Genotypic and phenotypic diversity within the neonatal HSV-2 population

Lisa N. Akhtar[1], Christopher D. Bowen[2], Daniel W. Renner[2], Utsav Pandey[2], Ashley N. Della Fera[3], David W. Kimberlin[4], Mark N. Prichard[4], Richard J. Whitley[4], Matthew D. Weitzman[3,5]*, Moriah L. Szpara[2]*

[1] Department of Pediatrics, Division of Infectious Diseases, Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine

[2] Department of Biochemistry and Molecular Biology, Center for Infectious Disease Dynamics, and the Huck Institutes of the Life Sciences, Pennsylvania State University

[3] Division of Protective Immunity and Division of Cancer Pathobiology, Children's Hospital of Philadelphia

[4] Department of Pediatrics, Division of Infectious Diseases, University of Alabama at Birmingham

[5] Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine

***Corresponding Authors**
Moriah L. Szpara
Dept. of Biochemistry & Molecular Biology
The Huck Institutes of the Life Sciences
W-208 Millennium Science Complex (MSC)
Pennsylvania State University
University Park, PA 16802 USA
Phone: 814-867-0008
Email: moriah@psu.edu

Matthew D. Weitzman
Division of Protective Immunity
The Children's Hospital of Philadelphia
4050 Colket Translational Research Building
3501 Civic Center Blvd
Philadelphia, PA 19104-4318
Phone: 267-425-2068
Email: weitzmanm@email.chop.edu

Author emails:
Lisa N. Akhtar: AkhtarL@email.chop.edu
Christopher D. Bowen: cdb5004@psu.edu
Daniel W. Renner: dwr19@psu.edu
Utsav Pandey: uup104@psu.edu
Ashley N. Della Fera: DellaFeraA@email.chop.edu
David W. Kimberlin: DKimberlin@peds.uab.edu
Mark N. Prichard: MPrichard@peds.uab.edu
Richard J. Whitley: RWhitley@peds.uab.edu
Matthew D. Weitzman: weitzmanm@email.chop.edu
Moriah L. Szpara: moriah@psu.edu

Keywords: Neonatal, herpes simplex virus 2, human herpesvirus 2, comparative genomics, viral spread, minority variants

Abstract *(currently 246 words)*

More than 14,000 neonates are infected with herpes simplex virus (HSV) annually. Approximately half display manifestations limited to the skin, eyes, or mouth (SEM disease). The rest develop invasive infections that spread to the central nervous system (CNS disease) or systemically (disseminated disease). This invasive HSV disease is associated with significant morbidity and mortality, but viral and host factors that predispose neonates to these forms are unknown. To define viral diversity within the neonatal population, we evaluated ten HSV-2 isolates from newborns with a range of clinical presentations. To assess viral fitness independent of host immune factors, we measured viral growth characteristics in cultured cells and found diverse *in vitro* phenotypes. Isolates from neonates with CNS disease were associated with larger plaque size and enhanced spread, with isolates from cerebrospinal fluid (CSF) exhibiting the most robust growth. We sequenced complete viral genomes of all ten neonatal viruses, providing new insights into HSV-2 genomic diversity in this clinical setting. We found extensive inter-host and intra-host genomic diversity throughout the viral genome, including amino acid differences in more than 90% of the viral proteome. The genes encoding glycoprotein G (gG, US4), gI (US7), gK (UL53), and viral proteins UL8, UL20, UL24, and US2 contained variants that were found only in association with CNS isolates. Many of these viral proteins are known to contribute to cell spread and neurovirulence in mouse models of CNS disease. This study represents the first application of comparative pathogen genomics to neonatal HSV disease.

Significance Statement *(currently 119 words)*

Herpes simplex virus (HSV) causes invasive disease in half of infected neonates, resulting in significant mortality and permanent cognitive morbidity. The factors that contribute to invasive disease are not understood. This study reveals diversity among HSV isolates from infected neonates, and makes the first associations between viral genetic variations and clinical disease manifestations. We found that viruses isolated from newborns with invasive brain disease show enhanced spread in culture. These viruses contain protein-coding variations not found in viruses causing non-invasive disease. Many of these variations are found in proteins known to impact neurovirulence and viral spread between cells. This work advances our understanding of HSV diversity in the neonatal population and will aid the development of neuroprotective anti-viral therapies.

Introduction

Each year an estimated 10,000 neonates are infected with HSV-2, and 4,000 infected with HSV-1, worldwide (1). Infants are typically infected at the time of birth by maternal genital shedding of HSV; most often mothers are not aware of their infection (2–4). The recent increase in genital HSV-1 incidence among women of childbearing age, particularly in developed nations, suggests that the burden of neonatal infection will continue to rise (1, 5). While some infected infants exhibit only superficial infection limited to the skin, eyes, or mouth (SEM disease; 45%), about half develop invasive systemic (disseminated disease; 25%) or central nervous system (CNS disease; 30%) infections associated with significant morbidity and mortality (6, 7). Currently, the

antiviral medication acyclovir is the standard therapy for all forms of neonatal HSV disease. Although this intervention has reduced mortality due to invasive disease, most survivors are left with permanent neurodevelopmental deficits (8, 9).

The factors that predispose a neonate to invasive HSV infection are not entirely known. Recent studies have found that a portion of individuals outside of the neonatal period who experience HSV infection of the brain have a host genetic defect within the Toll-like receptor-3 (TLR3) pathway (10, 11). HSV encephalitis outside of the neonatal period is uncommon, and its association with the rare host defects in the TLR3-pathway has not been examined. By contrast, half of HSV-infected neonates experience invasive CNS or disseminated disease, making it less likely that host genetic defects alone could account for all of the observed cases of invasive infection in neonates. Prior clinical data on mother-to-infant transmission of HSV indicate that most cases of neonatal disease, including invasive forms of disease, result from newly acquired HSV infection in the absence of maternal immune protection (2–4, 12). This suggests a window of opportunity where the contributions of viral genetic variation to the progression of invasive infection and disease may be greater than in adults.

Prior studies have identified viral genetic factors that influence virulence or disease for reoviruses, influenza virus, HIV, and others (13–18). In contrast to these RNA viruses, HSV was presumed to have lower genetic diversity and potential for variation in virulence, due to its relatively stable DNA genome and long co-evolutionary history with humans (19). The assumption of limited HSV heterogeneity was supported by early studies that utilized low-resolution restriction fragment length polymorphism (RFLP) or single-gene analyses to compare multiple HSV isolates (20–22). However Rosenthal and colleagues used RFLP and PCR analysis

of a single locus to demonstrate that a heterogeneous HSV population can exist in an invasive neonatal infection, and provided proof of principle that natural genetic variation can impact neurovirulence (23, 24). More recently, advances in high-throughput sequencing (HTSeq) have enabled a re-evaluation of herpesvirus genome-wide variation, which suggests that herpesviruses harbor extensive diversity both between strains or individuals (inter-host variation) as well as within a single individual (intra-host variation) (25–27). These minor genetic variants may become clinically important if a variant within the viral population becomes the new dominant allele or genotype as a result of a bottleneck at transmission, entry into a new body compartment, or selective pressure such as antiviral therapy (28, 29).

Several recent examples have demonstrated the potential for new insights gained by applying HTSeq approaches to herpesvirus infections in a clinical setting. In HTSeq studies of congenital infection by the beta-herpesvirus human cytomegalovirus (HCMV), Renzette *et al.* found evidence for heterogeneous viral populations both within and between hosts (30–35). The levels of diversity observed in congenital HCMV infections far exceeded those observed in adult infections (36–38). HTSeq-based examination of vaccine-associated rashes due to the alpha-herpesvirus varicella zoster virus (VZV) demonstrated that adult skin vesicles contain a subset of the viral population introduced during vaccination, and found at least 11 VZV genomic loci that were linked to rash formation (26). Recent HTSeq comparisons of adult genital HSV-2 revealed the first evidence of a single individual shedding two distinct strains (39), demonstrated changes in the viral genome over time in a recently infected host (40), and provided the first evidence of ancient recombination between HSV-1 and HSV-2 (41, 42). However, to date there has been no evaluation of genome-wide variation in neonatal HSV isolates to determine the levels of diversity in this population, or the potential impact(s) of viral genetic variants on disease.

Until recently, several technical barriers prevented thorough assessment of neonatal HSV genomes. A key constraint on studies of neonatal disease has been the availability of cultured viral samples associated with clinical information that have also been maintained in a low-passage state appropriate for sequencing and further experimental studies. Historically viral culture was part of the HSV diagnostic workflow, but this has been superseded in clinical laboratory settings by the speed and sensitivity of viral detection by PCR (43, 44, 6). This change limits neonatal HSV sample availability for *in vitro* and animal model studies. In addition, many previously archived neonatal HSV isolates have been passaged extensively, allowing them to acquire mutations that enhance viral growth in culture (23, 24). Additional challenges for HTSeq approaches to neonatal HSV include the large size of the viral genome (~152 kb), its high G+C content (~70%), and a large number of variable-number tandem repeats in the viral genome (> 240 mini/micro-satellite repeats and >660 homopolymers of ≥6 base pairs) (45, 27). Therefore, many studies of HSV diversity (21, 46–49) or the effect of HSV genetic variation on disease (50–52) have relied on low-resolution restriction fragment length polymorphism (RFLP) or single-gene PCR analyses, due to the speed and ease of analysis in comparison to whole-genome approaches (53–58). To overcome these challenges, we combined our expertise in HSV comparative genomics and phenotypic analysis (59, 53, 60, 58) with a unique resource of low-passage, well-annotated neonatal specimens (8, 61, 9).

Here we analyzed a set of ten low-passage clinical HSV-2 isolates collected from neonates with HSV infection, enrolled in one of two clinical studies that spanned three decades of patient enrollment (1981-2008) (8, 61, 9). These samples represented a wide range of clinical manifestations including SEM, CNS, and disseminated disease, and each sample was associated with a robust set of de-identified clinical information. We defined the level of diversity in this

population using comparative genomics and an array of cell-based phenotypic assays. We found that HSV-2 isolates displayed diverse *in vitro* phenotypes, as well as extensive inter- and intra-host diversity distributed throughout the HSV-2 genome. Finally, we found coding variations in several HSV-2 proteins associated with CNS disease. This study represents the first-ever application of comparative pathogen genomics to neonatal HSV disease and provides a basis for further exploration of genotype-phenotype links in this clinically vulnerable patient population.

Results

Neonatal HSV-2 samples represent a diverse clinical population

We utilized samples collected from ten HSV-2-infected neonates enrolled by the National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group (CASG) for clinical trials between 1981 and 2008 (8, 61, 9). These infants encompassed a range of clinical disease manifestations (see **Table 1**), with about half experiencing invasive CNS disease (5 patients) or disseminated (DISS) disease with CNS involvement (2 patients), and the remainder experiencing non-invasive SEM disease (3 patients). Extensive clinical information was available for each patient, including long-term neurocognitive and motor outcomes (**Table 1**). This population was also diverse with respect to sex, race, gestational age, and enrollment center (**Table 1**; enrollment center data not shown). All samples were collected at the time of diagnosis, prior to initiation of acyclovir therapy. Each isolate was cultured once as part of the diagnostic process, with expansion only for the experiments shown here. Although the sample size was constrained by the rarity of neonatal HSV infection and availability of appropriately maintained isolates, our group is similar in size to prior HTSeq comparisons of congenital HCMV samples

(31, 32, 34, 35), and is the largest group of neonatal HSV samples ever subjected to comparative genomic and phenotypic analysis.

Neonatal HSV-2 isolates have different fitness in culture

To determine whether the viruses isolated from this neonatal population **(Table 1)** were intrinsically different, we assessed viral growth in culture, which provides a consistent environment that is independent of host genetic variation. To minimize the impact of immune pressure we selected Vero monkey kidney cells, which lack an interferon response (62, 63). Each viral isolate was applied to a confluent monolayer of cells *in vitro*, and allowed to form plaques for 100 hours (**Figure 1A**). Average plaque size differed between isolates, with large plaques being more frequent among viruses derived from neonatal CNS disease (**Figure 1B**). Viruses isolated directly from the cerebrospinal fluid (CSF; isolates CNS11 and DISS14) produced plaques with an average size that was statistically larger than those isolated from the skin (one-way ANOVA followed by Holm-Sidak's multiple comparisons test, $p<0.05$). Plaque size was assessed throughout passage in culture and remained constant from the time the isolates were received in our laboratory (passage 2) through their genetic and phenotypic analysis (passage 4). The low-passage HSV-2 clinical isolate SD90e, which was isolated from an adult patient, was used as a control for comparison (64). The variance in plaque sizes produced by a given isolate was not statistically different between isolates **(Figure 1B)**. The differences in average plaque size between isolates suggested that the HSV-2 populations found in each neonatal isolate are indeed intrinsically different.

Entry kinetics, DNA replication, protein expression, and virus production do not account for differences in plaque size

Plaque formation is a complex endpoint that involves the ability of the virus to enter the cell, replicate its double stranded DNA genome, produce viral proteins and assemble new virions that spread to adjacent cells. Therefore, we explored whether the differences in plaque formation observed in Vero cells reflected inherent differences in the ability of isolates to complete each of these stages of the viral life cycle. For these comparisons, two large-plaque-forming isolates associated with CNS disease were selected, including one isolated from the CSF (CNS11) and one from the skin (CNS03). These were compared to two small-plaque-forming isolates associated with either CNS (CNS12) or SEM disease (SEM02), both of which were isolated from the skin. First, we compared each isolate's rate of cell entry. Virus was applied to chilled cells, followed by warming to synchronize cell entry **(Figure 2A).** A low pH solution was applied at various points over the first hour of cell entry to inactivate any virus that had not yet entered a cell, and plaque formation was then allowed to proceed for 100 hours. We found no difference in rates of cell entry between these four representative viral isolates (**Figure 2B**), suggesting that large plaques did not result from increased rates of virus entry into cells.

We next infected Vero cells at high multiplicity of infection (MOI=5) to compare the outcome of a single round of viral replication. We found that all four isolates produced similar numbers of genome copies (as measured by qPCR for the gB gene; see Methods for details) **(Figure 2C**), indicating that differences in viral DNA replication did not influence plaque size. We quantified the production of infectious virus by counting plaque-forming-units (PFU) on Vero cell monolayers, as well as on the highly-permissive U2OS human bone osteosarcoma epithelial cell

line (65). No differences in virus production were noted between the four isolates when quantified on either cell type (**Figure 2D, E**). U2OS cells lack innate sensing of viral infection through the STING pathway (66) and can even support the growth of highly-defective HSV-isolates that lack ICP0 function (65). All isolates formed large plaques on the highly-permissive U2OS cell monolayers (**Figure S1**), allowing us to rule out the possibility that very small foci of infection were missed during titering of the small plaque-forming isolates on Vero cells (**Figure 2D, E**). Finally, we compared viral protein production for these isolates, and found no differences in the expression levels of a panel of HSV-2 viral proteins at either early (6 hours post infection, hpi) or late (24 hpi) time points of a single round of high MOI infection (**Figure 2F, G**; see legend for list of proteins). Taken together, these results suggested that large- and small-plaque-forming neonatal HSV-2 isolates did not differ significantly in viral entry, DNA replication, infectious virus production, or protein production over a single round of infection.

Large-plaque-forming isolates exhibit enhanced cell-to-cell spread

We next assessed the ability of representative large- and small-plaque-forming isolates to spread from cell-to-cell. Vero cell monolayers were infected at low MOI (MOI=0.001) in order to assess differences over multiple rounds of viral replication and spread throughout the cell monolayer (**Figure 3A**). The contribution of indirect cell-free spread was minimized by including 0.1% human serum in the media, and changing the media every 24 hours to remove released virus and refresh serum levels. Over a 72-hour time course, cell monolayers were assessed to measure the extent of cell-to-cell spread, either by harvesting and titering PFU production (**Figure 3B-C**), or by fixing infected cell monolayers and evaluating the distribution of virus by immunofluorescence (**Figure 3D**, **E**, and **Figure S2**). Viral titers recovered from harvested cells

were similar at 2 hpi, confirming that equivalent amounts of each virus were present following initial infection (**Figure 3B**). By 72 hpi, after multiple rounds of replication and spread, the large-plaque-forming isolates CNS11 and CNS03 had achieved viral titers significantly greater than those of the small-plaque-forming isolates CNS12 and SEM02 (**Figure 3B, C**; two-way ANOVA followed by Tukey's multiple comparison test, $p<0.0001$ at 72h). Isolate CNS11, which was obtained directly from the CSF, produced titers statistically greater than all isolates collected from the skin **(Figure 3B, C)**.

We also directly evaluated cell-to-cell spread by immunostaining and quantifying the distribution of infected cells around each infectious focus. Infected Vero cell monolayers were fixed at 24, 48, and 72 hpi, and subjected to fluorescent immunocytochemistry with an antibody directed against total HSV (**Figure 3D** and **Figure S2**). The region of HSV-positive cells surrounding a single initial infection was greater for the large-plaque-forming neonatal HSV-2 isolates at 24 hpi, and increased by 48 and 72 hpi. The central cytolytic clearings seen in the CNS11- and CNS03-infected monolayers were approximately two-fold greater than those for CNS12 and SEM02, reflecting the average two-fold increase in plaque size observed in methylene-blue-stained monolayers in **Figure 1**. However, the region of infected cells surrounding the central cytolytic clearing for CNS11 and CNS03 was dramatically larger than for CNS12 and SEM02, suggesting that large-plaque-forming neonatal HSV-2 isolates show greater spread from cell-to-cell than could have been predicted by measuring plaque size alone. To quantify this increase in the area of infected cells after a low-MOI infection, each coverslip was imaged **(Figure S2)** and the total number of immunofluorescent pixels was quantified (**Figure 3E**). By 72 hpi, the area of HSV-infected cells was statistically greater for the large-plaque-forming isolates CNS11 and CNS03, as compared to the small-plaque-forming isolates CNS12 and SEM02 (two-way

ANOVA followed by Tukey's multiple comparison test, $p<0.05$ at 72h). Together these data indicated that large-plaque-forming isolates associated with CNS disease shared an enhanced ability to spread cell-to-cell through culture, in comparison to small-plaque-forming isolates.

