

PAPER ID: PMC5347006

TITLE:

Worldwide circulation of HSV-2×HSV-1 recombinant strains

ABSTRACT:

*Homo sapiens* harbor two distinct, medically significant species of simplexviruses, herpes simplex virus (HSV)-1 and HSV-2, with estimated divergence 6–8 million years ago (MYA). Unexpectedly, we found that circulating HSV-2 strains can contain HSV-1 DNA segments in three distinct genes. Using over 150 genital swabs from North and South America and Africa, we detected recombinants worldwide. Common, widely distributed gene UL39 genotypes are parsimoniously explained by an initial >457 basepair (bp) HSV-1 × HSV-2 crossover followed by back-recombination to HSV-2. Blocks of >244 and >539 bp of HSV-1 DNA within genes UL29 and UL30, respectively, have reached near fixation, with a minority of strains retaining sequences we posit as ancestral HSV-2. Our data add to previous in vitro and animal work, implying that in vivo cellular co-infection with HSV-1 and HSV-2 yields viable interspecies recombinants in the natural human host.

Participants and specimens ::: Results:

We studied genital specimens from Seattle and Portland, USA, and sites in Peru and western, eastern, and southern Africa. The source natural history and clinical trials are detailed Methods. Initially, we newly analyzed swab specimens from unique persons by NGS (ref. 13, Genbank KX574860 through KX574908). After observing evidence of HSV-1 × HSV-2 IRV and confirmation by manual sequencing (below), we newly obtained UL39 genotyping data from 111 other persons (Table 1). The criteria for additional specimen selection was to increase coverage to about 50 persons each in North and South American and Africa, and to maintain sex and HIV co-infection status distribution within geographic areas. Persons with HIV co-infection were relatively underrepresented in the US compared to South American and Africa. We used digital droplet PCR (ddPCR) (Supporting Information Methods) to genotype additional specimens. ddPCR was validated with specimens pre-sequenced by NGS and/or Sanger methods. The absolute fluorescence intensities for HSV-1 and HSV-2 genotypes varied between runs, but inclusion of samples with known genotypes in each ddPCR run permitted genotype assignment (examples, SI Fig. 1). One or both UL39 ddPCR assays failed for 9% of genital swab DNA specimens, correlating with low amounts of HSV-2 DNA in a quantitative PCR test. Overall, we obtained unambiguous UL39 genotyping data from 152 (91%) of 167 persons on whom it was attempted. Our final data set (Table 1) incorporated all previously available UL39 sequences (SI Table 1) with the NGS and ddPCR samples newly analyzed for this report. These sequences included sequence from 31 separate persons studied<sup>14</sup> by NGS from cultured HSV-2 isolates, all with geographic information, but some lacking sex and/or HIV infection data. We added analysis of 3 laboratory strains, 2 strains from South Africa<sup>15</sup>, and 5 reported by Kolb et al.<sup>16</sup> with limited demographics. Overall, the UL39 analyses combined newly and previously reported from specimens from 193 persons, amongst whom 28%, 26%, and 37%, respectively, are from North America, South America, and Africa, with a few from Asia or of unknown origin. Men comprised 38% of the cohort with 50% women and 12% unknown, while 44% were HIV-1 infected, 45% HIV-1 uninfected, and 11% unknown.

Detection of HSV-1×HSV-2 IRV ::: Results:

UL39 encodes the HSV large subunit of ribonucleotide reductase. We aligned 85 full-length HSV-2 UL39 sequences, including present and prior NGS data<sup>14</sup> 1516, Genbank, and in-house resequencing. The maximum nucleotide distance was 0.0183. Length varied from 1,141 to 1,146 AA/3,423 to 3,438 nucleotides (SI Table 2). Relative to the commonest 1,444 AA length in strain 186, strains SD90e<sup>15</sup> and G had identical 3 AA deletions, strain HG52 had a 2 AA deletion, and Ugandan strain D3965014 had a 2 AA insertion. Indels were in the UL39 N-terminus. HSV-1 UL39 sequences (*n* = 82, Genbank) were each 1,137 AA long and had lower 0.006 aximum nucleotide distance. HSV-1 17+<sup>17</sup> was used for further analyses.

