
632 that differ among the various genotypes and vertical lines above the HSV-1 row denote the
633 location of nucleotides that differ among the various genotypes ("genotype-defining" SNPs, see
634 Supplementary Note 2). The top bar represents the entire HSV-2 genome while the second bar
635 represents the region of the genome containing UL29. Of note, the UL29 gene is oriented in the
636 reverse direction. The orientation of the second bar and the genotype bars are flipped relative
637 to the full genome bar. B) Number of samples from each geographic region with the NR, CR,
638 and rare genotypes. MD denotes samples with missing data. Note that samples LT797622 –
639 LT797636, LT797682, LT797786, LT799380 (samples from Burrel et al., 2017) were not included
640 in these counts. C) Crystal protein structure of the UL29 gene product, ICP8, the HSV single
641 stranded DNA binding protein. The part of the protein encoded by the recombinant region is
642 highlighted in blue.
643 **Figure 4: UL30 Recombinant Region.** A) Schematic showing the HSV-1 sequence blocks for each
644 UL30 recombinant genotype. HSV-1 sequence is colored in blue. ChHV/non-recombinant HSV-2
645 coding sequence is colored in green. Intergenic regions are purple. Letters denote amino acids
646 that differ among the various genotypes and vertical lines above the HSV-1 row denote the
647 location of nucleotides that differ among the various genotypes. The top bar represents the
648 entire HSV-2 genome while the second bar represents the region of the genome containing
649 UL30. B) Number of samples from each geographic region with the NR, CR, and rare genotypes.
650 MD denotes samples with missing data. Note that samples LT797622 – LT797636, LT797682,

LT797786, LT799380 (samples from Burrel et al., 2017) were not included in these counts. C) Crystal protein structure of the UL30 gene product, the HSV DNA polymerase. The part of the protein encoded by the recombinant region is highlighted in blue.

Figure 5: UL39 Recombinant Region. A) Schematic showing the HSV-1 sequence blocks for each UL39 recombinant genotype. HSV-1 sequence is colored in blue. ChHV/non-recombinant HSV-2 coding sequence is colored in green. Intergenic regions are purple. Letters denote amino acids that differ among the various genotypes and vertical lines above the HSV-1 row denote the location of nucleotides that differ among the various genotypes. The top bar represents the entire HSV-2 genome while the second bar represents the region of the genome containing UL39. Asterix (*) in NR genotype can be amino acid S or C depending on the sample. B) Number of sequences from each geographic region with the NR, CR, and rare genotypes. MD denotes sequences with missing data. Note that samples LT797622 – LT797636, LT797682, LT797786, LT799380 (sample from Burrel et al., 2017) were not included in these counts. C) Crystal protein structure of the UL39 gene product, the ribonuclease reductase (structure is of yeast ribonucleotide reductase). The part of the protein encoded by the recombinant region is highlighted in blue.

Figure 6: Differential recognition of HSV-1 x HSV-2 interspecies recombinants by HSV-specific T-cells. A) Polyclonal HSV-1-reactive CD4 T-cells from HSV-1-infected persons recognize UL30 expressed in either Cos-7 or IVTT systems from HSV-1 and HSV-2 UL30 CR, but not from HSV-2 UL30 NR genotype. Readout is T-cell proliferation. B) UL47 C-Terminal epitope-specific CD8 T-cell clone recognizes HSV-2 UL47-UL50 recombinant and non-recombinant HSV-2 but not HSV-1. C) UL47 N-Terminal epitope-specific CD8 T-cell clone recognizes non-recombinant HSV-2 but

not HSV-1 or HSV-2 UL47-UL50 recombinant. D) UL48 N-terminal epitope-specific CD8 T-cell line recognizes HSV-1 and HSV-2 UL47-UL50 recombinant but not non-recombinant HSV-2. E) UL48 C-terminal epitope-specific CD8 T-cell line clone recognizes HSV-1 and HSV-2 UL47-UL50 recombinant but not non-recombinant HSV-2. F) UL49-specific CD4 T-cell clone recognizes HSV-1, non-recombinant HSV-2, and the HSV-2 UL47-UL50 recombinant. Data are secreted IFN-γ. All assays are triplicate.

References

Anderson CA, August MJ, Hsiung GD. 1980. Pathogenicity of wild-type and temperature-sensitive mutants of herpes simplex virus type 2 in guinea pigs. *Infect Immun* **30**:159–169.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res* **28**:235–242.

Boni MF, Posada D, Feldman MW. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**:1035–1047. doi:10.1534/genetics.106.068874

Burrel S, Boutolleau D, Ryu D, Agut H, Merkel K, Leendertz FH, Calvignac-Spencer S. 2017. Ancient Recombination Events between Human Herpes Simplex Viruses. *Mol Biol Evol* **34**:1713–1721. doi:10.1093/molbev/msx113

Burrel S, Désiré N, Marlet J, Dacheux L, Seang S, Caumes E, Bourhy H, Agut H, Boutolleau D. 2015. Genetic Diversity within Alphaherpesviruses: Characterization of a Novel Variant of Herpes Simplex Virus 2. *J Virol* **89**:12273–12283. doi:10.1128/JVI.01959-15

Chiam C-W, Chan Y-F, Sam I-C. 2010. Changing trends of genital herpes in Kuala Lumpur, Malaysia, 1982-2008. *Int J STD AIDS* **21**:450–451. doi:10.1258/ijsa.2010.009569

Colgrove R, Diaz F, Newman R, Saif S, Shea T, Young S, Henn M, Knipe DM. 2014. Genomic sequences of a low passage herpes simplex virus 2 clinical isolate and its plaque-purified derivative strain. *Virology* **450–451**:140–145. doi:10.1016/j.virol.2013.12.014

Colgrove RC, Liu X, Griffiths A, Raja P, Deluca NA, Newman RM, Coen DM, Knipe DM. 2016. History and genomic sequence analysis of the herpes simplex virus 1 KOS and KOS1.1 sub-strains. *Virology* **487**:215–221. doi:10.1016/j.virol.2015.09.026

Conner J, Cross A, Murray J, Marsden H. 1994. Identification of structural domains within the large subunit of herpes simplex virus ribonucleotide reductase. *J Gen Virol* **75 (Pt 12)**:3327–3335. doi:10.1099/0022-1317-75-12-3327

Corey L, Adams HG, Brown ZA, Holmes KK. 1983. Genital herpes simplex virus infections: clinical manifestations, course, and complications. *Ann Intern Med* **98**:958–972.