Comparative genomics reveals genetic diversity in neonatal HSV-2 isolates

The differences identified in cell-to-cell spread between neonatal isolates in culture indicated the existence of intrinsic differences between these viruses. To reveal how genetic variation may contribute to viral phenotypes in culture, and ultimately to clinical disease manifestations, we sequenced the complete viral genome of all ten neonatal HSV-2 isolates. For each isolate, we sequenced purified viral nucleocapsid DNA and assembled a consensus genome, which represents the most common genotype at each nucleotide locus in the viral population. The clinical trials utilized in this study enrolled HSV-infected infants from multiple sites across the United States (8, 9, 61). Therefore, we first assessed the overall degree of relatedness between these viral genomes to understand whether any similarities in viral or geographic origin might have contributed to *in vitro* or clinical phenotype patterns. In light of the known potential for recombination in the phylogenetic history of HSV (49, 67, 53, 42), we used a graph-based network to investigate the phylogenetic relationship between these isolates. We found a similar degree of divergence among all ten neonatal HSV-2 consensus genomes (**Figure 4A**), which was further corroborated by their distribution in a network graph of available HSV-2 genomes from GenBank (**Figure 4B**, see **Table S1** for HSV-2 GenBank accessions). This suggested that similarities in viral genetic origin were not responsible for determining the cellular or clinical outcomes of neonatal HSV-2 infection.

Overall protein-coding diversity in neonatal HSV-2 isolates is similar to that observed in adult HSV-2 strains

We next asked whether overt defects in any single HSV-2 protein might be associated with clinical or *in vitro* spread phenotypes. In examining the coding potential of all ten neonatal HSV-2 isolates, we found no protein deletions or truncations encoded by any of these viral genomes. These comparisons excluded the viral proteins ICP34.5 (RL1) and ICP4 (RS1), whose coding sequence is not fully determined in these isolates due to sequencing gaps and/or incomplete assembly at G+C-rich tandem repeats in these regions. Issues in sequencing and/or assembly of these two genes, as well as the nearby open reading frame (ORF) for ICP0 (RL2), have been observed in all prior HTSeq studies of HSV-2 genomes (68–70). For all other HSV-2 proteins, we found a total of 784 nucleotide differences in 71 genes (see **Table S2**), resulting in 342 non-synonymous AA changes in 65 proteins. This led to an average of 1% coding variation (range 0.0 : 2.6%). This variation was spread widely throughout the HSV-2 genome, without concentration in a particular genomic region or category of protein function **(Figure 5).** We found a similar level of coding diversity in this set of neonatal isolates as in a comparable set of 10 adult HSV-2 isolates **(Figure 5, Table S2).** We also compared these neonatal genomes to the full set of 58 annotated adult HSV-2 genomes from GenBank (listed **Table S1**). As expected from the difference in sample size, there was more overall diversity found across 58 adult HSV-2 genomes than in 10 neonatal genomes. A comparison of the dN/dS ratio in neonatal vs. adult HSV-2 genomes revealed a similar trend with a few outliers visible on each axis, e.g. UL38 and US8A had a higher dN/dS ratio in neonates than in adult isolates (see **Table S2** and **Figure S3** for full comparison). These data indicated that at the consensus level, neonatal HSV-2 isolates display substantial inter-host coding diversity spread throughout the genome, but do not possess

strikingly more diversity or an excess of genetic drift as compared to adult isolates.

Minority variants expand the potential coding diversity of neonatal HSV-2 isolates

We next focused our attention on differences below the consensus level in each intra-host viral population. The amino acid (AA) variations described above exist in the consensus genomes of each isolate. Since viral replication creates a population of genomes, we next assessed whether minor allelic variants existed within the viral population of any neonatal HSV-2 isolate, thereby expanding the viral genetic diversity within each host. The significant depth of coverage from deep-sequencing of each isolate allowed us to search for minority variants at every nucleotide position of each genome. We defined a minor variant as any nucleotide allele (single nucleotide polymorphism, or SNP) or insertion/deletion (INDEL) with frequency below 50%, but above 2% (our limit of detection; see Methods for additional criteria). We found minor variants in the viral genome population of all 10 neonatal HSV-2 isolates, albeit to a different degree in each isolate (**Figure 6A**). In total, there were 1,821 minor variants, distributed across all genomic regions (**Figure S4**). For both SNPs and INDELs, intergenic minority variants outnumbered those in genes (genic), likely reflecting the higher selective pressures against unfavorable mutations in coding regions. The neonatal isolate DISS29 had 8-10-fold higher levels of minority variants than other neonatal isolates (**Figures 6A and S4**), and these variants were often present at a higher frequency or penetrance of the minor allele than observed in other neonatal isolates (**Figures 6B and S4**). We further examined the distribution of minor variants that occurred in genes, and found that nearly every HSV-2 protein harbored minority variants in at least one neonatal isolate (**Figure 6C**). Only UL3, UL11, UL35, and UL55 were completely devoid of minority variants. Three of these genes (UL3, UL11, UL35) were also devoid of AA variations at

the consensus level (**Figure 5**). These data revealed the breadth of potential contributions of minority variants to neonatal HSV-2 biology, which could undergo selection over time or in specific niches.

Coding variations identified between neonatal HSV-2 isolates are associated with CNS disease

To understand how viral genetic variants might relate to clinical disease, we assessed whether any of the identified coding variations segregated with clinical disease features. We found that a number of AA variations were found only in viral isolates obtained from the cerebrospinal fluid (CSF) of neonates with CNS disease (CNS11 and DISS14), but none of those collected from the skin. These included AA variants at the level of the consensus genome in the HSV-2 genes encoding glycoprotein K (gK, UL53 gene: V323M), glycoprotein I (gI, US7 gene: R159L and P215S), UL8 (R221S), US2 (F137L), and glycoprotein G (gG, US4 gene: R338L, S442P, and E574D) (**Figure 7**). The gI variants exist individually in other viral isolates obtained from the skin of neonates with CNS disease, but are only found together in those HSV-2 isolates obtained from the CSF (**Table S3**). Isolates collected from infants with CNS disease and disseminated involvement also shared a variant in the HSV-2 UL20 gene (P129L) (**Figure 7**). One variant in the HSV-2 UL24 gene (V93A) was shared only by the SEM isolates, with all CNS isolates containing a valine at this position **(Figure 7).** The sample size of these comparisons was constrained by the overall limits of neonatal HSV-2 availability. While CNS11 and DISS14 appeared together frequently in the variants listed in **Figure 7**, these viral genomes were not genetically similar overall at the consensus genome level (**Figure 4**), and these isolates harbored unique variants in several of the proteins that also had specific shared variants (e.g. gK, UL8,

gG). This indicated that their similarities in these residues were not due to a common viral strain origin. Many of the coding variations that correlate with neonatal CNS disease phenotypes impact viral proteins known to modulate cell-to-cell spread (71–75) and/or contribute to neurovirulence in mouse models of CNS infection (73, 76–82) (**Figure 7** and **Table S2**), however, their role in human disease has not yet been assessed.

Discussion

Host factors have not been identified to explain the >50% of neonates experiencing invasive CNS or disseminated forms of HSV infection. There is growing evidence that most herpesviruses contain significant genetic variation, including HSV-1 and HSV-2 (53, 27, 56, 68–70). The potential contributions of viral genetic variation to clinical disease in neonates therefore warrants exploration. Here, we analyzed genetic and phenotypic diversity for HSV-2 isolated from ten neonatal patients spanning two clinical studies (8, 61, 9). We found that neonatal HSV-2 isolates exhibited diverse growth characteristics in culture, with isolates obtained from neonates with CNS disease being associated with larger average plaque size and enhanced viral spread through culture. Using comprehensive comparative genomics, we further demonstrated that these neonatal HSV-2 isolates contained extensive genetic diversity both within and between hosts. These data revealed several specific viral genetic variations that were associated with cases of CNS disease, in proteins known to contribute to cell-to-cell spread and/or neurovirulence in mouse models of CNS disease. Further studies are required to determine the impact of these variations on HSV-2 neurovirulence and progression to CNS disease.

Genomic comparison of these neonatal isolates revealed a wide range of genetic diversity. At the

consensus genome level, which reflects the most common allele in each viral population, we found that coding differences between strains were as numerous as between previously described sets of adult HSV-2 isolates **(Figure 5)** (68–70). Furthermore, the specific genetic variations associated with neonatal CNS disease in our study can also be found in genital HSV-2 isolates which are not associated with CNS disease in adults. This suggests that the dramatic differences in clinical manifestations following HSV-2 infection in neonates, with significantly higher rates of invasive CNS infection, are not due to unique neonatal HSV-2 strains. It is more likely that genetic variations occurring naturally in the adult HSV-2 population, when present in the specific context of neonatal infection, may confer neurovirulence and progression to CNS disease.

At the level of minority variants (MV), which represent rare alleles that exist within each intra-host viral population, we found that the DISS29 harbored 8-10 fold and CNS15 harbored 3-4 fold more MV than other neonatal virus genomes (**Figures 6 and S4**). This could be indicative of a mixed viral population (e.g. a multi-strain infection) or of decreased polymerase fidelity (28, 27). Diversity in viral populations has also been observed in congenital HCMV infection (31, 32, 35, 34). However, this is the first time that evidence has been found for this level of viral population diversity with HSV-2. These minor genotypes may be selected or genetically isolated in particular niches (e.g. CSF), as observed in a comparison of VZV skin vesicles (26), or by antiviral drug selection, as was recently demonstrated in two adult genital HSV-2 study participants (83). All of the isolates sequenced in this study were collected at the initial time of diagnosis (≤19 days old), prior to acyclovir treatment and prior to the development of an HSV-specific immune response. It would therefore be compelling to examine the viral genome population from serial patient isolates over time to identify shifts in the frequency of MV due to antiviral or immune selection. Isolates from different body sites of the same patient could

likewise be compared to determine whether particular genotypes are enriched in different body niches.

Interestingly, isolates obtained directly from the CSF shared the most unique genetic variations, and exhibited significantly enhanced spread through cultured cells, as compared to isolates obtained from the skin. This may be an artifact of the small number of CSF samples in this group, however, the known relevance of the identified proteins to neurovirulence suggest these variations warrant further exploration. The accumulation of neurovirulence-associated variations in isolates obtained from the CSF, but not all isolates causing CNS disease (including isolates obtained from the skin), may indicate that virus able to enter and/or establish infection in the CNS is genetically distinct from virus contained in other neonatal body compartments. For example, both viral isolates obtained directly from the CSF contained the unique combination of two US7 variants (R159L and P215S) at the level of the consensus genome. CNS isolates obtained from the skin often contained only one of these gI variants, at the consensus or minority variant level (see **Figure 7, Table S2**). A comparison of skin vs. CSF isolates from the same infant would reveal whether the sequential acquisition of these genetic variants correlates with viral spread from the skin into the CSF.

This comparison of viral genotype to clinical phenotype revealed associations between neonatal CNS disease and several viral protein variants that may impact neurovirulence through modulation of cell-to-cell spread. Although the sample set in this proof of concept study is small, we observed potential patterns that warrant exploration in a larger dataset. It is important to acknowledge that there is limited availability of samples from neonatal infection, both due to the rarity of these infections, and the fragility and limited body size of the infected infants. These

natural circumstances lead to minimal sample collection from infected neonates. The finding that CNS-associated isolates exhibit enhanced spread between cells in culture, particularly those derived directly from the CSF, suggests that one or more of these variants could be functionally significant. Coding differences in viral proteins not known to contribute to neurovirulence were also found to be associated with neonatal CNS disease, and represent potential novel contributions to invasive infection. These promising results warrant exploration in a larger study, ideally with isolates from multiple time points and/or body sites from each infected infant. This would enable a better understanding of how overall viral genetic diversity contributes to neuroinvasion.

Methods

Viruses

Viruses were collected from neonates enrolled in clinical studies (8, 9, 61) by the Collaborative Antiviral Study Group (CASG) at the University of Alabama Birmingham (UAB). Samples were collected from either the cerebrospinal fluid (CSF) or skin. Enrollment in the original studies was evenly split between males and females, and included both black and white patients. Clinical morbidity score was determined by individual CASG enrollment sites at 12 months of life as previously defined (84, 12). Initial collection of samples, use of samples in this study, and use of de-identified clinical information was approved by the UAB Institutional Review Board.

Cell culture

Human lung fibroblast MRC-5 cells (ATCC®, CCL-171) were cultured in minimum essential

medium Eagle (MEME, Sigma-Aldrich; M5650) supplemented with 10% fetal bovine serum (FBS, Hyclone; SH30071.03), 2mM L-glutamine (Gibco, A2916801) and 1X penicillin-streptomycin (Gibco; 15140122). African green monkey kidney Vero cells (ATCC®, CCL-81) and human epithelial bone osteosarcoma U2OS cells (ATCC®, HTB-96) were cultured in Dulbecco's modified Eagle's medium with high glucose (DMEM, Hyclone; SH30081.02) supplemented with 10% FBS, 2mM L-glutamine, and 1X penicillin-streptomycin. Cell lines were authenticated by ATCC prior to purchase, and were confirmed to be mycoplasma free throughout experiments by periodic testing (LookOut Mycoplasma, Sigma).

Virus culture

Viruses were cultured at the time of diagnosis and snap frozen at UAB, before shipment to the Children's Hospital of Philadelphia (CHOP). Each viral isolate was then passaged 3 times on MRC-5 cells at a MOI of 0.01, with harvest at the time of complete cytopathic effect (between 50-70 hours). Viral stocks were titered on either Vero cells (100 hours) or U2OS cells (48 hours) under a methylcellulose overlay. Plaque size and morphology were monitored carefully and did not change for any viral isolates over the course of virus stock expansion.

Plaque measurements

Plaques were stained with 0.5% methylene blue. Serial 4X brightfield images were collected on an EVOS FL Auto Imaging System and stitched by EVOS software (University of Pennsylvania Cell and Developmental Biology Microscopy Core). No processing was performed. Plaque area was measured using ImageJ software.

Genome copy number estimation by quantitative PCR for UL27

DNA was extracted using a PureLink genomic DNA mini kit (ThermoFisher Scientific). Real-time PCR was performed with primers specific to the viral glycoprotein B gene (gB; UL27) (43), on a ViiA 7 Real-Time PCR System. Absolute quantification was calculated based on a standard curve of HSV-1 strain F nucleocapsid DNA (59).

Viral entry assay

Monolayers of Vero cells were cooled to 4°C for 30 minutes prior to infection with 100 PFU of each viral isolate. Cells remained at 4°C for 1 hour to allow virus binding, after which unbound virus was removed by washing. Cells were then moved to 37°C to allow virus entry. At 0, 10, 20, 30, 45, or 60 minutes, a low-pH citrate buffer was applied to infected cells to inactivate virus that had not penetrated the cellular membrane. For each condition, parallel infections were performed without the addition of citrate solution (control). Cell monolayers were washed and allowed to plaque under methylcellulose for 100 hours. Viral entry was quantified as the fraction of plaques formed following citrate buffer application, where 100% is the number of plaques formed on a monolayer not treated with citrate buffer (control).

Single-step and multi-step growth curves

Growth curves of HSV infection were performed in Vero cells. At 2 hpi, 0.1% human serum was added to reduce cell-free spread of virus. Single-step growth curves were performed at MOI=5, and multi-step growth curves at MOI=0.001, as defined by titering viral stocks on U2OS cells. Every 24 hours, the supernatant was removed and media containing 0.1% human serum reapplied.

Immunocytochemistry

Immunocytochemistry was performed as previously described (85). Primary antibodies used were raised against total HSV (Agilent Dako, B0114), and binding was visualized with fluorophore-conjugated anti-Rabbit secondary antibodies (Invitrogen, A-11008). Cells were counterstained with DAPI to mark cell nuclei. 5X images were collected with a Leica DM6000 wide field microscope equipped with a Photometrics HQ2 high resolution monochrome CCD camera, and processed with LAS AF software (UPenn Cell and Developmental Microscopy Core). 10X images were collected on a EVOS FL Auto Imaging System and stitched using EVOS software (UPenn Cell and Developmental Microscopy Core). Exposure and gain were optimized within each experiment for one virus at the 72-hour time point and applied identically to each image within that experiment. Subsequent image processing (ImageJ) was applied equally to all images in a given experiment.

Immunoblotting

Immunoblotting was performed as previously described (85). Equal amounts of whole cell lysate were separated by SDS-PAGE. Membranes were immunoblotted with antibodies raised against total HSV (Agilent Dako, B0114); glycoproteins (g)C, gD, gE, gH, and VP5 (all gifts from Gary Cohen); ICP8 (gift from David Knipe); and GAPDH (GeneTex, GTX100118).

Viral DNA isolation and Illumina sequencing

Viral nucleocapsid DNA for genome sequencing was prepared by infecting MRC-5 cells at an MOI ≥5 as previously described (86, 87). Viral nucleocapsid gDNA was sheared using a Covaris M220 sonicator/disruptor under the following conditions: 60s duration, peak power 50, 10% duty

cycle, at 4°C. Barcoded sequencing libraries were prepared using the Illumina TruSeq low-throughput protocol according to manufacturer's specifications and as previously described (60, 88). The quality of sequencing libraries was evaluated by Qubit (Invitrogen, CA), Bioanalyzer (Agilent), and qPCR (KAPA Biosystems). Paired-end sequencing (2 × 300bp length) was performed on an Illumina MiSeq, according to manufacturer's recommendations (17pM input).

De novo genome assembly

A consensus genome was assembled for each viral isolate using a previously described Viral Genome Assembly (VirGA) bioinformatics workflow (60). Annotation of new genome sequences was guided by the HSV-2 reference genome (strain HG52; GenBank NC_001798) based on sequence homology (89).

