Comparison of HSV-1 strains circulating in Finland demonstrates the uncoupling of whole-genome relatedness and phenotypic outcomes of viral infection

[1]Christopher D. Bowen, [2#]Henrik Paavilainen, [1#]Daniel W. Renner, [2]Jussi Palomäki, [2]Jenni Lehtinen, [3]Tytti Vuorinen, [4]Peter Norberg, [2*]Veijo Hukkanen, [1*]Moriah L. Szpara

[1]Department of Biochemistry and Molecular Biology, Center for Infectious Disease Dynamics, and the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA
[2] Institute of Biomedicine, University of Turku, Kiinamyllynkatu 10 D, FIN-20520 Turku, Finland
[3]Clinical Microbiology, Turku University Hospital and Institute of Biomedicine, University of Turku, Kiinamyllynkatu 13, FIN-20520 Turku, Finland
[4]Department of Virology, University of Gothenburg, Gothenburg, Sweden

#equal contribution of DWR and HP
***equal contribution of MLS and VH**

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

Author emails:
Christopher D. Bowen: cdb5004@psu.edu
Henrik Paavilainen: hojpaa@utu.fi
Daniel W. Renner: dwr19@psu.edu
Jussi Palomäki: jmmpal@utu.fi
Jenni Lehtinen: jen_ni_90@hotmail.com
Tytti Vuorinen: tyvuori@utu.fi
Peter Norberg: peter.norberg@microbio.gu.se
Veijo Hukkanen: veijo.hukkanen@utu.fi
Moriah L. Szpara: moriah@psu.edu

40 # Abstract

41 A majority of adults in Finland are seropositive carriers of herpes simplex viruses (HSV).
42 Infection occurs at epithelial or mucosal surfaces, after which virions enter innervating nerve
43 endings, eventually establishing lifelong infection in neurons of the sensory or autonomic
44 nervous system. Recent data have highlighted the genetic diversity of HSV-1 strains, and
45 demonstrated apparent geographic patterns in strain similarity. Though multiple HSV-1 genomes
46 have been sequenced from Europe to date, there is a lack of sequenced genomes from Nordic
47 countries. Finland's history includes at least two major waves of human migration, suggesting
48 the potential for diverse viruses to persist in the population. Here we used HSV-1 clinical isolates
49 from Finland to test the relationship between viral phylogeny, genetic variation, and phenotypic
50 characteristics. We found that Finnish HSV-1 isolates separated into two distinct phylogenetic
51 groups, potentially reflecting historical waves of human (and viral) migration into Finland. Each
52 HSV-1 isolate harbored a distinct set of phenotypes in cell culture, including differences in virus
53 production, extracellular virus release, and cell-type-specific fitness. Importantly, the
54 phylogenetic clusters were not predictive of any detectable pattern in phenotypic differences,
55 demonstrating that whole-genome relatedness is not a proxy for overall viral phenotype. Instead,
56 we highlight specific gene-level differences that may contribute to observed phenotypic
57 differences, and we note that strains from different phylogenetic groups contain the same genetic
58 variations.

59 # Importance

60 Herpes simplex viruses (HSV) infect a majority of adults. Recent data have highlighted the
61 genetic diversity of HSV-1 strains, and demonstrated apparent genomic relatedness between
62 strains from the same geographic region. We use HSV-1 clinical isolates from Finland to test the
63 relationship between viral genomic and geographic relationships, differences in specific genes,
64 and characteristics of viral infection. We found that viral isolates from Finland separated into
65 two distinct groups of genomic and geographic relatedness, potentially reflecting historical
66 patterns of human and viral migration into Finland. These Finland HSV-1 isolates had distinct
67 infection characteristics in multiple cell types tested, which were specific to each isolate and did
68 not group according to genomic and geographic relatedness. This demonstrates that HSV-1 strain
69 differences in specific characteristics of infection are set by a combination of host cell type and
70 specific viral gene-level differences.