Corey L, Huang M-L, Selke S, Wald A. 2005. Differentiation of herpes simplex virus types 1 and 2 in clinical samples by a real-time taqman PCR assay. *J Med Virol* **76**:350–355. doi:10.1002/jmv.20365

710 Corey L, Wald A. 2009. Maternal and neonatal herpes simplex virus infections. *N Engl J Med* **361**:1376–
711 1385. doi:10.1056/NEJMra0807633
712 Davison AJ. 2011. Evolution of sexually transmitted and sexually transmissible human herpesviruses. *Ann*
713 *N Y Acad Sci* **1230**:E37-49. doi:10.1111/j.1749-6632.2011.06358.x
714 Dong L, Li P, Oenema T, McClurkan CL, Koelle DM. 2010. Public TCR use by herpes simplex virus-2-
715 specific human CD8 CTLs. *J Immunol Baltim Md 1950* **184**:3063–3071.
716 doi:10.4049/jimmunol.0903622
717 Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals
718 in recombinant sequences. *Bioinforma Oxf Engl* **16**:573–582.
719 Gilbert M, Li X, Petric M, Krajden M, Isaac-Renton JL, Ogilvie G, Rekart ML. 2011. Using centralized
720 laboratory data to monitor trends in herpes simplex virus type 1 and 2 infection in British
721 Columbia and the changing etiology of genital herpes. *Can J Public Health Rev Can Sante*
722 *Publique* **102**:225–229.
723 Gottlieb SL, Deal CD, Giersing B, Rees H, Bolan G, Johnston C, Timms P, Gray-Owen SD, Jerse AE,
724 Cameron CE, Moorthy VS, Kiarie J, Broutet N. 2016. The global roadmap for advancing
725 development of vaccines against sexually transmitted infections: Update and next steps. *Vaccine*
726 **34**:2939–2947. doi:10.1016/j.vaccine.2016.03.111
727 Greninger AL, Roychoudhury P, Xie H, Casto A, Cent A, Pepper G, Koelle DM, Huang M-L, Wald A,
728 Johnston C, Jerome KR. 2018. Ultrasensitive Capture of Human Herpes Simplex Virus Genomes
729 Directly from Clinical Samples Reveals Extraordinarily Limited Evolution in Cell Culture. *mSphere*
730 **3**. doi:10.1128/mSphereDirect.00283-18
731 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and
732 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML
733 3.0. *Syst Biol* **59**:307–321. doi:10.1093/sysbio/syq010
734 Halliburton IW. 1980. Intertypic recombinants of herpes simplex viruses. *J Gen Virol* **48**:1–23.
735 doi:10.1099/0022-1317-48-1-1
736 Harland J, Brown SM. 1985. Isolation and characterization of deletion mutants of herpes simplex virus
737 type 2 (strain HG52). *J Gen Virol* **66 (Pt 6)**:1305–1321. doi:10.1099/0022-1317-66-6-1305
738 Harpaz R, Dahl RM, Dooling KL. 2016. Prevalence of Immunosuppression Among US Adults, 2013. *JAMA*
739 **316**:2547–2548. doi:10.1001/jama.2016.16477
740 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinforma Oxf*
741 *Engl* **17**:754–755.
742 Jing L, Haas J, Chong TM, Bruckner JJ, Dann GC, Dong L, Marshak JO, McClurkan CL, Yamamoto TN, Bailer
743 SM, Laing KJ, Wald A, Verjans GMGM, Koelle DM. 2012. Cross-presentation and genome-wide
744 screening reveal candidate T cells antigens for a herpes simplex virus type 1 vaccine. *J Clin Invest*
745 **122**:654–673. doi:10.1172/JCI60556
746 Johnson KE, Redd AD, Quinn TC, Collinson-Streng AN, Cornish T, Kong X, Sharma R, Tobian AAR, Tsai B,
747 Sherman ME, Kigozi G, Serwadda D, Wawer MJ, Gray RH. 2011. Effects of HIV-1 and herpes
748 simplex virus type 2 infection on lymphocyte and dendritic cell density in adult foreskins from
749 Rakai, Uganda. *J Infect Dis* **203**:602–609. doi:10.1093/infdis/jiq091
750 Johnston C, Magaret A, Roychoudhury P, Greninger AL, Reeves D, Schiffer J, Jerome KR, Sather C, Diem
751 K, Lingappa JR, Celum C, Koelle DM, Wald A. 2017. Dual-strain genital herpes simplex virus type
752 2 (HSV-2) infection in the US, Peru, and 8 countries in sub-Saharan Africa: A nested cross-
753 sectional viral genotyping study. *PLoS Med* **14**:e1002475. doi:10.1371/journal.pmed.1002475
754 Johnston C, Zhu J, Jing L, Laing KJ, McClurkan CM, Klock A, Diem K, Jin L, Stanaway J, Tronstein E, Kwok
755 WW, Huang M-L, Selke S, Fong Y, Magaret A, Koelle DM, Wald A, Corey L. 2014. Virologic and
756 immunologic evidence of multifocal genital herpes simplex virus 2 infection. *J Virol* **88**:4921–
757 4931. doi:10.1128/JVI.03285-13

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780. doi:10.1093/molbev/mst010

Koelle DM, Chen HB, Gavin MA, Wald A, Kwok WW, Corey L. 2001. CD8 CTL from genital herpes simplex lesions: recognition of viral tegument and immediate early proteins and lysis of infected cutaneous cells. *J Immunol Baltim Md 1950* **166**:4049–4058.

Koelle DM, Norberg P, Fitzgibbon MP, Russell RM, Greninger AL, Huang M-L, Stensland L, Jing L, Magaret AS, Diem K, Selke S, Xie H, Celum C, Lingappa JR, Jerome KR, Wald A, Johnston C. 2017. Worldwide circulation of HSV-2□×□HSV-1 recombinant strains. *Sci Rep* **7**:44084. doi:10.1038/srep44084

Koelle DM, Reymond SN, Chen H, Kwok WW, McClurkan C, Gyaltsong T, Petersdorf EW, Rotkis W, Talley AR, Harrison DA. 2000. Tegument-specific, virus-reactive CD4 T cells localize to the cornea in herpes simplex virus interstitial keratitis in humans. *J Virol* **74**:10930–10938.