Comparative genomics and phylogenetic analysis

The genomes of all 10 neonatal HSV-2 isolates were combined with all annotated HSV-2 genomes available in GenBank (see **Table S1** for full list; all derived from adults) and aligned using MAFFT (90). The genome-wide alignment used a trimmed genome format (lacking the terminal repeats) to avoid giving undue weight to these duplicated sequences. The MAFFT alignment was used to generate a NeighborNet phylogenetic network in SplitsTree with Uncorrected P distances (49, 91, 92). A diverse subset of ten adult HSV-2 isolates was selected for protein-level comparisons with the ten neonatal isolates (indicated in **Table S1** with an asterisk). ClustalW2 was used to construct pairwise nucleotide alignments between whole genomes and pairwise amino acid alignments for each gene and protein (93). Custom Python scripts were used on these alignments to identify nucleotide and AA differences between

samples.

Minority variant detection & quantification

Minority variants (MV) were detected using VarScan v2.2.11 (mpileup2snp and mpileup2indel commands) (94), using the following parameters to differentiate true MV from technical artifacts (26): minimum allele frequency ≥ 0.02 (2%); base call quality ≥ 20; read depth ≥ 100; independent reads supporting minor allele ≥5. MV with directional strand bias ≥ 90% were excluded. The genomic location and potential impact of each MV was assessed using SnpEff and SnpSift (95, 96).

Acknowledgements

We thank Penny Jester and the UAB clinical support team for their assistance in retrieving archival records. The UPenn Cell and Developmental Biology Microscopy Core provided imaging assistance, and Elise M. Peaurori provided technical assistance during early stages of this project. We thank members of the Szpara and Weitzman labs for helpful discussions. Research in the Szpara lab was supported by the startup funds from the Pennsylvania State University (MLS) and a grant with the Pennsylvania Department of Health using Tobacco CURE funds (MLS). Research in the Weitzman lab was supported in part by a grant from the National Institutes of Health (NS082240 to MDW). LNA was supported the Pediatric Scientist Development Program through grant K12-HD000850 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The published NIAID CASG clinical studies of neonatal HSV infection were funded by a contract (N01-AI-62554) with the NIAID Development and Applications Branch and by grants from the General Clinical Research Center

Program (RR-032) and the state of Alabama.

Author Contributions

We describe contributions to the paper using the CRediT taxonomy (97). Writing – Original Draft: LNA, CB, AND, and MLS; Writing – Review & Editing: LNA, MDW, and MLS; Conceptualization: LNA and MLS; Investigation: LNA, CB, AND, DWR, UP; Resources: DWK, MNP, and RJW; Funding Acquisition: MLS, RJW, and MDW.

Declaration of Interests

The authors declare no competing interests.

Tables

Table 1: Clinical characteristics associated with HSV-2 isolates from ten patients.

Clinical Isolate*	Clinical Disease at Diagnosis	Sample Source	Morbidity Score - Mental	Morbidity Score - Motor	Patient Age at Disease Onset (days)	Gestational Age at Time of Birth (weeks)	Patient Sex & Race
CNS11	CNS	CSF	4	4	12	37	M, W
DISS14	DISS + CNS	CSF	2	2	7	39	M, W
CNS03	CNS	SKIN	4	4	17	37	M, W
CNS15	CNS	SKIN	3	4	19	36	M, W
CNS17	CNS	SKIN	4	4	17	40	M, W
DISS29	DISS + CNS	SKIN	3	3	5	38	F, B
CNS12	CNS	SKIN	4	4	16	41	F, W
SEM02	SEM	SKIN	1	1	5	38	F, W
SEM13	SEM	SKIN	4	4	11	27	F, B
SEM18	SEM	SKIN	2	2	17	37	F, W

* Clinical isolate order based on **Figure 1**.

Table 2. Genome sequencing statistics for neonatal HSV-2 strains.

Clinical Isolate*	Average Coverage	Raw Sequence Reads	Used for Assembly**	% viral	Depth > 100	GenBank Accession
CNS11	6,081X	3.5 million	2.8 million	79%	96%	In process^
DISS14	6,267X	3.8 million	3.0 million	77%	97%	In process^
CNS03	6,269X	3.9 million	3.1 million	78%	96%	In process^
CNS15	7,486X	6.1 million	4.4 million	73%	97%	In process^
CNS17	5,059X	3.1 million	2.3 million	73%	96%	In process^
DISS29	2,588X	1.4 million	1.1 million	78%	99%	In process^
CNS12	5,839X	3.9 million	2.8 million	73%	96%	In process^
SEM02	6,455X	4.8 million	3.7 million	77%	95%	In process^
SEM13	5,291X	3.4 million	2.4 million	72%	89%	In process^
SEM18	7,514X	16.9 million	12.2 million	72%	98%	In process^

* Clinical isolate order based on **Figure 1**.
** Numbers reflect the count of sequence read 1 of paired-end reads.
^ GenBank accessions are in process. Please see uploaded “Material for Review” for concatenated file of all 10 GenBank submissions in flatfile format.

Figures

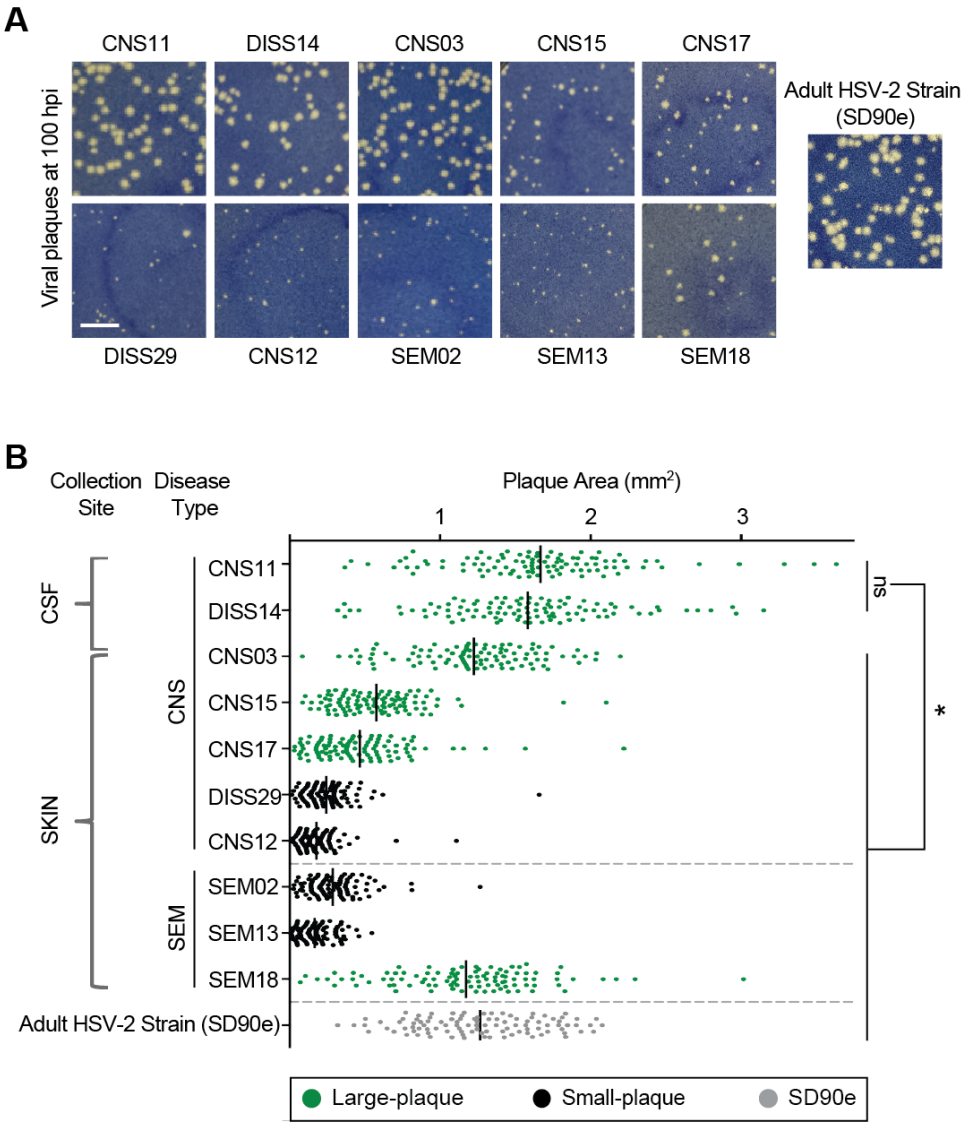

Figure 1. Neonatal HSV-2 isolates generate different sized plaques in culture. (**A**) Representative plaques are shown after neonatal viruses were allowed to incubate for 100 hours on Vero cells. Plaques formed by the low-passage adult HSV-2 strain SD90e are shown for comparison. Cell monolayers are stained with methylene blue. Scale bar = 5mm. (**B**) Quantification of plaque area on Vero cells for each isolate. Dots represent 100 individually measured plaques and black bar represents the mean. Largest average plaque areas are indicated in green, and smallest plaque areas indicated in black. SD90e is indicated in gray. Clinical disease is indicated as CNS (includes CNS only and DISS + CNS disease) or SEM, and site of virus collection indicated as CSF or SKIN. Each green isolate is statistically larger than each black isolate. Black isolates are not statistically different from one another. Additionally, each CSF-derived isolate is statistically larger than all other isolates shown. For all statistics, $p<0.05$ by one-way ANOVA followed by Holm-Sidak's multiple comparisons test.

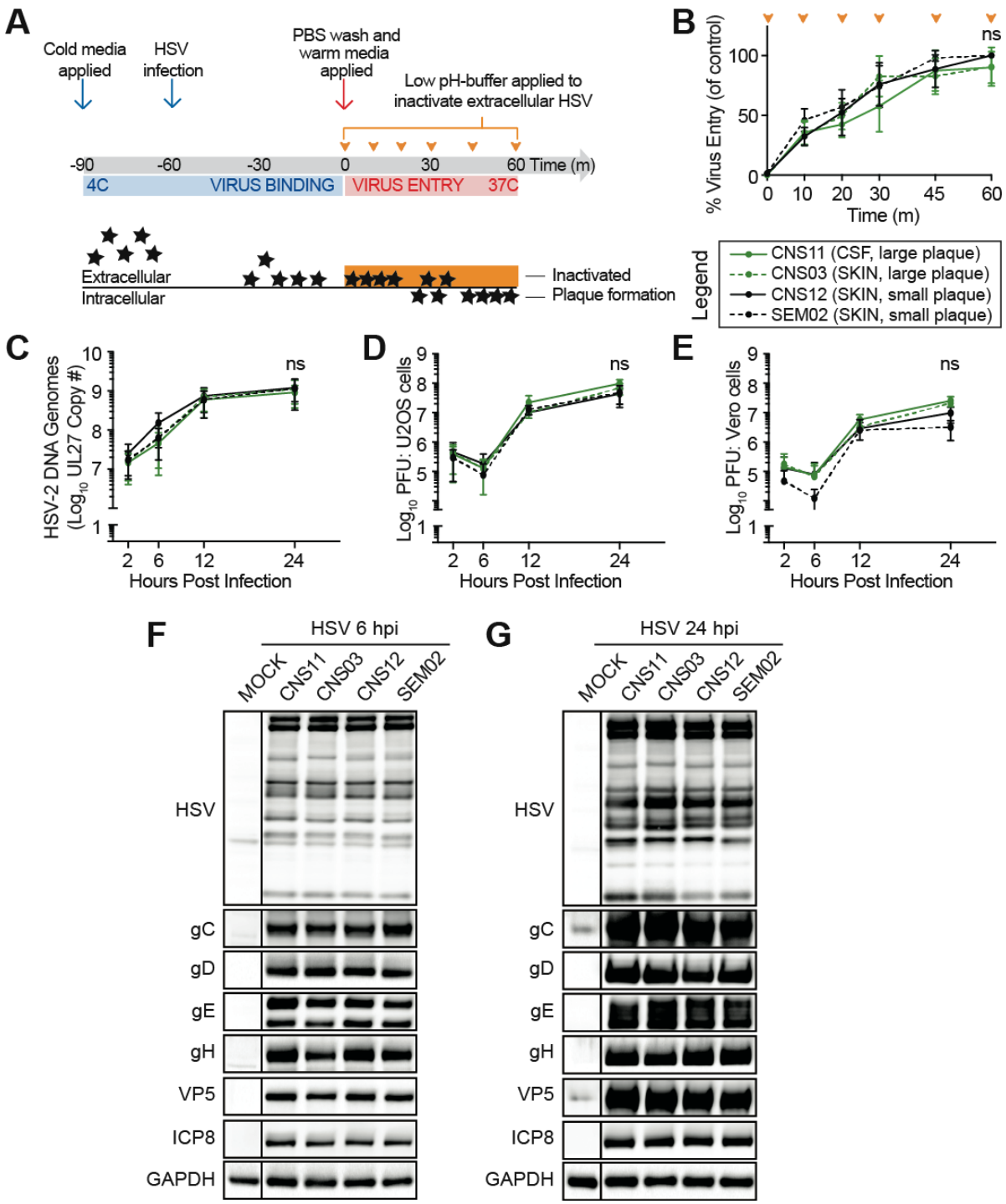

Figure 2. Increased plaque size in culture is not determined by viral entry, DNA replication, protein expression, or infectious virus production.

Viral growth characteristics were compared for several representative neonatal HSV-2 isolates, including large-plaque-formers (green) and small-plaque-formers (black). **(A-B)** Viral entry kinetics. The rate of virus penetration into Vero cells was compared. **(A)** Cells were cooled to

4°C for 30 minutes prior to infection. Each viral isolate was applied to cell monolayers at 4°C for 1 hour to allow virus binding, then moved to 37°C to allow virus entry. Extracellular virus was inactivated by the application of a low-pH citrate buffer at the times indicated (orange arrowheads). Cell monolayers were washed and overlaid with methylcellulose. Plaques were scored after 100 hours of incubation. **(B)** Viral entry is quantified as the fraction of plaques formed following citrate buffer application, where 100% is the number of plaques formed on a monolayer not treated with citrate buffer (control). These data represent three independent experiments. Two-way ANOVA followed by Tukey's multiple comparison test was applied. **(C-E)** Single-cycle viral replication kinetics. Confluent Vero cell monolayers were infected at MOI=5 and incubated in the presence of 0.1% human serum for the time points indicated. The cell monolayer was harvested at each time point. **(C)** The rate of viral DNA replication was compared. The quantity of viral genomes present was evaluated by qPCR for UL27. Infectious virion production (titer) was evaluated by plaque formation on U2OS **(D)** or Vero **(E)** cells. These data represent three independent experiments. Two-way ANOVA followed by Tukey's multiple comparison test was applied. **(F-G)** Protein production. Confluent Vero cell monolayers were infected at MOI=5 for 6 hours **(F)** or 24 **(G)** hours. Whole cell lysates were subjected to immunoblot analysis with the following antibodies: gC/gD/gE/gH, four virion glycoproteins; VP5 (UL19), capsid protein; ICP8 (UL29), viral single-strand DNA-binding protein; HSV, viral antibody against whole HSV-1; and GAPDH, cellular glyceraldehyde-3 phosphate dehydrogenase as a loading control.

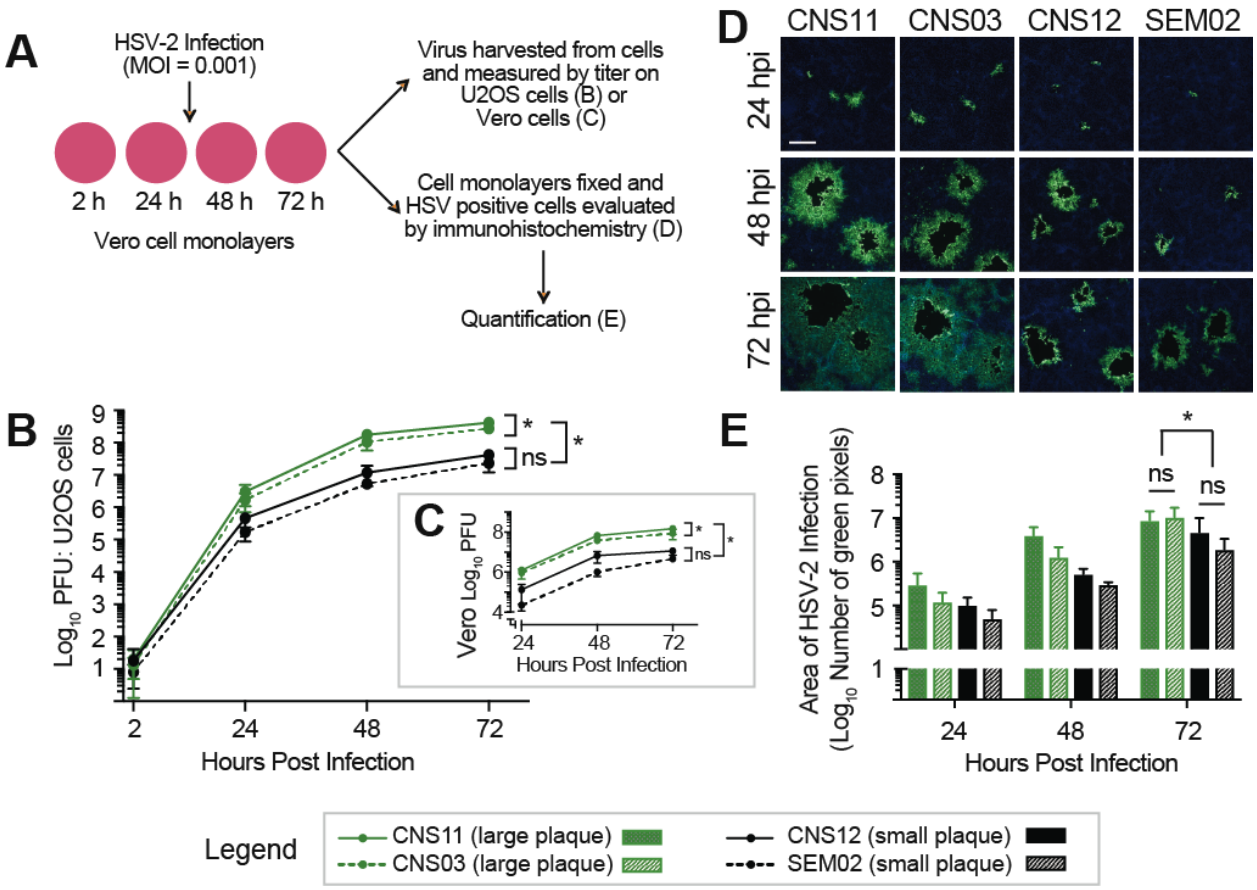

Figure 3. Enhanced viral cell-to-cell spread contributes to increased plaque size in culture.

