Full title:

Herpes simplex virus 2 (HSV-2) evolves faster in cell culture than HSV-1 by generating greater genetic diversity

Short title:

Differential *in vitro* generation of genetic diversity of HSV

Alberto Domingo López-Muñoz[1,#a], Alberto Rastrojo[1,#b], Rocío Martín[1], Antonio Alcamí[1]*

[1] Centro de Biología Molecular Severo Ochoa (Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid), Madrid, Spain

[#a]Current address: Cellular Biology Section, Laboratory of Viral Diseases, NIAID, NIH, United States

[#b]Current address: Genetic Unit, Department of Biology, Universidad Autónoma de Madrid, Madrid, Spain

* Corresponding author

E-mail: aalcami@cbm.csic.es

ABSTRACT

Herpes simplex virus type 1 and 2 (HSV-1 and HSV-2, respectively) are prevalent human pathogens of clinical relevance that establish long-life latency in the nervous system. They have been considered, along with the *Herpesviridae* family, to exhibit a low level of genetic diversity during viral replication. However, the high ability shown by these viruses to rapidly evolve under different selective pressures does not correlates with that presumed genetic stability. High-throughput sequencing has revealed that heterogeneous or plaque-purified populations of both serotypes contain a broad range of genetic diversity, in terms of number and frequency of minor genetic variants, both *in vivo* and *in vitro*. This is reminiscent of the quasispecies phenomenon traditionally associated with RNA viruses. Here, by plaque-purification of two selected viral clones of each viral subtype, we reduced the high level of genetic variability found in the original viral stocks, to more genetically homogeneous populations. After having deeply characterized the genetic diversity present in the purified viral clones as a high confidence baseline, we examined the generation of *de novo* genetic diversity under culture conditions. We found that both serotypes gradually increased the number of *de novo* minor variants, as well as their frequency, in two different cell types after just five and ten passages. Remarkably, HSV-2 populations displayed a much higher raise of nonconservative *de novo* minor variants than the HSV-1 counterparts. Most of these minor variants exhibited a very low frequency in the population, increasing their frequency over sequential passages. These new appeared minor variants largely impacted the coding diversity of HSV-2, and we found some genes more prone to harbor higher variability. These data show that herpesviruses generate *de novo* genetic diversity differentially under equal *in vitro* culture conditions. This might have contributed to the evolutionary divergence of HSV-1 and HSV-2 adapting to different anatomical niche, boosted by selective pressures found at each epithelial and neuronal tissue.

AUTHOR SUMMARY

Herpesviruses are highly human pathogens that establish latency in neurons of the peripheral nervous system. Colonization of nerve endings is required for herpes simplex virus (HSV) persistence and pathogenesis. HSV-1 global prevalence is much higher than HSV-2, in addition to their preferential tendency to infect the oronasal and genital areas, respectively. How these closely related viruses have been adapting and evolving to replicate and colonize these two different anatomical areas remains unclear. Herpesviruses were presumed to mutate much less than viruses with RNA genomes, due to the higher fidelity of the DNA polymerase and proofreading mechanisms when replicating. However, the worldwide accessibility and development of high-throughput sequencing technologies have revealed the heterogenicity and high diversity present in viral populations clinically isolated. Here we show that HSV-2 mutates much faster than HSV-1, when compared under similar and controlled cell culture conditions. This high mutation rate is translated into an increase in coding diversity, since the great majority of these new mutations lead to nonconservative changes in viral proteins. Understanding how herpesviruses differentially mutate under similar selective pressures is critical to prevent resistance to anti-viral drugs.

INTRODUCTION

Herpes simplex virus (HSV) is well-known for being one of the most prevalent neurotropic pathogens worldwide, causing a broad range of diseases in humans. The latest epidemiological studies estimate around 66% and 13.2% of global seroprevalence for HSV-1 and HSV-2, respectively, depending on age, sex, and geographical region [1]. HSV-1 infects the oral mucosa preferentially, causing characteristic minor skin lesions and eventually, encephalitis. Genital herpes is more often caused by HSV-2. Nevertheless, both viruses can infect either mucosa, frequently as a consequence of oral-genital sex [2-5]. After initial replication and dissemination at epithelial tissue, HSV infects sensory neuronal endings innervating the tissue [6, 7]. Once the virus reaches the trigeminal ganglia or the dorsal root ganglia, the virus establishes latency for the host's lifetime. The virus periodically reactivates from latent to replicative stage, traveling anterogradely back to the epithelial tissue, where recurrent infection and transmission occur, exhibiting associated disease symptoms [8]. During these multiple cycles of latency and reactivation, the virus may reinfect the nervous system or be transmitted to a new host, finding numerous chances to expand its genetic repertoire for subsequent reactivation cycles [9, 10]. The genetic diversity of the reactivating viral population can evolve by genetic drift, allowing the virus to respond to the host selective pressures that it faces when replicating. Other than the selective pressures due to immunological surveillance at the replication site, the virus may undergo selection in response to the differences between epithelial and neuronal environments [11-13]. Since HSV-1 and HSV-2 exhibit a preference for infecting different anatomical areas, it is reasonable to think that the selective pressures during each cycle of latency and reactivation might have contributed differentially to their evolution in humans, adapting their life cycle to each epithelial and neuronal niche.

The HSV genome is 152-155 Kilobase pairs long of double-stranded DNA, varying slightly between subtypes and strains. The genome is organized as unique long (UL)

113 and unique short (US) coding regions, flanked by terminal/internal long/short structural
114 repeats (i.e., T/I-Repeat(R)-L/S). The “a” sequence is present at the IRL-IRS border, but
115 also at the termini of the TRL and TRS, enabling the inversion of the unique fragments
116 orientation and producing four genomic isomers in equal ratios and functionality [14, 15].
117 Genomic replication generates long concatemers of viral DNA, which are processed into
118 unit-length genomes after cellular endonuclease G cleavage in the “a” sequence [16].
119 Those concatemers are highly branched, promoting recombination events between
120 repeated regions and resulting in the inversion of the UL and US segments [17].
121 Interestingly, high recombination and inversion rates in the HSV genome have been
122 extensively described, assisted by repetitive regions with a high G + C content [18-20].
123 Fluctuations in copy number, tandem repeats, and homopolymeric areas have been
124 described as a frequent source of genetic variability in HSV-1 [21]. These mechanisms
125 are critical components in the generation of variability within the large structural repeats
126 of the HSV genome, which in fact contain immediate-early expressed genes essential to
127 productive replication [8]. Nonetheless, genetic variability can be generated by different
128 mechanisms other than recombination and copy number/length fluctuations, where
129 genetic drift driven by polymerase error plays a critical role. Misincorporation of
130 nucleotides during genome replication leads to the appearance of single nucleotide
131 polymorphisms (SNPs) and insertions/deletions (InDels) both *in vivo* and *in vitro*.
132 Previous studies proposed a very low mutation rate for HSV-1 polymerase (1×10^{-7} to 1
133 $\times 10^{-8}$ mutations per base per infectious cycle), despite being performed on a single gene
134 analysis located at a unique coding region [11, 22, 23]. However, that mutation rate does
135 not correlate with the rapid ability observed when selecting HSV variants under drug
136 pressures, either by selection of preexisting minor variants (MVs) in the population or by
137 *de novo* mutations [22, 24-27].
138 Herpesviruses were presumed to generate lower genetic diversity when replicating and
139 then evolve with a slower rate than RNA viruses, due to their relatively more stable DNA
140 genome and proofreading mechanisms [12, 28, 29]. However, over the last decade, the

141 idea that herpesviruses exist as heterogenic and highly diverse populations *in vivo* has
142 gained increasing supportive evidence, aided by worldwide accessibility and cost
143 reduction of high-throughput sequencing technologies [30-33]. Nonetheless, how this
144 increasingly evident generation of genetic variability is generated remains not well
145 characterized. Shipley et al. reported that the genetic diversity of HSV-1 is able to change
146 over multiple cycles of latency and reactivation in the genital context [34]. Clinical isolates
147 of HSV-2 from a neonatal population showed extensive intra-host diversity, displaying
148 different *in vitro* phenotypes [31]. A similar level of generation of genetic diversity may
149 happen *in vitro* as well, where the method of preparation/propagation of HSV stocks
150 plays a key role [35]. In this context, multiple rounds of amplification in cell culture
151 increase the heterogenicity of viral stocks by genetic drift, where plaque-isolation has
152 been classically used to reduce the impact of this evolutionary force. HSV has been
153 routinely propagated in African green monkey kidney (Vero) cells, which are interferon
154 incompetent [36]. This, together with the fact that these cells are not of human origin and
155 just represent only one cell type among all the ones that HSV faces when infecting
156 humans, very likely impact the generation of genetic diversity *in vitro*. On the other hand,
157 *in vitro* studies of viral evolution can bring an ideal scenario with more controlled and
158 reduced selective pressures, where gaining insights into the mechanisms governing the
159 generation of genetic variability can be easier. Kuny et al. [37] have recently reported
160 that a heterogeneous population of HSV-1 increased its genetic diversity and changed
161 its phenotype more dramatically than a plaque-purified population of the same strain,
162 after ten passages in Vero cells. Certain genetic MVs already present in the
163 heterogeneous population appeared to be positively selected, whereas the purified
164 population did not appear to increase its MVs repertoire. However, the number and
165 frequency of the *de novo* mutations that appeared over those ten sequential passages
166 in Vero cells were not assessed. As far as we know, similar studies reporting how HSV-
167 2 *in vitro* evolves in terms of generation of genetic diversity have not been reported yet.
168 Thus, how differentially fast this genetic diversity is *de novo* generated *in vitro*, and how

169 it impacts the potential coding diversity (and thus, evolution) between HSV-1 and HSV-
170 2, remains unclear.
171 Here we performed sequential passage of two plaque-purified populations of both HSV-
172 1 and HSV-2 to characterize the differential ability and speed of these closely related
173 viruses to *in vitro* evolve in two different cell lines. We first assessed the genetic diversity
174 present in our original viral stocks by high-throughput sequencing, isolating plaque-
175 purified clones, and sequencing them. After nonconservative variant analysis, two
176 purified clones were ultra-deep re-sequenced, characterized in cell culture, and tested
177 in animal models of infection; in order to ensure that the genetic bottleneck of the plaque-
178 purification procedure did not alter the viral infectiousness due to a deleterious variant
179 unintentionally selected [38-40]. After establishing a high confident baseline of the
180 genetic diversity present in each purified population at passage 0 (P0), each of them was
181 subjected to ten serial passages in Vero and HaCaT cells, and ultra-deep sequenced at
182 passage 5 (P5) and passage 10 (P10). We detected the frequency and distribution of
183 preexisting and *de novo* genetic MVs, examining their impact on the coding capacity of
184 both HSV serotypes. These results have helped us to better understand how each HSV
185 subtype differentially evolves, depending on the selective pressures behind a given
186 cellular environment, and brings new insights into the different generation of genetic
187 diversity of the closely related HSV-1 and HSV-2.
188
189
190
191
192
193
194
195
196

197 **RESULTS**

198

199 **The viral phenotype of plaque-purified clones may dramatically change just in ten**
200 **passages in cell culture**

201 To better understand the generation of genetic diversity in HSV-1 and HSV-2, we first
202 isolated five viral clones from both HSV-1 strain SC16 and HSV-2 strain 333 original
203 stocks after five rounds of sequential plaque purification in Vero cells (Fig 1A). This was
204 done in order to reduce the preexisting genetic variability present in the original viral
205 populations, which was used as the baseline to assess the generation of genetic
206 variability.

207 Both original stock and purified clones, for each serotype, were deep-sequenced (S1-S3
208 Figs, see S1 Table for genome sequence statistics). After variant analysis, two purified
209 clones for each subtype were selected: clones 2 and 3 for HSV-1, and clones 1 and 5
210 for HSV-2. This selection was based on the lowest degree of nonconservative variability
211 compared to their corresponding reference sequence, previously described as *de novo*
212 assembled consensus genomes of the same original stocks used in this study [41, 42].
213 We compared the virus growth kinetics in cell culture, as well as the infectivity and
214 pathogenesis of these selected purified clones to their parental populations in mouse
215 models of infection, in order to ensure that the MVs detected in the purified populations
216 did not affect its viral fitness significantly (S4 Fig and S1 Text). These selected purified
217 clones were then ultra-deep re-sequenced (referred to as P0, Fig 1A), allowing us to
218 accurately identify very low frequency MVs present in these purified populations (S5 Fig).
219 That was a critical step to establish a high confidence baseline in order to be able to
220 discriminate between preexisting or *de novo* generated MVs, due to the intrinsic
221 variability generation rate in each HSV serotype replication cycle.