Kolb AW, Adams M, Cabot EL, Craven M, Brandt CR. 2011. Multiplex sequencing of seven ocular herpes simplex virus type-1 genomes: phylogeny, sequence variability, and SNP distribution. *Invest Ophthalmol Vis Sci* **52**:9061–9073. doi:10.1167/iovs.11-7812

Kolb AW, Larsen IV, Cuellar JA, Brandt CR. 2015. Genomic, phylogenetic, and recombinational characterization of herpes simplex virus 2 strains. *J Virol* **89**:6427–6434. doi:10.1128/JVI.00416-15

Liu S, Knafels JD, Chang JS, Waszak GA, Baldwin ET, Deibel MR, Thomsen DR, Homa FL, Wells PA, Tory MC, Poorman RA, Gao H, Qiu X, Seddon AP. 2006. Crystal structure of the herpes simplex virus 1 DNA polymerase. *J Biol Chem* **281**:18193–18200. doi:10.1074/jbc.M602414200

Looker KJ, Magaret AS, May MT, Turner KME, Vickerman P, Gottlieb SL, Newman LM. 2015a. Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. *PloS One* **10**:e0140765. doi:10.1371/journal.pone.0140765

Looker KJ, Magaret AS, Turner KME, Vickerman P, Gottlieb SL, Newman LM. 2015b. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PloS One* **10**:e114989. doi:10.1371/journal.pone.0114989

Macdonald SJ, Mostafa HH, Morrison LA, Davido DJ. 2012a. Genome sequence of herpes simplex virus 1 strain KOS. *J Virol* **86**:6371–6372. doi:10.1128/JVI.00646-12

Macdonald SJ, Mostafa HH, Morrison LA, Davido DJ. 2012b. Genome sequence of herpes simplex virus 1 strain McKrae. *J Virol* **86**:9540–9541. doi:10.1128/JVI.01469-12

Mapelli M, Panjikar S, Tucker PA. 2005. The crystal structure of the herpes simplex virus 1 ssDNA-binding protein suggests the structural basis for flexible, cooperative single-stranded DNA binding. *J Biol Chem* **280**:2990–2997. doi:10.1074/jbc.M406780200

Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinforma Oxf Engl* **16**:562–563.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1**:vev003. doi:10.1093/ve/vev003

Martin DP, Posada D, Crandall KA, Williamson C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* **21**:98–102. doi:10.1089/aid.2005.21.98

Masese L, Baeten JM, Richardson BA, Bukusi E, John-Stewart G, Graham SM, Shafi J, Kiarie J, Overbaugh J, McClelland RS. 2015. Changes in the contribution of genital tract infections to HIV acquisition among Kenyan high-risk women from 1993 to 2012. *AIDS Lond Engl* **29**:1077–1085. doi:10.1097/QAD.0000000000000646

Morse LS, Buchman TG, Roizman B, Schaffer PA. 1977. Anatomy of herpes simplex virus DNA. IX. Apparent exclusion of some parental DNA arrangements in the generation of intertypic (HSV-1 X HSV-2) recombinants. *J Virol* **24**:231–248.

Newman RM, Lamers SL, Weiner B, Ray SC, Colgrove RC, Diaz F, Jing L, Wang K, Saif S, Young S, Henn M, Laeyendecker O, Tobian AAR, Cohen JI, Koelle DM, Quinn TC, Knipe DM. 2015. Genome Sequencing and Analysis of Geographically Diverse Clinical Isolates of Herpes Simplex Virus 2. *J Virol* **89**:8219–8232. doi:10.1128/JVI.01303-15

Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**:218–225. doi:10.1006/viro.1999.0056

Parsons LR, Tafuri YR, Shreve JT, Bowen CD, Shipley MM, Enquist LW, Szpara ML. 2015. Rapid genome assembly and comparison decode intrastrain variation in human alphaherpesviruses. *mBio* **6**. doi:10.1128/mBio.02213-14

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605–1612. doi:10.1002/jcc.20084

Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**:13757–13762. doi:10.1073/pnas.241370698

Ryder N, Jin F, McNulty AM, Grulich AE, Donovan B. 2009. Increasing role of herpes simplex virus type 1 in first-episode anogenital herpes in heterosexual women and younger men who have sex with men, 1992-2006. *Sex Transm Infect* **85**:416–419. doi:10.1136/sti.2008.033902

Salvucci LA, Bonneau RH, Tevethia SS. 1995. Polymorphism within the herpes simplex virus (HSV) ribonucleotide reductase large subunit (ICP6) confers type specificity for recognition by HSV type 1-specific cytotoxic T lymphocytes. *J Virol* **69**:1122–1131.

Sauerbrei A, Bohn-Wippert K, Kaspar M, Krumbholz A, Karrasch M, Zell R. 2016. Database on natural polymorphisms and resistance-related non-synonymous mutations in thymidine kinase and DNA polymerase genes of herpes simplex virus types 1 and 2. *J Antimicrob Chemother* **71**:6–16. doi:10.1093/jac/dkv285

Severini A, Tyler SD, Peters GA, Black D, Eberle R. 2013. Genome sequence of a chimpanzee herpesvirus and its relation to other primate alphaherpesviruses. *Arch Virol* **158**:1825–1828. doi:10.1007/s00705-013-1666-y

Smith JM. 1992. Analyzing the mosaic structure of genes. *J Mol Evol* **34**:126–129.

Spruance SL, Chow FS. 1980. Pathogenesis of herpes simplex labialis. I. Replication of herpes simplex virus in cultures of epidermal cells from subjects with frequent recurrences. *J Infect Dis* **142**:671–675.

St Leger AJ, Peters B, Sidney J, Sette A, Hendricks RL. 2011. Defining the herpes simplex virus-specific CD8+ T cell repertoire in C57BL/6 mice. *J Immunol Baltim Md 1950* **186**:3927–3933. doi:10.4049/jimmunol.1003735

Szpara ML, Tafuri YR, Parsons L, Shreve JT, Engel EA, Enquist LW. 2014. Genome Sequence of the Anterograde-Spread-Defective Herpes Simplex Virus 1 Strain MacIntyre. *Genome Announc* **2**. doi:10.1128/genomeA.01161-14

Tigges MA, Koelle D, Hartog K, Sekulovich RE, Corey L, Burke RL. 1992. Human CD8+ herpes simplex virus-specific cytotoxic T-lymphocyte clones recognize diverse virion protein antigens. *J Virol* **66**:1622–1634.

Topalis D, Gillemot S, Snoeck R, Andrei G. 2016. Distribution and effects of amino acid changes in drug-resistant α and β herpesviruses DNA polymerase. *Nucleic Acids Res* **44**:9530–9554. doi:10.1093/nar/gkw875

Tronstein E, Johnston C, Huang M-L, Selke S, Magaret A, Warren T, Corey L, Wald A. 2011. Genital shedding of herpes simplex virus among symptomatic and asymptomatic persons with HSV-2 infection. *JAMA* **305**:1441–1449. doi:10.1001/jama.2011.420

van Diemen FR, Kruse EM, Hooykaas MJG, Bruggeling CE, Schürch AC, van Ham PM, Imhof SM, Nijhuis M, Wiertz EJHJ, Lebbink RJ. 2016. CRISPR/Cas9-Mediated Genome Editing of Herpesviruses

Limits Productive and Latent Infections. *PLoS Pathog* **12**:e1005701. doi:10.1371/journal.ppat.1005701