(**A**) The rate of viral spread in Vero cells was compared for several representative neonatal isolates, including large-plaque-formers (green) and small-plaque-formers (black). Confluent Vero cell monolayers were infected at MOI=0.001 for the time points indicated, in the presence of 0.1% human serum, which was replenished every 24 hours. (**B-C**) Samples were harvested at each time point and viral titer was evaluated by plaque formation on U2OS cells **(B)** or Vero cells **(C).** These data represent three independent experiments. Two-way ANOVA followed by Tukey's multiple comparison test, *p<0.0001 at 72 hours. **(D)** In parallel experiments, the Vero cell monolayer was fixed with 4% PFA, and HSV positive cells (green) were evaluated at each time point by immunofluorescence. Cell nuclei are counterstained with DAPI (blue). Scale bar = 200 µm. Exposure, gain, and other image capture settings were optimized for the 72 hour time point of CNS11 and then applied uniformly for each image. Images are representative of three independent experiments. Images of the entire 10 mm coverslips were then captured and stitched to create a composite image (see **Figure S2**). (**E**) The total number of immunofluorescent (green) pixels was quantified for each coverslip. Two-way ANOVA followed by Tukey's multiple comparison test, *p<0.05 at 72 hours.

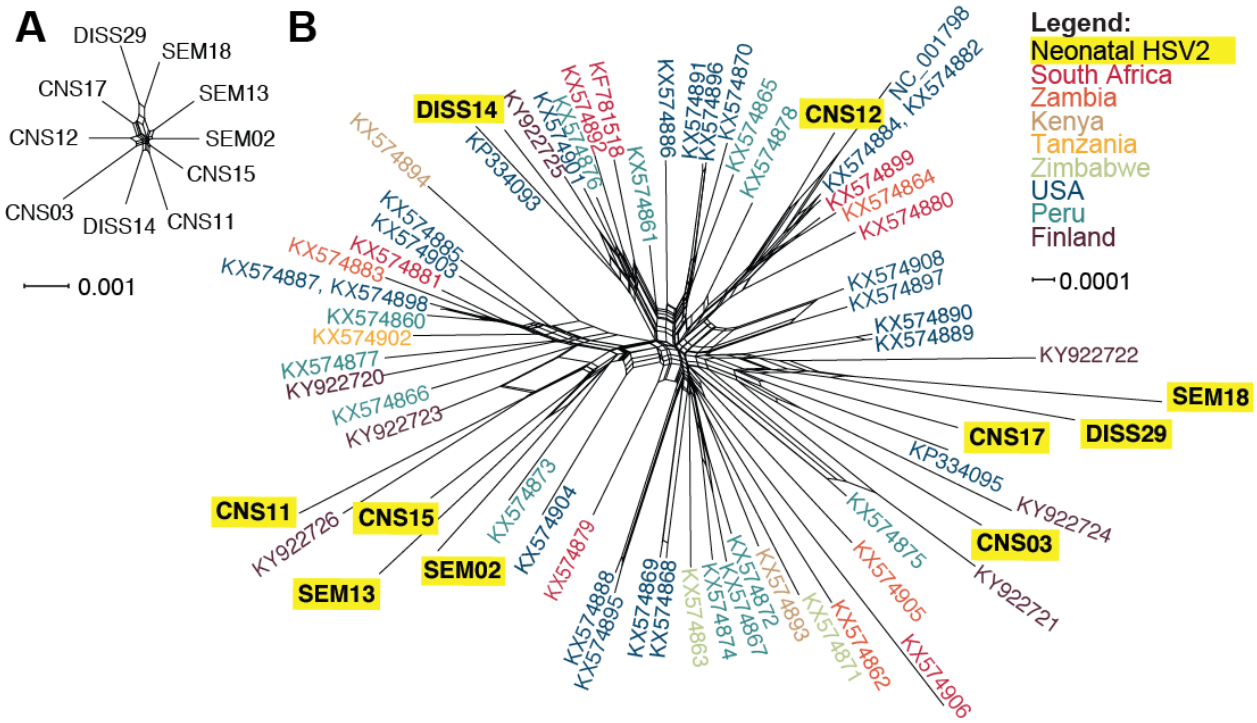

Figure 4. Neonatal HSV-2 genomes are genetically distinct from one another and encompass a broad range of known HSV-2 genetic diversity

A phylogenetic network constructed among neonatal HSV-2 genomes (**A**), or neonatal and adult HSV-2 genomes (**B**), reveals a lack of geographic clustering and the wide genetic distribution of these unrelated isolates. Unlike HSV-1, HSV-2 genomes have been previously noted to lack geographic separation into clades. The network graph was created using SplitsTree4, from a MAFFT trimmed genome alignment. See **Table S1** for complete list of accessions and geographic origins of adult HSV2 strains.

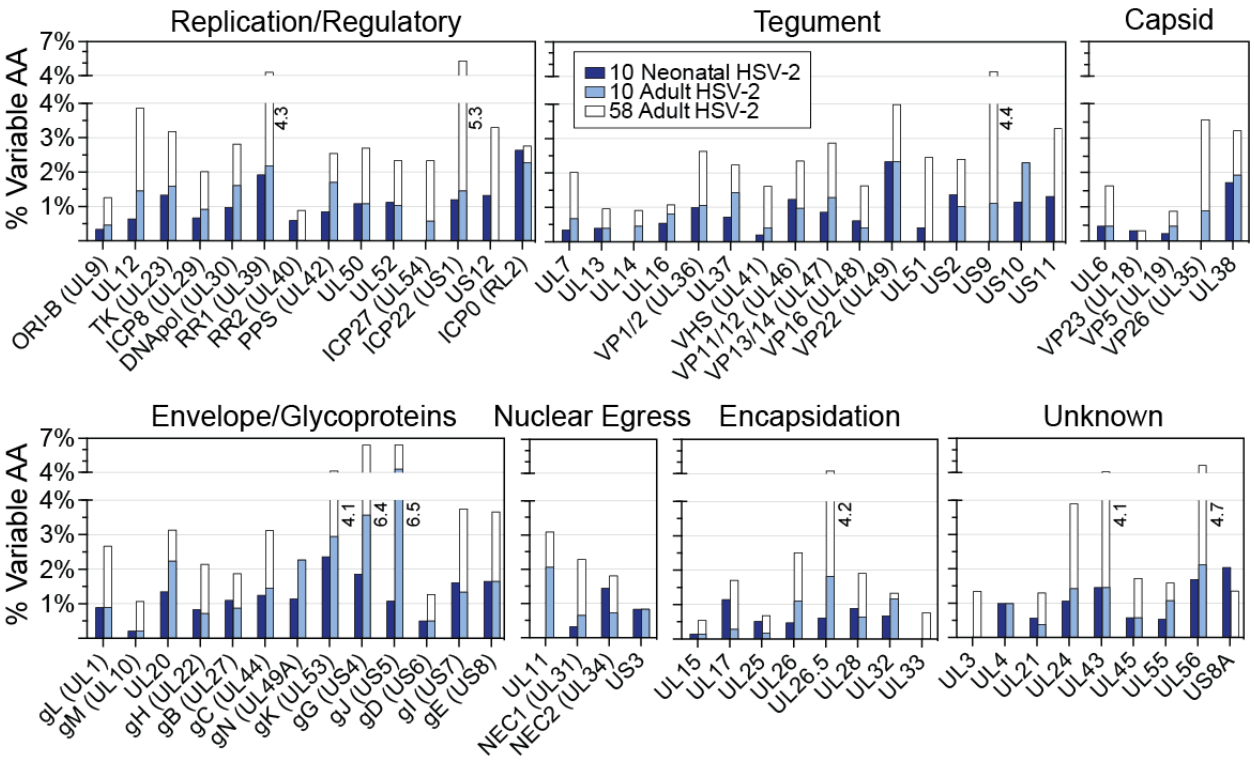

Figure 5: Neonatal HSV-2 proteins contain a similar degree of consensus-level amino acid (AA) differences as adult HSV-2 isolates.

HSV-2 proteins are grouped by function, and the percentage of variable AAs in each protein (#AA differences divided by protein length) are plotted for the 10 neonatal isolates (dark blue), for 10 representative adult HSV-2 isolates (light blue), and for all 58 annotated adult HSV-2 genomes in GenBank (clear outlines behind light blue boxes). All adult HSV-2 genomes used for this comparison are listed in **Table S1**. The ratio of synonymous nucleotide differences in each open reading frame (ORF) was also compared to the number of nonsynonymous coding differences; these data are summarized in **Table S2** and **Figure S3**. Proteins such as ICP27 (UL54) and US9 display an absence of coding variation in neonatal HSV-2 genomes (see also **Figure S3**).

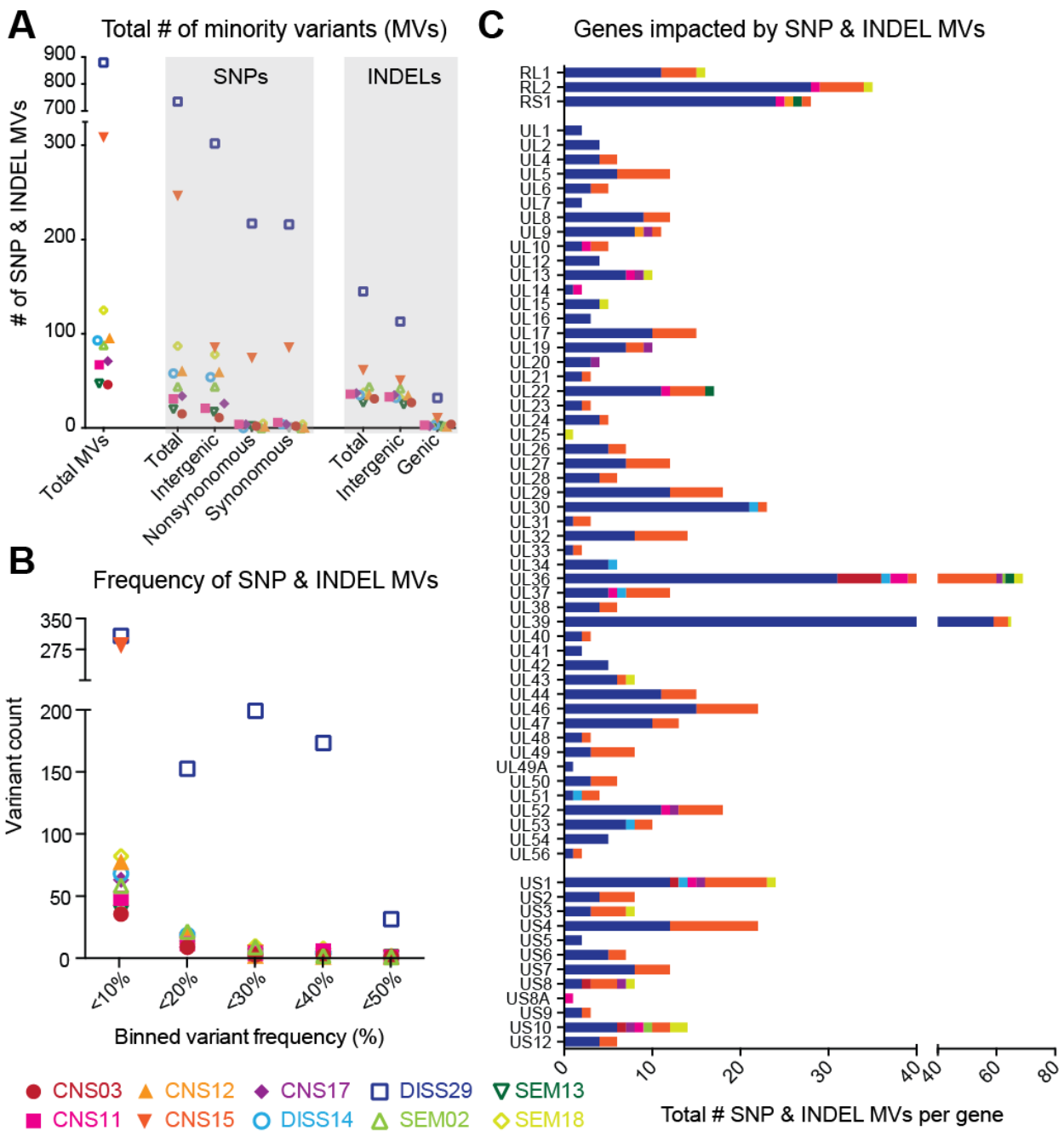

Figure 6: Minority variants expand the range of neonatal HSV-2 coding diversity.

Minority variants (MV) are rare alleles that exist within each viral population, which are not recorded as part of the consensus genome because they exist at <50% frequency in each viral population. (**A**) This scatter plot indicates the total number of minor variants (MV; y-axis) observed in each neonatal isolate. The total number of MV on the left is separated into single-nucleotide polymorphism (SNP) vs. insertion/deletion (INDEL) variants on the right (x-axis). The genomic location of each SNP or INDEL variant is also summarized: intergenic vs. inside of

genes (genic) for INDELS, and intergenic vs. nonsynonymous or synonymous SNPs inside of genes. Neonatal isolate DISS29 had an 8-10 fold higher number of MV than most other neonatal isolates (>2-fold higher than CNS15). (**B**) The frequency, or penetrance, of each minority variant was examined for each isolate. Data (x-axis) was binned in increments of 5% (e.g. 0-<5% frequency, 5-<10% frequency, and so on), and plotted according to the number of MV observed at each frequency (y-axis). SNP and INDEL variants were combined for this analysis. The large overall number of MV in neonatal isolate DISS29 also correlates with these MV existing at a higher frequency than observed in other isolates, raising the possiblity of a mixed viral population or reduced polymerase fidelity. (**C**) Stacked histograms show the number of genic MV (x-axis) located in each HSV-2 coding sequence (gene; y-axis). SNP and INDEL variants were combined for this analysis. UL3, UL11, UL35, and UL55 lacked any minority variants and are not included in the histogram.

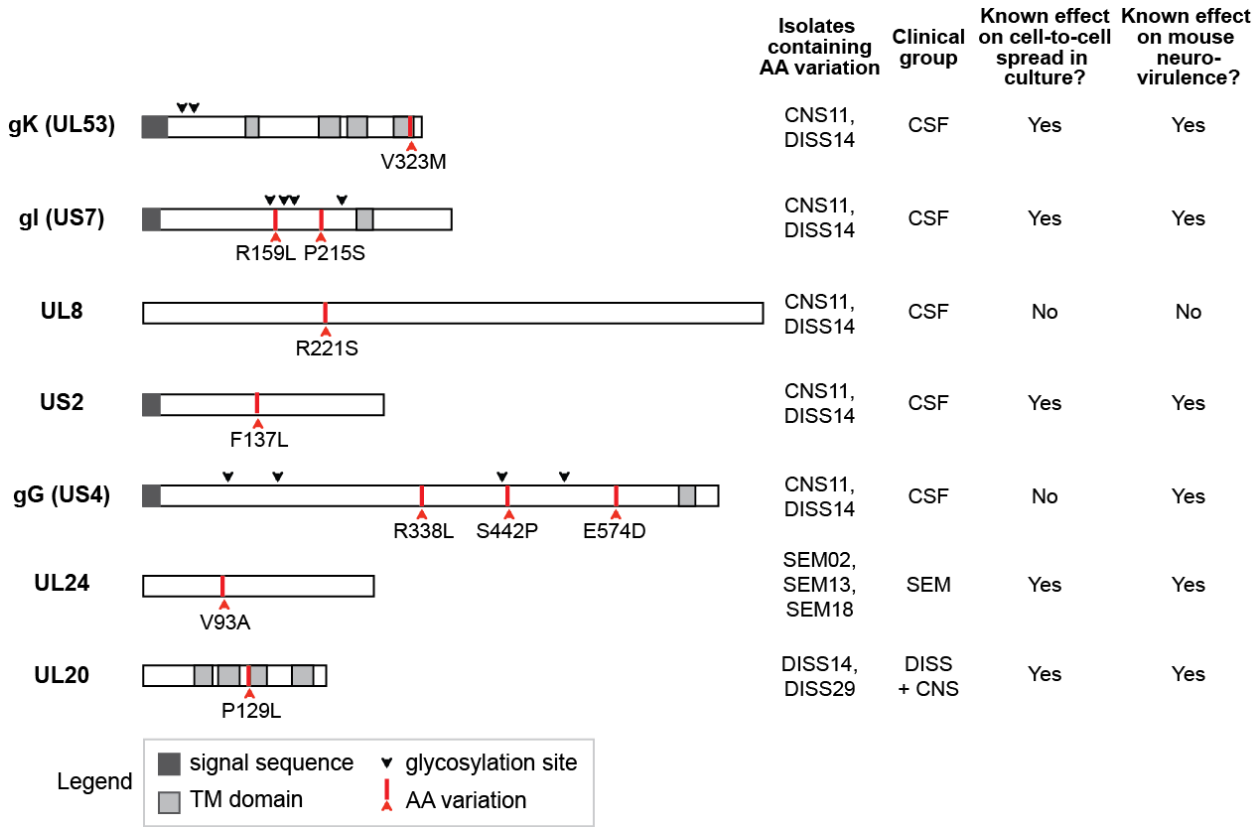

Figure 7. Genetic variations associated with neonatal CNS disease identified in proteins known to contribute to cell-to-cell spread and neurovirulence. The domain structure of each HSV protein is based on published literature for both HSV-1 and -2. More detailed information regarding cell-to-cell spread and neurovirulence can be found in **Table S3**.