71

72 **Keywords:** Herpes simplex virus (HSV), comparative genomics, phylogeny, genotype,
73 phenotype, genetic variation

74

75 # Introduction

76 Human herpes simplex virus 1 (HSV-1; Family: *Herpesviridae,* Genus: *Simplexvirus,* species:
77 *Human herpesvirus 1*) has been recognized as a cause of major human disease since the era of

78 Hippocrates (1), and to this day there is no effective vaccine (2). Approximately 50% of adults in
79 Finland are seropositive for HSV-1 (3–5), and 13% are seropositive for HSV-2 (3, 6). HSV
80 infects via epithelial or mucosal surfaces, after which it is taken up by nerve endings and
81 establishes a lifelong infection in neurons of the sensory or sympathetic nervous system. From
82 these neuronal sites the virus can reactivate, transit via nerve endings back to the skin surface,
83 and re-initiate skin or mucosal shedding at the same site as the original infection. In addition to
84 recurrent epithelial lesions, HSV also causes infectious keratitis, worsens acquisition and
85 shedding rates for human immunodeficiency virus (HIV), and can progress to cause rare but life-
86 threatening encephalitis (1).

87 Recent comparative genomics studies of HSV-1, from our lab and others, have demonstrated that
88 HSV-1 strains from unrelated individuals can differ in 2-4% of the viral genome (7–13). HSV-1
89 has a large DNA genome of ~152 kilobases, encoding >75 proteins (14, 15). Prior comparisons
90 of HSV-1 strains based on single-gene or whole-genome sequencing have found at least three
91 major clades that cluster geographically, with evidence of recombination between these groups
92 (16–18, 9). However these studies have focused almost exclusively on the quantification of
93 sequence diversity, without connecting these observations to experimental measures of viral
94 fitness and/or virulence. Several recent examples have demonstrated the usefulness of comparing
95 closely related virus strains to explore the effects of minor sequence differences on phenotypic
96 outcomes (19–22, 10). To date, publications examining the genome-wide diversity of HSV-1 in
97 diverse clinical isolates have not connected these observations to any classic measures of viral
98 fitness, such as viral replication in cells, or antiviral drug resistance (23–25). Based on the
99 decades of prior research leading to all current antiviral approaches for HSV-1, we anticipate that

100 a connection between viral genetic data, cell-based measures of viral fitness, and in vivo
101 measures of viral pathogenesis will be crucial to drive the next generation of viral therapeutics.
102 Examining the phenotypic differences displayed by HSV-1 strains in culture provides an
103 opportunity to explore the scope of these differences and test whether or not they are linked to
104 previously observed patterns of geographic diversity.

105 Finland has a unique history that has led to a relatively homogeneous and stable population (26–
106 28), providing a unique view on the evolution of viruses that are persistent in the population.
107 There has been gene flow contributing to human population of Finland: both from the east, and
108 from the south (southwest) (26–28). The Eastern migrations from Siberia began over 3500 years
109 ago, and allele sharing with modern East Asian populations can be observed even in present day
110 Finns (27, 28). Finland also has one of the best genealogical databases in the world, which in
111 combination with computerized medical records and a high rate of patient participation, has led
112 to many recent advances in human medical genetics (29, 30). This may enable future studies to
113 explore the co-evolution of human and viral genetic variation in this population.

114 We have characterized ten HSV-1 strains isolated from a random subset of Finland clinic visits,
115 and compared their growth properties, drug resistance, and other phenotypic features. Full
116 genome sequences of these viruses were compared to each other and to other previously
117 described strains of HSV-1, to reveal patterns of inter-host and intra-host variability. While two
118 phylogenetic sub-groups were detected at a genomic level, these sub-groups were not predictive
119 of any detectable pattern in the observed phenotypic differences, demonstrating that whole-
120 genome relatedness is not a proxy for viral phenotype. We anticipate that these data will aid in
121 efforts to develop improved sequence-based antiviral therapies, by providing additional data on

122 conserved vs. divergent areas of the HSV-1 genome, and contribute to development of a vaccine
123 against HSV infections based on attenuation of these or related clinical isolates (31). These data
124 present an opportunity to explore the diversity of chronic herpesviruses in the Finland
125 population, and to lay the foundation for future studies that explore the connections between
126 viral genetic differences, host genetic predispositions, and their potential relationship(s) to
127 clinical outcomes of HSV-1 infection.

128

129 # Methods

130 **Virus isolates and virus stock propagation**

131 Clinical isolates were obtained from anonymous coded diagnostic samples from herpes lesions
132 representing currently circulating viral strains in Finland (**Table 1**). The virus diagnostic unit of
133 the Turku University Hospital receives viral culture samples from the whole country except for
134 the Helsinki-Uusimaa region (capital Helsinki area and the South-Eastern Finland). Recent
135 immigrations to Finland were not represented in the sample material. Approval for the study of
136 anonymous HSV isolates was provided by the Turku University Central Hospital (permit #
137 J10/17).