We constructed UL39 phylogenetic trees from HSV-2, HSV-1 17+ , and the ChHV homolog. Across the entire gene, the larger of the two major branches (SI Fig. 2) contained HSV-2 SD90e and laboratory strain 186. These data support the recent suggestion<sup>15</sup> 18 to use SD90e as the

prototype HSV-2 strain, rather than strain HG52 (see below), and we named the largest UL39 group SD90e/186. We did use strain 186 rather than SD90e for recombination analyses herein, because strain SD90e has a very rare 9 nucleotide deletion in UL39 (above), while overall 81/85 strains (95%), including strain 186 had the same, prevalent length for UL39. Laboratory strain HG52 clearly fell in the minor UL39 branch. HSV-1 and ChHV segregated outside of these major HSV-2 groups, with ChHV closer to HSV-2 than HSV-1 from HSV-2. Aside from low prevalence SNPs and short indels, we noticed that the prevalent differences between UL39 sequences were concentrated in the 3' region of the ORF. We therefore re-analyzed from base pair (bp) 950 to the stop codon (Fig. 1) to focus on the divergent region. To simplify, we present data for the 26 HSV-2 unique sequences in this region, excluding identical sequences. The HSV-2 SD90e/186 group was again the largest. The minor group now resolved more clearly into two subgroups. One subgroup, designated HG52, included 8 of 85 strains (9.4% of total). The second subgroup contained 15 of 85 (17.6%) sequences and was designated G19080 (SI Table 1) after a representative member. G19080 was isolated from a 45 year old HIV-1 co-infected US woman (Genbank KR13530814). Notably, both the HG52 and G19080 HSV-2 subgroups segregated with HSV-1, while ChHV again segregated separately.

Segregation of some HSV-2 UL39 genes with HSV-1 (Fig. 1) could reflect continuous or discontinuous HSV-1 UL39 alleles. Sequence alignments revealed blocks of contiguous HSV-1 markers in the HG52 and G19080 subgroups (Fig. 2A, SI Table 3). The G19080 subgroup contained a HSV-1 DNA segment of at least 457 bp as defined by flanking 5' and 3' HSV-1 alleles (Table 2). The flanking HSV-2 type-specific alleles were located 12 and 72 bp further in the 5' and 3' directions, respectively. The exact crossover loci were ambiguous given the zones of HSV-1/HSV-2 sequence identity between type-specific alleles.

Every HG52 and G19080 subgroup virus had identical 5' and 3' crossover boundaries.

Remarkably, the 5' crossover marks for the HG52 and G19080 IRV subgroups were also identical (Table 2). The HG52 subgroup included bp 2,572 of HSV-1 as the 3' flanking HSV-1 allele, for a minimum 224 bp HSV-1 insert. These IRV had considerable UL39 amino acid variation from strain 186 (Fig. 2B) in addition to nucleotide differences.

We performed formal recombination analyses on HSV genome segments from UL37 through UL42 using HSV-1 17+ and HSV-2 186, G19080, and HG52 as representative strains. The bootscan algorithm readily identified crossover points with highly significant shifting (Fig. 3) coinciding with sequence inspection (SI Table 3). The three phylogenetic trees based on the recombinant fragment and the two flanking regions clearly demonstrated a shift in topology, where the recombinant HSV-2 strain 19080 clustered closely to HSV-1 in the tree representing the recombinant fragment, while clustering closely to the HSV-2 strain 186 in the flanking regions (Fig. 3). The high similarity between the HSV-2 strain 19080 and HSV-1 in the recombinant fragment was also shown by Simplot analysis. Similar results for a shorter recombinant fragment were observed for HSV-2 strain HG52 (Fig. 3). The phi-statistics for the presence of recombination had p-values of 0.0 for both HSV-2 strains G19080 and HG52, and the algorithms implemented in RDP419 (i.e. RDP, GENECONV, bootScan, MaxChi, Chimaera, and SiScan) all yielded p-values supporting recombination (ranging from  $2.7 \times 10^{-71}$  to  $3.2 \times 10^{-13}$ , and from  $2.0 \times 10^{-41}$  to  $2.0 \times 10^{-07}$  for G19080 and HG52, respectively). ChHV was included and no areas of significant recombination included this strain (Fig. 3).