References

1. Looker KJ, Magaret AS, May MT, Turner KM, Vickerman P, Newman LM, Gottlieb SL. 2017. First estimates of the global and regional incidence of neonatal herpes infection. Lancet Glob Health 5:e300–e309.

2. Stephenson-Famy A, Gardella C. 2014. Herpes Simplex Virus Infection During Pregnancy. Obstet Gynecol Clin North Am 41:601–614.

3. Brown ZA, Wald A, Morrow RA, Selke S, Zeh J, Corey L. 2003. Effect of serologic status and cesarean delivery on transmission rates of herpes simplex virus from mother to infant. Jama 289:203–209.

4. Corey L, Wald A. 2009. Maternal and Neonatal HSV Infections. N Engl J Med 361:1376–1385.

5. Bradley H, Markowitz LE, Gibson T, McQuillan GM. 2014. Seroprevalence of herpes simplex virus types 1 and 2--United States, 1999-2010. J Infect Dis 209:325–333.

6. James SH, Kimberlin DW. 2014. Neonatal Herpes Simplex Virus Infection. Clin Perinatol.

7. Pinninti SG, Kimberlin DW. 2014. Preventing Herpes Simplex Virus in the Newborn. Clin Perinatol 41:945–955.

8. Kimberlin DW, Lin CY, Jacobs RF, Powell DA, Frenkel LM, Gruber WC, Rathore M, Bradley JS, Diaz PS, Kumar M, Arvin AM, Gutierrez K, Shelton M, Weiner LB, Sleasman JW, de Sierra TM, Soong SJ, Kiell J, Lakeman FD, Whitley RJ, National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group. 2001. Natural history of neonatal herpes simplex virus infections in the acyclovir era. Pediatrics 108:223–229.

9. Kimberlin DW, Whitley RJ, Wan W, Powell DA, Storch G, Ahmed A, Palmer A, Sánchez PJ, Jacobs RF, Bradley JS. 2011. Oral acyclovir suppression and neurodevelopment after neonatal herpes. N Engl J Med 365:1284–1292.

10. Casanova J-L. 2015. Human genetic basis of interindividual variability in the course of infection. Proc Natl Acad Sci 201521644.

11. Zhang S-Y, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, Segal D, Sancho-Shimizu V, Lorenzo L, Puel A, Picard C, Chapgier A, Plancoulaine S, Titeux M, Cognet C, von Bernuth H, Ku C-L, Casrouge A, Zhang X-X, Barreiro L, Leonard J, Hamilton C, Lebon P, Heron B, Vallee L, Quintana-Murci L, Hovnanian A, Rozenberg F, Vivier E, Geissmann F, Tardieu M, Abel L, Casanova J-L. 2007. TLR3 deficiency in patients with herpes simplex encephalitis. Science 317:1522–1527.

12. Whitley R, Arvin A, Prober C, Corey L, Burchett S, Plotkin S, Starr S, Jacobs R, Powell D, Nahmias A, Sumaya C, Edwards K, Alford C, Caddell G, Soong S-J, the National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group*. 1991. Predictors of Morbidity and Mortality in Neonates with Herpes Simplex Virus Infections. N Engl J

Med 324:450–454.

13. Fields BN, Byers K. 1983. The Genetic Basis of Viral Virulence. Philos Trans R Soc Lond B Biol Sci 303:209–218.

14. Gonzalez-Scarano F, Beaty B, Sundin D, Janssen R, Endres MJ, Nathanson N. 1988. Genetic determinants of the virulence and infectivity of La Crosse virus. Microb Pathog 4:1–7.

15. Tscherne DM, García-Sastre A. 2011. Virulence determinants of pandemic influenza viruses. J Clin Invest 121:6–13.

16. Müller V, Fraser C, Herbeck JT. 2011. A Strong Case for Viral Genetic Factors in HIV Virulence. Viruses 3:204–216.

17. Robinson CM, Jesudhasan PR, Pfeiffer JK. 2014. Bacterial Lipopolysaccharide Binding Enhances Virion Stability and Promotes Environmental Fitness of an Enteric Virus. Cell Host Microbe 15:36–46.

18. Yoshii K, Sunden Y, Yokozawa K, Igarashi M, Kariwa H, Holbrook MR, Takashima I. 2014. A Critical Determinant of Neurological Disease Associated with Highly Pathogenic Tick-Borne Flavivirus in Mice. J Virol 88:5406–5420.

19. Roizman B, Whitley RJ. 2001. The nine ages of herpes simplex virus. Herpes 8:23–27.

20. Sakaoka H, Kawana T, Grillner L, Aomori T, Yamiguchi T, Saito H, Fujinaga K. 1987. Genome Variations in Herpes Simplex Virus Type 2 Strains Isolated in Japan and Sweden. J Gen Virol 68:2105–16.

21. Sakaoka H, Kurita K, Iida Y, Takada S, Umene K, Kim YT, Ren CS, Nahmias AJ. 1994. Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. J Gen Virol 75:513–27.

22. Bowden R, Sakaoka H, Ward R, Donnelly P. 2006. Patterns of Eurasian HSV-1 molecular diversity and inferences of human migrations. Infect Genet Evol 6:63–74.

23. Bower JR, Mao H, Durishin C, Rozenbom E, Detwiler M, Rempinski D, Karban TL, Rosenthal KS. 1999. Intrastrain variants of herpes simplex virus type 1 isolated from a neonate with fatal disseminated infection differ in the ICP34.5 gene, glycoprotein processing, and neuroinvasiveness. J Virol 73:3843–53.

24. Mao H, Rosenthal KS. 2003. Strain-Dependent Structural Variants of Herpes Simplex Virus Type 1 ICP34.5 Determine Viral Plaque Size, Efficiency of Glycoprotein Processing, and Viral Release and Neuroinvasive Disease Potential. J Virol 77:3409–3417.

25. Renzette N, Gibson L, Jensen JD, Kowalik TF. 2014. Human cytomegalovirus intrahost evolution—a new avenue for understanding and controlling herpesvirus infections. Curr Opin Virol 8:109–115.

26. Depledge DP, Kundu S, Jensen NJ, Gray ER, Jones M, Steinberg S, Gershon A, Kinchington PR, Schmid DS, Balloux F, Nichols RA, Breuer J. 2014. Deep Sequencing of Viral

Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans. Mol Biol Evol 31:397–409.

27. Renner DW, Szpara ML. 2018. The Impacts of Genome-wide Analyses on our Understanding of Human Herpesvirus Diversity and Evolution. J Virol 92:e00908-17.

28. Houldcroft CJ, Bryant JM, Depledge DP, Margetts BK, Simmonds J, Nicolaou S, Tutill HJ, Williams R, Worth AJJ, Marks SD, Veys P, Whittaker E, Breuer J. 2016. Detection of Low Frequency Multi-Drug Resistance and Novel Putative Maribavir Resistance in Immunocompromised Pediatric Patients with Cytomegalovirus. Front Microbiol 7.

29. Houldcroft CJ, Beale MA, Breuer J. 2017. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol 15:183–192.

30. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. 2011. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog 7:1–14.

31. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, Jensen JD, Kowalik TF. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. PLoS Genet 9:1–14.

32. Renzette N, Pokalyuk C, Gibson L, Bhattacharjee B, Schleiss MR, Hamprecht K, Yamamoto AY, Mussi-Pinhata MM, Britt WJ, Jensen JD, Kowalik TF. 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci 201501880.

33. Renzette N, Kowalik TF, Jensen JD. 2016. On the relative roles of background selection and genetic hitchhiking in shaping human cytomegalovirus genetic diversity. Mol Ecol 25:403–413.

34. Renzette N, Pfeifer SP, Matuszewski S, Kowalik TF, Jensen JD. 2017. On the Analysis of Intrahost and Interhost Viral Populations: Human Cytomegalovirus as a Case Study of Pitfalls and Expectations. J Virol 91:e01976-16.

35. Pokalyuk C, Renzette N, Irwin KK, Pfeifer SP, Gibson L, Britt WJ, Yamamoto AY, Mussi-Pinhata MM, Kowalik TF, Jensen JD. 2017. Characterizing human cytomegalovirus reinfection in congenitally infected infants: an evolutionary perspective. Mol Ecol.

36. Sijmons S, Van Ranst M, Maes P. 2014. Genomic and Functional Characteristics of Human Cytomegalovirus Revealed by Next-Generation Sequencing. Viruses 6:1049–1072.

37. Sijmons S, Thys K, Mbong Ngwese M, Van Damme E, Dvorak J, Van Loock M, Li G, Tachezy R, Busson L, Aerssens J, Van Ranst M, Maes P. 2015. High-Throughput Analysis of Human Cytomegalovirus Genome Diversity Highlights the Widespread Occurrence of Gene-Disrupting Mutations and Pervasive Recombination. J Virol 89:7673–7695.

38. Lassalle F, Depledge DP, Reeves MB, Brown AC, Christiansen MT, Tutill HJ, Williams RJ, Einer-Jensen K, Holdstock J, Atkinson C, Brown JR, van Loenen FB, Clark DA, Griffiths PD, Verjans GMGM, Schutten M, Milne RSB, Balloux F, Breuer J. 2016. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human

cytomegalovirus genomes. Virus Evol 2:vew017.

39. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Reeves D, Schiffer J, Jerome KR, Sather C, Diem K, Lingappa JR, Celum C, Koelle DM, Wald A. 2017. Dual-strain genital herpes simplex virus type 2 (HSV-2) infection in the US, Peru, and 8 countries in sub-Saharan Africa: A nested cross-sectional viral genotyping study. PLOS Med 14:e1002475.

40. Minaya MA, Jensen TL, Goll JB, Korom M, Datla SH, Belshe RB, Morrison LA. 2017. Molecular evolution of herpes simplex virus 2 complete genomes: comparison between primary and recurrent infections. J Virol 91:e00942–17.

41. 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.

42. 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.

43. Ryncarz AJ, Goddard J, Wald A, Huang M-L, Roizman B, Corey L. 1999. Development of a high-throughput quantitative assay for detecting herpes simplex virus DNA in clinical samples. J Clin Microbiol 37:1941–1947.

44. Wald A, Huang M-L, Carrell D, Selke S, Corey L. 2003. Polymerase chain reaction for detection of herpes simplex virus (HSV) DNA on mucosal surfaces: comparison with HSV isolation in cell culture. J Infect Dis 188:1345–1351.

45. Szpara ML, Tafuri YR, Parsons L, Shamim SR, Verstrepen KJ, Legendre M, Enquist LW. 2011. A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. PLoS Pathog 7:1–23.

46. Brown J. 2004. Effect of gene location on the evolutionary rate of amino acid substitutions in herpes simplex virus proteins. Virology 330:209–20.

47. Bowden R, Sakaoka H, Donnelly P, Ward R. 2004. High recombination rate in herpes simplex virus type 1 natural populations suggests significant co-infection. Infect Genet Evol 4:115–23.

48. Norberg P, Bergstrom T, Rekabdar E, Lindh M. 2004. Phylogenetic Analysis of Clinical Herpes Simplex Virus Type 1 Isolates Identified Three Genetic Groups and Recombinant Viruses. J Virol 78:10755–10764.

49. Norberg P, Kasubi MJ, Haarr L, Bergstrom T, Liljeqvist J-A. 2007. Divergence and recombination of clinical herpes simplex virus type 2 isolates. J Virol 81:13158–13167.

50. Lonsdale DM, Brown SM, Lang J, Subak-Sharpe JH, Koprowski H, Warren KG. 1980. Variations in herpes simplex virus isolated from human ganglia and a study of clonal variation in HSV-1. Ann N Y Acad Sci 291–308.

51. Liljeqvist JA, Tunback P, Norberg P. 2009. Asymptomatically shed recombinant herpes simplex virus type 1 strains detected in saliva. J Gen Virol 90:559–566.

52. Haarr L, Nilsen A, Knappskog PM, Langeland N. 2014. Stability of glycoprotein gene sequences of herpes simplex virus type 2 from primary to recurrent human infection, and diversity of the sequences among patients attending an STD clinic. BMC Infect Dis 14:63.

53. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, Enquist LW, Legendre M, Davison AJ. 2014. Evolution and diversity in human herpes simplex virus genomes. J Virol 88:1209–27.

54. 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:e02213-14.

55. Kolb AW, Lee K, Larsen I, Craven M, Brandt CR. 2016. Quantitative Trait Locus Based Virulence Determinant Mapping of the HSV-1 Genome in Murine Ocular Infection: Genes Involved in Viral Regulatory and Innate Immune Networks Contribute to Virulence. PLOS Pathog 12:e1005499.

56. Pfaff F, Groth M, Sauerbrei A, Zell R. 2016. Genotyping of herpes simplex virus type 1 (HSV-1) by whole genome sequencing. J Gen Virol.

57. Bowen CD, Renner DW, Shreve JT, Tafuri Y, Payne KM, Dix RD, Kinchington PR, Gatherer D, Szpara ML. 2016. Viral forensic genomics reveals the relatedness of classic herpes simplex virus strains KOS, KOS63, and KOS79. Virology 492:179–186.

58. Pandey U, Renner DW, Thompson RL, Szpara ML, Sawtell NM. 2017. Inferred father-to-son transmission of herpes simplex virus results in near-perfect preservation of viral genome identity and in vivo phenotypes. Sci Rep 7:13666.

59. Szpara ML, Parsons L, Enquist LW. 2010. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. J Virol 84:5303–13.

60. 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:e02213-14.

61. Kimberlin DW, Lin CY, Jacobs RF, Powell DA, Corey L, Gruber WC, Rathore M, Bradley JS, Diaz PS, Kumar M, Arvin AM, Gutierrez K, Shelton M, Weiner LB, Sleasman JW, de Sierra TM, Weller S, Soong SJ, Kiell J, Lakeman FD, Whitley RJ, National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group. 2001. Safety and efficacy of high-dose intravenous acyclovir in the management of neonatal herpes simplex virus infections. Pediatrics 108:230–238.

62. Desmyter J, Melnick JL, Rawls WE. 1968. Defectiveness of interferon production and of rubella virus interference in a line of African green monkey kidney cells (Vero). J Virol 2:955–961.

63. Emeny JM, Morgan MJ. 1979. Regulation of the interferon system: evidence that Vero cells

have a genetic defect in interferon production. J Gen Virol 43:247–252.

64. 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.

65. Yao F, Schaffer PA. 1995. An activity specified by the osteosarcoma line U2OS can substitute functionally for ICP0, a major regulatory protein of herpes simplex virus type 1. J Virol 69:6249–6258.

66. Deschamps T, Kalamvoki M. 2017. Impaired STING Pathway in Human Osteosarcoma U2OS Cells Contributes to the Growth of ICP0-Null Mutant Herpes Simplex Virus. J Virol 91:e00006-17.

67. Norberg P. 2010. Divergence and genotyping of human alpha-herpesviruses: an overview. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 10:14–25.

68. 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 JVI.01303-15.

69. 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.

70. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Cheng A, Diem K, Fitzgibbon MP, Huang M, Selke S, Lingappa JR, Celum C, Jerome KR, Wald A, Koelle DM. 2017. Highly conserved intragenic HSV-2 sequences: Results from next-generation sequencing of HSV-2 U L and U S regions from genital swabs collected from 3 continents. Virology 510:90–98.

71. David AT, Saied A, Charles A, Subramanian R, Chouljenko VN, Kousoulas KG. 2012. A Herpes Simplex Virus 1 (McKrae) Mutant Lacking the Glycoprotein K Gene Is Unable To Infect via Neuronal Axons and Egress from Neuronal Cell Bodies. mBio 3.

72. Dingwell KS, Brunetti CR, Hendricks RL, Tang Q, Tang M, Rainbow AJ, Johnson DC. 1994. Herpes simplex virus glycoproteins E and I facilitate cell-to-cell spread in vivo and across junctions of cultured cells. J Virol 68:834–845.

73. Weber PC, Levine M, Glorioso JC. 1987. Rapid identification of nonessential genes of herpes simplex virus type 1 by Tn5 mutagenesis. Science 236:576–579.

74. Jacobson JG, Martin SL, Coen DM. 1989. A conserved open reading frame that overlaps the herpes simplex virus thymidine kinase gene is important for viral growth in cell culture. J Virol 63:1839–1843.

75. Foster TP, Melancon JM, Baines JD, Kousoulas KG. 2004. The herpes simplex virus type 1 UL20 protein modulates membrane fusion events during cytoplasmic virion morphogenesis and virus-induced cell fusion. J Virol 78:5347–5357.

76. David AT, Baghian A, Foster TP, Chouljenko VN, Kousoulas KG. 2008. The herpes simplex virus type 1 (HSV-1) glycoprotein K(gK) is essential for viral corneal spread and neuroinvasiveness. Curr Eye Res 33:455–467.

77. Dingwell KS, Johnson DC. 1998. The herpes simplex virus gE-gI complex facilitates cell-to-cell spread and binds to components of cell junctions. J Virol 72:8933–8942.

78. Meignier B, Longnecker R, Mavromara-Nazos P, Sears AE, Roizman B. 1988. Virulence of and establishment of latency by genetically engineered deletion mutants of herpes simplex virus 1. Virology 162:251–254.

79. Rochette P-A, Bourget A, Sanabria-Solano C, Lahmidi S, Lavallée GO, Pearson A. 2015. Mutation of UL24 impedes the dissemination of acute herpes simplex virus 1 infection from the cornea to neurons of trigeminal ganglia. J Gen Virol 96:2794–2805.