222 Each purified clone was used to infect a separate monolayer of Vero and HaCaT cells
223 at a multiplicity of infection (MOI) of 0.1 PFU/cell. Once cytopathic effect was observed,
224 cells and supernatant were harvested. This cycle of infection is considered as a passage

225 in the generation of variability experiments. The viral stock from the first passages of
226 each selected purified clone was used to infect the next monolayer of Vero and HaCaT
227 cells at the same estimated MOI, being this process repeated for ten sequential
228 passages (Fig 1A, lower part). The viral population of each purified clone in each cell line
229 was ultra-deep sequenced after five and finally ten sequential passages (S6-S9 Figs).
230 An MOI of 0.1 PFU/cell was selected to allow for multiple rounds of replication in each
231 passage, favoring the generation of genetic variability in the viral population.
232 Kuny et al. [37] did not find any changes in plaque phenotype in an HSV-1 purified
233 population passaged ten times in Vero cells. However, we observed dramatic changes
234 in plaque morphology in the purified HSV-1 clone 3 after ten passages in Vero cells (Fig
235 1B). After ultra-deep sequencing and variant analysis, we found a nonsynonymous SNP
236 previously described as the cause of the syncytial plaque phenotype [43-45]. The
237 frequency of the R858H variant in the UL27 gene (encoding glycoprotein B, gB) was
238 inexistent at P0 and P5, reaching 46.52% at P10, just in 5 passages in Vero cells (S9
239 Table, variant #38). In addition, the purified HSV-2 clone 1 also exhibited the same
240 variant at P5 in Vero cells, with a frequency of 1.75% and being undetected at P0 (S11
241 Table, variant #64). In contrast with HSV-1 clone 3, the frequency of this variant did not
242 increase at P10 for HSV-2 clone 1, not acquiring the syncytial plaque phenotype. Neither
243 this variant nor other syncytia-inducing MVs in gB, such as L817P [37], were detected
244 when purified clones were passaged in HaCaT cells, for both HSV subtypes. These
245 results suggested that even low genetically diverse purified viral populations are able to
246 quickly change or evolve just in a few passages in cell culture, being more prone to
247 happen in some cell lines than in others.
248
249 **The high genetic diversity found in HSV-1 and HSV-2 original stocks was**
250 **significantly reduced in five rounds of plaque isolation in Vero cells**
251 Genetic diversity can be defined as nucleotide alleles or variants present in a given
252 percent of the sequencing reads, at a given locus in a sequenced viral population. With

253 enough deep sequencing coverage, these MVs can be confidently detected, revealing
254 the genetic diversity present in the viral population. HSV-1 and HSV-2 original stocks
255 had an average coverage depth of 1783 and 1320 reads/position, respectively (S1
256 Table). We identified SNPs and InDels that were present in greater than 1 percent of the
257 sequencing reads (1 percent cut-off as the threshold of detection, plus additional
258 coverage-dependent filters, see "Material and methods" for detailed criteria). Both viral
259 populations from original stocks had MVs at different sites and frequencies, being evenly
260 distributed across highly repetitive areas and coding regions into their reference genome
261 (S1 Fig). As previously described for other mixed populations of HSV [31, 37], both HSV-
262 1 and HSV-2 original stocks displayed a significant number of MVs (Fig 2A), when
263 sequencing reads were aligned to each corresponding reference genome (*de novo*
264 assembled consensus genome of a purified clone from each same parental stock [41,
265 42]). HSV-1 original stock registered a total of 712 MVs, whereas HSV-2 stock, 1044
266 (Fig 2A, see S3 and S4 Tables for full list). Detailed analysis after variant calling showed
267 that the higher fraction of MVs corresponded to nonsynonymous SNPs for both HSV-1
268 and HSV-2, accounting for 452 and 701, respectively (S2 Table). In terms of frequency,
269 the major fraction of total detected MVs showed to be between 1% - 10% for both HSV
270 serotypes (Fig 2B).
271 Because we found a high level of genetic variability in sequenced viral population from
272 each original stock, we plaque-purified five viral clones from each parental stock in order
273 to use low genetically diverse viral populations as a baseline to evaluate the generation
274 of *de novo* variability in both HSV subtypes. After sequential plaque-isolation of
275 independent viral clones from their parental stocks, sequencing data from these purified
276 viral populations revealed a significant reduction in the number of total detected MVs, for
277 both HSV-1 and HSV-2 (Fig 2C, see S3 and S4 Tables for full list). In fact, every purified
278 clone showed a significant reduction in the total number of MVs, proportionally reflected
279 in the number of nonsynonymous SNPs, for both serotypes. With the exception of HSV-
280 1 clone 4, as well as HSV-2 clone 3, every other purified clone reduced its genetic

281 variability by 10-fold or higher, when the total detected number of MVs was compared to
282 the number registered for their original stocks. The total number of MVs detected from
283 each clone was grouped by HSV subtype and compared between them, showing no
284 statistically significant differences (Fig 2D), despite the differences in average coverage
285 depth among them (S1 Table). Based on the lowest number of nonconservative changes
286 (i.e., nonsynonymous SNPs and InDels in coding regions) observed among the five
287 purified clones for each HSV subtype, as well as their frequencies in the viral populations,
288 two purified clones of each serotype were selected for further characterization and *in*
289 *vitro* evolution studies. HSV-1 clones 2 and 3 were selected and tracked by a differential
290 SNP in the UL14 CDS (#162, S3 Table), while HSV-2 clones 1 and 5, by a differential
291 SNP in the UL13 CDS (#180, S4 Table).
292 The selected purified clones were used to perform replication kinetics in Vero cells, as
293 well as to infect mice as described in S1 Text, in order to confirm whether the MVs
294 detected could cause a deleterious effect in terms of infectivity and pathogenesis,
295 compared to their parental stocks (S4 Fig). No significant differences were found in the
296 replication and infectivity of the purified clones (S4A Fig). Despite finding some variability
297 in terms of survival when compared to original stocks, all four selected clones were able
298 to successfully infect and cause disease in both mouse models of infection tested (S4B
299 Fig), understanding that none of the unintentionally selected MVs during the plaque-
300 purification genetic bottleneck caused a significant deleterious effect. Thus, plaque-
301 isolation proved to be a successful approach to decrease the genetic diversity in the
302 purified viral populations before studying the generation of genetic variability in cell
303 culture. These purified viral populations of more uniformed genetic diversity constituted
304 the key starting point to determine with high confidence the generation of genetic
305 variability, particularly in terms of very low frequency MVs.
306
307

308 **Depth of sequencing is critical to establish a high confidence baseline in order to**
309 **detect very low frequency genetic diversity**

310 Depth of sequencing coverage is instrumental in detecting MVs with high reliance and
311 accuracy. Because the number of sequencing reads correlates directly with the
312 frequency of alleles in the viral population, a higher depth of coverage allows a better
313 resolution of the genetic diversity present in a given viral population. Notably, when
314 identifying *de novo* appeared MVs, a high depth of coverage is crucial to detect the
315 genetic diversity represented with a very low frequency in the viral population (i.e.,
316 theoretically at least an average coverage of 200 reads/position, a hundred paired-end
317 reads, would be required in order to be able to detect a 1 percent frequency variant,
318 supported by two reads contained the same allele).

319 Nonetheless, having an average coverage depth of 200X from a sequenced viral stock
320 containing 10^6 PFU, it would only represent 0.02 percent of the viral population. To
321 surmount the fact that the sequencing of a minute fraction of the whole viral population
322 might seriously bias our ability to detect a given variant, and therefore to determine its
323 novelty, we ultra-deep re-sequenced the previously selected purified viral clones. The
324 average coverage depth increased approx. 2 logs for each purified viral clone, from 10^2
325 (standard-deep sequencing, SDS) to 10^4 (ultra-deep sequencing, UDS) (Fig 3A, S1
326 Table). In this context, this depth of coverage would represent a theoretical 1 percent of
327 the viral population (assuming 10^6 PFU), having increased a hundred times the actual
328 genetic diversity sampled from each viral population. Based on this, we understood this
329 level of coverage depth constituted a high confidence representation of the genetic
330 diversity present in each viral population.

331 A more detailed look at the sequencing statistics of the ten purified clones
332 aforementioned showed that among each serotype, HSV-1 clone 4 and HSV-2 clone 3
333 showed the highest depth of average coverage, as well as the highest number of total
334 MVs (Fig 2C, S1 Table). This could reasonably lead to think that the lower number of
335 detected MVs in the selected purified clones was due to their lower average coverage.

336 Nevertheless, we did not find a dramatic increment in the total number of MVs obtained
337 from SDS versus UDS for any of the four selected purified viral clones (Fig 3B, S2 Table).
338 The total number of MVs from SDS versus UDS were grouped and compared, showing
339 no statistically significant differences (Fig 3C left graph), while grouped preexisting MVs
340 did (Fig 3C center graph). The number of preexisting MVs (i.e., variants already detected
341 in the original parental stock) increased in HSV-1 purified clones, being this increment
342 higher in HSV-2 clones (Fig 3B). However, when *de novo* MVs were grouped, there was
343 no statistically significant difference between them (Fig 3C right graph), despite all but
344 HSV-1 clone 3 showing an increment in the total accounted *de novo* MVs (Fig 3B, S2
345 Table). When we grouped these *de novo* MVs based on their frequency, from SDS
346 versus UDS data, we observed a consistent increase in the number of very low frequency
347 MVs (1% to 2%), as well as a reduction in the number of low frequency MVs (>2% to
348 <10%), across every purified clone (Fig 3D, S2 Table). This reduction in the number of
349 *de novo* low frequency MVs was particularly pronounced in the case of HSV-1 clone 3,
350 explaining the observed decrease in the total number of *de novo* MVs foresaid. Because
351 of the higher depth of average coverage from UDS, we were able to detect a significantly
352 higher number of *de novo* MVs with a very low frequency in the viral populations (Fig
353 3E), but also to better discriminate the actual frequency of these MVs in the viral
354 population. These improvements of UDS may explain the reduction in the number of low
355 frequency MVs, switching to be detected with a lower frequency when the deep
356 sequencing coverage increases. The fact that the number of preexisting MVs, but not
357 the number of de novo MVs, increased dramatically between SDS vs. UDS indicated
358 that higher depth of coverage helps to gain accuracy and resolution characterizing the
359 existing genetic diversity. Since all four purified clones showed a comparable level of
360 genetic diversity, in terms of total and *de novo* MVs (ranging from 98 to 139, S2 Table),
361 we determined this as a suitable, high confidence starting point to study the generation
362 of genetic variability of both HSV subtypes in cell culture.
363

364 **HSV-2 *in vitro* evolves dramatically faster than HSV-1 in both Vero and HaCaTs**
365 **cells**

366 Having characterized the genetic diversity present in each purified viral clone with high
367 confidence, we next investigated how HSV-1 and HSV-2 differentially evolve in cell
368 culture. We conducted ten sequential passages of each purified viral clone separately in
369 Vero and HaCaT cells, ultra-deep sequencing each viral population after 5 and 10
370 passages in each cell line. We obtained a broad range of average coverage depth,
371 ranging from 872 to 25892 reads/position (S1 Table). Nonhuman primate kidney-derived
372 epithelial (Vero) cells are widely and routinely used for HSV propagation, whereas
373 human keratinocyte (HaCaT) cells are closer to the natural physiology of HSV infection
374 in the skin. HSV plaque formation, cell-to-cell spread, and cell migration were reported
375 to be significantly different when compared HaCaT to Vero cell infections [46]. Based on
376 that, we sought to examine the effects of the differential selective pressures present in
377 each cultured cell line to the generation of genetic diversity of each HSV subtype.

378 After variant analysis, we analyzed the total number and types of MVs that were present
379 in each viral population at P5 and P10, for every purified clone passaged in each cell
380 type. Most of the detected MVs corresponded to mutations impacting coding regions,
381 including SNPs and InDels (see S2 Table). Nonetheless, both HSV-1 and HSV-2
382 populations displayed a similar fraction of the total number of MVs impacting coding
383 regions, around 56 percent (+/- 0.094% SD) on average (i.e., nonsynonymous SNPs
384 plus InDels in coding regions, divided by the total number of detected MVs, see S2
385 Table). We detected a consistently higher total number of MVs among HSV-2 than in
386 HSV-1 viral populations, as well as a higher and increasing number of *de novo* MVs after
387 sequential passages, not previously detected in the purified viral populations at P0 (Fig
388 4A, S1-S4 Animations). Preexisting variability in HSV-1 populations remained constant
389 over the ten passages in both cell types, but neither a relevant increase in the total
390 number of MVs nor in *de novo* generated MVs were observed (Fig 4A left graph, see S2
391 Table for more details). However, despite remaining the preexisting number of MVs

392 constant in HSV-2 populations over sequential passages, we found a consistent
393 increment in the appearance of *de novo* generated MVs after five, and even greater,
394 after ten passages for every purified clone (Fig 4A right graph). Both HSV-2 purified
395 clones showed the most drastic increment in the number of *de novo* MVs after being
396 subjected to ten sequential passages in HaCaT cells, dramatically higher than that found
397 when sequentially passaged in Vero cells (Fig 4A right graph). These results showed
398 that HSV-2 generates *de novo* genetic diversity faster than HSV-1, where the selective
399 pressures present in each cell type used for viral propagation may differentially affect
400 how the genetic diversity is generated. In a more detailed analysis of these *de novo* MVs,
401 we classified them as substitutions or SNPs and InDels. We found that the *de novo* MVs
402 generated over sequential passage of both HSV-1 and HSV-2 purified populations were
403 predominately nonconservative changes (Figs 4B and 4C). That was particularly
404 remarkable in the case of HSV-2 populations, where the number of detected
405 nonsynonymous *de novo* SNPs and InDels impacting coding regions gradually and
406 consistently increased, over the ten passages in both cell types and for both purified
407 clones (Figs 4B and 4C, right graphs). These data suggest that the generation of *de novo*
408 genetic diversity in HSV promotes predominantly nonconservative changes.
409 For each *de novo* variant, we also examined its frequency in the population. They were
410 clustered, based on their frequency, in very low frequency MVs (1% to 2%), low
411 frequency MVs (>2% to <10%), medium frequency MVs (10% to <50%), and high
412 frequency variants (equal to or greater than 50%). We observed that the major fraction
413 of *de novo* generated variability corresponded to very low frequency MVs, consistently
414 displayed by almost every purified clone after ten passages in both cell lines (Fig 4D).
415 These very low frequency MVs barely increased in HSV-1 populations after sequential
416 passages, even slightly decreasing as displayed by HSV-1 clone 3, whereas HSV-2
417 clones showed a consistent and significant increment of those. As each HSV-2
418 population was passaged, the very low frequency MVs increased their proportion in the
419 population as the predominant group of MVs. Additionally, we also observed that there

420 was a gradual and systematic increment of low-medium frequency MVs (taken together)
421 across every viral population of both HSV serotypes (Fig 4D). In this regard, we found
422 that both HSV-2 purified clones remarkably showed the highest increment of low-
423 medium frequency MVs when passaged in HaCaT cells, where HSV-2 clone 5 increased
424 the proportion of these low frequency MVs over the total number of *de novo* variants,
425 from 30% (at P5) to 67% (at P10) (Fig 4D right graph, see S2 Table for details). These
426 data identify these very low frequency MVs as the main source of generation of genetic
427 variability in HSV, gradually increasing as a percentage of the viral population over
428 sequential passage, where each HSV subtype changes with a different speed in
429 response to the same pressures of a given environment.
430