#### Rare UL39 IRV genotypes ::: Results:

We used a ddPCR assay (loci schematized in Fig. 2) to rapidly genotype additional samples (Supplementary Methods, SI Table 4 for methods; sample data in SI Fig. 1). Strains HSV-1 17+ , HSV-2 G19080, HSV-2 HG52, and HSV-2 186 had the predicted HSV1:HSV1, HSV1:HSV1, HSV2:HSV1, and HSV2:HSV2 genotypes at the 5' and 3' UL39 loci respectively. Most swab-derived DNA specimens had one of these genotype patterns. A new HSV1:HSV2 ddPCR pattern, consistent with additional UL39 IRV, was detected in in two HIV-uninfected participants (samples 2010\_29297 and 2008\_35742) (Fig. 2, SI Fig. 1). For participant 5060 who donated sample 2010\_29297, 11 swabs collected over 37 days had this ddPCR genotype.

Sanger sequencing (SI tables 5 and 6) confirmed the ddPCR data and defined two additional UL39 IRV genotypes. Each had 3' crossover markers identical to G19080 (SI Table 3, Fig. 2A) and novel 5' crossovers internal to the long G19080 HSV-1 insert. Sample 2008\_35742 had a longer HSV-1 insert than sample 2010\_29297, containing 11 additional HSV-1 type-specific allelic variant markers and extending a minimum of 77 bp further in the 5' direction within UL39. Bootscan, Simplot and phylogenetic analyses again were highly consistent with recombination (SI Fig. 3). Likely due to the relatively short region analyzed, the SiScan method failed to detect

recombination in specimen 2010\_29297 (participant 5060), although all other methods in RDP4 ( $p < 8.0 \times 10^{-5}$ ) and the phi-statistics ( $p = 4.1 \times 10^{-7}$ ) presented high statistical significance. Each RDP4 method ( $p < 1.5 \times 10^{-6}$ ) and the phi statistics ( $p = 4.3 \times 10^{-9}$ ) were significant for recombination for 2008\_35742 (participant 6376). Taken together, the most parsimonious explanation for both the common HG52 group and these rare, short HSV-1 UL39 fragments is an ancestral long recombination event giving rise to G19080, followed by back-recombination events wherein a part of the HSV-1 fragment in G19080 was expelled by recombination with a 186-like HSV-2 strain.

We noted additional unique sequences in the C-terminal UL39 phylogeny (Fig. 1). Sequencing revealed three additional probable additional recombinants, each with HSV-1-like sequences remaining internal to the long G19080 HSV-1 insert (Supplementary Table 3), but with complex patterns. Sample 2009\_2198 from an HIV-infected Zambian man and sample G19083 (ref. 14, from an HIV-uninfected US man) each had distinct sequence with four contiguous HSV-1 markers (blue bars in Fig. 2A). The 3' crossover markers of the HSV-1 insert in 2009\_2198 match the HG52 subgroup, suggesting common ancestry and a backcross. Sample 2009\_4556 from an HIV co-infected Kenyan man has three regions of contiguous HSV-1 allelic markers. The 5' end of 5'-most HSV-1 insert and the 3' end of the 3'-most HSV-1 insert were identical to the HSV-1 flanking markers of G19080, suggesting common ancestry and backcrosses. Overall, we detected 6 distinct backcross patterns of HSV-1  $\times$  HSV-2 UL39 recombination within the longest G19080 recombination event (Fig. 2A).

#### Distribution of UL39 variants ::: Results:

Evidence for geographically restricted HSV-2 genotypes is limited<sup>1420</sup>. From 193 persons in this report and published data (Table 1)<sup>141516</sup>, we excluded 8 published strains with uncertain geographic origin and one HIV co-infected and one HIV non-infected participant with evidence for dual strain infection (below), leaving 183 participants (Table 3). Excluding the 5 rare UL39 genotypes noted above, 178 had mono-infection with a prevalent UL39 variant (51 from North America, 49 from South America, 70 from Africa, and 8 from Asia). Each major UL39 genotype was detected in Africa, North America, and South America in roughly equal proportions (Table 3). Specifically amongst 51 participants from North America with single-strain infection, 31 (61%), 16 (31%) and 4 (8%) had the SD90e, G19080, and HG52 genotypes, respectively. For the 49 participants from South American, the distribution was 31 (63%), 16 (33%), and 2 (4%) respectively for these genotypes, and for the 70 donors from Africa with single-strain infection, the distribution was 45 (57%), 17 (24%), and 8 (11%). Asian participants, limited to 8 from Japan<sup>14</sup>, each had the SD90e genotype. Amongst these 178 participants with the 3 prevalent UL39 genotypes, there was no association between continent and UL39 genotype ( $p = 0.30$  or  $p = 0.58$  excluding or including Japanese specimens, respectively).