80. Jacobson JG, Chen SH, Cook WJ, Kramer MF, Coen DM. 1998. Importance of the herpes simplex virus UL24 gene for productive ganglionic infection in mice. Virology 242:161–169.

81. Leiva-Torres GA, Rochette P-A, Pearson A. 2010. Differential importance of highly conserved residues in UL24 for herpes simplex virus 1 replication in vivo and reactivation. J Gen Virol 91:1109–1116.

82. Liu J, Lewin AS, Tuli SS, Ghivizzani SC, Schultz GS, Bloom DC. 2008. Reduction in severity of a herpes simplex virus type 1 murine infection by treatment with a ribozyme targeting the UL20 gene RNA. J Virol 82:7467–7474.

83. Minaya MA, Jensen T, Goll J, Korom M, Datla SH, Belshe RB, Morrison LA. 2017. Molecular evolution of herpes simplex virus 2 complete genomes: Comparison between primary and recurrent infections. J Virol JVI.00942-17.

84. Whitley R, Arvin A, Prober C, Burchett S, Corey L, Powell D, Plotkin S, Starr S, Alford C, Connor J, Jacobs R, Nahmias A, Soong S-J, the National Institute of Allergy, Infectious Diseases Collaborative Antiviral Study Group*. 1991. A Controlled Trial Comparing Vidarabine with Acyclovir in Neonatal Herpes Simplex Virus Infection. N Engl J Med 324:444–449.

85. Avgousti DC, Della Fera AN, Otter CJ, Herrmann C, Pancholi NJ, Weitzman MD. 2017. Adenovirus Core Protein VII Downregulates the DNA Damage Response on the Host Genome. J Virol 91:e01089-17.

86. Szpara ML, Tafuri YR, Enquist LW. 2011. Preparation of viral DNA from nucleocapsids. J Vis Exp JoVE 2–7.

87. Szpara M. 2014. Isolation of herpes simplex virus nucleocapsid DNA, p. 31–41. *In* Diefenbach, RJ, Fraefel, C (eds.), Herpes Simplex Virus. Springer New York, New York, NY.

88. Bowen CD, Renner DW, Shreve JT, Tafuri Y, Payne KM, Dix RD, Kinchington PR, Gatherer D, Szpara ML. 2016. Viral forensic genomics reveals the relatedness of classic

herpes simplex virus strains KOS, KOS63, and KOS79. Virology 492:179–186.

89. Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ. 1998. The genome sequence of herpes simplex virus type 2. J Virol 72:2010–2021.

90. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066.

91. Strimmer K, Wiuf C, Moulton V. 2001. Recombination analysis using directed graphical models. Mol Biol Evol 18:97–99.

92. Norberg P, Tyler S, Severini A, Whitley R, Liljeqvist J-A, Bergstrom T. 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. PloS One 6:1–8.

93. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinforma Appl Note 23:2947–2948.

94. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

95. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6:80–92.

96. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 3.

97. Brand A, Allen L, Altman M, Hlava M, Scott J. 2015. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Publ 28:151–155.

Supplementary Information for

Genotypic and phenotypic diversity within the neonatal HSV-2 population

Lisa N. Akhtar, Christopher D. Bowen, Daniel W. Renner, Utsav Pandey[2], Ashley N. Della Fera, David W. Kimberlin, Mark N. Prichard, Richard J. Whitley, Matthew D. Weitzman, Moriah L. Szpara*

***Corresponding Authors**
Moriah L. Szpara
Email: moriah@psu.edu
Matthew D. Weitzman
Email: weitzmanm@email.chop.edu

This PDF file includes:

Supplementary text
Figs. S1 to S4
Tables S1 to S3
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Table S4 in separate Excel file

Supporting Information (SI)

Virus source

Viruses analyzed were collected from neonates that were culture- or PCR-confirmed to have HSV-2 infection, enrolled in three published clinical studies (1–3) by the Collaborative Antiviral Study Group (CASG) at the University of Alabama Birmingham (UAB). These HSV isolates originated from patients with SEM, CNS, or disseminated (DISS) disease with CNS involvement (**Table 1**). Samples were collected at the time of diagnosis, prior to initiation of acyclovir therapy, from either the cerebrospinal fluid (CSF) or skin. Viral isolates are associated with de-identified clinical information including age, sex, race, and clinical morbidity scores, as approved by the UAB Institutional Review Board (IRB). Enrollment in the original studies was evenly split between males and females, and included both black and white patients. Clinical morbidity score was determined by the CASG after 12 months of life as (1) normal; (2) mild impairment which includes ocular sequelae (keratoconjunctivitis), speech delay, or mild motor delay; (3) moderate impairment which includes hemiparesis, persistent seizure disorder, or developmental delay of less than or equal to 3 months adjusted developmental age; (4) severe impairment which includes microcephaly, spastic quadriplegia, chorioretinitis or blindness, or a serious developmental delay of least 3 months according to the Denver Developmental Assessment Scale (4, 5). These assessments were performed at each collaborating site. Initial collection of samples was approved by the UAB IRB. Use of these samples in this study was approved by the UAB IRB.

Cell culture

Human lung fibroblast MRC-5 cells (ATCC®, CCL-171) were cultured in minimum essential medium Eagle (MEME, Sigma-Aldrich; M5650) supplemented with 10% fetal bovine serum (FBS, Hyclone; SH30071.03), 2mM L-glutamine (Gibco, A2916801) and 1X penicillin-streptomycin (Gibco; 15140122). African green monkey kidney Vero cells (ATCC®, CCL-81) and human epithelial bone osteosarcoma U2OS cells (ATCC®, HTB-96) were cultured in Dulbecco's modified Eagle's medium with high glucose (DMEM, Hyclone; SH30081.02) supplemented with 10% FBS, 2mM L-glutamine, and 1X penicillin-streptomycin. Cell lines were authenticated by ATCC prior to purchase, and were confirmed to be mycoplasma free throughout experiments by periodic testing (LookOut Mycoplasma, Sigma).

Virus culture

Initial viral isolates (passage 1-2) were obtained from CASG samples stored at the UAB. Each isolate had been cultured at the time of diagnosis, aliquoted after one passage, and snap frozen. To conduct viral genome sequencing and phenotypic assays, each stock was expanded over three serial passages by infecting monolayers of human MRC-5 cells at a MOI of 0.01. Each infection was allowed to progress to complete cytopathic effect (CPE) before harvest (between 50-70 hours), with increasing cell volume at each passage to produce sufficient virions for subsequent phenotypic and genomic comparisons. Viral stocks were titered on monolayers of either Vero cells or U2OS cells, for 100 or 48 hours respectively, to allow plaques to develop. Plaque formation was facilitated by limiting viral diffusion with a methylcellulose overlay. Plaque size and morphology were monitored carefully and did not change for any viral isolates over the course of virus stock expansion.

Plaque measurements

After appropriate incubation plaques were stained with 0.5% methylene blue and allowed to dry. Serial 4X brightfield images were collected on an EVOS FL Auto Imaging System and stitched by EVOS software to create an image of the entire well (University of Pennsylvania Cell and Developmental Biology Microscopy Core). No processing was performed. The area of 100 plaques was measured for each viral isolate using ImageJ software.

Genome copy number estimation by quantitative PCR for UL27

DNA was extracted using a PureLink genomic DNA mini kit (ThermoFisher Scientific). Viral genome copy number was determined using an established assay based on real-time PCR using primers and dual-fluorescent probe specific to viral glycoprotein B gene (gB; UL27) . Samples were assayed alongside a standard curve of HSV-1 strain F nucleocapsid DNA (6), on a ViiA 7 Real-Time PCR System.

Viral entry assay

Monolayers of Vero cells were cooled to 4°C for 30 minutes prior to infection and 100 PFU of each viral isolate was applied to cell monolayers at 4°C for 1 hour to allow virus binding, after which unbound virus was washed from the cells. Cells were then moved to 37°C to allow virus entry. At 0, 10, 20, 30, 45, or 60 minutes, a low-pH citrate buffer was applied to infected cells to inactivate virus that had not penetrated the cellular membrane. At each time point and for each virus, parallel infections were performed without the addition of citrate solution. These served as controls to determine the maximum number of plaques formed. Cell monolayers were washed and overlaid with methylcellulose. Plaques were scored after 100 hours of incubation. Viral entry

was quantified as the fraction of plaques formed following citrate buffer application, where 100% is the number of plaques formed on a monolayer not treated with citrate buffer (control).

Single-step and multi-step growth curves

Virus diluted in basic growth media containing 2% FBS was applied to near confluent monolayers of Vero cells and allowed to adsorb for 1 hour. Virus was removed and cells were incubated at 37C for the duration of infection. At 2 hpi, media containing 0.1% human serum was added to reduce the contribution of cell-free spread of virus. Single-step growth curves were performed at MOI=5 as defined by titering viral stocks on U2OS cells, and monolayers harvested at 2, 6, 12, and 24 hpi. Multi-step growth curves were performed at MOI=0.001 as defined by titering viral stocks on U2OS cells, and monolayers were harvested at 2, 24, 48, and 72 hpi. Every 24 hours, the supernatant was removed and media containing 0.1% human serum reapplied. At the conclusion of each infection, cells were washed two times with PBS and collected by scraping into an equal volume of media.

Immunocytochemistry

Multi-step growth curves were performed as described above in Vero cell monolayers grown on glass coverslips. To terminate each infection, coverslips were washed with PBS and fixed in 4% paraformaldehyde for 15 minutes at room temperature. Cells were permeabilized with 0.5% Triton-X, blocked in 3% BSA, and incubated with polyclonal rabbit antibodies raised against total HSV (Agilent Dako, B0114). After washing, cells were incubated with fluorophore-conjugated anti-Rabbit secondary antibodies (Invitrogen, A-11008) to mark HSV infected cells, and counterstained with DAPI to mark cell nuclei. Coverslips were initially visualized with a

Leica DM6000 wide field microscope (UPenn Cell and Developmental Microscopy Core). The 5X images were collected on a Photometrics HQ2 high resolution monochrome CCD camera, and processed with LAS AF software. The 10X images were collected on a EVOS FL Auto Imaging System (UPenn Cell and Developmental Microscopy Core) and stitched using EVOS software. Exposure and gain were optimized within each experiment for one virus at the 72-hour time point and applied identically to each image within that experiment. Any subsequent image processing (ImageJ) was applied equally to all images in a given experiment.

Immunoblotting

Whole cell lysates were prepared with 1X LDS sample buffer (NuPage) and equal amounts of lysate were separated by SDS-PAGE. Membranes were immunoblotted with polyclonal rabbit antibodies raised against total HSV (Agilent Dako, B0114); glycoproteins (g)C, gD, gE, gH, and VP5 (all gifts from Gary Cohen); ICP8 (gift from David Knipe); and GAPDH (GeneTex, GTX100118). Proteins were visualized with Pierce ECL Western Blotting Substrate (ThermoFisher Scientific) and detected using a G:Box imaging system (Syngene).

Viral DNA isolation and Illumina sequencing

Viral nucleocapsid DNA for genome sequencing was prepared by infecting MRC-5 cells at an MOI ≥5 as previously described (7, 8). Viral nucleocapsid gDNA was sheared using a Covaris M220 sonicator/disruptor under the following conditions: 60s duration, peak power 50, 10% duty cycle, at 4°C. Barcoded sequencing libraries were prepared using the Illumina TruSeq low-throughput protocol according to manufacturer's specifications and as previously described (9, 10). The quality of sequencing libraries was evaluated by Qubit (Invitrogen, CA), Bioanalyzer

(Agilent), and qPCR (KAPA Biosystems). Paired-end sequencing (2 × 300bp length) was performed on an Illumina MiSeq, according to manufacturer's recommendations (17pM input).

De novo **genome assembly**

A consensus genome was assembled for each viral isolate using a previously described Viral Genome Assembly (VirGA) bioinformatics workflow (9). VirGA begins by quality-filtering the MiSeq sequence reads and removing sequences that match the host (human) genome. Thereafter VirGA uses a combination of SSAKE *de novo* assemblies run with differing parameters, which are then combined into a single draft genome using Celera and GapFiller (11–13). After quality-checking and iterative improvement of the genome assembly, annotations were transferred from the HSV-2 reference genome (strain HG52; GenBank NC_001798) to each new genome based on sequence homology (14).

Comparative genomics and phylogenetic analysis

The genomes of all 10 neonatal HSV-2 isolates were combined with all annotated HSV-2 genomes available in GenBank (see **Table S1** for full list) and aligned using MAFFT (15). All previously published HSV-2 genomes were derived from infected adults. The genome-wide alignment used a trimmed genome format (lacking the terminal repeats) to avoid giving undue weight to these duplicated sequences. The MAFFT alignment was used to generate a NeighborNet phylogenetic network in SplitsTree with Uncorrected P distances (16–18). A diverse subset of ten adult HSV-2 isolates was selected for protein-level comparisons with the ten neonatal isolates; these are indicated in **Table S1** with an asterisk. GenBank genomes that lacked open reading frame (ORF) and protein annotations were excluded from all comparisons

(19, 20). ClustalW2 was used to construct pairwise nucleotide alignments between whole genomes and pairwise amino acid alignments for each gene and protein (21). Custom Python scripts were used on these alignments to identify nucleotide and AA differences between samples.

Minority variant detection & quantification

Minority variants (MV) were detected using VarScan v2.2.11 (mpileup2snp and mpileup2indel commands) (22), using the following parameters to differentiate true MV from technical artifacts (23): minimum allele frequency ≥ 0.02 (2%); base call quality ≥ 20; read depth ≥ 100; independent reads supporting minor allele ≥5. MV with directional strand bias ≥ 90% were excluded. The genomic location and potential impact of each MV was assessed using SnpEff and SnpSift (24, 25).

Supporting Figures & Tables

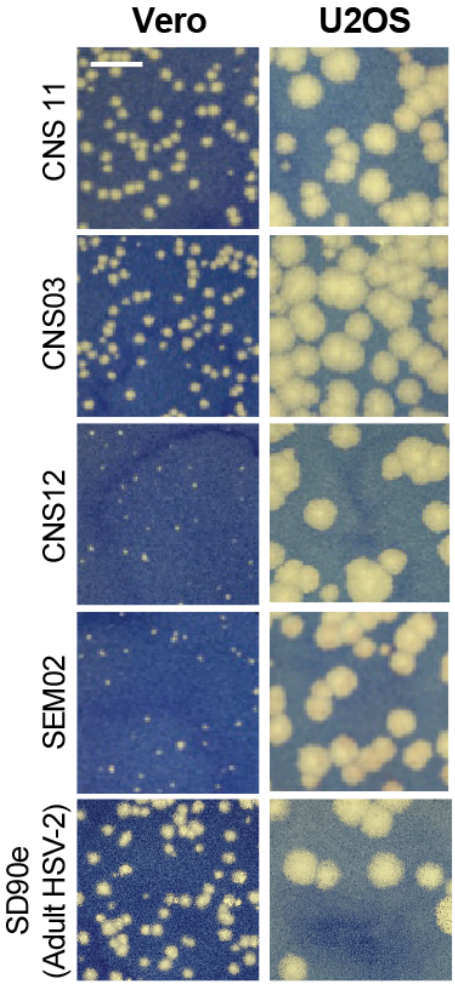

Figure S1. U2OS cells support large plaque formation by all isolates.

Representative plaques are shown after neonatal viruses were allowed to incubate for 100 hours on Vero **(A)** or U2OS cells **(B)**. All neonatal isolates were capable of forming large plaques on U2OS cells, which lack innate sensing of viral infection through the STING pathway (26).

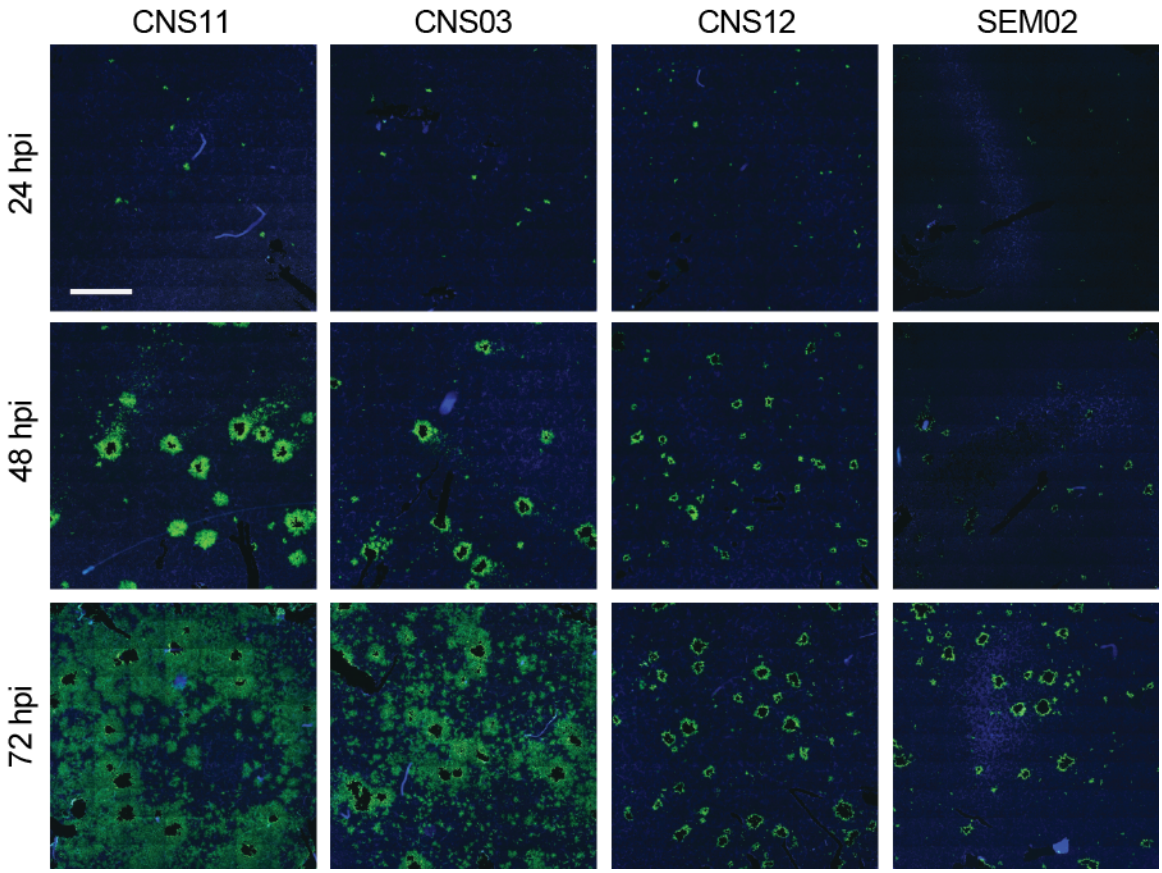

Figure S2. Cell-to-cell spread is enhanced in certain neonatal HSV-2 isolates.