138 An immunoperoxidase-rapid culture assay (32) was used to type the viruses as HSV-1, which
139 was confirmed by a HSV type-specific gD (US6) gene-based PCR test (33). Viruses were
140 initially propagated on Vero cells (African green monkey kidney cells; ATCC), maintained in
141 Dulbecco's MEM with 2% fetal calf serum and gentamycin. A stock was made by addition of 3
142 ml of 9% autoclaved skimmed milk (Valio, Finland) on to 5 ml of the culture medium and

143 subsequent freezing. The cells and the medium were collected and combined upon thawing, and
144 were frozen and thawed for two further rounds before aliquoting. The viral titer was determined
145 by plaque titration on Vero cells as described before (34). Parallel aliquots were used for further
146 viral culture studies and for preparation of viral nucleocapsid DNA.

147 **Statistical methods**

148 For viral production data, SPSS Statistics 20 (IBM, Armonk NY, USA) software was use to
149 perform statistical analyses. Non-parametric Mann-Whitney U-test was used to calculate
150 statistical significances. The threshold for significance was set to $p<0.05$.

151 **Viral genomic DNA isolation**

152 The viral genomic DNA was prepared from isolated viral nucleocapsids as described previously
153 (7, 35). In brief, viral stock collected from the first or second passage in Vero cells was used to
154 infect ~$1x10^8$ HaCaT cells (Department of Dentistry, University of Turku (36)) at a multiplicity
155 of infection (m.o.i.) of 0.1-1 plaque-forming units (PFU)/cell, and the infection was allowed to
156 proceed to completion at +35 °C (1-3 days). The cells were collected, washed with PBS and
157 suspended to LCM buffer (0.125 M KCl, 30 mM Tris pH 7.4, 5 mM $MgCl_2$, 0.5 mM EDTA,
158 0.5% Nonidet P-40; with 0.6 mM beta-mercaptoethanol). After two successive extractions with
159 Freon (1,1,2-trichloro-1,2,2-trifluoroethane, Sigma-Aldrich), the extracts were added on top of
160 layers in LCM buffer with 45% and 5% glycerol and ultracentrifuged at 77 000 x g for 1 h at
161 +4°C (Beckman Coulter SW41Ti rotor). Viral nucleocapsids were recovered from the bottom of
162 the ultracentrifuge tube, and the DNA was prepared by treatment with Proteinase K and SDS,
163 followed by repeated extractions with phenol-chloroform and ethanol precipitation. The DNA

164 content and purity were observed by spectrophotometry and by agarose gel electrophoresis after
165 restriction enzyme digestions.

166 **Cell culture**

167 Vero cells (ATCC; Manassas, VA) were propagated in M199 medium supplemented with 5%
168 fetal bovine serum and gentamycin. HaCaT cells (Department of Dentistry, University of Turku;
169 (36)) were propagated in DMEM medium with HEPES buffer, supplemented with 7% fetal
170 bovine serum and gentamycin. SH-SY5Y neuroblastoma cells (K. Åkerman, Åbo Akademi
171 University, Turku, Finland) were propagated in DMEM (high glucose) medium supplemented
172 with 10% fetal bovine serum, 2 mM L-glutamine and gentamycin. Initial differentiation of SH-
173 SY5Y cells involved culture for 10 days in DMEM/F-12 medium containing 5% fetal bovine
174 serum, 10 µM All-trans retinoic acid (Sigma), 2 mM L-glutamine and gentamycin. Thereafter
175 SH-SY5Y cells were transferred on Matrigel-coated (B-D) 96-well plates and the medium was
176 changed to serum-free DMEM/F-12 medium containing 10 µM all-trans retinoic acid (Sigma),
177 0.5 µg/ml of brain-derived neurotrophic factor (Millipore), 2 mM L-glutamine and gentamycin.