#### Distribution of recombinant strains in persons of known HIV status ::: Results:

Amongst the 183 persons donating specimens with a single UL39 genotype and with known geographic origin, we examined the distribution of UL39 genotypes and HIV infection status (Table 3). Amongst these persons, 13 had unknown HIV status, leaving 170 participants, and within these, 5 were infected with HSV-2 strains with rare UL39 genotypes, leaving 165 persons (83 HIV-uninfected and 82 HIV co-infected) infected with one of the three prevalent UL39 genotypes. Each major UL39 genotype were detected in both HIV-uninfected and HIV-infected persons in roughly equal proportions. Amongst the 83 HIV-uninfected persons, 50 (60%), 25 (30%) and 8 (10%) were infected with the SD90e, G19080, and HG52 genotypes, respectively. Among 82 the HIV co-infected persons, 53 (65%), 23 (28%), and 6 (7%) were infected with SD90e, G19080, and HG52 genotypes, respectively. There was no apparent association between HIV infection status and the three prevalent UL39 genotypes ( $p = 0.51$ ). For UL30 (discussed below), 7 of 72 (9.7%) HIV co-infected and 2 of 66 (3%) HIV-uninfected persons with available data had the variant ( $p = 0.29$ ). For UL29 (also discussed below), only one person had the rare variant such that HIV co-infection correlates were not meaningful.

#### Co-infection with HSV-2 UL39 variants ::: Results:

Within-person dual-strain HSV-2 infection is required for HSV-2  $\times$  HSV-2 recombination, such as those implied by the data discussed above. A rectal swab from a HIV-infected Peruvian man, R-103-1010 had multiple C-terminal heterozygous UL39 loci detected by NGS, manual sequencing of bulk UL39 PCR product, and clonal analysis of amplicons cloned from PCR product (SI Fig. 2, Supplementary Table 6). The pattern indicated a simultaneous, dual HSV-2

SD90e-group and G19080-group infection pattern. UL39 bulk PCR amplicon sequencing of 10 additional swabs collected over a 14 day genital HSV-2 outbreak were also each heterozygous, indicating a prolonged episode of simultaneous dual-strain infection. Separately, we found that HIV un-infected North America participant 11848 had either a pure SD90e group- or a pure G19080-group UL39 ddPCR pattern in swab DNA samples collected 4 years apart (SI Table 6), consistent with dual strain infection but not simultaneous reactivation.

#### HSV-2 UL30 and UL29 genes with HSV-1×HSV-2 recombination ::: Results:

UL30 encodes DNA polymerase with mutations conferring drug resistance<sup>21</sup>. It is required for viral replication, has a proof-reading exonuclease function, and cooperates with UL42 processivity factor. Burrell et al. sequenced a C-terminal UL30 variant, termed HSV-2v, from cultured isolates from HIV co-infected West Africans<sup>5</sup>. We sequenced UL30 in Ugandan specimen 5073\_9333, using target capture to enrich HSV DNA (Greninger et al. in preparation), and obtained sequence essentially identical to HSV-2v (SI Table 6). We found HSV-2v in 9 (6%) of 146 persons using a ddPCR assay data (SI Table 4). These included 6 of 49 Eastern or Southern African, 1 of 46 North American, and 2 of 51 South American subjects. HIV co-infection was present in 7 persons, but absent in one Ugandan, and in the US subject. The American participants with HSV-2v UL30 had the G19080 HSV-1 UL39 insert, while each African had the SD90e group UL39 gene.