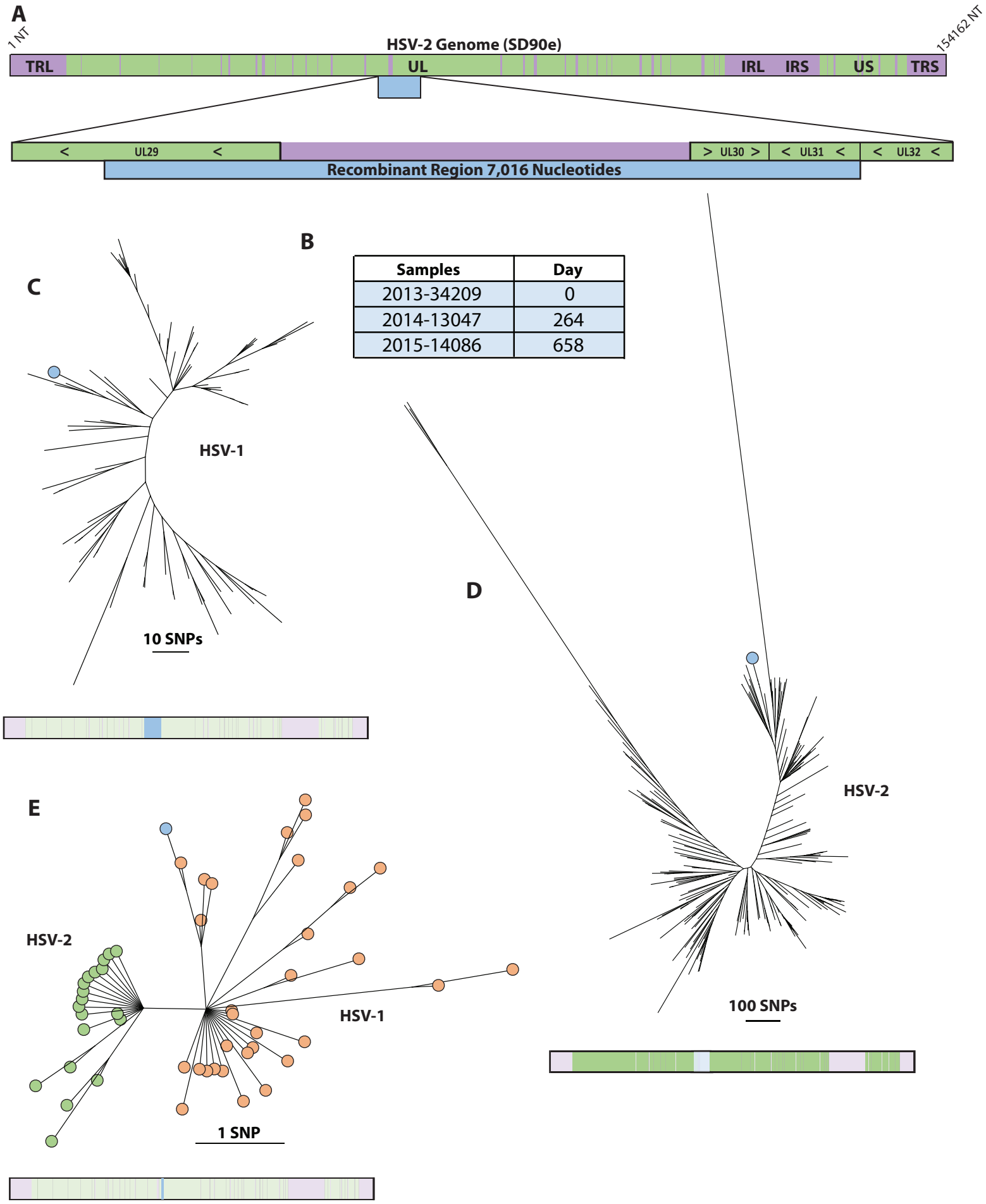
van Velzen M, Jing L, Osterhaus ADME, Sette A, Koelle DM, Verjans GMGM. 2013. Local CD4 and CD8 T-cell reactivity to HSV-1 antigens documents broad viral protein expression and immune competence in latently infected human trigeminal ganglia. *PLoS Pathog* **9**:e1003547. doi:10.1371/journal.ppat.1003547

Watson G, Xu W, Reed A, Babra B, Putman T, Wick E, Wechsler SL, Rohrmann GF, Jin L. 2012. Sequence and comparative analysis of the genome of HSV-1 strain McKrae. *Virology* **433**:528–537. doi:10.1016/j.virol.2012.08.043

Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL. 2014. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol Biol Evol* **31**:2356–2364. doi:10.1093/molbev/msu185

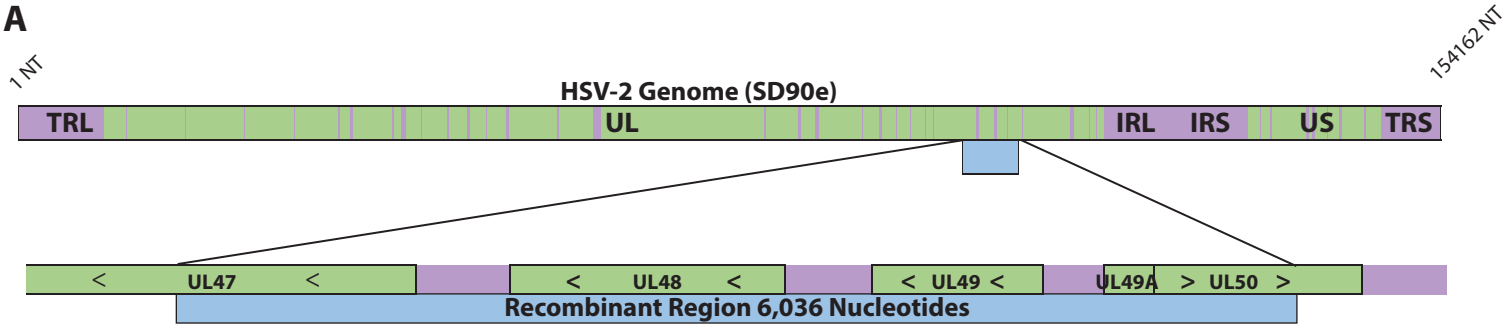
Xu H, Faber C, Uchiki T, Fairman JW, Racca J, Dealwis C. 2006. Structures of eukaryotic ribonucleotide reductase I provide insights into dNTP regulation. *Proc Natl Acad Sci U S A* **103**:4022–4027. doi:10.1073/pnas.0600443103

Zhu J, Hladik F, Woodward A, Klock A, Peng T, Johnston C, Remington M, Magaret A, Koelle DM, Wald A, Corey L. 2009. Persistence of HIV-1 receptor-positive cells after HSV-2 reactivation is a potential mechanism for increased HIV-1 acquisition. *Nat Med* **15**:886–892. doi:10.1038/nm.2006