178 **Acyclovir resistance testing**

179 The sensitivity of each HSV strain to acyclovir was tested in a microplate format. Vero cells
180 grown on 96-well cell culture plates were treated with cell culture medium (DMEM with 5%
181 fetal bovine serum) supplemented with acyclovir (ACV; Sigma), in concentrations of 128 µg/ml
182 to 0.03125 µg/ml (1:4 serial dilutions). Duplicate wells containing each ACV dilution, and wells
183 without ACV, were infected with 100 PFU of each virus. As a control, the HSV-1 delta305 virus
184 was included; it is resistant to ACV due to deletion of its thymidine kinase gene (37). Infected

185 cells were incubated at +37ºC, at 5% CO_2, for 72 hours before fixation with methanol and
186 staining with crystal violet. The reduction of plaque numbers at each ACV dilution was observed
187 in comparison with wells infected without ACV. A logistic fit curve was used for determination
188 of IC_{50} values. A virus with an IC_{50} value of over 2 μg/ml (of ACV) was considered resistant
189 (38).

190 **Image acquisition**

191 Photomicrographic images of viral plaques were obtained using a Zeiss Primovert inverted
192 microscope with Plan-Achromat 4x and 10x objectives, recorded using an AxioCam ERc 5s
193 camera, and analyzed using Zeiss ZEN 2012 software.

194 **Next generation sequencing**

195 Viral nucleocapsid DNA was sheared on a Covaris M220 (parameters: 60 seconds duration, peak
196 power 50, 10% duty cycle, 4C). We used the Illumina TruSeq DNA Sample Prep kit to prepare
197 barcoded sequencing libraries, according to the manufacturer's protocol for low-throughput
198 sample handling. Libraries were quantified and assessed by Qubit (Invitrogen, CA), Bioanalyzer
199 (Agilent), and library adapter qPCR (KAPA Biosystems). Illumina MiSeq paired-end sequencing
200 (2 x 300 bp) was completed according to manufacturer's recommendations, using 17 pM library
201 concentration.

202 A consensus viral genome for each strain was assembled using a *de novo* viral genome assembly
203 (VirGA) workflow (10). This approach begins with quality control, including removal of
204 contaminating host sequences, adapters from library preparation, and imaging artifacts. Next
205 VirGA iterates through multiple *de novo* assemblies using SSAKE, which we then combine into

206 longer blocks of sequence (contigs) using Celera and GapFiller. VirGA uses Mugsy alignment to
207 match these contigs to the HSV-1 reference genome (strain 17, GenBank accession JN555585).
208 The best-matched contigs are stitched into a single consensus genome, which is then annotated
209 and subjected to additional quality control measures. These include an examination of coverage
210 (sequence read) depth, detection of minority variants within each consensus genome, and manual
211 inspection of gaps and low coverage areas. See **Table 2** for GenBank accessions.

212 **Intra-strain minority-variant detection**

213 Minority-variant (MV) positions within each *de novo* assembled genome were determined using
214 VarScan v2.2.11 as previously described (19, 39). Conservative variant calling parameters to
215 eliminate sequencing-induced errors were set at: minimum allele frequency ≥0.02; base call
216 quality ≥20; read depth at the position ≥100; independent reads supporting minor allele ≥5. MVs
217 containing ≥90% unidirectional strand support were excluded from further analyses (40, 41), as
218 were those occurring within 10 base pairs of one another. MVs passing quality control were
219 mapped to respective genomes and assessed for mutational effect using SnpEff and SnpSift (42,
220 43).

221 **Phylogenetic and recombination analyses**

222 DNA sequences were aligned using the Kalign algorithm included in eBioTools. To avoid
223 inferences caused by false phylogenetic signals, all gap and repeat regions were excluded prior to
224 this analysis. Repeat regions not leading to gaps were also excluded since these may contain
225 single nucleotide differences that have been shuffled into new positions by random aspects of
226 tandem-repeat alignment. Furthermore, nucleotides that were marked as "N" in GenBank-

227 derived strains were excluded, along with the corresponding aligned nucleotides in remaining
228 sequences. Complete genomes in GenBank harbouring long stretches of "N" were not included
229 in the analysis (e.g. strain B3x_1_5). The network is based on an alignment of 116,610
230 nucleotides, which is more than 75% of the entire HSV-1 genome.

231 The phylogenetic network was constructed by using SplitsTree4. It is a NeighborNet with
232 OrdinaryLeastSquares Variance depicted as a RootedEqualAngle splitstree. A recombination test
233 on the complete dataset was performed by using the Phi test for recombination implemented in
234 SplitsTree4. Bootstrapping values greater than 80 are shown for clades with three or more
235 strains.