Confluent Vero cell monolayers were infected at MOI=0.001 for the time points indicated, in the presence of 0.1% human serum. HSV positive cells (green) were evaluated at each time point by immunofluorescence. Cell nuclei are counterstained with DAPI (blue). Serial 10X images were obtained on an EVOS FL Auto Imaging System and stitched together to create an image of the entire experimental coverslip. Scale bar = 1mm. These images were quantified in **Figure 3E**.

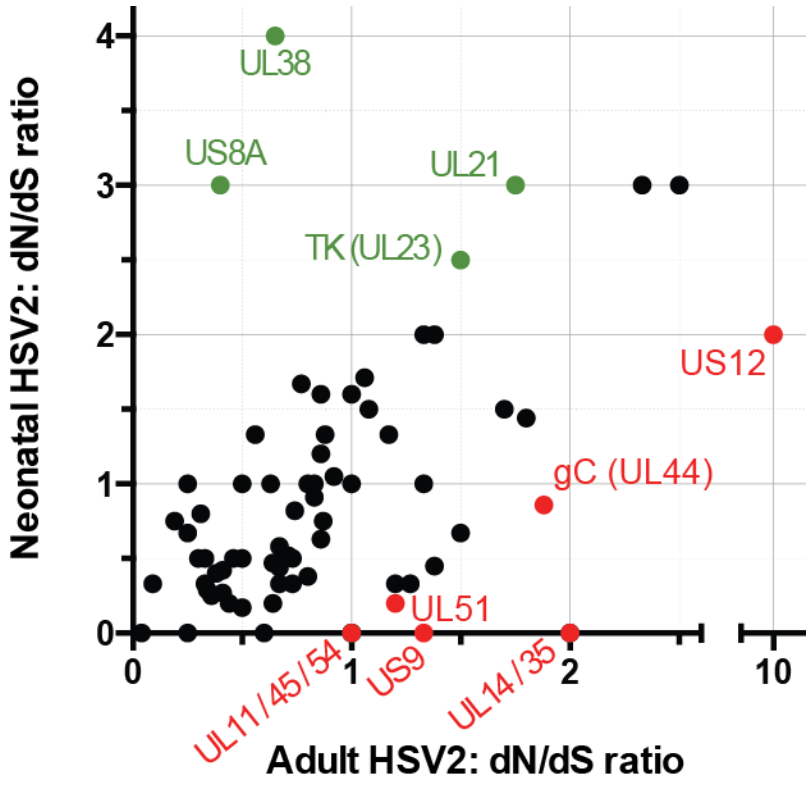

Figure S3. Ratio of non-synonymous to synonymous coding variations in neonatal HSV-2 versus adult HSV-2 strains.

The ratio of non-synonymous (dN) to synonymous (dS) coding variations were plotted for each HSV-2 protein. The x-axis value represents the average dN/dS ratio for each protein in 58 adult HSV-2 strains, while the y-axis value represents the average dN/dS ratio in 10 neonatal isolates. Proteins with a difference in average dN/dS ratios ≥1 in neonatal vs. adult HSV-2 genomes are labeled (green indicates a higher average dN/dS in neonatal HSV-2 genomes; red indicates a higher dN/dS in adult HSV-2 genomes). The average dN/dS ratios for all proteins are listed in **Table S2**.

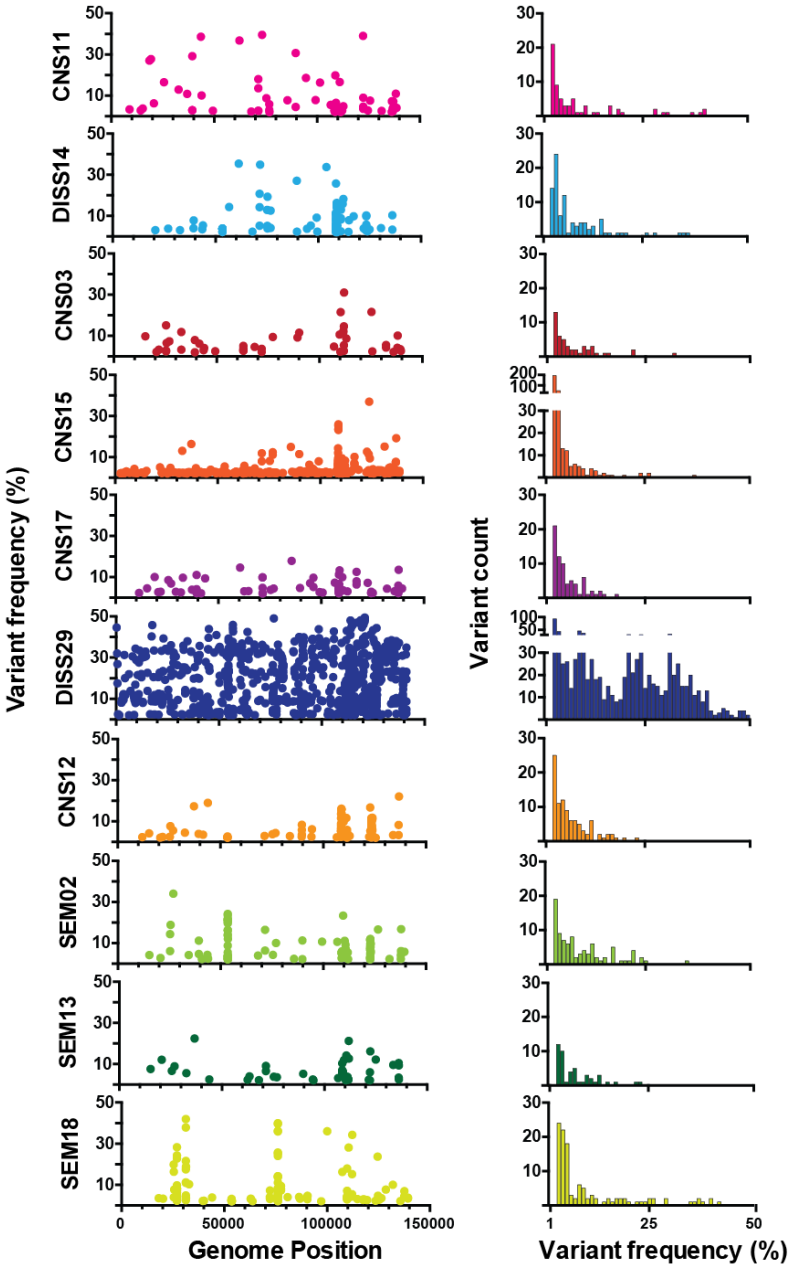

Figure S4. Genome-wide distribution and frequency of minority variants in neonatal HSV-2 isolates.

For each neonatal HSV-2 isolate, the graph on the left plots spatial location in the genome (x-axis) against the frequency at which each minority variant was observed. The plot on the right summarizes the number of minor variants (y-axis height) in binned increments of 1% (x-axis). The data reveal the distinctly different distribution of minority variants in DISS29, and to a lesser extent CNS15, than in other isolates. Color code matches that used in **Figure 6**.

1137 **Table S1. Adult HSV-2 genomes used for comparative genomic analyses.**

Virus Isolate	Country (with location, if available)	GenBank Accession #	References
H1227*	Finland	KY922721	In process
H1229*	Finland	KY922722	In process
H12212*	Finland	KY922726	In process
Uganda.2*	Uganda	KX574906	(27)
Peru.8*	Peru	KX574874	(27)
Zambia.4*	Zambia	KX574883	(27)
HG52*	Scotland	NC_001798	(28)
CtSF-R*	USA	KP334093	(29)
1192*	Wisconsin, USA	KP334095	(29)
SD90e*	Carletonville, South Africa	KF781518	(30, 31)
Peru.1	Peru	KX574860	(27, 32)
Peru.12	Peru	KX574861	(27, 32)
Zambia.3	Zambia	KX574862	(27, 32)
Zimbabwe.2	Zimbabwe	KX574863	(27, 32)
Zambia.2	Zambia	KX574864	(27, 32)
Peru.9	Peru	KX574865	(27, 32)
Peru.11	Peru	KX574866	(27, 32)
Peru.10	Peru	KX574867	(27, 32)
USA.8a	USA	KX574868	(27, 32)
USA.8b	USA	KX574869	(27, 32)
USA.1b	USA	KX574870	(27, 32)
Zimbabwe.1	Zimbabwe	KX574871	(27, 32)
Peru.7	Peru	KX574872	(27, 32)
Peru.2	Peru	KX574873	(27, 32)
Peru.6	Peru	KX574875	(27, 32)
Peru.4	Peru	KX574876	(27, 32)
Peru.3	Peru	KX574877	(27, 32)
Peru.5	Peru	KX574878	(27, 32)
South Africa.2	South Africa	KX574879	(27, 32)
South Africa.3	South Africa	KX574880	(27, 32)
South Africa.4	South Africa	KX574881	(27, 32)
USA.3a	USA	KX574882	(27, 32)
USA.3b	USA	KX574884	(27, 32)
USA.2b	USA	KX574885	(27, 32)
USA.10	USA	KX574886	(27, 32)
USA.4a	USA	KX574887	(27, 32)
USA.7a	USA	KX574888	(27, 32)
USA.6a	USA	KX574889	(27, 32)

Virus Isolate	**Country (with location, if available)**	**GenBank Accession #**	**References**
USA.6b	USA	KX574890	(27, 32)
USA.5a	USA	KX574891	(27, 32)
South Africa.1	South Africa	KX574892	(27, 32)
Kenya.2	Kenya	KX574893	(27, 32)
Kenya.1	Kenya	KX574894	(27, 32)
USA.7b	USA	KX574895	(27, 32)
USA.5b	USA	KX574896	(27, 32)
USA.11	USA	KX574897	(27, 32)
USA.4b	USA	KX574898	(27, 32)
Uganda.1	Uganda	KX574899	(27, 32)
USA.9	USA	KX574901	(27, 32)
Tanzania.1	Tanzania	KX574902	(27, 32)
USA.1a	USA	KX574903	(27, 32)
USA.2a	USA	KX574904	(27, 32)
Zambia.1	Zambia	KX574905	(27, 32)
2011-21400	USA	KX574908	(27, 32)
H12211	Finland	KY922725	In process
H1226	Finland	KY922720	In process
H1421	Finland	KY922723	In process
H1526	Finland	KY922724	In process

*Indicates 10 adult HSV2 strains used for AA comparisons in **Figure 5**.

Table S2. Number of nucleotide and amino acid (AA) differences observed in each set of neonatal or adult HSV2 genomes.

Gene*	ORF length in NT^	protein length (#AA)	10 Neonatal # NT diff.^	10 Neonatal # AA changes	10 Neonatal %AA diff.	10 Neonatal dN/dS ratio	10 Adult # NT diff.	10 Adult # AA changes	10 Adult %AA diff.	10 Adult dN/dS ratio	58 Adult # NT diff.	58 Adult # AA diff.	58 Adult %AA diff.	58 Adult dN/dS ratio
RL2	2,503	834	43	22	2.6	**1.0**	33	19	2.3	**1.4**	48	23	2.8	**0.9**
UL1	676	225	5	2	0.9	**0.7**	3	2	0.9	**2.0**	10	6	2.7	**1.5**
UL2	1,006	335	7	4	1.2	**1.3**	10	3	0.9	**0.4**	14	5	1.5	**0.6**
UL3	670	223	1	0	0.0	**0.0**	2	0	0.0	**0.0**	8	3	1.3	**0.6**
UL4	607	202	5	2	1.0	**0.7**	6	2	1.0	**0.5**	10	2	1.0	**0.3**
UL5	2,647	882	18	4	0.5	**0.3**	21	3	0.3	**0.2**	43	11	1.2	**0.3**
UL6	2,038	679	9	3	0.4	**0.5**	6	3	0.4	**1.0**	26	11	1.6	**0.7**
UL7	892	297	4	1	0.3	**0.3**	5	2	0.7	**0.7**	15	6	2.0	**0.7**
UL8	2,260	753	13	5	0.7	**0.6**	18	7	0.9	**0.6**	39	18	2.4	**0.9**
UL9	2,617	872	9	3	0.3	**0.5**	14	4	0.5	**0.4**	35	11	1.3	**0.5**
UL10	1,405	468	5	1	0.2	**0.3**	6	1	0.2	**0.2**	19	5	1.1	**0.4**
UL11	292	97	1	0	0.0	**0.0**	4	2	2.1	**1.0**	6	3	3.1	**1.0**
UL12	1,864	621	6	4	0.6	**2.0**	14	9	1.4	**1.8**	42	24	3.9	**1.3**
UL13	1,558	519	8	2	0.4	**0.3**	6	2	0.4	**0.5**	20	5	1.0	**0.3**
UL14	661	220	3	0	0.0	**0.0**	2	1	0.5	**1.0**	3	2	0.9	**2.0**
UL15	2,206	735	2	1	0.1	**1.0**	4	1	0.1	**0.3**	20	4	0.5	**0.3**
UL16	1,120	373	4	2	0.5	**1.0**	5	3	0.8	**1.5**	8	4	1.1	**1.0**
UL17	2,110	703	13	8	1.1	**1.6**	9	2	0.3	**0.3**	24	12	1.7	**1.0**
UL18	958	319	4	1	0.3	**0.3**	4	0	0.0	**0.0**	12	1	0.3	**0.1**
UL19	4,126	1,375	14	3	0.2	**0.3**	21	6	0.4	**0.4**	41	12	0.9	**0.4**
UL20	670	223	4	3	1.3	**3.0**	6	5	2.2	**5.0**	10	7	3.1	**2.3**
UL21	1,600	533	4	3	0.6	**3.0**	3	2	0.4	**2.0**	11	7	1.3	**1.8**
UL22	2,518	839	22	7	0.8	**0.5**	20	6	0.7	**0.4**	46	18	2.1	**0.6**
UL23	1,132	377	7	5	1.3	**2.5**	11	6	1.6	**1.2**	20	12	3.2	**1.5**
UL24	847	282	12	3	1.1	**0.3**	14	4	1.4	**0.4**	26	11	3.9	**0.7**
UL25	1,759	586	7	3	0.5	**0.8**	8	1	0.2	**0.1**	25	4	0.7	**0.2**
UL26.5	991	330	8	2	0.6	**0.3**	11	6	1.8	**1.2**	25	14	4.2	**1.3**
UL26	1,915	638	11	3	0.5	**0.4**	18	7	1.1	**0.6**	36	16	2.5	**0.8**
UL27	2,716	905	16	10	1.1	**1.7**	17	8	0.9	**0.9**	39	17	1.9	**0.8**
UL28	2,359	786	14	7	0.9	**1.0**	11	5	0.6	**0.8**	39	15	1.9	**0.6**