431 ***De novo* variants increase the potential coding diversity of HSV-2**

432 After observing that the *de novo* genetic diversity detected after sequential passages in
433 culture was predominantly translated into nonconservative changes for both HSV
434 subtypes, we further examined the distribution of *de novo* MVs impacting coding regions.
435 In accordance with the total number of *de novo* MVs aforementioned (Fig 4A), HSV-1
436 viral populations displayed an anecdotic low number of MVs impacting just a few genes,
437 while nearly every HSV-2 gene harbored *de novo* MVs (Fig 5). Despite finding an even
438 distribution of *de novo* MVs across the coding genome, some HSV-2 genes were
439 identified as hotspots of novel coding variability (e.g., UL16, UL27, UL28, UL29, UL36).
440 It is also remarkable to observe how the sequential passage of both HSV-2 purified
441 clones in HaCaT cell impacted much more heavily the coding capacity of the HSV-2
442 genome, as shown in Fig 5, where yellow and orange colors (purified clones passaged
443 in HaCaT cells) dominate over blue tones (passages in Vero cells). Although RL1, RL2,
444 and RS1 genes are located into repetitive areas of the genome, which have been
445 demonstrated to be regions of high variability [38, 40], we also identified a higher number
446 of *de novo* MVs present in these genes after sequential passage of HSV-2 clones in
447 HaCaT cells, compared to Vero cells (Fig 5). As shown in Figs 4B and 4C, for both *de*

448 *novo* appeared SNPs and InDels, intragenic MVs outnumbered those in intergenic
449 regions, despite thinking that higher selective pressures would reduce the appearance
450 of unfavorable mutations in coding regions. These data revealed the impact of very low
451 (1%-2%) and low frequency (<10%) *de novo* MVs in expanding the coding genetic
452 variability in HSV-2, which might have differentially contributed to HSV-2 evolution
453 depending on specific niches.
454
455 **Frequency increase of *de novo* variants may depend on positive selective**
456 **pressure in cell culture**
457 Finally, we examined how the frequency of nonconservative *de novo* variants changed
458 over sequential passages of each purified viral population in both cell types. We found
459 that most of these nonconservative *de novo* MVs did not increase their frequency in the
460 population, seeming to reach a stationary equilibrium or just disappearing after ten
461 passages (S7-S14 Tables). We did not detect any *de novo* InDel in coding regions
462 gradually increasing its frequency in the passaged viral populations, likely reflecting the
463 stronger selective pressure against missense mutations in coding regions. However, a
464 few nonsynonymous *de novo* SNPs were found increasing their frequency significantly
465 in the viral populations, which suggested that those might be conferring a selective
466 advantage. The clearest example for this outcome is illustrated by the syncytia-forming
467 MVs in UL27, as previously described by others [37, 43-45], also observed in the purified
468 HSV-1 clone 3 passaged in Vero cells (Fig 1B). That variant in UL27 (R858H) was not
469 high-confidently detected neither at P0 nor at P5 in Vero cells, *de novo* appearing after
470 P5 and reaching almost a frequency of 50% in the viral population at P10 (Fig 6 left
471 graph, S9 Table). On the other hand, we also observed an interesting positive selection
472 example of a *de novo* variant not conferring a selective advantage in cell culture. We
473 found a nonsynonymous *de novo* variant in the UL13 gene with a frequency of 67%
474 within the HSV-2 clone 1 population after five passages in Vero cells, reaching 90% of
475 the population at P10 (Fig 6 right graph, S11 Table). It has been described that UL13

476 kinase activity is required for axonal transport *in vivo* [47], but it is dispensable *in vitro*
477 [48]. Missense mutations in the UL13 gene have been reported to increase in frequency
478 in different strains of HSV-1 [19, 21, 37, 38]. Thus, it is reasonable to think that the UL13
479 variant found in HSV-2 clone 1 was then not selected based on a selective advantage,
480 but because of the previously observed high tolerance of UL13 inactive kinases. Other
481 MVs, such as those impacting UL14 and UL24 in HSV-2 populations (Fig 6), also
482 reached almost 80-90% of the population at P10. Different variants impacting those
483 genes in HSV-1 have been described, increasing their frequency over sequential
484 passage in cell culture [37], which suggests that nonconservative MVs affecting those
485 genes might confer a selective advantage, or just be more tolerable when HSV replicates
486 *in vitro*.
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503

504 **DISCUSSION**

505

506 In this study, we assessed the *de novo* evolution of HSV-1 and HSV-2 generated over
507 sequential passage in two different cell types, using the neurovirulent viral strains SC16
508 and 333 as a model for each HSV subtype. We characterized for the first time the whole-
509 genome generation of new genetic diversity during viral replication *in vitro*, after having
510 set a high confidence baseline of the preexisting genetic diversity by ultra-deep
511 sequencing of plaque-purified viral populations. This approach allowed us to identify very
512 low frequency *de novo* mutations within genetically homogeneous viral populations, and
513 then examine how viral populations of both HSV serotypes drifted under equal cell
514 culture conditions, in two different cell types. We found that both HSV-1 and HSV-2
515 increased the number and frequency of *de novo* MVs after five and ten passages, being
516 most of those low frequency nonconservative mutations impacting coding regions.
517 Interestingly, we observed that purified HSV-2 populations were much more prone to
518 generate genetic diversity during passaging than HSV-1, despite displaying a similar
519 number of total preexisting MVs before sequential passages. While the genetic diversity
520 of HSV-1 clonal populations remained similarly stable after being passaged in both Vero
521 and HaCaT cells, HSV-2 purified clones evolved significantly faster when passaged in
522 HaCaT than in Vero cells. HaCaT cells are human skin keratinocytes, an epithelial cell
523 type closer to the natural physiology of HSV infection in the skin. HSV plaque formation,
524 cell-to-cell spread, and cell migration were reported to be significantly different when
525 compared HaCaT to Vero cell infections [46]. In addition, HSV-1 clinical isolates
526 circulating in Finland replicated to lower titers and produced fewer extracellular viral
527 particles in Vero than in HaCaT cells [30]. Since HSV propagation in Vero cells has been
528 the traditional method for viral stock production and *in vitro* studies in virology labs, the
529 evolutionary dynamic shift shown here by HSV-2 in each cell type highlights how critical
530 it is to understand and characterize viral evolution and adaptation in specific cell type
531 cultures. It is instrumental to determine how specific cell-type-associated selective

532 pressures affect our experimental understanding of viral evolution and population
533 dynamics *in vitro*.
534 From these *in vitro* studies, we monitored how genetic drift happened faster in HSV-2
535 than in HSV-1, emphasizing the differential contribution to generate genetic diversity
536 under equal *in vitro* controlled conditions. This lower genetic variability showed by HSV-
537 1 seems to be strain-independent, as recently reported for HSV-1 strain F [37]. Both
538 heterogeneous and purified populations of this HSV-1 strain were sequentially passaged
539 ten times in Vero cells. We quantified the number of *de novo* mutations that appeared
540 over five and ten sequential passages using their available data, finding 19 (P5) and 7
541 (P10) *de novo* MVs (≥2%) not listed at P0 for the heterogeneous population (Mixed as
542 referred in the original article), while 3 (P5) and 4 (P10) *de novo* MVs were detected in
543 the purified population. These numbers are similar to the number of *de novo* mutations
544 (≥2%) that we found for HSV-1 strain SC16 purified populations in both cell types. Before
545 the high-throughput sequencing era, it was also reported that the spontaneous mutation
546 rate in laboratory strains of HSV-2 (including the 333 strain) was 9- to 16-fold more
547 frequent than that in HSV-1 SC16, when selecting drug-resistant mutants in cell culture
548 [25]. These data correlate with our findings observed from the genome-wide analysis of
549 genetic diversity, where HSV-2 populations generated a 5-fold higher number of variants
550 than HSV-1 counterparts in Vero cells, but 10-fold higher when passaged in HaCaT cells.
551 On the other hand, different studies describing the genetic diversity found in clinical
552 isolates of both serotypes, by high-throughput sequencing technologies, also support
553 that HSV-2 is more prone to generate nonconservative genetic diversity than HSV-1 also
554 *in vivo*. Seven uncultured swab specimens of genital HSV-1 showed a total of 114
555 summed variants (≥2%) [34], while 10 HSV-1 clinical isolates from Finland exhibited less
556 than 150 grouped variants, mainly localized at repeated regions [30]. However, when ten
557 neonatal HSV-2 isolates were examined, a total of 1,821 variants was found, of which
558 784 were found across 71 genes [31]. The degree of coding diversity was also similar

559 when compared to a different set of 10 adult HSV-2 isolates. Additionally, it was also
560 reported that HSV-2 isolates generated drug-resistant mutants 30 times faster than HSV-
561 1 clinical isolates [25]. It seems reasonable to attribute the higher mutation frequency of
562 HSV-2 to a lower fidelity of its polymerase during viral replication. Nevertheless, HSV-1
563 recombinants expressing HSV-2 polymerase were reported to have similar error rates to
564 HSV-1 parental homologs, suggesting that the polymerase would not be solely
565 responsible for these serotype-specific differences in mutation frequency, and other viral
566 proteins and secondary structures of the genome might contribute to explain it [49].
567 These experiments emphasized the critical value and usefulness of using *de novo*
568 assembled consensus genomes generated from the actual stocks under study [41, 42],
569 as well as of using purified homogeneous viral populations from the exact parental
570 stocks. These previously described consensus genomes for strains SC16 and 333
571 represented with higher accuracy the structural and genetic heterogenicity contained into
572 each parental stock, rather than having used reference strain genomes commonly used
573 for comparative genomics [35]. The reduction of the initial genetic variability contained
574 into each original viral stock by plaque-picked isolation of subclones, contributed to set
575 a more uniform baseline to identify new generated MVs during viral replication. Some
576 genome-wide studies of HSV-1 revealed approximately 3-4% of nucleotide variation
577 genome-wide, being reduced up to 1-2% after plaque-purification of subclones [21, 40,
578 50]. We reduced this genome-wide nucleotide variation rate from 0.47% (712 MVs /
579 150,000 bp) and 0.7% (1044 MVs) to an average of 0.08% (118 MVs) and 0.1% (145
580 MVs), between five times plaque-purified clones and their parental stocks for HSV-1 and
581 HSV-2, respectively. Nonetheless, it has been reported that the extreme genetic
582 bottleneck induced by this isolation technique can cause severe attenuation of mortality
583 in mice, even the complete loss of *in vitro* replicative capacity [39]. Characterizing the
584 biological phenotype of these plaque-isolated subclones is essential to ensure that the
585 genetic bottleneck exerted by the experimental approach has not altered the virus
586 biology. The measurement of these phenotypes may include classical measures, such

587 as plaque morphology or replication kinetics, and *in vivo* measures of pathogenesis. This
588 connection between comparative genomic studies to the measurement of biological
589 phenotypes is critical to integrate previous phenotypic effects of over-expression,
590 deletion, and modifications of defined loci with the new insights from comparative
591 genomic studies.
592 From a technical perspective, this study highlights the importance of ultra-deep
593 sequencing for the identification of *de novo* genetic diversity, since greater coverage
594 depth is directly translated into higher confidence when identifying MVs. Only a few
595 complete rounds of viral replication occur in a single passage in cell culture, where *de*
596 *novo* MVs would be represented with a very low frequency within the replicating
597 population. If these new variants are not beneficial or are just tolerable in terms of viral
598 fitness, they might not increase their frequency enough in the population to be detectable
599 by standard-deep sequencing, which average coverage ranges between 100-1000X [30,
600 31, 33, 34, 37]. Those and other studies generally reported only MVs above 2%,
601 accepted as the minimum frequency threshold cut-off. If so, with a 100X coverage, a 2%
602 threshold for MVs calling would mean that only two sequencing reads were required to
603 detect the minor variant. In order to be able to lower this threshold to 1%, we increased
604 the coverage cut-off greater than 200X. On top of these, we implemented an additional
605 filter to allow the call of those MVs with high frequency but below the 200X coverage
606 threshold, proportionally increasing the coverage needed for positive filtering as lower
607 the frequency was. These comprehensive quality controls, together with the benefits
608 provided by an ultra-deep sequence coverage, are instrumental in detecting very low
609 MVs and minimizing the chance of calling false positives. The integration of comparative
610 genomic and reverse genetic approaches will improve our understanding of fundamental
611 aspects of HSV biology, where studying the phenotypic effect of *in vitro* and *in vivo*
612 generated variants can complement previous discoveries on gene roles, as well as
613 explaining or predicting clinical outcomes.