Samples	Day
2013-34209	0
2014-13047	264
2015-14086	658

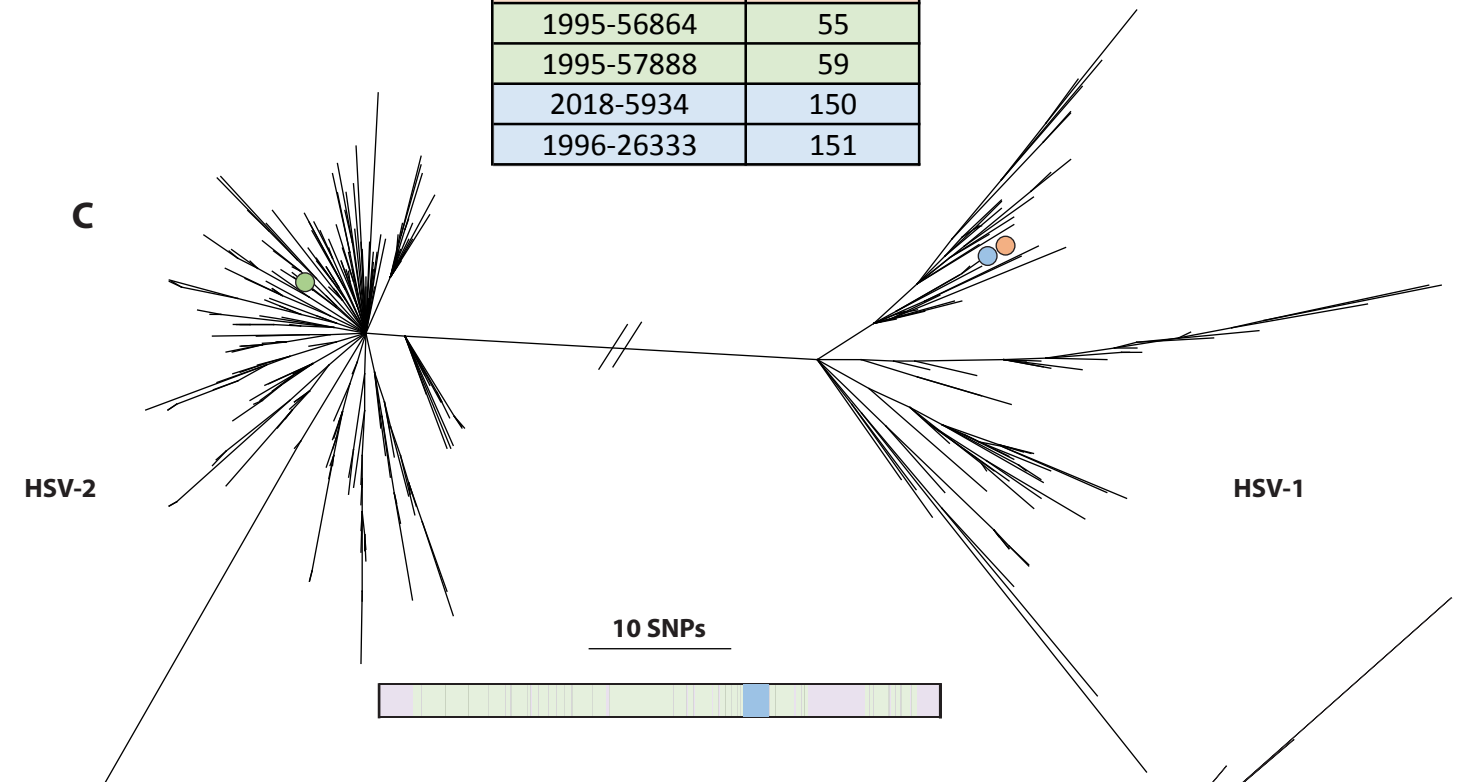
A



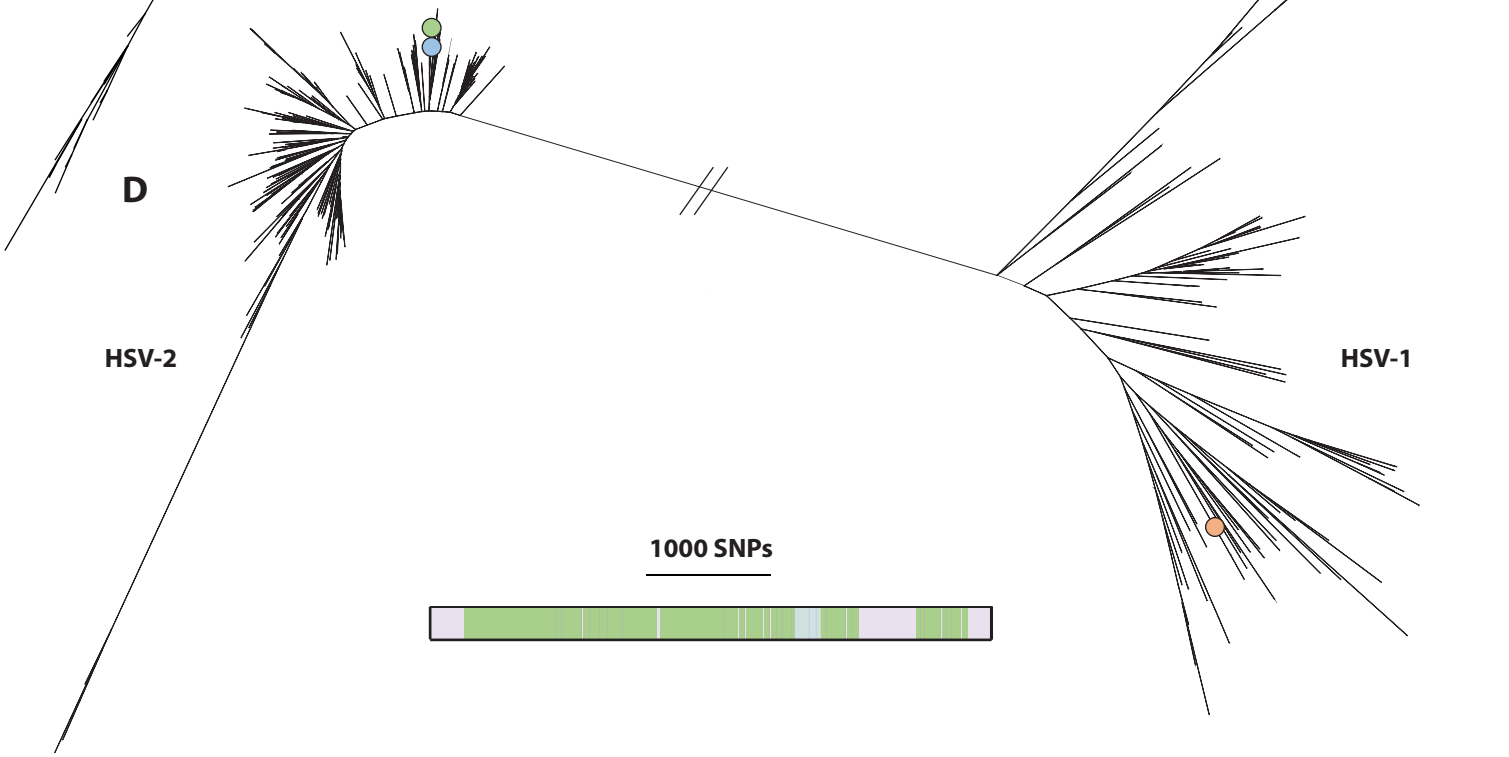
B

Samples	Day
1995-63175	0
1995-56864	55
1995-57888	59
2018-5934	150
1996-26333	151