236 **Restriction fragment length polymorphism (RFLP) analysis**

237 A cytoplasmic viral DNA preparation was modified from the protocol described by Igarashi et al
238 (44). Subconfluent Vero cell cultures were infected at an m.o.i. of 0.1, and incubated at +34ºC
239 for 2 days until the cytopathic effect was complete. The cells were collected in 150 mM NaCl, 10
240 mM Tris pH 7.6, 1.5 mM MgCl buffer, with 0.1% Nonidet P-40. The nuclei were pelleted, and
241 DNA was extracted from the supernatant by two successive extractions with phenol-chloroform.
242 DNA was recovered by ethanol precipitation. 12.5 µg of each viral DNA was diluted in 15 µl
243 solution containing sterile H_2O, FastDigest 10X green buffer and FastDigest BamHI or SalI
244 restriction enzyme (Thermo Scientific). Electrophoresis was run in 0.8 % TBE-agarose gels, with
245 GeneRuler mix DNA ladder, for 24 h at 45 V before imaging. In order to separate large DNA
246 fragments after SalI digestion, electrophoresis was continued for additional 36 h at 30 V.

247

248 # Results

249 ## Comparison of growth properties of Finnish HSV-1 strains in vitro

250 A random set of ten circulating Finnish HSV-1 isolates were selected from residual laboratory
251 diagnostic specimens (**Table 1)**. We first examined *in vitro* phenotypic characteristics of these
252 HSV-1 isolates, by comparing their overall titer and rate of intracellular virus vs. extracellular
253 (released) virus production at 24 hours post infection (hpi) of a range of cell types (**Figure 1**).
254 These included non-human primate kidney-derived epithelial (Vero) cells (**Figure 1A**), human
255 keratinocyte (HaCaT) cells (**Figure 1B**), and human neuroblastoma (SH-SY5Y) cells in a mixed,
256 undifferentiated state (**Figure 1C**), as well as a differentiated, neuronal state (**Figure 1D**).
257 Compared to epithelial and keratinocyte cells, overall viral production was markedly reduced in
258 neuronal precursor cells and differentiated neuronal cells (compare **Figure 1AB** vs. **CD**). The
259 wild-type HSV-1 reference strain 17+ replicated to significantly higher titers than any of the
260 circulating clinical isolates in Vero cells, which are standardly used for HSV propagation, and to
261 a lesser extent strain 17+ replicated at higher titer in keratinocytes as well (**Figure 1**; $p<0.05$, 10-
262 100 fold higher viral amounts). Different clinical isolates excelled at producing virions in each
263 cell type, with no clear patterns of most- or least-efficient viral production or release across all
264 cell types. The amount of extracellular released virus was less than 10% of the total virion
265 production, in all cell types except the undifferentiated neuronal precursor cells (**Figure 1,**
266 **insets**). Each virus strain was also tested for acyclovir (ACV) resistance. Despite some variation
267 in ACV susceptibility (IC50 values), none of the circulating Finnish viruses was considered
268 resistant to ACV (**Table 1**). There were also no significant differences among the 10 clinical

269 isolates in the plaque morphology or the type of cytopathic effect induced in Vero cell cultures
270 (data not shown).

271 Genetic and genomic analysis of Finnish HSV-1 strains

272 Based on prior data suggesting the effects of successive waves of migration on the human
273 population in Finland (45, 46), we next assessed the overall genetic diversity of these ten HSV-1
274 isolates. We first performed a restriction fragment length polymorphism (RFLP) analysis on viral
275 genomic DNA, which revealed at least two broad patterns of variation (**Figure 2**). The diversity
276 of bands led us to examine these genetic differences with greater precision using high-
277 throughput, deep genome sequencing (HTS) and comparative genomics analysis (see **Methods**
278 for details). A previously described bioinformatics pipeline was used to *de novo* assemble a full-
279 length consensus genome for each strain **(Table 2)**. A viral consensus genome represents the
280 most common nucleotide detected at each nucleotide position in that viral population. All viral
281 genomes had an average coverage depth between 1,000-2,800-fold, with >96% of the genome
282 covered at a depth exceeding 100-fold **(Table 2)**. As observed in prior publications, the few
283 areas with <100X coverage correlated with regions of high G+C-content, and/or highly repetitive
284 sequences (47, 9, 10). The >1,000X average coverage depth enabled us to compare the full
285 genetic complement of the viral genome for each strain, to analyze differences both between and
286 within each viral strain population, and to compare them to previously described viral genomes.