614 The human herpesviruses literature shows how researchers have used cell lines,
615 commonly Vero cells, to generate high titer stocks of both laboratory-adapted strains but
616 also to amplify scarce clinical samples to decipher the *in vitro* and *in vivo* aspects of HSV
617 biology. Most of the HSV comparative genomic studies have also used *in vitro*
618 amplification to generate a high yield of viral genomic DNA for the preparation of high-
619 quality sequencing libraries. Other authors pointed out that growing HSV in cell culture
620 is clearly different from how the virus would replicate within its human host, where
621 selective pressures and genetic bottlenecks must be substantially different between
622 these two replicative scenarios [37, 40, 51]. However, here for the first time, we have
623 identified and characterized how the genetic diversity is differentially generated between
624 human herpesviruses when serially passaged in cell culture. The effects that this
625 differentially generated genetic diversity may have had on each aspect of HSV biology,
626 as well as on the clinical outcomes of infections, is currently an active area of research
627 [52].
628 Alphaherpesviruses are no longer seen as a static and homogeneous population but as
629 such presenting *in vivo* heterogeneous diversity [34, 53]. HSV-1 and HSV-2 exhibit a
630 remarkably unequal global seroprevalence and preference for infecting different
631 anatomical areas, where each cellular environment may have exerted differential
632 selective pressures over viral replication, latency and reactivation. These unique
633 selective pressures, found at each epithelial and neuronal tissue of the oronasal and
634 genital areas, might have contributed to their evolutionary divergence and differential
635 genetic drift rate. A better understanding of how human herpesviruses mutate during
636 each phase of their life cycle will provide a better knowledge on sequence determinants
637 of virulence factors and will help to monitor resistance to anti-viral drugs.
638
639
640
641

642 **MATERIALS AND METHODS**

643

644 **Cells and viruses**

645 Vero (*Cercopithecus aethiops* kidney epithelial) cells (ATCC, CCL-81) and HaCaT
646 (human epithelial keratinocytes) cells (Section of Virology, Department of Infectious
647 Disease, Imperial College London [54]) were maintained at 37ºC with 5 percent CO_2.
648 Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with
649 5% (v/v) Fetal Bovine Serum (FBS), 2 mM L-glutamine, and antibiotics (75 µg/ml
650 penicillin, 75 U/ml streptomycin, and 25 µg/ml gentamycin). Cells were regularly tested
651 for mycoplasma contaminations by standard PCR with primers Myco_Fw
652 (GGCGAATGGCTGAGTAACACG) and Myco_Rv (CGGATAACGCTTGCGACCTAT).
653 HSV-1 strain SC16 and HSV-2 strain 333 original stocks were kindly provided by Dr.
654 Helena Browne, University of Cambridge (UK). The genome sequence of plaque purified
655 viral clones from original stocks are GenBank available under accession no. KX946970
656 for HSV-1 strain SC16 [42], and under accession no. LS480640 for HSV-2 strain 333
657 [41].

658

659 **High-throughput sequencing**

660 High-throughput sequencing was performed in a similar manner as previously described
661 [41, 42]. Briefly, viral DNA was prepared by infection of one confluent P150-cm^2 plate of
662 Vero cells (MOI = 5 PFU/cell). Cells and supernatant were collected when reaching 90-
663 100% of cytopathic effect. Viral nucleocapsids were extracted by mechanical disruption
664 of the cellular pellet and clarified by cellular debris after 10 min of centrifugation at 300 x
665 *g*. Viral particles were treated with DNAse I, RNAse A, and nuclease S7 to eliminate
666 remaining cellular DNA/RNA, and nuclease activity was then inactivated with EDTA-
667 EGTA. Nucleocapsids were then lysed using sodium dodecyl sulfate and Proteinase K,
668 and viral genomic DNA was purified using phenol-chloroform-isoamyl alcohol. Potential
669 contaminating DNA was checked by PCR against mycoplasma, prokaryotic 16S rRNA

670 (primers 16S_Fw: CCTACGGGNBGCASCAG, and 16S_Rv:
671 GACTACNVGGGTATCTAATCC) and eukaryotic 18S rRNA (primers 18S_Fw:
672 GCCAGCAVCYGCGGTAAY, and 18S_Rv: CCGTCAATTHCTTYAART) [55, 56].
673 Finally, viral DNA was tested by PCR to determine HSV-type cross-contamination
674 (primers Up_US4(1)_Fw: AGCGCCGTTGACTACATTCAC, and Dw_US4(1)_Rv:
675 GCGCACCGGTGATTTATACCA, for HSV-1; Up_US4(2)_Fw:
676 TCTTGAGCGCCATCGACTACG, and Dw_US4(2)_Rv:
677 CCGCTCCATAGCTGCTGTACC, for HSV-2). An aliquot of viral genomic DNA (100 ng)
678 was submitted to MicrobesNG, University of Birmingham (UK), to prepare barcode
679 sequencing libraries, according to the NEBNext Ultra DNA Library Prep kit instructions
680 (New England Biolabs). Libraries were quantified by Qubit (Invitrogen, CA), assessed by
681 Bioanalyzer (Agilent), and library adapter qPCR (KAPA Biosystems). Sequencing was
682 performed on an Illumina MiSeq device as paired-end reads (2 x 250 bp), according to
683 the manufacturer's recommendations. Sequencing statistics for every sample used in
684 this study can be found in S1 Table.
685
686 **Selection of viral clones from original stocks**
687 Five plaque-purified viral clones from both HSV-1 (SC16) and HSV-2 (333) original
688 stocks were isolated by plaque isolation. Vero cell monolayers with 5 x 10^5 cells/well in
689 6-well plates were infected at an MOI of 0.01 PFU/cell. After 48 hours post-infection (hpi),
690 defined and single viral plaques were carefully isolated by fine pipetting 10 µl of media
691 containing the selected plaque. Then, 30 µl of fresh media were added to each isolated
692 plaque, followed by three rounds of freezing and thawing. An aliquot of 1 µl was three
693 times serially diluted to infect fresh Vero cells monolayer in 6-well plates. After five
694 subsequent rounds, one confluent P150-cm^2 plate of Vero cells was infected in order to
695 produce a viral stock for sequencing and subsequent infections.
696
697 ***In vitro* generation of genetic variability experiments**

698 Two selected plaque-purified clones from HSV-1 and HSV-2 original stocks were used
699 to infect a P60-cm^2 plate of Vero and HaCaT cells separately at an MOI of 0.1 PFU/cell.
700 After 2 hpi, the viral inoculum was removed and fresh DMEM with 2% FBS was added.
701 Forty-eight hpi viruses were harvested by collecting both cells and supernatant, followed
702 by three freezing and thawing cycles. Each cycle of infection and harvest was considered
703 a passage. The harvested viruses were then used to infect the next fresh plate of Vero
704 or HaCaT cells (estimated MOI of 0.1 PFU/cell). Each selected clone was passaged ten
705 times in each cell line. Passages from 4th to 5th and 9th to 10th were made by infecting
706 one confluent P150-cm^2 plate of corresponding cells (adjusting the MOI), in order to
707 obtain higher yields of viral DNA for sequencing.
708
709 **Genetic variant analysis and identification of *de novo* variants**
710 Reads from each sequenced sample were trimmed using Trimmomatic v0.36 [57],
711 quality-filtered with PrinSeq v1.2 [58], and aligned against the reference sequence for
712 each case, by using Bowtie 2 v2.3.4.1 [59], with default settings. Alignments were
713 visualized using Integrative Genomics Viewer v2.8.2 [60] to detect large gaps and
714 rearrangements. MVs present in each sequenced viral population were detected by
715 using VarScan v2.4.3 [61], with settings intended to minimize sequencing-induced errors
716 from the raw calling of MVs: minimum variant allele frequency ≥0.01 (1%); minimum
717 coverage ≥20, base call quality ≥20, exclusion of variants supported on one strand by
718 >90 percent. Detected MVs from VarScan calling were then annotated onto their
719 corresponding genome to determine their mutational effects. MVs were then additionally
720 filtered by coverage >200. An additional filter was implemented in order to detect those
721 MVs with high frequency but low coverage, where read depth at the given position had
722 to be greater than the product obtained from dividing 200 (coverage threshold) by the
723 variant frequency (0-100) at the given position:

724 $$read\ depth\ at\ the\ given\ position > \left(\frac{200\ (Coverage\ threshold)}{frequency\ (0-100)}\right)$$

725 MVs were considered as *de novo* appearance when, after coverage filtering, its
726 frequency in the previous parental viral population was inexistent or < 0.01. Coverage
727 plots for each alignment, as well as detected MVs in each viral population, were
728 represented across their corresponding genome, according to their location and
729 frequency (S1-S3 and S5-S9 Figs). For a summary list of detected, filtered, and
730 categorized MVs for every sample sequenced in this study, see S2 Table. For a full list
731 of MVs detected in each viral population, see S3 Table (HSV-1 original stock and isolated
732 clones), S4 Table (HSV-2 original stock and isolated clones), S5 Table (ultra-deep
733 sequencing of HSV-1 clones 2 and 3), S6 Table (ultra-deep sequencing of HSV-2 clones
734 2 and 3), S7 and S8 Tables (HSV-1 clone 2 in Vero and HaCaT cells, respectively), S9
735 and S10 Tables (HSV-1 clone 3 in Vero and HaCaT cells, respectively), S11 and S12
736 Tables (HSV-2 clone 1 in Vero and HaCaT cells, respectively), and S13 and S14 Tables
737 (HSV-2 clone 5 in Vero and HaCaT cells, respectively).
738
739 **Statistical analysis**
740 All statistical analyses were performed using GraphPad Prism 8 (v8.4.3) software. Two-
741 tailed Mann–Whitney *U*-test was used with the number of MVs analyses ($p < 0.05$).
742
743 **Data availability**
744 Consensus genomes of a plaque purified viral clone from each original stock were
745 retrieved from GenBank as follows: KX946970 for HSV-1 strain SC16; LS480640 for
746 HSV-2 strain 333. Raw sequence reads are available at the European Bioinformatics
747 Institute (EMBL-EBI) European Nucleotide Archive (ENA) as Bioproject ID PRJEB32133
748 and PRJEB32148.
749
750
751
752

ACKNOWLEDGMENTS

We thank Anthony Minson and Helena Brown (University of Cambridge, UK) for kindly providing the HSV-1 strain SC16 and HSV-2 strain 333 viral stocks, respectively. We are grateful to the Genomics and Next Generation Sequencing Service at Centro de Biología Molecular Severo Ochoa for their support and advice. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant no. BB/L024209/1).

781 **REFERENCES**

782

783 1. James C, Harfouche M, Welton NJ, Turner KM, Abu-Raddad LJ, Gottlieb SL, et
784 al. Herpes simplex virus: global infection prevalence and incidence estimates, 2016. Bull
785 World Health Organ. 2020;98(5):315-29. doi: 10.2471/BLT.19.237149. PubMed PMID:
786 32514197.

787 2. Forward KR, Lee SH. Predominance of herpes simplex virus type 1 from patients
788 with genital herpes in Nova Scotia. Can J Infect Dis. 2003;14(2):94-6. doi:
789 10.1155/2003/168673. PubMed PMID: 18159431.

790 3. Wald A, Ericsson M, Krantz E, Selke S, Corey L. Oral shedding of herpes simplex
791 virus type 2. Sex Transm Infect. 2004;80(4):272-6. doi: 10.1136/sti.2003.007823.
792 PubMed PMID: 15295123.

793 4. Gupta R, Warren T, Wald A. Genital herpes. Lancet. 2007;370(9605):2127-37.
794 Epub 2007/12/25. doi: 10.1016/S0140-6736(07)61908-4. PubMed PMID: 18156035.

795 5. Ayoub HH, Chemaitelly H, Abu-Raddad LJ. Characterizing the transitioning
796 epidemiology of herpes simplex virus type 1 in the USA: model-based predictions. BMC
797 Med. 2019;17(1):57. doi: 10.1186/s12916-019-1285-x. PubMed PMID: 30853029.

798 6. Knipe DM, Cliffe A. Chromatin control of herpes simplex virus lytic and latent
799 infection. Nat Rev Microbiol. 2008;6(3):211-21. doi: 10.1038/nrmicro1794. PubMed
800 PMID: 18264117.

801 7. Kawaguchi Y, Mori Y, Kimura H. Human Herpesviruses: Springer; 2018.

802 8. Roizman B, D.M. Knipe, and R. Whitley. Herpes Simplex Viruses. Fields
803 Virology. 6th ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2013. p. 1823-97.

804 9. Kennedy PG, Rovnak J, Badani H, Cohrs RJ. A comparison of herpes simplex
805 virus type 1 and varicella-zoster virus latency and reactivation. J Gen Virol. 2015;96(Pt
806 7):1581-602. doi: 10.1099/vir.0.000128. PubMed PMID: 25794504.

807 10. Lieberman PM. Epigenetics and Genetics of Viral Latency. Cell Host Microbe.
808 2016;19(5):619-28. doi: 10.1016/j.chom.2016.04.008. PubMed PMID: 27173930.