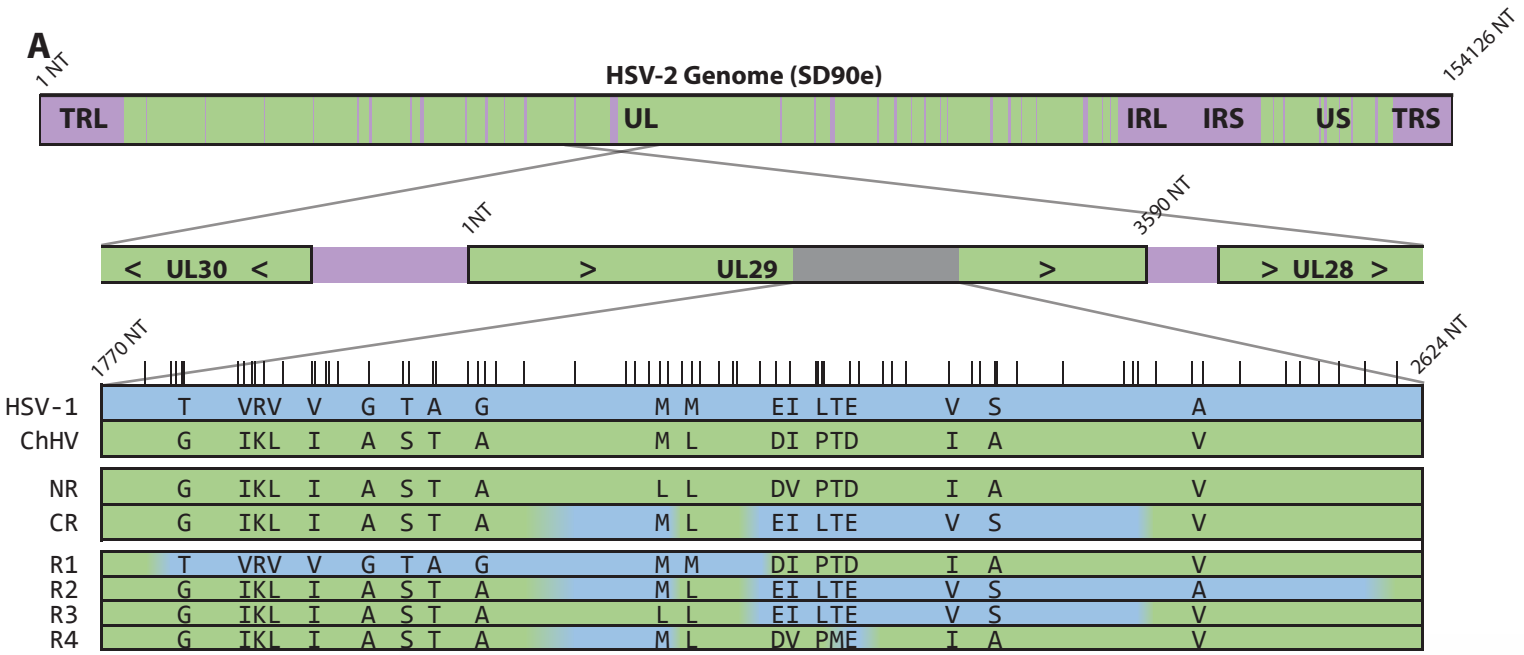
C



D



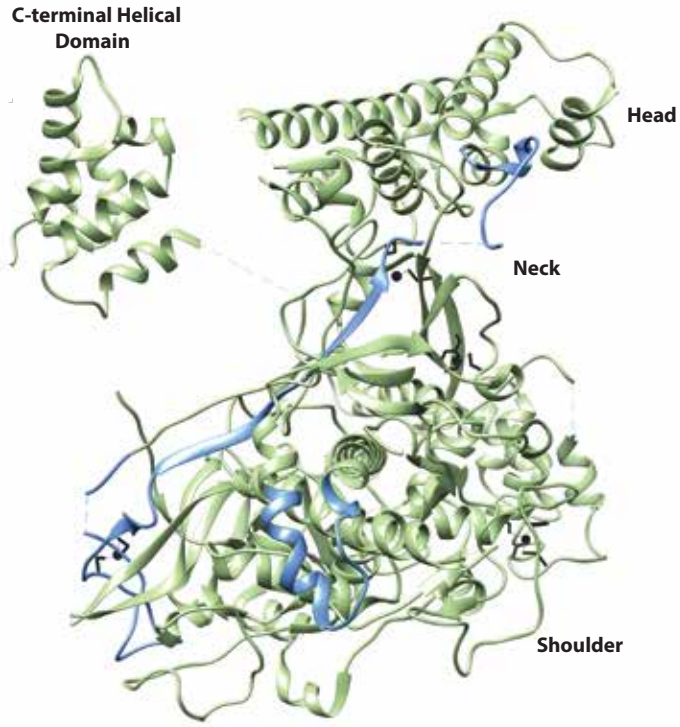
A

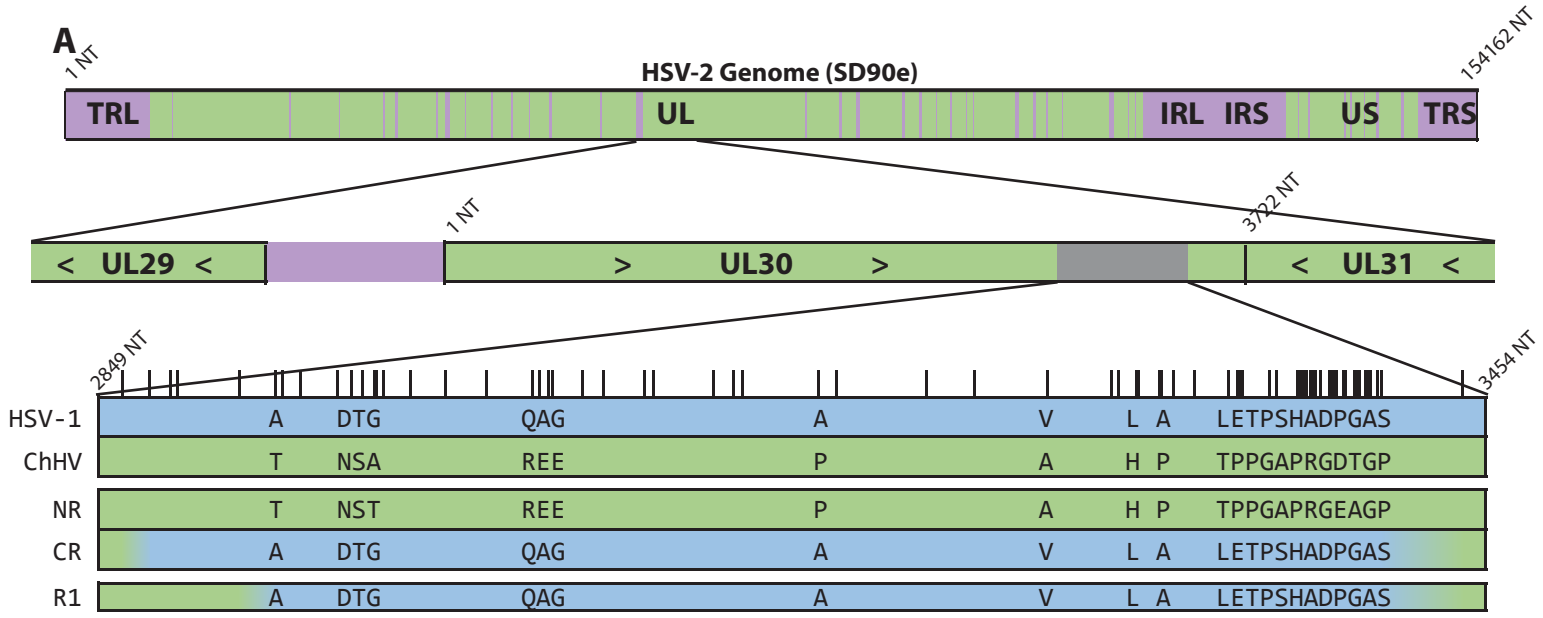


B

	Total Samples	UL29			
		NR	CR	Rare	MD
Total Samples	213	1	209	3	
Central Africa	19	1	17	1	
East Africa	26		25	1	
Southern Africa	19		19		
West Africa	6		6		
Asia	8		8		
Europe	8		8		
North America	110		109	1	
South America	16		16		
Geography Unknown	1		1		