Burrell et al. proposed that HSV-2v UL30 resulted from HSV-2-ChHV recombination<sup>5</sup>. We aligned full length UL30 from HSV-1 (n = 98), ChHV22, HSV-2v, HSV-2 NGS sequences<sup>13141516</sup>, laboratory strains, and HSV-2 sequences in Genbank. HSV-1 UL30 were closely related, with maximum pairwise uncorrected nucleotide distance of 0.01 and little C-terminal variation. HSV-1 17+ and HSV-2 186 were chosen as the species genotypes for recombination analyses for consistency with our analyses of UL39. For both UL30 and UL29, strains 186 and the proposed prototype strain SD90e were extremely similar. We noted that most HSV-2 UL30 sequences have a 539 bp zone of identity to HSV-1 (Table 2, Fig. 4C). This zone contains over 70 consecutive HSV-1-specific allelic SNPs and an HSV-1-specific 12 bp insert. Lateral to this zone, all HSV-2 UL30 genes, including HSV-2v, are similar, and quite dissimilar from HSV-1 (Fig. 4D). Contravening the ChHV hypothesis, many SNPs and a 3 bp indel in the 539 bp region separate HSV-2v from ChHV. Bootscan (Fig. 4D) confirmed a HSV-1 × HSV-2 crossover event with highly shifting bootstrap values supporting different phylogenetic topologies. Simplot analysis further supported recombination by significant shifts in sequence similarity. Phylogenetic analyses of discrete regions of UL30 demarcated by the putative crossover zones associated with the tree clustering patterns (Fig. 4D). The phi-statistics ( $p = 0.0$ ) and RDP4 algorithms demonstrated high significance for recombination ( $p < 7.6 \times 10^{-18}$  for all methods). Most contemporary HSV-2, exemplified by strains SD90e and 186, have an HSV-1 UL30 insert. We hypothesize that UL30 HSV-2v may retain an older sequence related to ChHV.

HSV UL29 encodes a DNA binding protein, ICP8, involved in DNA unwinding and origin of replication processes, interacting with multiple viral and host proteins. HSV-1 UL29 sequences (n = 81) were very similar, with a maximum nucleotide distance of 0.0097 and no indels. Alignments of HSV-2 UL29 disclosed one divergent sequence, 2009\_4556, from a 30 year old HIV-infected Kenyan (Fig. 4A). Phylogenetic analyses (Fig. 4B) paralleled UL30. Between bp 2186 and bp 2439, HSV-2 186 and SD90e, and other HSV-2 strains show near-identity to HSV-1, including 19 consecutive HSV-1-like alleles. This 254 bp segment placed HSV-1 in the same group as the majority HSV-2 genotype (Fig. 4B). This single unique strain contained 10 differences from ChHV in this region, arguing against origination through recombination with ChHV. Lateral to this region, all HSV-2 were similar but not identical to ChHV and quite divergent from HSV-1. Similar to UL39 and UL30, the bootscan, Simplot, and segmental phylogenetic analyses of UL29 showed distinct shifts in phylogenetic topologies and sequence similarity, consistent with HSV-1 × HSV-2 recombination. Recombination was statistically supported by the phi-statistics ( $p = 4.4 \times 10^{-10}$ ) and the RDP4 program ( $p < 5.9 \times 10^{-6}$  for all methods except RDP, which did not detect recombination). We propose that an ancestral HSV-1 × HSV-2 UL29 crossover has become near-fixed in HSV-2, with 2009\_4556 as the only currently known strain perhaps retaining an ancestral sequence.

#### Discussion:

We document HSV-2 strains containing regions of HSV-1 sequence identity within three distinct nucleotide metabolism genes, UL29, UL30, and UL39, involved in HSV virulence, replication, and drug activity. The strains likely arose from interspecies recombination. The principle of HSV-1 ×

HSV-2 recombination in vivo in the natural host has now been established. These findings have implications for viral evolution, HSV drug therapy, vaccines, and viral vector safety. UL30 is targeted by HSV drugs<sup>23</sup> and UL29 and UL39 are also antiviral targets<sup>24,25,26,27,28</sup>. While UL30 HSV-2v strains are drug-sensitive<sup>23</sup>, our data suggest that resistance alleles could disseminate via recombination, including across the HSV-1/HSV-2 divide. Attenuated HSV-2 strains<sup>29</sup> used as vaccines are also of potential concern. These could recombine with HSV-1 to yield replication-competent variants with unexpected properties. Similarly, a replication competent HSV-1 strain expressing GM-CSF is US-licensed for intratumoral injection as oncolytic therapy, at up to  $4 \times 10^8$  infectious units. Recurrent HSV-2 replication can occur at widespread anatomic areas, including within the central nervous system to cause recurrent meningitis<sup>30</sup>. The possibility of recombinants bearing HSV-2 virulence and replication genes and GM-CSF may be possible has been previously discussed<sup>31</sup>. There is no reason to think that gain of GM-CSF by a wild-type HSV-2 strain would lead to a selective advantage or increase pathogenicity. Of potentially higher concern, some of the genes deleted in the cancer therapy HSV-1 strain to reduce virulence could be regained, as their HSV-2 versions, via interspecies recombination.