809 11. Drake JW, Hwang CB. On the mutation rate of herpes simplex virus type 1.
810 Genetics. 2005;170(2):969-70. doi: 10.1534/genetics.104.040410. PubMed PMID:
811 15802515.
812 12. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J
813 Virol. 2010;84(19):9733-48. doi: 10.1128/JVI.00694-10. PubMed PMID: 20660197.
814 13. McCrone JT, Lauring AS. Genetic bottlenecks in intraspecies virus transmission.
815 Curr Opin Virol. 2018;28:20-5. doi: 10.1016/j.coviro.2017.10.008. PubMed PMID:
816 29107838.
817 14. Mahiet C, Ergani A, Huot N, Alende N, Azough A, Salvaire F, et al. Structural
818 variability of the herpes simplex virus 1 genome in vitro and in vivo. J Virol.
819 2012;86(16):8592-601. doi: 10.1128/JVI.00223-12. PubMed PMID: 22674981.
820 15. Chou J, Roizman B. Isomerization of herpes simplex virus 1 genome:
821 identification of the cis-acting and recombination sites within the domain of the a
822 sequence. Cell. 1985;41(3):803-11. doi: 10.1016/s0092-8674(85)80061-1. PubMed
823 PMID: 2988789.
824 16. Huang KJ, Zemelman BV, Lehman IR. Endonuclease G, a candidate human
825 enzyme for the initiation of genomic inversion in herpes simplex type 1 virus. J Biol
826 Chem. 2002;277(23):21071-9. doi: 10.1074/jbc.M201785200. PubMed PMID:
827 11912214.
828 17. Muylaert I, Tang KW, Elias P. Replication and recombination of herpes simplex
829 virus DNA. J Biol Chem. 2011;286(18):15619-24. doi: 10.1074/jbc.R111.233981.
830 PubMed PMID: 21362621.
831 18. Norberg P, Kasubi MJ, Haarr L, Bergstrom T, Liljeqvist JA. Divergence and
832 recombination of clinical herpes simplex virus type 2 isolates. J Virol.
833 2007;81(23):13158-67. doi: 10.1128/jvi.01310-07. PubMed PMID: 17881457.
834 19. Lee K, Kolb AW, Sverchkov Y, Cuellar JA, Craven M, Brandt CR. Recombination
835 Analysis of Herpes Simplex Virus 1 Reveals a Bias toward GC Content and the Inverted

836 Repeat Regions. J Virol. 2015;89(14):7214-23. doi: 10.1128/JVI.00880-15. PubMed
837 PMID: 25926637.
838 20. Koelle DM, Norberg P, Fitzgibbon MP, Russell RM, Greninger AL, Huang ML, et
839 al. Worldwide circulation of HSV-2 x HSV-1 recombinant strains. Sci Rep. 2017;7:44084.
840 doi: 10.1038/srep44084. PubMed PMID: 28287142.
841 21. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, et al.
842 Evolution and diversity in human herpes simplex virus genomes. J Virol.
843 2014;88(2):1209-27. doi: 10.1128/JVI.01987-13. PubMed PMID: 24227835.
844 22. Hall JD, Almy RE. Evidence for control of herpes simplex virus mutagenesis by
845 the viral DNA polymerase. Virology. 1982;116(2):535-43. doi: 10.1016/0042-
846 6822(82)90146-5. PubMed PMID: 6278726.
847 23. Brown J. Effect of gene location on the evolutionary rate of amino acid
848 substitutions in herpes simplex virus proteins. Virology. 2004;330(1):209-20. doi:
849 10.1016/j.virol.2004.09.020. PubMed PMID: 15527847.
850 24. Parris DS, Harrington JE. Herpes simplex virus variants restraint to high
851 concentrations of acyclovir exist in clinical isolates. Antimicrob Agents Chemother.
852 1982;22(1):71-7. doi: 10.1128/aac.22.1.71. PubMed PMID: 6289742.
853 25. Sarisky RT, Nguyen TT, Duffy KE, Wittrock RJ, Leary JJ. Difference in incidence
854 of spontaneous mutations between herpes simplex virus types 1 and 2. Antimicrob
855 Agents Chemother. 2000;44(6):1524-9. doi: 10.1128/aac.44.6.1524-1529.2000.
856 PubMed PMID: 10817703.
857 26. Burrel S, Deback C, Agut H, Boutolleau D. Genotypic characterization of UL23
858 thymidine kinase and UL30 DNA polymerase of clinical isolates of herpes simplex virus:
859 natural polymorphism and mutations associated with resistance to antivirals. Antimicrob
860 Agents Chemother. 2010;54(11):4833-42. doi: 10.1128/AAC.00669-10. PubMed PMID:
861 20733037.

862 27. Sauerbrei A, Deinhardt S, Zell R, Wutzler P. Phenotypic and genotypic
863 characterization of acyclovir-resistant clinical isolates of herpes simplex virus. Antiviral
864 Res. 2010;86(3):246-52. doi: 10.1016/j.antiviral.2010.03.002. PubMed PMID: 20211650.
865 28. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive
866 genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS
867 Pathog. 2011;7(5):e1001344. doi: 10.1371/journal.ppat.1001344. PubMed PMID:
868 21625576.
869 29. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci.
870 2016;73(23):4433-48. doi: 10.1007/s00018-016-2299-6. PubMed PMID: 27392606.
871 30. Bowen CD, Paavilainen H, Renner DW, Palomaki J, Lehtinen J, Vuorinen T, et
872 al. Comparison of Herpes Simplex Virus 1 Strains Circulating in Finland Demonstrates
873 the Uncoupling of Whole-Genome Relatedness and Phenotypic Outcomes of Viral
874 Infection. J Virol. 2019;93(8). doi: 10.1128/JVI.01824-18. PubMed PMID: 30760568.
875 31. Akhtar LN, Bowen CD, Renner DW, Pandey U, Della Fera AN, Kimberlin DW, et
876 al. Genotypic and Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected
877 Neonatal Population. mSphere. 2019;4(1). doi: 10.1128/mSphere.00590-18. PubMed
878 PMID: 30814317.
879 32. Depledge DP, Gray ER, Kundu S, Cooray S, Poulsen A, Aaby P, et al. Evolution
880 of cocirculating varicella-zoster virus genotypes during a chickenpox outbreak in Guinea-
881 Bissau. J Virol. 2014;88(24):13936-46. doi: 10.1128/JVI.02337-14. PubMed PMID:
882 25275123.
883 33. Hage E, Wilkie GS, Linnenweber-Held S, Dhingra A, Suarez NM, Schmidt JJ, et
884 al. Characterization of Human Cytomegalovirus Genome Diversity in
885 Immunocompromised Hosts by Whole-Genome Sequencing Directly From Clinical
886 Specimens. J Infect Dis. 2017;215(11):1673-83. doi: 10.1093/infdis/jix157. PubMed
887 PMID: 28368496.
888 34. Shipley MM, Renner DW, Ott M, Bloom DC, Koelle DM, Johnston C, et al.
889 Genome-Wide Surveillance of Genital Herpes Simplex Virus Type 1 From Multiple

890 Anatomic Sites Over Time. J Infect Dis. 2018;218(4):595-605. doi: 10.1093/infdis/jiy216.
891 PubMed PMID: 29920588.
892 35. Renner DW, Szpara ML. Impacts of Genome-Wide Analyses on Our
893 Understanding of Human Herpesvirus Diversity and Evolution. J Virol. 2018;92(1). doi:
894 10.1128/jvi.00908-17. PubMed PMID: 29046445.
895 36. Mosca JD, Pitha PM. Transcriptional and posttranscriptional regulation of
896 exogenous human beta interferon gene in simian cells defective in interferon synthesis.
897 Mol Cell Biol. 1986;6(6):2279-83. doi: 10.1128/mcb.6.6.2279. PubMed PMID: 3785197.
898 37. Kuny CV, Bowen CD, Renner DW, Johnston CM, Szpara ML. In vitro evolution
899 of herpes simplex virus 1 (HSV-1) reveals selection for syncytia and other minor variants
900 in cell culture. Virus Evol. 2020;6(1):veaa013. doi: 10.1093/ve/veaa013. PubMed PMID:
901 32296542.
902 38. Szpara ML, Parsons L, Enquist LW. Sequence variability in clinical and laboratory
903 isolates of herpes simplex virus 1 reveals new mutations. J Virol. 2010;84(10):5303-13.
904 doi: 10.1128/jvi.00312-10. PubMed PMID: 20219902.
905 39. Jaramillo N, Domingo E, Munoz-Egea MC, Tabares E, Gadea I. Evidence of
906 Muller's ratchet in herpes simplex virus type 1. J Gen Virol. 2013;94(Pt 2):366-75. doi:
907 10.1099/vir.0.044685-0. PubMed PMID: 23100362.
908 40. Parsons LR, Tafuri YR, Shreve JT, Bowen CD, Shipley MM, Enquist LW, et al.
909 Rapid genome assembly and comparison decode intrastrain variation in human
910 alphaherpesviruses. MBio. 2015;6(2). doi: 10.1128/mBio.02213-14. PubMed PMID:
911 25827418.
912 41. Lopez-Munoz AD, Rastrojo A, Alcami A. Complete Genome Sequence of Herpes
913 Simplex Virus 2 Strain 333. Microbiol Resour Announc. 2018;7(9). doi:
914 10.1128/MRA.00870-18. PubMed PMID: 30533931.
915 42. Rastrojo A, Lopez-Munoz AD, Alcami A. Genome Sequence of Herpes Simplex
916 Virus 1 Strain SC16. Genome Announc. 2017;5(4). doi: 10.1128/genomeA.01392-16.
917 PubMed PMID: 28126930.

918 43. Bzik DJ, Fox BA, DeLuca NA, Person S. Nucleotide sequence of a region of the
919 herpes simplex virus type 1 gB glycoprotein gene: mutations affecting rate of virus entry
920 and cell fusion. Virology. 1984;137(1):185-90. doi: 10.1016/0042-6822(84)90022-9.
921 PubMed PMID: 6089415.
922 44. Engel JP, Boyer EP, Goodman JL. Two novel single amino acid syncytial
923 mutations in the carboxy terminus of glycoprotein B of herpes simplex virus type 1 confer
924 a unique pathogenic phenotype. Virology. 1993;192(1):112-20. doi:
925 10.1006/viro.1993.1013. PubMed PMID: 8390747.
926 45. Diakidi-Kosta A, Michailidou G, Kontogounis G, Sivropoulou A, Arsenakis M. A
927 single amino acid substitution in the cytoplasmic tail of the glycoprotein B of herpes
928 simplex virus 1 affects both syncytium formation and binding to intracellular heparan
929 sulfate. Virus Res. 2003;93(1):99-108. doi: 10.1016/s0168-1702(03)00070-4. PubMed
930 PMID: 12727347.
931 46. Abaitua F, Zia FR, Hollinshead M, O'Hare P. Polarized cell migration during cell-
932 to-cell transmission of herpes simplex virus in human skin keratinocytes. J Virol.
933 2013;87(14):7921-32. doi: 10.1128/jvi.01172-13. PubMed PMID: 23658449.
934 47. Coller KE, Smith GA. Two viral kinases are required for sustained long distance
935 axon transport of a neuroinvasive herpesvirus. Traffic. 2008;9(9):1458-70. doi:
936 10.1111/j.1600-0854.2008.00782.x. PubMed PMID: 18564370.
937 48. Gershburg S, Geltz J, Peterson KE, Halford WP, Gershburg E. The UL13 and
938 US3 Protein Kinases of Herpes Simplex Virus 1 Cooperate to Promote the Assembly
939 and Release of Mature, Infectious Virions. PLoS One. 2015;10(6):e0131420. doi:
940 10.1371/journal.pone.0131420. PubMed PMID: 26115119.
941 49. Duffy KE, Quail MR, Nguyen TT, Wittrock RJ, Bartus JO, Halsey WM, et al.
942 Assessing the contribution of the herpes simplex virus DNA polymerase to spontaneous
943 mutations. BMC Infect Dis. 2002;2:7. doi: 10.1186/1471-2334-2-7. PubMed PMID:
944 12019036.

945 50. Bowen CD, Renner DW, Shreve JT, Tafuri Y, Payne KM, Dix RD, et al. Viral
946 forensic genomics reveals the relatedness of classic herpes simplex virus strains KOS,
947 KOS63, and KOS79. Virology. 2016;492:179-86. doi: 10.1016/j.virol.2016.02.013.
948 PubMed PMID: 26950505.
949 51. Wilkinson GW, Davison AJ, Tomasec P, Fielding CA, Aicheler R, Murrell I, et al.
950 Human cytomegalovirus: taking the strain. Medical microbiology and immunology.
951 2015;204(3):273-84. doi: 10.1007/s00430-015-0411-4. PubMed PMID: 25894764.
952 52. Kuny CV, Szpara ML. Alphaherpesvirus Genomics: Past, Present and Future.
953 Curr Issues Mol Biol. 2020;42:41-80. doi: 10.21775/cimb.042.041. PubMed PMID:
954 33159012.
955 53. Minaya MA, Jensen TL, Goll JB, Korom M, Datla SH, Belshe RB, et al. Molecular
956 Evolution of Herpes Simplex Virus 2 Complete Genomes: Comparison between Primary
957 and Recurrent Infections. J Virol. 2017;91(23). doi: 10.1128/jvi.00942-17. PubMed
958 PMID: 28931680.
959 54. Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig
960 NE. Normal keratinization in a spontaneously immortalized aneuploid human
961 keratinocyte cell line. J Cell Biol. 1988;106(3):761-71. PubMed PMID: 2450098.
962 55. Hugerth LW, Muller EE, Hu YO, Lebrun LA, Roume H, Lundin D, et al. Systematic
963 design of 18S rRNA gene primers for determining eukaryotic diversity in microbial
964 consortia. PLoS One. 2014;9(4):e95567. doi: 10.1371/journal.pone.0095567. PubMed
965 PMID: 24755918.
966 56. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a
967 prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using
968 next-generation sequencing. PLoS One. 2014;9(8):e105592. doi:
969 10.1371/journal.pone.0105592. PubMed PMID: 25144201.
970 57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
971 sequence data. Bioinformatics. 2014;30(15):2114-20. doi:
972 10.1093/bioinformatics/btu170. PubMed PMID: 24695404.

58. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863-4. doi: 10.1093/bioinformatics/btr026. PubMed PMID: 21278185.

59. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286.

60. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. doi: 10.1038/nbt.1754. PubMed PMID: 21221095.

61. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76. doi: 10.1101/gr.129684.111. PubMed PMID: 22300766.

1001 **FIGURE CAPTIONS**

1002

1003 **Fig 1. Experimental design of the *in vitro* generation of genetic variability studies**
1004 **for HSV-1 and HSV-2 subtypes and detected changes in plaque phenotype.** (A)
1005 Five viral clones from each original stock were five times plaque-purified in Vero cells
1006 and then deep sequenced. Two clones were re-sequenced at ultrahigh depth, whose
1007 replication, infectivity, and pathogenesis were compared to their corresponding
1008 parental stocks in cell culture and animal models of infection (S4 Fig). Those two
1009 clones were used to infect Vero and HaCaT cells at an MOI of 0.1 PFU/cell. After 48
1010 hpi, viral progenies were harvested, referring to this infection cycle as a passage. Viral
1011 populations from each plaque-purified clone were ultra-deep sequenced after five and
1012 ten passages in each cell line. (B) Four representative pictures are showing the plaque
1013 morphology phenotype of HSV-1 clone 3 in Vero cells (48 hpi), before (passage 0, P0),
1014 and after ten passages (P10) in Vero cells. This syncytial plaque phenotype is due to
1015 the previously well-described syncytia-inducing mutation in UL27 CDS (see S9 Table,
1016 variant #38: R858H) [37]. Tiled images (4 x 4) were taken using a Leica DM IL LED
1017 inverted microscope equipped with a Leica DFC3000-G digital camera. Scale bars
1018 indicate 100 μm.