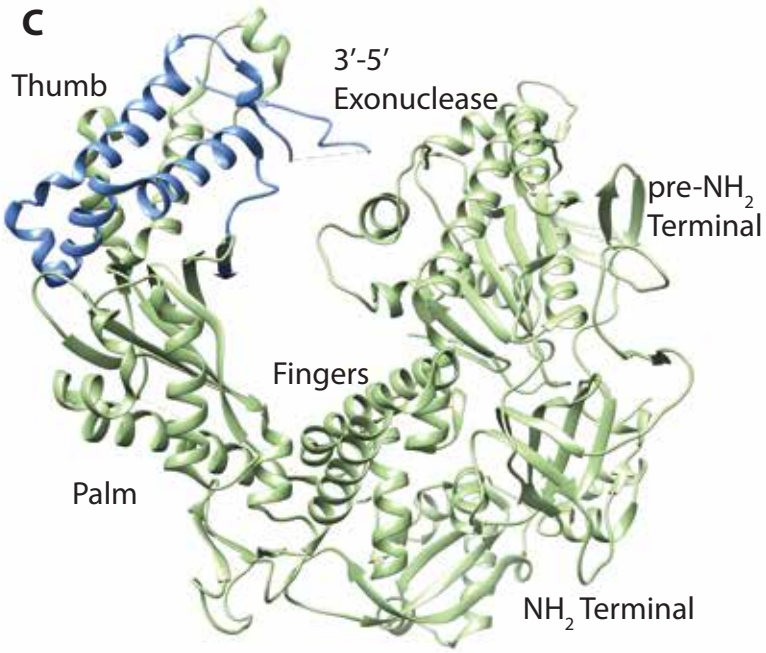
C



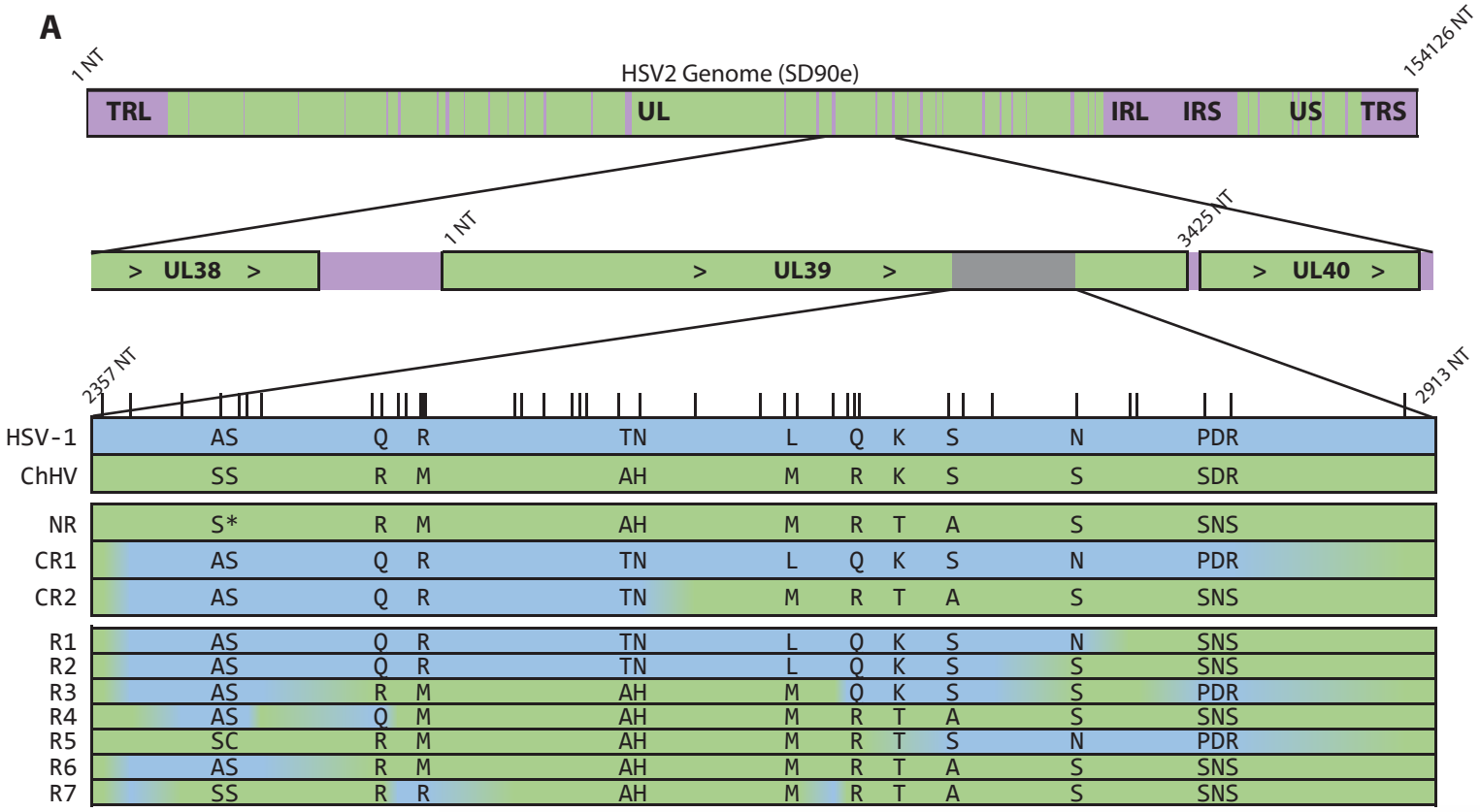


B

	Total Samples	UL30			
		NR	CR	Rare	MD
Total Samples	213	8	202	1	2
Central Africa	19	2	17		
East Africa	26	4	22		
Southern Africa	19		19		
West Africa	6		6		
Asia	8		8		
Europe	8		8		
North America	110	2	106	1	1
South America	16		15		1
Geography Unknown	1		1		



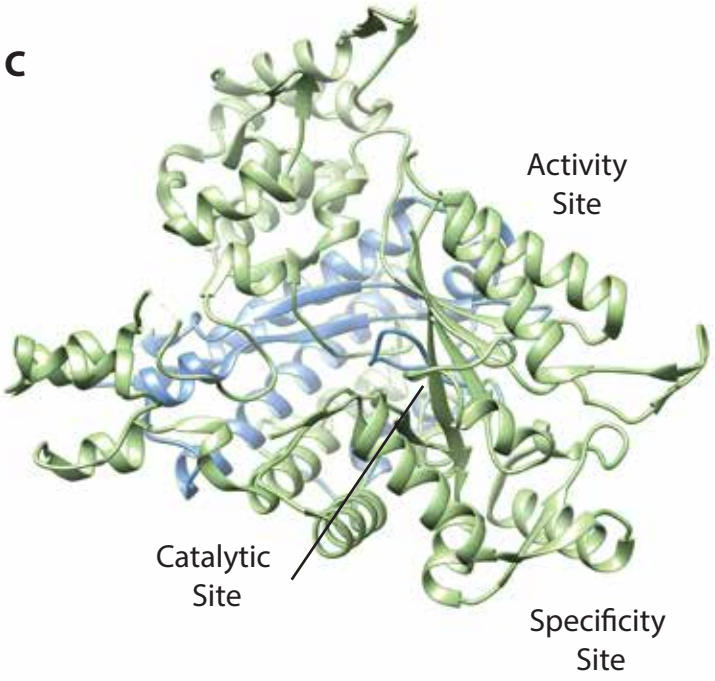
A



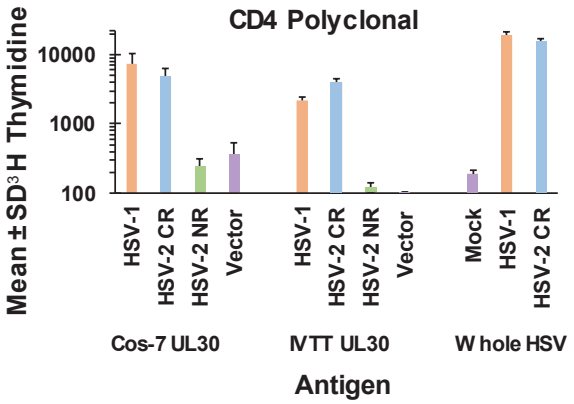
B

	Total Samples	UL39				