HSV-1 and HSV-2 have separate, overlapping ecologic niches. HSV-1 lytic infection typically involves oral or ocular regions, with latency in trigeminal ganglia. HSV-2 lytic replication is commonest in the anogenital region, with latency in lumbosacral ganglia. Alpha herpesviruses in non-human primates typically show one of these anatomic preferences. The genes responsible for tropism are poorly understood. Remarkably, primary HSV-1 infection now accounts for the majority of cases of clinical first episode genital herpes in the US and northern Europe, while in earlier decades, HSV-2 predominated<sup>32,33</sup>. An RFLP search for HSV-1  $\times$  HSV-2 IRV amongst HSV-1 strains recovered from human recurrent genital infections did not reveal insertions of HSV-2 sequence<sup>34</sup>. Thus far, analysis of wild-type, low-passage HSV-1 strains by NGS has not revealed insertion of HSV-2 sequences<sup>35,36,37</sup>, but NGS-based examination of genital HSV-1 strains has not been performed.

HSV-1  $\times$  HSV-2 IRV can occur after cell co-infection in vitro. Complex patterns of multiple crossovers are typically observed within progeny strains. There is some sequence preference for inclusion of origins of replication<sup>338</sup> in crossed-over segments. Both these spontaneous IRV, and engineered IRV that are crossed over at desired loci, have been widely used to map species-specific phenotypes (for example<sup>39,40</sup>). The large number of such studies cannot be reviewed in detail. Regarding anatomic tropism, in vivo research using engineered HSV-1  $\times$  HSV-2 IRV suggests segregation of tropism with DNA encoding latency-associated transcripts<sup>41</sup>. Thus far, we have not detected naturally occurring IRV in these loci. Studies in vivo suggest that IRV generally have attenuated virulence<sup>42,43</sup>. In contrast the IRV described in this report by definition are virulent for infection of humans. IRV can occur in vivo after experimental simultaneous HSV-1 and HSV-2 co-inoculation of animals<sup>44</sup>. All animal models have a fundamental differences in pathogenesis compared to HSV infection of humans, for example acute lethality in some systems, and lack of recurrences in almost all. Given the exquisite host-pathogen co-evolution of HSV with humans, demonstration the HSV-1  $\times$  HSV-2 IRV occur in the natural host establish a new mechanism for HSV evolution.

We hypothesize that all variants of HSV-2, as well as ChHV, are descended from a common ancestor of HSV-2/ChHV, estimated to have infected a primate around 2 million years ago, rather than by HSV-2/ChHV recombination<sup>5</sup>. A recent estimate places human/chimp divergence at 13 million years ago<sup>45</sup>. We cannot date the HSV-2  $\times$  HSV-1 recombination events, but based on the relatively equal geographic distribution of the major UL39 variants, each appears to predate human emigration from Africa (now estimated at about 70,000 years ago<sup>46</sup>). Recent horizontal transmission of UL39 variants is also possible. Resolution will require larger, curated specimen sets including those from isolated populations, and optimally archival DNA from before the era of widespread travel. In preliminary analyses of phylogenetic networks, we studied the HSV-1 UL39 insert found in HSV-2 strain G19080 and the corresponding region from known HSV-1 strains. No clustering into clades was observed as seen for HSV-1 US7 or US8. Interestingly, the HSV-2 strain 19080 UL39 HSV-1 fragment does not cluster more distantly or closely to any particular HSV-1 strain than do HSV-1 strains cluster with each other. This implies that the HSV-1 UL39 fragment in HSV-2 G19080 was incorporated into this strain through homologous recombination after emergence of the most recent common ancestor of circulating HSV-1 strains, estimated to have occurred 700,000 years ago<sup>4</sup>.

After their origination, IRV have had different fates in the population. Within UL29 and UL30, small minorities of strains lack the HSV-1-like regions, likely related to fitness. For UL39, in contrast, the relatively equal prevalence of the three major genotypes implies that none has a large fitness

advantage. It is also possible, especially for UL29 and UL30, that a bottleneck occurred during migration from Africa, as the rare variants are thus far found mostly in Africa.