287 Genetic relatedness of Finland HSV-1 samples

288 First, we used the viral consensus genomes to compare how overall variation in these ten Finnish
289 HSV-1 genomes related to the patterns observed by RFLP analysis (**Figure 2**). Whole genome

290 alignments were created using the ten Finnish HSV-1 genomes, as well as these ten in
291 conjunction with a diverse set of previously published HSV-1 genomes (see Methods and
292 **Supplemental Table S1** for a full list). We then used SplitsTree to create a phylogenetic
293 network that revealed the relatedness of the ten Finnish viral genomes to each other (**Figure 3A**)
294 and to previously described HSV-1 genomes (**Figure 3B**). This approach revealed that the
295 Finland HSV-1 genomes separated into two sub-groups, which appear to relate to previously
296 recognized Asian and European / North American clades. These data echo those of the initial
297 RFLP analysis.

298 **Protein-coding variation in Finland HSV-1 samples**

299 Next, we examined more fine-scale genetic differences in these HSV-1 isolates, which are
300 invisible at the level of RFLP analysis. Here we examined DNA- and amino acid (AA)-
301 alignments for each protein-coding region of the HSV-1 genome (**Supplemental Table S2-S3**).
302 We found that from zero to 8% of each coding region differed at the AA level between the ten
303 Finnish viral genomes (**Figure 4**). Only a handful of small coding regions (UL20, VP26 (UL35),
304 UL49A, UL55) showed no AA variation at all between viral isolates. Consistent with previous
305 findings (9), gL (UL1), UL11, UL43, gG (US4), and gJ (US5) were among the most divergent
306 proteins. Together these inter-strain genetic differences in 70 HSV-1 proteins in **Figure 4**
307 provide ample opportunities to generate the phenotypic diversity observed in **Figure 1**. Overall,
308 these levels of AA coding diversity reflect those seen in previous analyses (9), and between other
309 known HSV-1 genomes.

310 **Minority allelic variants in Finland HSV-1 viral populations**

311 Each viral consensus genome represents the most common nucleotide detected at each position
312 in the viral genome. In contrast, minority variants (MV) within each viral population are rare
313 alleles, presented in <50% of the sequenced reads. These MV may rise to greater prevalence
314 during viral spread to new niches or hosts, or under selective evolutionary pressure. For each
315 viral genome, we examined the number of MVs detected in each genome. Stringent quality-
316 control criteria were used to reduce the number of false-positive MVs (see **Methods**), and only
317 those MV detected at ≥2% prevalence were used in this analysis. MVs were found dispersed
318 across each genome (**Figure 5A**), mostly in intergenic regions, and were mostly of low
319 frequency (**Figure 5B**). A concentration of MVs was found in the internal-repeat region, which
320 may result from stochasticity in tandem repeat alignment in this area, and/or from the reduced
321 selective pressures on intergenic sequences found in this region. The MVs detected in US7 (gI)
322 occur at the same tandem-repeat site as the consensus-level AA variations in this gene (**Figure**
323 **6**). Taken together, these results define the level of inter-host (**Figure 4**) and intra-host (**Figure**
324 **5**) variation seen in the Finnish HSV-1 strains described herein.

325 **Patterns of variation in Finnish HSV-1 strains**

326 Finally, we considered whether any of the observed patterns of genomic or phenotypic variation
327 could be linked to fine-scale coding variations in these ten Finnish HSV-1 isolates. The genomic
328 sub-groups detected by RFLP (**Figure 2**) and phylogenetic network analysis (**Figure 3**) were
329 reflected in several coding variations that correlated with phylogenetic group. For example,
330 coding differences in the secreted and virion surface glycoprotein G (gG; US4) correlated with
331 the two Finnish phylogenetic sub-groups (**Figure 6A**; see **Figure 3** for phylogenetic sub-groups).

332 Other impacts on coding variation may result from changes in copy number at tandem repeats.
333 Variation in tandem repeat length in glycoprotein I (gI; US7) leads to changes in the length of a
334 repeating tract of mucin-type O-linked glycosylation sites (**Figure 6B)**, akin to that previously
335 described in clinical isolates of HSV-1 (16, 48). Finally, there are detectable patterns of AA
336 variation that correlate with phenotypic differences in viral fitness in specific cell types. Coding
337 differences in the nuclear egress complex protein NEC2 (UL34) (**Figure 6C**) were observed in
338 the two strains (F-12g, M-15) which showed no detectable extracellular virion release in
339 differentiated neuronal cells (**Figure 1D**). The terminally differentiated and non-dividing state of
340 these neuronal cells may constitute a sensitized background to detect impacts on virion egress.
341 Taken together, these examples illustrate the type and degree of variation present in these strains,
342 and highlight potential genetic insights into viral phenotype.