1019

1020 **Fig 2. Genetic diversity in viral populations from original stocks and five times**
1021 **plaque-isolated HSV-1 and HSV-2 clones.** (A) Total number of MVs observed in each
1022 original stock, at a frequency equals or above a 1% limit of detection (see Materials
1023 and Methods for details). The total number of MVs (*y*-axis) is separated into variant
1024 type and genomic location (*x*-axis). Variant type distinguishes between SNPs and
1025 InDels, discriminating nonsynonymous SNPs. The genomic location of each variant is
1026 categorized as non-coding or coding region. (B) Histograms show the number of MVs
1027 in each frequency range for each original stock. The frequency of each variant was
1028 examined and grouped in shown ranges (e.g., 10% to <20% frequency, 20% to <30%

1029 frequency and so on). SNPs and InDels were combined for this analysis. (C) Total
1030 number of MVs observed in each HSV-1 (left, blue bars) and HSV-2 (right, red bars)
1031 clones after five rounds of plaque-isolation, categorized by nonconservative changes
1032 (SNPs and InDels). See S3 and S4 Tables for a full list of MVs position and frequency
1033 data. (D) Total number of variants detected in each clone are grouped by subtype and
1034 graphed showing mean +/- SEM (ns = not significant $p > 0.05$ by two-tailed Mann–
1035 Whitney *U*-test).
1036
1037 **Fig 3. Comparative variant analysis from standard- and ultra-deep sequencing**
1038 **data of two plaque-isolated HSV-1 and HSV-2 clones.** (A) Histograms show the
1039 average depth of coverage per genomic position of reads alignments from standard-
1040 deep sequencing (SDS, white bars) and ultra-deep sequencing (UDS, black bars) for
1041 each viral clone. (B) Histograms bar plot total number of MVs observed after variant
1042 analysis of SDS and UDS data in each viral clone, discriminating between preexisting
1043 MVs found in the corresponding original stock (white) and *de novo* appearance (red).
1044 (C) The number of total, preexisting, and *de novo* MVs detected in each clone from
1045 SDS, and UDS data are grouped and graphed (blue shapes for HSV-1, red for HSV-2
1046 clones), showing mean +/- SEM (ns $p > 0.05$, * $p = 0.029$ by two-tailed Mann–Whitney
1047 *U*-test). (D) Number of *de novo* MVs observed after variant analysis of SDS and UDS
1048 data in each viral clone, stacked by frequency ranges. SNPs and InDels were
1049 combined for this analysis. (E) *de novo* MVs with a frequency between 1% to 2%
1050 detected in each clone from SDS, and UDS data are grouped and graphed (blue
1051 shapes for HSV-1, red for HSV-2 clones), showing mean +/- SEM (* $p = 0.029$ by two-
1052 tailed Mann–Whitney *U*-test). See S5 and S6 Tables for a full list of MVs position and
1053 frequency data.
1054
1055 **Fig 4. Comparison of *de novo* generation of total and nonconservative genetic**
1056 **diversity between HSV-1 and HSV-2 purified clones, after five (P5) and ten (P10)**

1057 **passages in Vero and HaCaT cells.** (A) Total number of MVs are plotted according to
1058 variant analysis data for each clone, passage, and cell line, differentiating between
1059 preexisting MVs observed in the corresponding passage zero (P0) in white, and *de*
1060 *novo* generated MVs in red. (B) Number of *de novo* SNPs detected in each viral
1061 population, categorized by nonsynonymous (black) and synonymous/non-coding
1062 changes (white). (C) Number of de novo InDels detected among each viral population
1063 are stacked by their location impacting coding regions (black) or non-coding regions
1064 (white). (D) *De novo* MVs are stacked by frequency ranges. SNPs and InDels were
1065 combined for this analysis. See S7-S14 Tables for a full list of MVs position and
1066 frequency data.
1067
1068 **Fig 5. Stacked histograms show the number of *de novo* genic MVs (*x*-axis)**
1069 **located in each HSV-1 (left) and HSV-2 (right) coding sequence (gene; *y*-axis),**
1070 **after five (P5) and ten (P10) passages in Vero and HaCaT cells.** Only coding
1071 sequences registering at least one variant are included in the histogram. MVs found in
1072 both copies of each RL1, RL2, and RS1 coding sequences are listed together. SNPs
1073 and InDels were combined for this analysis. See S7-S14 Tables for a full list of MVs
1074 position and frequency data.
1075
1076 **Fig 6. Dynamics of nonconservative *de novo* variants in each HSV-1 (left) and**
1077 **HSV-2 purified populations (right), whose frequency increased over sequential**
1078 **passages in cell culture.** Nonsynonymous *de novo* SNPs were plotted by their
1079 frequency in the sequenced viral population after five (P5) and ten (P10) passages in
1080 Vero and HaCaT cells. SNPs, their encoded proteins (bold), as well as the change that
1081 they would cause in the translated protein (italic), are listed in the legend according to
1082 their frequency at P5 and P10.
1083
1084

1085 **SUPPORTING INFORMATION CAPTIONS**

1086

1087 **S1 Fig. Schematic of the HSV-1 strain SC16 (A) and HSV-2 strain 333 (B)**
1088 **sequenced genomes from original stocks.** Each CDS is presented in forward (red)
1089 or reverse (blue) orientation. Detected MVs are mapped as black (not *de novo*) or red
1090 (*de novo*) dots across the genome, according to their location (*x*-axis) and frequency
1091 (*y*-axis). GC% plots (purple lines) and coverage plots from data alignments
1092 (blue/orange profiles) have also been mapped across each genome.

1093

1094 **S2 Fig. Variant analysis of HSV-1 plaque-purified clones.** Coverage plots from data
1095 alignments are represented in blue, for each individual case. Detected MVs are
1096 mapped as black (not *de novo*) or red (*de novo*) dots across the genome, according to
1097 their location (*x*-axis) and frequency (*y*-axis). MVs were considered as *de novo* when
1098 these were not previously found in the original stock (see Materials and Methods for
1099 details).

1100

1101 **S3 Fig. Variant analysis of HSV-2 plaque-purified clones.** Coverage plots from data
1102 alignments are represented in orange, for each individual case. Detected MVs are
1103 mapped as black (not *de novo*) or red (*de novo*) dots across the genome, according to
1104 their location (*x*-axis) and frequency (*y*-axis). MVs were considered as *de novo* when
1105 these were not previously found in the original stock.

1106

1107 **S4 Fig. Replication kinetics and pathogenesis of HSV-1 and HSV-2 plaque-**
1108 **isolated clones compared to their corresponding original stocks.** (A) Vero cells
1109 were infected with the indicated viruses at high MOI (5 PFU/cell) for one-step growth
1110 curves, and at low MOI (0.01 PFU/cell) for multi-step growth curves. Virus titers from
1111 fractions containing cell-associated virus were determined by plaque assay at 24 hpi in
1112 the one-step curves, and at the indicated times in the multi-step curves. (B) Female

1113 BALB/c mice (n=5) were infected with the indicated virus and dose, by intranasal (i.n.)
1114 or intravaginal (i.v.) inoculations. Mice were monitored daily for survival, body weight,
1115 and signs of illness. Weight data are expressed as the mean +/- SEM of the five animal
1116 weights compared to their original weight on the day of inoculation. Signs of illness, as
1117 a score ranged from 1 to 4, is also expressed as the mean +/- SEM of the five animals.
1118 A colored "1" indicates thereafter only one animal remained in that group. Statistical
1119 analysis was performed for bodyweight data, using multiple *t*-tests with Sidak-
1120 Bonferroni correction ($p < 0.05$).
1121
1122 **S5 Fig. Variant analysis of HSV-1 plaque-purified clones 2 and 3 (A) and HSV-2**
1123 **clones 1 and 5 (B) from high-depth sequencing data.** Coverage plots from
1124 alignments are represented in blue or orange, for each case. Detected MVs are
1125 mapped as black (not *de novo*) or red (*de novo*) dots across the genome, according to
1126 their location (*x*-axis) and frequency (*y*-axis). MVs were considered as *de novo* when
1127 these were not previously found in the corresponding original stock.
1128
1129 **S6 Fig. Variant analysis of HSV-1 plaque-purified clone 2, after 5 and 10**
1130 **passages in Vero and HaCaT cells.** Coverage plots from high-depth sequencing data
1131 alignments are represented in blue. Detected MVs are mapped as black (not *de novo*)
1132 or red (*de novo*) dots across the genome, according to their location (*x*-axis) and
1133 frequency (*y*-axis). Mutations from passage 0 were considered as *de novo* when these
1134 were not previously found in the original stock, whereas those from passage 5 and 10,
1135 regarding passage 0.
1136
1137 **S7 Fig. Variant analysis of HSV-1 plaque-purified clone 3, after 5 and 10**
1138 **passages in Vero and HaCaT cells.** Coverage plots from high-depth sequencing data
1139 alignments are represented in blue. Detected MVs are mapped as black (not *de novo*)
1140 or red (*de novo*) dots across the genome, according to their location (*x*-axis) and

1141 frequency (*y*-axis). Mutations from passage 0 were considered as *de novo* when these
1142 were not previously found in the original stock, whereas those from passage 5 and 10,
1143 regarding passage 0.
1144
1145 **S8 Fig. Variant analysis of HSV-2 plaque-purified clone 1, after 5 and 10**
1146 **passages in Vero and HaCaT cells.** Coverage plots from high-depth sequencing data
1147 alignments are represented in orange. Detected MVs are mapped as black (not *de*
1148 *novo*) or red (*de novo*) dots across the genome, according to their location (*x*-axis) and
1149 frequency (*y*-axis). Mutations from passage 0 were considered as *de novo* when these
1150 were not previously found in the original stock, whereas those from passage 5 and 10,
1151 regarding passage 0.
1152
1153 **S9 Fig. Variant analysis of HSV-2 plaque-purified clone 5, after 5 and 10**
1154 **passages in Vero and HaCaT cells.** Coverage plots from high-depth sequencing data
1155 alignments are represented in orange. Detected MVs are mapped as black (not *de*
1156 *novo*) or red (*de novo*) dots across the genome, according to their location (*x*-axis) and
1157 frequency (*y*-axis). Mutations from passage 0 were considered as *de novo* when these
1158 were not previously found in the original stock, whereas those from passage 5 and 10,
1159 regarding passage 0.
1160
1161 **S1 Animation. Variant analysis of HSV-1 plaque-purified clone 2, after 5 and 10**
1162 **passages in Vero and HaCaT cells.** See S5 Fig for additional details.
1163
1164 **S2 Animation. Variant analysis of HSV-1 plaque-purified clone 3, after 5 and 10**
1165 **passages in Vero and HaCaT cells.** See S6 Fig for additional details.
1166
1167 **S3 Animation. Variant analysis of HSV-2 plaque-purified clone 1, after 5 and 10**
1168 **passages in Vero and HaCaT cells.** See S7 Fig for additional details.

1169

1170 **S4 Animation. Variant analysis of HSV-2 plaque-purified clone 5, after 5 and 10**

1171 **passages in Vero and HaCaT cells.** See S8 Fig for additional details.

1172

1173 **S1 Table. Genome sequencing statistics for each sample sequenced in this**

1174 **study.**

1175 Legend: SRA (Sequence Read Archive), QF (quality-filtered).

1176

1177 **S2 Table. Categorized number of MVs for each sample sequenced in this study.**

1178 Legend: NCR (non-coding region), CR (coding region).

1179

1180 **S3 Table. List of detected MVs from deep sequencing of HSV-1 original stock and**

1181 **plaque-purified clones 1-5.**

1182 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1183 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1184 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1185

1186 **S4 Table. List of detected MVs from deep sequencing of HSV-2 original stock and**

1187 **plaque-purified clones 1-5.**

1188 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1189 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1190 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1191

1192 **S5 Table. List of detected MVs from ultra-deep sequencing of HSV-1 plaque-**

1193 **purified clones 2 and 3.**

1194 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1195 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1196 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1197

1198 **S6 Table. List of detected MVs from ultra-deep sequencing of HSV-2 plaque-**

1199 **purified clones 1 and 5.**

1200 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1201 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1202 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1203

1204 **S7 Table. List of detected MVs from ultra-deep sequencing of HSV-1 clone 2 at**

1205 **P0, P5, and P10 in Vero cells.**

1206 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1207 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1208 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1209

1210 **S8 Table. List of detected MVs from ultra-deep sequencing of HSV-1 clone 2 at**

1211 **P0 in Vero cells, and P5 and P10 in HaCaT cells.**

1212 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1213 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1214 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1215

1216 **S9 Table. List of detected MVs from ultra-deep sequencing of HSV-1 clone 3 at**

1217 **P0, P5, and P10 in Vero cells.**

1218 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS

1219 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total

1220 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).