The grouping of IRV signatures in the 3' region of HSV UL39, including evidence for multiple back-crosses, suggests an ancestral recombination leading to the G19080 group, and susceptibility to back-recombination in this area. It is possible that the recombination points that flank the HSV-1 insert within the HSV-2 G19080 group arose multiple times due to strong site preferences for recombination. Sequence bias for HSV IRV formation has been noted *in vitro*<sup>3</sup>, but the mechanism is unknown. In this earlier work, crossovers were noted near UL39 but not mapped precisely. The HSV origin of replication in the HSV unique long (UL) region is between UL29 and UL30, but it is not known if this is related to IRV formation. A correlation was found between high GC content and intratypic recombination for HSV-147. HSV-1 UL29, UL30, and UL39 GC content very near the average of 66.8% for HSV-1 UL, disfavoring GC content as an explanation. Overall, we consider multiple, independent crossover events less likely than persistence of a relatively ancient recombinant G19080 UL39 variant in the human population as well as the persistence of the progeny of occasional back-crosses to HSV-2.

Strengths of our approach include the use of direct sequencing, international specimens, the availability of multiple specimens from some subjects to document multi-strain infection, and sequence confirmation with several technologies. HSV sequences can shift *in vitro*<sup>48</sup> with mutation rates 10–100 fold higher for HSV-2 than HSV-1, and may favor certain genes<sup>49</sup>. UL39 appears stable *in vitro*: our lab's strain 333 is identical to that from 3 other labs. The complementary genotyping platforms each have strengths and weaknesses. In-lab progress in HSV-2 target enrichment has recently made direct NGS sequencing practical and economical for swab specimens with scarce HSV-2 DNA and will likely be used in the future.

Our study has several caveats. We may have missed additional IRV within UL39, as ddPCR rather than sequencing was used for some specimens. For example, specimen 2008\_35742 and specimen 2010\_29297 had the same ddPCR patterns (Fig. 2A), but sequencing revealed different 5' recombination zones. We used a relatively limited set of HSV-2 sequences to detect IRV. Additional full-length HSV-2 sequences including specimens from other continents will likely identify other sites of recombination. Of special interest are detailed studies of specimens from persons with both HSV-1 and HSV-2 infections at the same anatomic site, which might reveal additional and possibly transitory IRV. Our cross-sectional specimen set likely failed to detect some interspecies recombination events that have occurred in the past due to decay of IRV with a fitness disadvantage, and similarly, the temporal sequence of events leading to the variants circulating today cannot be assigned with certainty. Long term, IRV with fitness advantages could reach 100% prevalence, appear as tracts of HSV-1/HSV-2 identity, and again be undetectable without longitudinal samples. The effects of the recombination events detected in this study can be tested *in vitro* or in animals, but it will be difficult to prove their effects on virulence in the natural host.

It has recently been suggested that HSV-2 strain SD90e be used as a prototype, as it is more virulent in animals, is lower passage, and has a functional copy of a gene disabled in HG5215. As discussed herein, we elected to use strain 186 as the representative of the most abundant UL39 genotype for our formal recombination analyses, as strain SD90e contains a rare 9 basepair deletion in UL39. Otherwise, the sequences of 186 and SD90e are very similar for the three genes in which we detected interspecies recombination and overall, our data partially support SD90e as an HSV-2 prototype. The selection of a HSV-2 strain as genotypic prototype is complex. Strains 333 and SD90e are identical in the C terminal region of UL39, but the 9 basepair deletion seen in SD90e is not present in 333. We propose that SD90e is a better rational candidate for prototype strain given the fact that 333 is highly passaged and has not been compared head to head with other strains for virulence or protection after administration of candidate vaccines, as has SD90e<sup>18</sup>. For HSV-1, the use of HSV-1 17+ for analyses of UL29, UL30, and UL39 is reasonable as there is no consensus as to prototype strain. For the three HSV genes in this report, strain 17+ is closely representative of every HSV-1 sequence data available in Genbank, including those recently analyzed by Szpara et al.<sup>353650</sup> and of any other HSV-1 strain would have no bearing on the conclusions of this study. This area will likely undergo further refinement as additional sequences and analyses become available.

Functional studies with variant UL29, UL30, and UL39 proteins and virus genotypes will be necessary to examine the biological consequences of variation. These genes are pivotal in HSV pathogenesis. Briefly, UL39 is non-essential *in vitro* but required for virulence *in vivo*, while UL29 and UL30 are essential regardless. UL39 variants could influence dNTP pools and neuronal reactivation, as UL39 is required for replication in neurons with low endogenous dNTP levels.