		NR	CR1	CR2	Rare	MD
Total Samples	213	135	54	16	6	2
Central Africa	19	8	3	6	1	1
East Africa	26	21	3	1	1	
Southern Africa	19	11	5	3		
West Africa	6	4	2			
Asia	8	8				
Europe	8	4	2	1	1	
North America	110	68	33	5	3	1
South America	16	10	6			
Geography Unknown	1	1				

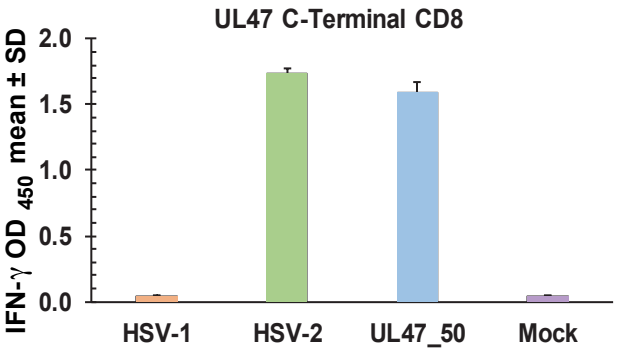
C



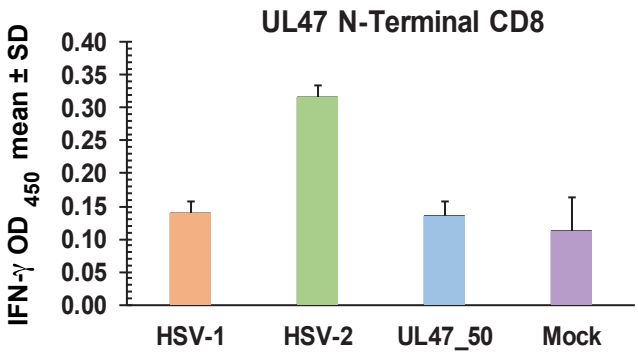
A



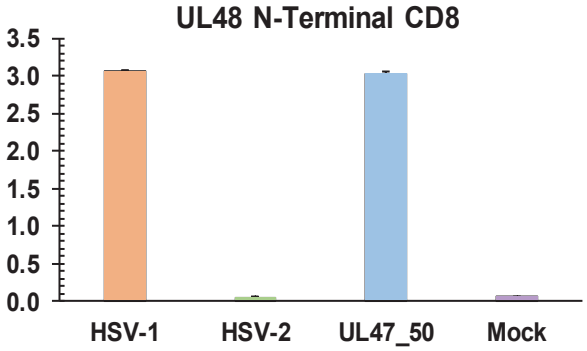
B



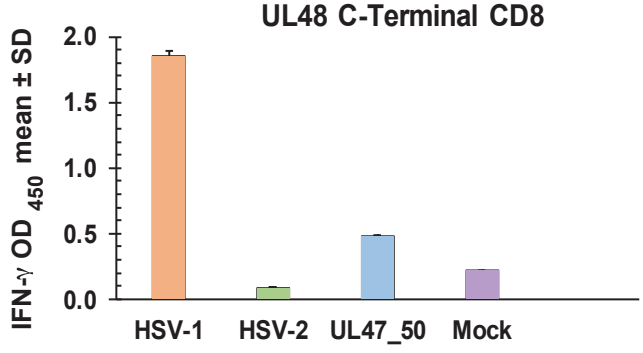
C



D



E



F