1221

1222 **S10 Table. List of detected MVs from ultra-deep sequencing of HSV-1 clone 3 at**

1223 **P0 in Vero cells, and P5 and P10 in HaCaT cells.**

1224 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS
1225 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total
1226 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).
1227
1228 **S11 Table. List of detected MVs from ultra-deep sequencing of HSV-2 clone 1 at**
1229 **P0, P5, and P10 in Vero cells.**
1230 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS
1231 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total
1232 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).
1233
1234 **S12 Table. List of detected MVs from ultra-deep sequencing of HSV-2 clone 1 at**
1235 **P0 in Vero cells, and P5 and P10 in HaCaT cells.**
1236 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS
1237 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total
1238 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).
1239
1240 **S13 Table. List of detected MVs from ultra-deep sequencing of HSV-2 clone 5 at**
1241 **P0, P5, and P10 in Vero cells.**
1242 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS
1243 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total
1244 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).
1245
1246 **S14 Table. List of detected MVs from ultra-deep sequencing of HSV-2 clone 5 at**
1247 **P0 in Vero cells, and P5 and P10 in HaCaT cells.**
1248 Legend: Ref (reference allele), Var (variant allele), NCR (non-coding region), INS
1249 (insertion), DEL (deletion), *name*_freq (variant allele frequency), *name*_cov (total
1250 coverage), *name*_Ref (reference allele coverage), *name*_Var (variant allele coverage).
1251

1252 **S1 Text. Supporting Material and Methods for HSV replication kinetics and**

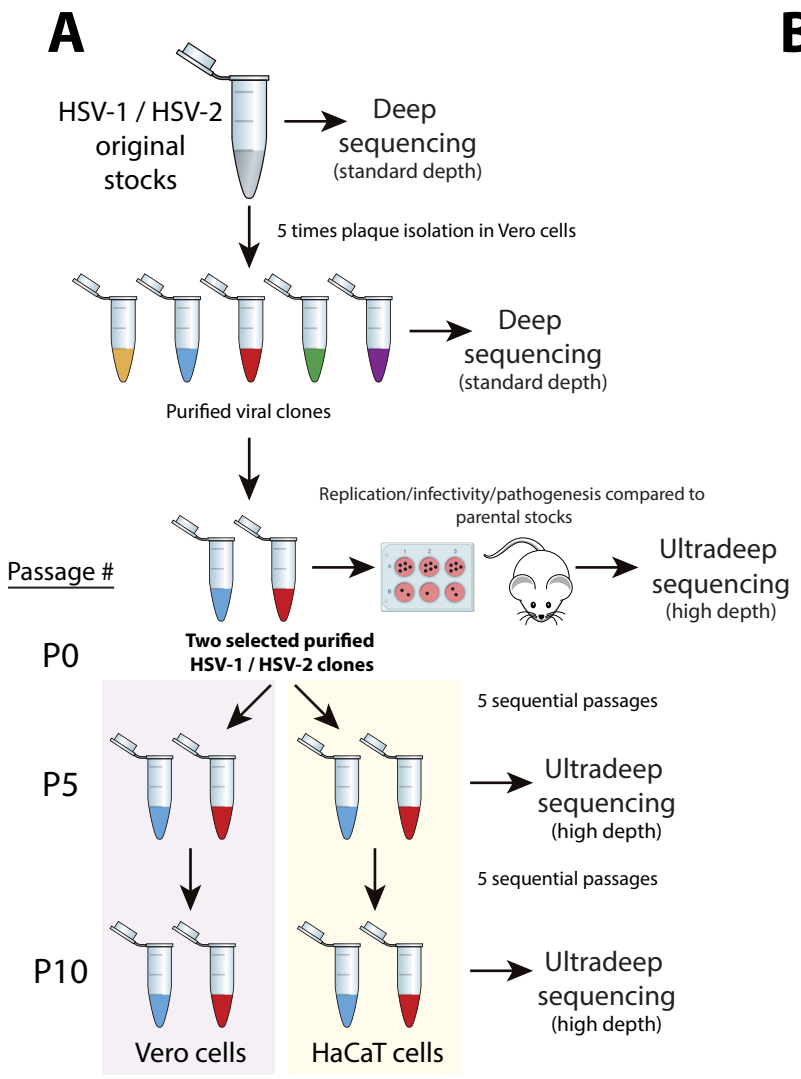
1253 **infection models shown in S4 Fig.**

1254 Virus growth curves protocol, ethical statement, description of procedures employed to

1255 infect and monitor mouse pathogenesis, and statistical analysis.

1256

1257


A
HSV-1 / HSV-2 original stocks
Deep sequencing
(standard depth)
5 times plaque isolation in Vero cells
Deep sequencing
(standard depth)
Purified viral clones
Replication/infectivity/pathogenesis compared to parental stocks
Passage #
Ultradeep sequencing
(high depth)
P0
Two selected purified HSV-1 / HSV-2 clones
5 sequential passages
P5
Ultradeep sequencing
(high depth)
5 sequential passages
P10
Ultradeep sequencing
(high depth)
Vero cells
HaCaT cells

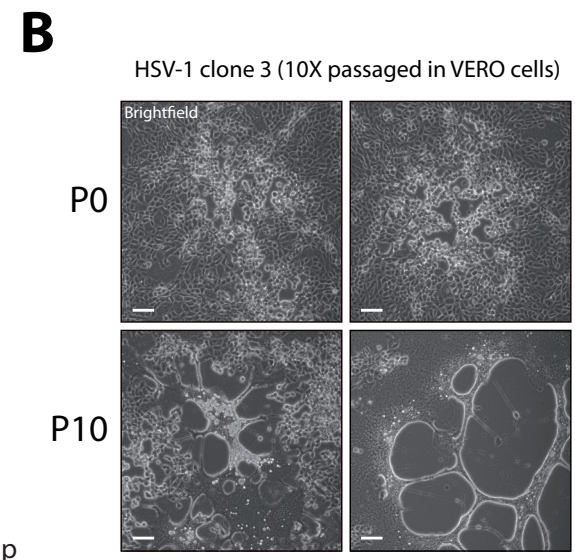

B
HSV-1 clone 3 (10X passaged in VERO cells)
Brightfield
P0
P10

Fig 1

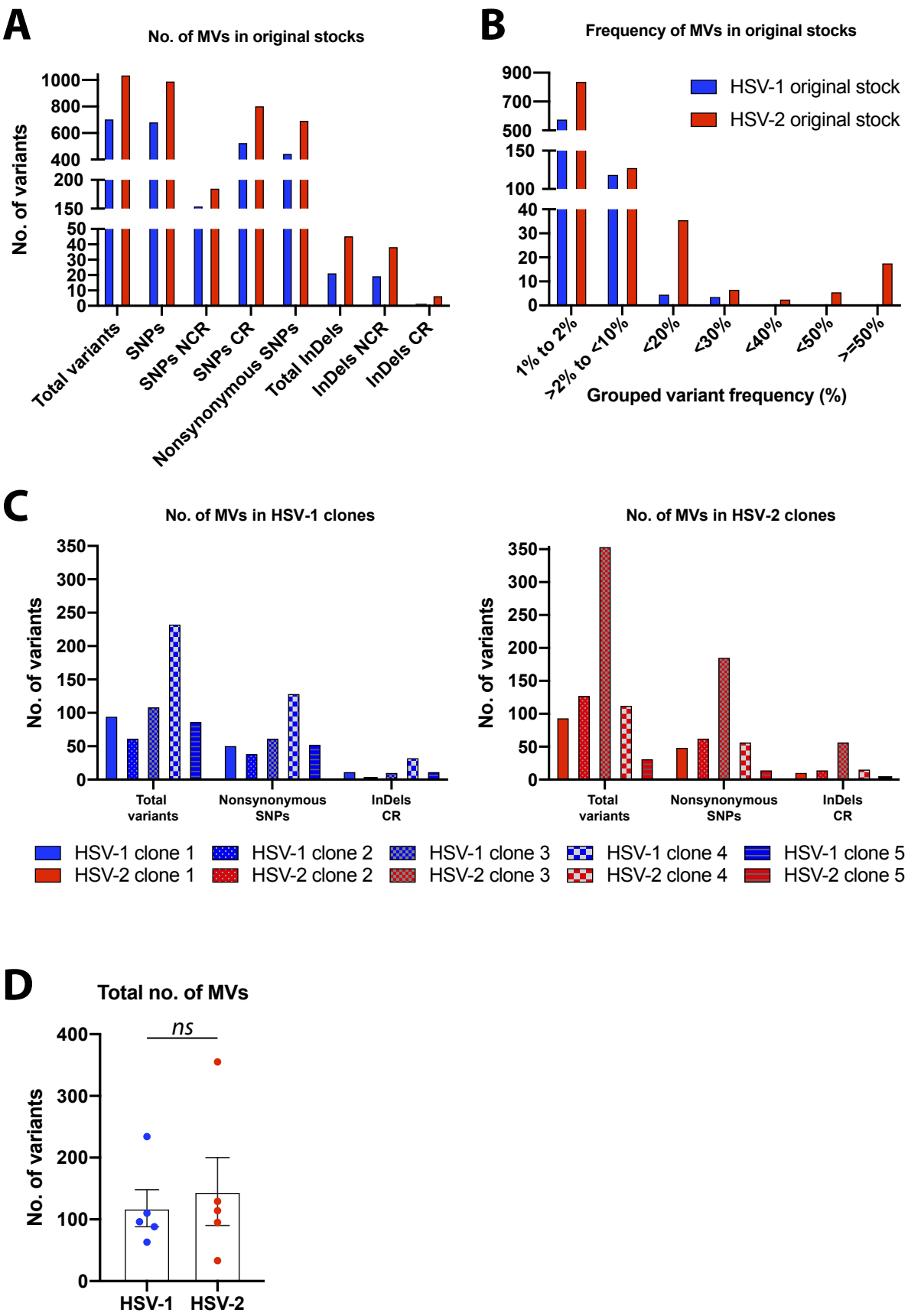

A
No. of MVs in original stocks
No. of variants
1000
800
600
400
200
150
50
40
30
20
10
0
Total variants
SNPs
SNPs NCR
SNPs CR
Nonsynonymous SNPs
Total InDels
InDels NCR
InDels CR
B
Frequency of MVs in original stocks
HSV-1 original stock
HSV-2 original stock
900
700
500
150
100
40
30
20
10
0
1% to 2%
>2% to <10%
<20%
<30%
<40%
<50%
>=50%
Grouped variant frequency (%)
C
No. of MVs in HSV-1 clones
No. of MVs in HSV-2 clones
No. of variants
350
300
250
200
150
100
50
0
Total variants
Nonsynonymous SNPs
InDels CR
HSV-1 clone 1
HSV-1 clone 2
HSV-1 clone 3
HSV-1 clone 4
HSV-1 clone 5
HSV-2 clone 1
HSV-2 clone 2
HSV-2 clone 3
HSV-2 clone 4
HSV-2 clone 5
D
Total no. of MVs
ns
No. of variants
400
300
200
100
0
HSV-1
HSV-2

Fig 2

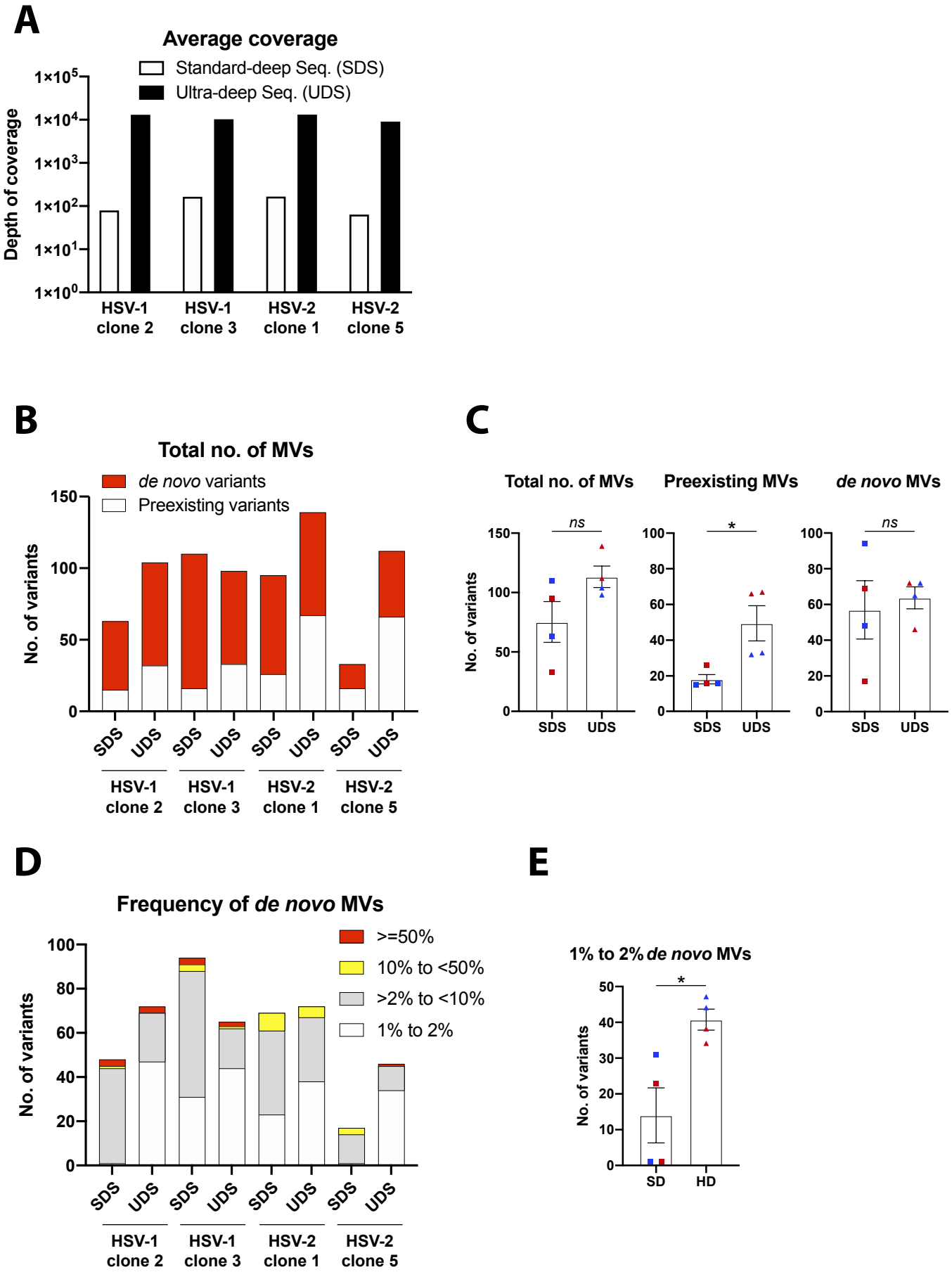

A
Average coverage
Standard-deep Seq. (SDS)
Ultra-deep Seq. (UDS)
Depth of coverage
1×10^5
1×10^4
1×10^3
1×10^2
1×10^1
1×10^0
HSV-1 clone 2
HSV-1 clone 3
HSV-2 clone 1
HSV-2 clone 5
B
Total no. of MVs
de novo variants
Preexisting variants
No. of variants
150
100
50
0
SDS
UDS
HSV-1 clone 2
HSV-1 clone 3
HSV-2 clone 1
HSV-2 clone 5
C
Total no. of MVs
Preexisting MVs
de novo MVs
ns
*
ns
No. of variants
SDS
UDS
D
Frequency of de novo MVs
>=50%
10% to <50%
>2% to <10%
1% to 2%
No. of variants
HSV-1 clone 2
HSV-1 clone 3
HSV-2 clone 1
HSV-2 clone 5
E
1% to 2% de novo MVs
*
No. of variants
SD
HD