Additional UL39 functions include an N-terminal RHIM domain that inhibits necroptosis<sup>51</sup>, a possible protein kinase<sup>52</sup>, and a C-terminal region has been implicated in inhibition of apoptosis. The UL29 ssDNA binding protein is involved in DNA unwinding and origin of replication processes, interacting with multiple cellular and viral proteins. The UL30 DNA polymerase has a proofreading exonuclease function and cooperates with the UL42 processivity factor; mutations can lead to mutator phenotypes and altered drug sensitivity. The variant proteins uncovered herein can be studied biochemically, in vitro, or in vivo in the isogenic virus context. Within the immunocompetent population there is enormous person-to-person variation in the frequency of HSV-2 symptomatic recurrence and asymptomatic shedding<sup>53</sup>. Quantitative study of reactivation parameters in persons with prevalent IRV genotypes might reveal phenotypic-genotype correlations.

In summary, contemporary circulating genital HSV-2 strains recovered from widely distributed geographic areas show evidence for recombination with HSV-1. The pattern of UL39 sequence variants, taken together, is consistent with an ancestral crossover event followed by backcrosses to HSV-2. Data for UL29 and UL30 are most consistent with an ancient HSV-1 crossover that has been successfully propagated in the human population, leaving a minority of residual strains with sequences similar to ChHV, the closest known relative of HSV-2. Overall, the principle of HSV-1 × HSV-2 interspecies recombination in vivo has now been established in the natural host.

#### Subjects and specimens ::: Methods:

Protocols were approved by the University of Washington Institutional Review Board and all participants provided informed consent. All methods were performed in accordance with the relevant guidelines and regulations for human subjects research. Sources included natural history and interventional protocols HPTN 039 and Partners in Prevention, concerning HSV-2 and HIV transmission<sup>54,55</sup>, natural history studies, and a drug treatment study<sup>56</sup>. Subjects and specimen details for NGS have been documented<sup>57</sup>.

#### Sequence determination, confirmation, and genotyping ::: Methods:

The next-generation sequencing (NGS) laboratory and analysis workflow has been reported<sup>57</sup> and data uploaded (Genbank accessions KX574861-KX574908 inclusive). BAM Illumina sequence files were assembled and visualized with NCBI Genome Workbench and Integrative Genomics Viewer<sup>58</sup>. Supporting Information Methods contains primers for PCR amplification and dideoxy sequencing of selected regions of HSV-2 DNA (SI Table 5), rapid genotyping by ddPCR (SI Table 4), and Genbank accession numbers (SI Table 6) of new sequences.

#### Nucleotide alignment and recombination detection ::: Methods:

We accessed full-length ChHV, HSV1 and HSV-2 UL39 sequences from recent NGS data and other Genbank-available sequences including BLAST searches using HSV-2 HG52 or HSV-1 strain 17 full length U39 sequences as search terms. Only full-length sequences without ambiguous base calls were used. UL39 sequences were analyzed full-length or starting at base 1,000 in HSV-2 strain 186 coordinates. Alignments were performed with MegAlign Pro (Lasergene, Madison, WI) using MUSCLE and default parameters. Genetic distances were calculated by the Tamura-Nei method<sup>59</sup> using Megalign Pro. Phylogenetic trees were generated in MegAlign Pro using default parameters and visualized with Figtree 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). Re-rooting for visualization was done within Figtree. Recombination crossovers were identified and visualized using the bootscan method included in the Simplot program. Sequence similarities between different strains were analyzed and visualized using Simplot. The parameter settings for bootscan and Simplot were as follows: Window size: 300 bp, step: 5 bp, gapstrip: on, Kimura (2-parameter), and T/t = 2.0. Detection of recombination, and calculation of statistical significance for recombination, was further performed using the RDP, GENECONV, bootScan, MaxChi, Chimaera, and SiScan algorithms implemented in the RDP4 program<sup>19</sup> using default settings, and the phi-statistics implemented in the SplitsTree program, after discarding all but one of each set of identical sequences. Phylogenetic trees (networks) based on recombinant fragments, and flanking regions, were constructed using the SplitsTree program and the uncorrected P characters transformation and default settings. Similar analyses were performed for UL30 and UL29.

#### Statistics ::: Methods:

The abundance of UL39 genotypes in different continents or among persons with different HIV status was compared with Fisher's exact test or Chi-squared test, two-tailed.

Additional Information:

How to cite this article: Koelle, D. M. et al. Worldwide circulation of HSV-2 × HSV-1 recombinant strains. *Sci. Rep.*

7, 44084; doi: 10.1038/srep44084 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.