Gene*	ORF length in NT^	protein length (#AA)	10 Neonatal # NT diff.^	10 Neonatal # AA changes	10 Neonatal %AA diff.	10 Neonatal dN/dS ratio	10 Adult # NT diff.	10 Adult # AA changes	10 Adult %AA diff.	10 Adult dN/dS ratio	58 Adult # NT diff.	58 Adult # AA diff.	58 Adult %AA diff.	58 Adult dN/dS ratio
UL29	3,592	1,197	27	8	0.7	**0.4**	32	11	0.9	**0.5**	83	24	2.0	**0.4**
UL30	3,724	1,241	19	12	1.0	**1.7**	39	20	1.6	**1.1**	68	35	2.8	**1.1**
UL31	919	306	6	1	0.3	**0.2**	9	2	0.7	**0.3**	18	7	2.3	**0.6**
UL32	1,798	599	13	4	0.7	**0.4**	13	7	1.2	**1.2**	20	8	1.3	**0.7**
UL33	394	131	2	0	0.0	**0.0**	1	0	0.0	**0.0**	5	1	0.8	**0.3**
UL34	832	277	8	4	1.4	**1.0**	7	2	0.7	**0.4**	15	5	1.8	**0.5**
UL35	340	113	0	0	0.0	**0.0**	2	1	0.9	**0.0**	6	4	3.5	**2.0**
UL36	9,370	3,123	65	31	1.0	**0.9**	73	33	1.1	**0.8**	181	82	2.6	**0.8**
UL37	3,346	1,115	13	8	0.7	**1.6**	27	16	1.4	**1.5**	54	25	2.2	**0.9**
UL38	1,402	467	10	8	1.7	**4.0**	22	9	1.9	**0.7**	38	15	3.2	**0.7**
UL39	3,430	1,143	64	22	1.9	**0.5**	72	25	2.2	**0.5**	118	49	4.3	**0.7**
UL40	1,015	338	6	2	0.6	**0.5**	4	0	0.0	**0.0**	13	3	0.9	**0.3**
UL41	1,480	493	6	1	0.2	**0.2**	13	2	0.4	**0.2**	26	8	1.6	**0.4**
UL42	1,414	471	8	4	0.8	**1.0**	16	8	1.7	**1.0**	21	12	2.5	**1.3**
UL43	1,246	415	10	6	1.4	**1.5**	8	6	1.4	**3.0**	27	17	4.1	**1.7**
UL44	1,444	481	13	6	1.2	**0.9**	12	7	1.5	**1.4**	23	15	3.1	**1.9**
UL45	520	173	1	1	0.6	**0.0**	3	1	0.6	**0.0**	6	3	1.7	**1.0**
UL46	2,170	723	20	9	1.2	**0.8**	18	7	1.0	**0.6**	40	17	2.4	**0.7**
UL47	2,092	697	14	6	0.9	**0.8**	13	9	1.3	**2.3**	43	20	2.9	**0.9**
UL48	1,474	491	6	3	0.6	**1.0**	7	2	0.4	**0.4**	18	8	1.6	**0.8**
UL49A	265	88	3	1	1.1	**0.5**	6	2	2.3	**0.5**	8	2	2.3	**0.3**
UL49	904	301	19	7	2.3	**0.6**	17	7	2.3	**0.7**	30	12	4.0	**0.7**
UL50	1,111	370	8	4	1.1	**1.0**	7	4	1.1	**1.3**	22	10	2.7	**0.8**
UL51	736	245	6	1	0.4	**0.2**	1	0	0.0	**0.0**	11	6	2.4	**1.2**
UL52	3,202	1,067	22	12	1.1	**1.2**	31	11	1.0	**0.6**	54	25	2.3	**0.9**
UL53	1,018	339	14	8	2.4	**1.3**	16	10	2.9	**1.7**	26	14	4.1	**1.2**
UL54	1,540	513	5	0	0.0	**0.0**	7	3	0.6	**0.8**	24	12	2.3	**1.0**
UL55	562	187	3	1	0.5	**0.5**	2	2	1.1	**0.0**	9	3	1.6	**0.5**
UL56	709	236	6	4	1.7	**2.0**	8	5	2.1	**1.7**	19	11	4.7	**1.4**
US1	1,246	415	16	5	1.2	**0.5**	14	6	1.4	**0.8**	38	22	5.3	**1.4**
US2	877	292	7	4	1.4	**1.3**	4	3	1.0	**3.0**	15	7	2.4	**0.9**
US3	1,447	482	9	4	0.8	**0.8**	7	4	0.8	**1.3**	17	4	0.8	**0.3**
US4	2,101	700	22	13	1.9	**1.4**	40	25	3.6	**1.7**	70	45	6.4	**1.8**
US5	280	93	4	1	1.1	**0.3**	8	4	4.3	**1.0**	11	6	6.5	**1.2**

Gene*	ORF length in NT^	protein length (#AA)	10 Neonatal # NT diff.^	10 Neonatal # AA changes	10 Neonatal %AA diff.	10 Neonatal dN/dS ratio	10 Adult # NT diff.	10 Adult # AA changes	10 Adult %AA diff.	10 Adult dN/dS ratio	58 Adult # NT diff.	58 Adult # AA diff.	58 Adult %AA diff.	58 Adult dN/dS ratio
US6	1,183	394	7	2	0.5	**0.4**	8	2	0.5	**0.3**	18	5	1.3	**0.4**
US7	1,120	373	10	6	1.6	**1.5**	11	5	1.3	**0.8**	27	14	3.8	**1.1**
US8A	442	147	4	3	2.0	**3.0**	2	0	0.0	**0.0**	7	2	1.4	**0.4**
US8	1,639	546	12	9	1.6	**3.0**	9	9	1.6	**0.0**	28	20	3.7	**2.5**
US9	271	90	2	0	0.0	**0.0**	3	1	1.1	**0.5**	7	4	4.4	**1.3**
US10	262	87	7	1	1.1	**0.2**	3	2	2.3	**2.0**	6	2	2.3	**0.5**
US11	457	152	2	2	1.3	**0.0**	0	0	0.0	**0.0**	136	5	3.3	**0.0**
US12	910	303	6	4	1.3	**2.0**	6	0	0.0	**0.0**	11	10	3.3	**10.0**

* See Table S1 for list of strain names and genome accessions for comparisons of 10 neonatal, 10 adult, and 58 adult HSV2 genomes.
^ Abbreviations: ORF, open reading frame; NT, nucleotide; AA, amino acid; diff., difference; %, percent; dN/dS, ratio of nonsynonymous to synonymous nucleotide changes

Table S3. Viral genetic variations potentially associated with clinical groups.

HSV-2 Gene: Variation	Clinical Group (Isolates w/ Variation)	Predicted effect^	Protein Function	Neurovirulence
gK (UL53): V323M	All CSF Isolates (CNS11, DISS14)	Loss of one TM domain	Glycoprotein at virion and cell surface; interacts with gB and UL20 (33); important for cytoplasmic envelopment, viral egress, and virus-induced cell fusion	gK null virus exhibits poor neuronal spread in culture including decreased retrograde and anterograde transport (34); reduced spread within eye and to nervous system, as well as increased survival, following ocular challenge in mice (35)
gI (US7): R159L + P215S	All CSF Isolates (Both: CNS11, DISS14 R159L only: CNS03, SEM13, CNS15*, DISS29* P215S only: CNS15*, DISS29)	R159L: adds one TM domain; P215S: no change	Glycoprotein at virion and cell surface; forms heterodimer with gE to form viral Fc receptor (36); required for cell-to-cell spread of virus in epithelial and neuronal cells (37, 38);	gI null virus exhibits poor neuronal spread in culture; reduced spread in retina and to retinorecipient areas of the brain following ocular challenge in mice (38)
UL8: R221S	All CSF Isolates (CNS11, DISS14)	No predicted change	Putative primase subunit of helicase primase complex; required for unwinding of viral DNA	None identified
US2: F137L	All CSF Isolates (CNS11, DISS14)	No predicted change	Tegument protein; binds and phosphorylates TAK1 to positively regulate NFkB signaling (39)	Virus containing mutated US2 forms small plaques in culture (40); Mutant virus containing a deletion of both US2 and PK result in decreased death following intracerebral injection in mice (41)
gG (US4): R338L, S442P, E574D	All CSF Isolates (CNS11, DISS14)	No predicted change	Binds and enhances function of chemokines (42); enhances neurotrophin-dependent axonal growth in culture (43)	Virus containing mutated gG results in decreased death following intracerebral injection in mice (40)
UL24: V93A	No CNS Isolates; only found in SEM isolates (SEM02, SEM13, SEM18)	No predicted change	Membrane-associated nuclear protein; associated with dispersal of nucleolin during infection (44)	UL24 null virus forms small plaques in culture (45); decreased viral load in eye and nervous system following ocular challenge in mice (46–48)
UL20: P129L	All DISS Isolates w/ CNS Involved (DISS14, DISS29)	Change in position of three TM domains	Membrane-associated protein interacts with gK and gB (33); required for gK glycosylation, cell surface expression (49), and virus-induced cell fusion (50); important for viral egress (51)	UL20 null virus forms small plaques in culture (50); RNA inhibition of UL20 results in decreased rates of encephalitis following HSV footpad infection in mice (52)

* present as minority variant ^ Prediction by SMART Analysis

Supporting Information (SI) References

1. Kimberlin DW, Lin CY, Jacobs RF, Powell DA, Frenkel LM, Gruber WC, Rathore M, Bradley JS, Diaz PS, Kumar M, Arvin AM, Gutierrez K, Shelton M, Weiner LB, Sleasman JW, de Sierra TM, Soong SJ, Kiell J, Lakeman FD, Whitley RJ, National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group. 2001. Natural history of neonatal herpes simplex virus infections in the acyclovir era. Pediatrics 108:223–229.

2. Kimberlin DW, Whitley RJ, Wan W, Powell DA, Storch G, Ahmed A, Palmer A, Sánchez PJ, Jacobs RF, Bradley JS. 2011. Oral acyclovir suppression and neurodevelopment after neonatal herpes. N Engl J Med 365:1284–1292.

3. Kimberlin DW, Lin CY, Jacobs RF, Powell DA, Corey L, Gruber WC, Rathore M, Bradley JS, Diaz PS, Kumar M, Arvin AM, Gutierrez K, Shelton M, Weiner LB, Sleasman JW, de Sierra TM, Weller S, Soong SJ, Kiell J, Lakeman FD, Whitley RJ, National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group. 2001. Safety and efficacy of high-dose intravenous acyclovir in the management of neonatal herpes simplex virus infections. Pediatrics 108:230–238.

4. Whitley R, Arvin A, Prober C, Burchett S, Corey L, Powell D, Plotkin S, Starr S, Alford C, Connor J, Jacobs R, Nahmias A, Soong S-J, the National Institute of Allergy, Infectious Diseases Collaborative Antiviral Study Group*. 1991. A Controlled Trial Comparing Vidarabine with Acyclovir in Neonatal Herpes Simplex Virus Infection. N Engl J Med 324:444–449.

5. Whitley R, Arvin A, Prober C, Corey L, Burchett S, Plotkin S, Starr S, Jacobs R, Powell D, Nahmias A, Sumaya C, Edwards K, Alford C, Caddell G, Soong S-J, the National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group*. 1991. Predictors of Morbidity and Mortality in Neonates with Herpes Simplex Virus Infections. N Engl J Med 324:450–454.

6. Szpara ML, Parsons L, Enquist LW. 2010. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. J Virol 84:5303–13.

7. Szpara ML, Tafuri YR, Enquist LW. 2011. Preparation of viral DNA from nucleocapsids. J Vis Exp JoVE 2–7.

8. Szpara M. 2014. Isolation of herpes simplex virus nucleocapsid DNA, p. 31–41. *In* Diefenbach, RJ, Fraefel, C (eds.), Herpes Simplex Virus. Springer New York, New York, NY.

9. 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:e02213-14.

10. Bowen CD, Renner DW, Shreve JT, Tafuri Y, Payne KM, Dix RD, Kinchington PR, Gatherer D, Szpara ML. 2016. Viral forensic genomics reveals the relatedness of classic herpes simplex virus strains KOS, KOS63, and KOS79. Virology 492:179–186.

11. Myers EW, Sutton G, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KHJ, Remington KA, Anson EL, Bolanos RA, Chou H-H, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter CJ. 2000. A Whole-Genome Assembly of Drosophila. Science 287:2196–2204.

12. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:1–9.

13. Warren RL, Sutton GG, Jones SJM, Holt R a. 2007. Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23:500–1.

14. Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ. 1998. The genome sequence of herpes simplex virus type 2. J Virol 72:2010–2021.

15. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066.

16. Strimmer K, Wiuf C, Moulton V. 2001. Recombination analysis using directed graphical models. Mol Biol Evol 18:97–99.

17. Norberg P, Kasubi MJ, Haarr L, Bergstrom T, Liljeqvist J-A. 2007. Divergence and recombination of clinical herpes simplex virus type 2 isolates. J Virol 81:13158–13167.

18. Norberg P, Tyler S, Severini A, Whitley R, Liljeqvist J-A, Bergstrom T. 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. PloS One 6:1–8.

19. 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 JVI.01303-15.

20. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Reeves D, Schiffer J, Jerome KR, Sather C, Diem K, Lingappa JR, Celum C, Koelle DM, Wald A. 2017. Dual-strain genital herpes simplex virus type 2 (HSV-2) infection in the US, Peru, and 8 countries in sub-Saharan Africa: A nested cross-sectional viral genotyping study. PLOS Med 14:e1002475.

21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinforma Appl Note 23:2947–2948.

22. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration

discovery in cancer by exome sequencing. Genome Res 22:568–576.

23. Depledge DP, Kundu S, Jensen NJ, Gray ER, Jones M, Steinberg S, Gershon A, Kinchington PR, Schmid DS, Balloux F, Nichols RA, Breuer J. 2014. Deep Sequencing of Viral Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans. Mol Biol Evol 31:397–409.

24. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6:80–92.

25. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 3.

26. Deschamps T, Kalamvoki M. 2017. Impaired STING Pathway in Human Osteosarcoma U2OS Cells Contributes to the Growth of ICP0-Null Mutant Herpes Simplex Virus. J Virol 91:e00006-17.

27. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Cheng A, Diem K, Fitzgibbon MP, Huang M, Selke S, Lingappa JR, Celum C, Jerome KR, Wald A, Koelle DM. 2017. Highly conserved intragenic HSV-2 sequences: Results from next-generation sequencing of HSV-2 U L and U S regions from genital swabs collected from 3 continents. Virology 510:90–98.

28. Davison AJ. 2012. Evolution of sexually transmitted and sexually transmissible human herpesviruses. Ann N Y Acad Sci 1230:E37–E49.

29. 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.

30. 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.

31. Lai W, Chen CY, Morse SA, Htun Y, Fehler HG, Liu H, Ballard RC. 2003. Increasing relative prevalence of HSV-2 infection among men with genital ulcers from a mining community in South Africa. Sex Transm Infect 79:202–207.

32. 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.

33. Foster TP, Chouljenko VN, Kousoulas KG. 2008. Functional and physical interactions of the herpes simplex virus type 1 UL20 membrane protein with glycoprotein K. J Virol 82:6310–6323.

34. David AT, Saied A, Charles A, Subramanian R, Chouljenko VN, Kousoulas KG. 2012. A Herpes Simplex Virus 1 (McKrae) Mutant Lacking the Glycoprotein K Gene Is Unable To Infect via Neuronal Axons and Egress from Neuronal Cell Bodies. mBio 3.

35. David AT, Baghian A, Foster TP, Chouljenko VN, Kousoulas KG. 2008. The Herpes Simplex Virus Type 1 (HSV-1) Glycoprotein K(gK) is Essential for Viral Corneal Spread and Neuroinvasiveness. Curr Eye Res 33:455–467.

36. Johnson DC, Frame MC, Ligas MW, Cross AM, Stow ND. 1988. Herpes simplex virus immunoglobulin G Fc receptor activity depends on a complex of two viral glycoproteins, gE and gI. J Virol 62:1347–1354.

37. Dingwell KS, Brunetti CR, Hendricks RL, Tang Q, Tang M, Rainbow AJ, Johnson DC. 1994. Herpes simplex virus glycoproteins E and I facilitate cell-to-cell spread in vivo and across junctions of cultured cells. J Virol 68:834–845.

38. Dingwell KS, Doering LC, Johnson DC. 1995. Glycoproteins E and I facilitate neuron-to-neuron spread of herpes simplex virus. J Virol 69:7087–7098.

39. Lu X, Huang C, Zhang Y, Lin Y, Wang X, Li Q, Liu S, Tang J, Zhou L. 2017. The Us2 Gene Product of Herpes Simplex Virus 2 modulates NF-κB activation by targeting TAK1. Sci Rep 7:8396.

40. Weber PC, Levine M, Glorioso JC. 1987. Rapid identification of nonessential genes of herpes simplex virus type 1 by Tn5 mutagenesis. Science 236:576–579.

41. Meignier B, Longnecker R, Mavromara-Nazos P, Sears AE, Roizman B. 1988. Virulence of and establishment of latency by genetically engineered deletion mutants of herpes simplex virus 1. Virology 162:251–254.

42. Viejo-Borbolla A, Martinez-Martín N, Nel HJ, Rueda P, Martín R, Blanco S, Arenzana-Seisdedos F, Thelen M, Fallon PG, Alcamí A. 2012. Enhancement of chemokine function as an immunomodulatory strategy employed by human herpesviruses. PLoS Pathog 8:e1002497.

43. Cabrera JR, Viejo-Borbolla A, Martinez-Martín N, Blanco S, Wandosell F, Alcamí A. 2015. Secreted herpes simplex virus-2 glycoprotein G modifies NGF-TrkA signaling to attract free nerve endings to the site of infection. PLoS Pathog 11:e1004571.

44. Lymberopoulos MH, Pearson A. 2007. Involvement of UL24 in herpes-simplex-virus-1-induced dispersal of nucleolin. Virology 363:397–409.

45. Jacobson JG, Martin SL, Coen DM. 1989. A conserved open reading frame that overlaps the herpes simplex virus thymidine kinase gene is important for viral growth in cell culture. J Virol 63:1839–1843.

46. Rochette P-A, Bourget A, Sanabria-Solano C, Lahmidi S, Lavallée GO, Pearson A. 2015. Mutation of UL24 impedes the dissemination of acute herpes simplex virus 1 infection

from the cornea to neurons of trigeminal ganglia. J Gen Virol 96:2794–2805.

47. Leiva-Torres GA, Rochette P-A, Pearson A. 2010. Differential importance of highly conserved residues in UL24 for herpes simplex virus 1 replication in vivo and reactivation. J Gen Virol 91:1109–1116.

48. Jacobson JG, Chen SH, Cook WJ, Kramer MF, Coen DM. 1998. Importance of the herpes simplex virus UL24 gene for productive ganglionic infection in mice. Virology 242:161–169.

49. Foster TP, Alvarez X, Kousoulas KG. 2003. Plasma membrane topology of syncytial domains of herpes simplex virus type 1 glycoprotein K (gK): the UL20 protein enables cell surface localization of gK but not gK-mediated cell-to-cell fusion. J Virol 77:499–510.

50. Foster TP, Melancon JM, Baines JD, Kousoulas KG. 2004. The herpes simplex virus type 1 UL20 protein modulates membrane fusion events during cytoplasmic virion morphogenesis and virus-induced cell fusion. J Virol 78:5347–5357.

51. Foster TP, Melancon JM, Olivier TL, Kousoulas KG. 2004. Herpes simplex virus type 1 glycoprotein K and the UL20 protein are interdependent for intracellular trafficking and trans-Golgi network localization. J Virol 78:13262–13277.

52. Liu J, Lewin AS, Tuli SS, Ghivizzani SC, Schultz GS, Bloom DC. 2008. Reduction in severity of a herpes simplex virus type 1 murine infection by treatment with a ribozyme targeting the UL20 gene RNA. J Virol 82:7467–7474.