Fig 3

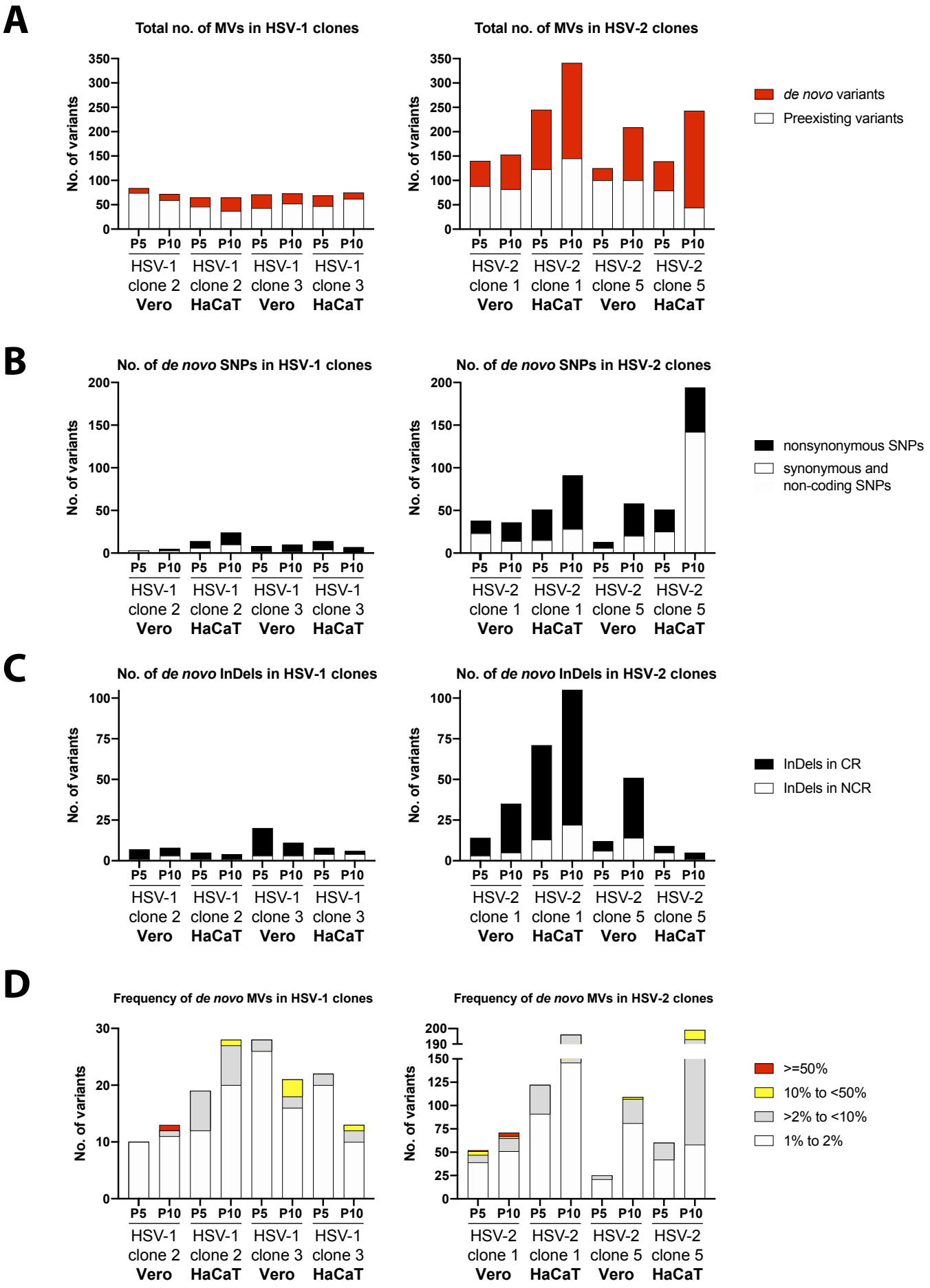

A
Total no. of MVs in HSV-1 clones
Total no. of MVs in HSV-2 clones
No. of variants
de novo variants
Preexisting variants
P5 P10 P5 P10 P5 P10 P5 P10
HSV-1 clone 2 Vero
HSV-1 clone 2 HaCaT
HSV-1 clone 3 Vero
HSV-1 clone 3 HaCaT
HSV-2 clone 1 Vero
HSV-2 clone 1 HaCaT
HSV-2 clone 5 Vero
HSV-2 clone 5 HaCaT
B
No. of de novo SNPs in HSV-1 clones
No. of de novo SNPs in HSV-2 clones
nonsynonymous SNPs
synonymous and non-coding SNPs
C
No. of de novo InDels in HSV-1 clones
No. of de novo InDels in HSV-2 clones
InDels in CR
InDels in NCR
D
Frequency of de novo MVs in HSV-1 clones
Frequency of de novo MVs in HSV-2 clones
>=50%
10% to <50%
>2% to <10%
1% to 2%

Fig 4

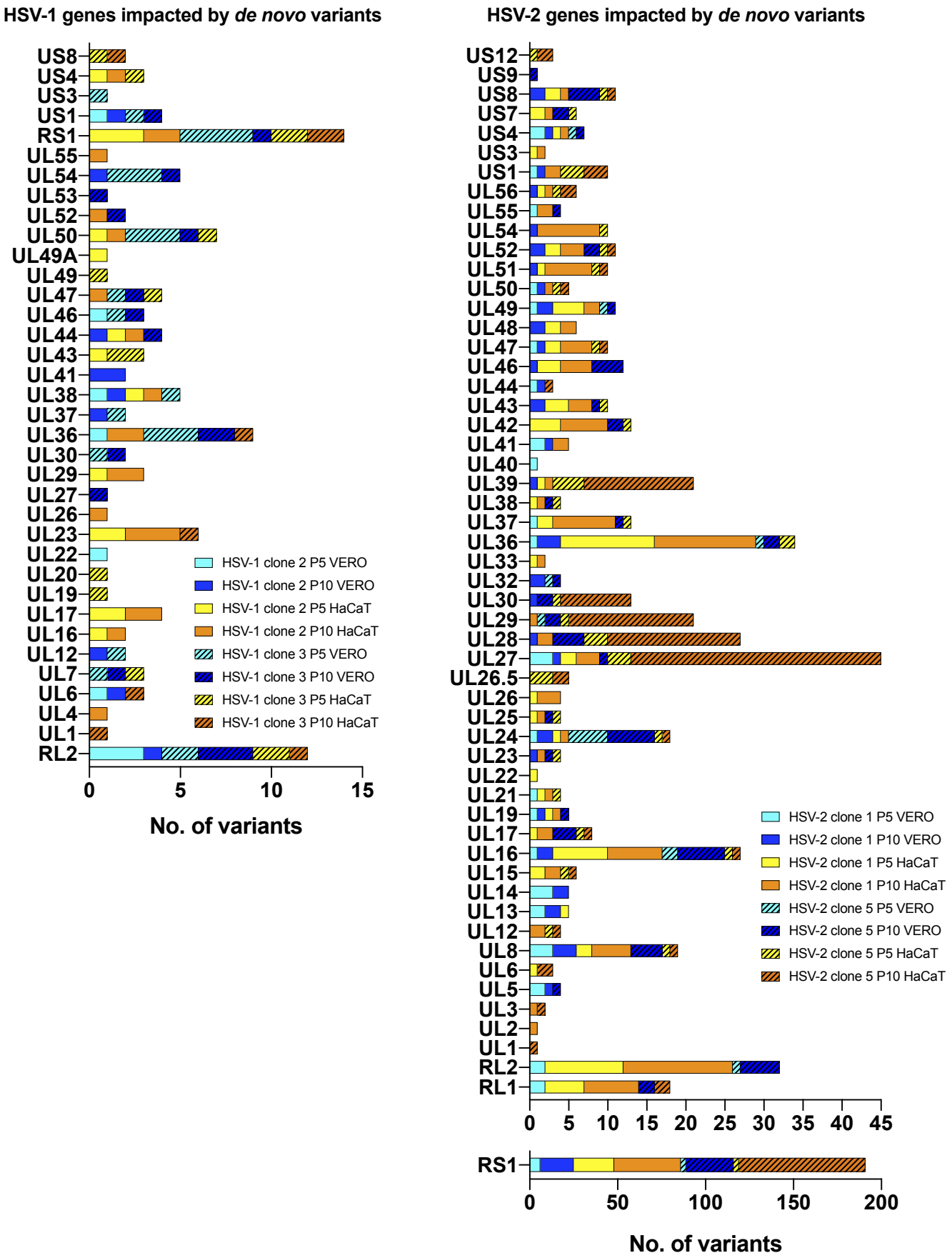

HSV-1 genes impacted by *de novo* variants
US8
US4
US3
US1
RS1
UL55
UL54
UL53
UL52
UL50
UL49A
UL49
UL47
UL46
UL44
UL43
UL41
UL38
UL37
UL36
UL30
UL29
UL27
UL26
UL23
UL22
UL20
UL19
UL17
UL16
UL12
UL7
UL6
UL4
UL1
RL2
0
5
10
15
No. of variants
HSV-1 clone 2 P5 VERO
HSV-1 clone 2 P10 VERO
HSV-1 clone 2 P5 HaCaT
HSV-1 clone 2 P10 HaCaT
HSV-1 clone 3 P5 VERO
HSV-1 clone 3 P10 VERO
HSV-1 clone 3 P5 HaCaT
HSV-1 clone 3 P10 HaCaT
HSV-2 genes impacted by *de novo* variants
US12
US9
US8
US7
US4
US3
US1
UL56
UL55
UL54
UL52
UL51
UL50
UL49
UL48
UL47
UL46
UL44
UL43
UL42
UL41
UL40
UL39
UL38
UL37
UL36
UL33
UL32
UL30
UL29
UL28
UL27
UL26.5
UL26
UL25
UL24
UL23
UL22
UL21
UL19
UL17
UL16
UL15
UL14
UL13
UL12
UL8
UL6
UL5
UL3
UL2
UL1
RL2
RL1
0
5
10
15
20
25
30
35
40
45
RS1
0
50
100
150
200
No. of variants
HSV-2 clone 1 P5 VERO
HSV-2 clone 1 P10 VERO
HSV-2 clone 1 P5 HaCaT
HSV-2 clone 1 P10 HaCaT
HSV-2 clone 5 P5 VERO
HSV-2 clone 5 P10 VERO
HSV-2 clone 5 P5 HaCaT
HSV-2 clone 5 P10 HaCaT

Fig 5

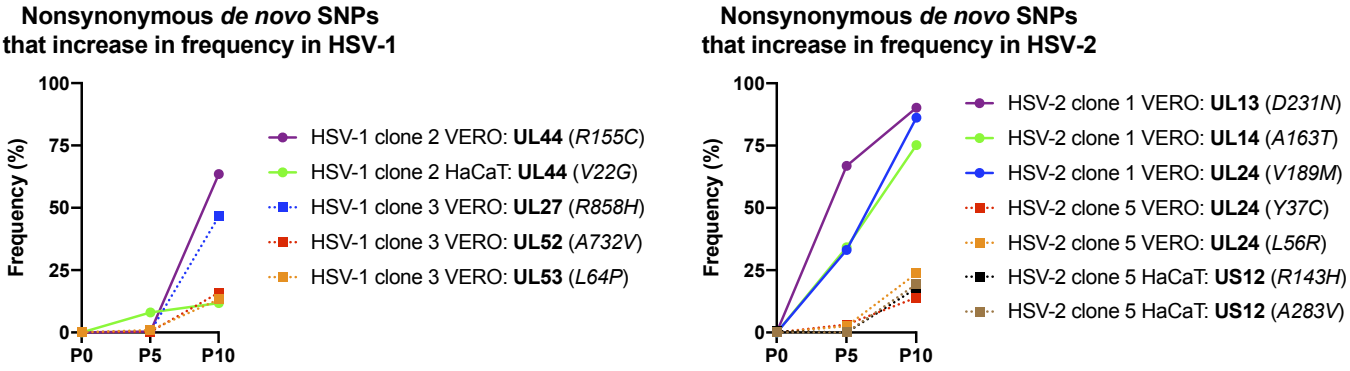

Nonsynonymous *de novo* SNPs
that increase in frequency in HSV-1
Frequency (%)
100
75
50
25
0
P0
P5
P10
HSV-1 clone 2 VERO: **UL44** (*R155C*)
HSV-1 clone 2 HaCaT: **UL44** (*V22G*)
HSV-1 clone 3 VERO: **UL27** (*R858H*)
HSV-1 clone 3 VERO: **UL52** (*A732V*)
HSV-1 clone 3 VERO: **UL53** (*L64P*)
Nonsynonymous *de novo* SNPs
that increase in frequency in HSV-2
Frequency (%)
100
75
50
25
0
P0
P5
P10
HSV-2 clone 1 VERO: **UL13** (*D231N*)
HSV-2 clone 1 VERO: **UL14** (*A163T*)
HSV-2 clone 1 VERO: **UL24** (*V189M*)
HSV-2 clone 5 VERO: **UL24** (*Y37C*)
HSV-2 clone 5 VERO: **UL24** (*L56R*)
HSV-2 clone 5 HaCaT: **US12** (*R143H*)
HSV-2 clone 5 HaCaT: **US12** (*A283V*)

Fig 6