343

344 # Discussion

345 In this study, we described the first-ever HSV-1 genomes from Finland. Two distinct sub-groups
346 or clades were observed in the ten Finnish clinical isolates, with four strains clustering in one
347 clade and six in another. These groups were detectable using both classical RFLP approaches
348 (**Figure 2**) and deep-sequencing methods for whole-genome analyses (**Figure 3**). These data
349 appear to corroborate with previous findings on colonization of Finland, with migration from
350 Europe in the south/southwest and from Asia in the northeast (26–28). However, we found that
351 phenotypic differences between these isolates were uncoupled from the overall genomic patterns
352 that grouped them into two geographic clades. While plaque morphology was consistent across

353 isolates (data not shown), the amount of virus production, extracellular virion release, and
354 acyclovir sensitivity differed between isolates and across cell types (**Figure 1** and **Table 1**). We
355 detected gene-specific patterns of genetic variation that may impact protein function. Future
356 studies will need to dissect individual genetic variations in each viral isolate in order to test their
357 precise impact(s) on viral fitness.

358 When analyzing the replication capabilities of these recently isolated clinical samples in multiple
359 cell types, we found that they replicated to lower overall titers and produced fewer extracellular
360 virions than the lab-adapted HSV-1 strain 17+, regardless of cell type (**Figure 1**). This difference
361 was more pronounced in non-human primate epithelial and human keratinocyte cells, and was
362 more subtle in neuronal precursors and differentiated neuronal cells. There did not appear to be
363 any clear differences in titer or extracellular virus production between geographic clades. There
364 also seemed to be no noticeable differences in plaque morphology among these samples, with
365 plaque size and cellular cytopathic effect being similar across all samples (data not shown).
366 Although we report findings in a number of cell lines in this study, it will be important for future
367 studies to examine each virus and cell-type pairing separately, since these data demonstrate that
368 maximal viral fitness in one cell type is not generally predictive of fitness in another.

369 We observed several potential connections between genetic differences and phenotypic
370 differences. For instance, we detected an AA difference in the UL34 (NEC2) binding domain in
371 F-12g and M-15 **(Figure 6C).** Previous work has shown that the viral proteins UL31 (NEC1) and
372 UL34 (NEC2) form a nuclear egress complex that localizes to the perinuclear space and plays a
373 critical role in egress of viral capsids from the nucleus (49, 50). In these two clinical isolates (F-
374 12g and M-15), there was no measurable amount of extracellular virus produced in differentiated

375 neuronal cells (**Figure 1D**). Further studies will be needed to examine if this phenotype is linked
376 to a disruption of NEC1 and NEC2 binding and impaired nuclear egress for isolates F-12g and
377 M-15. Variation in the copy-number of tandem repeats is yet another way that HSV-1 isolates
378 can generate genetic and potential phenotypic variation (9, 16, 51–53). Glycoprotein I (gI; US7)
379 contains a mucin-like domain with repeating units of serine, threonine, and proline (**Figure 6B**).
380 This domain varies in length across these isolates. This domain has been previously shown to
381 serve as a site of O-linked glycosylation, with longer tracts (e.g. 8 repeating blocks of 7 AA
382 each) being more heavily glycosylated than short tracts (e.g. 2 repeating blocks) (16, 48). A prior
383 study of over 80 Swedish HSV-1 isolates found roughly equal distributions of isolates with two,
384 three, or four repeating blocks in this mucin-like gI domain (16, 48). Half of the Finland HSV-1
385 isolates analyzed here have at least four repeating units, suggesting the potential for gI to be
386 more heavily glycosylated in these isolates. These data generate hypotheses for future
387 investigations of genotype-phenotype links in these and other new clinical isolates.

388 This study provides one of the first combinations of phenotypic analyses that span multiple cell
389 types, alongside full comparative genomic analyses that span both phylogenetic and gene-
390 specific analyses. One comparison not yet explored in this or prior studies is the specific pairing
391 of each virus with cells derived from the human source of each isolate. This approach is now
392 technically feasible, using induced pluripotent stem cell technology to generate specific cell
393 types from a single source such as a buccal swab. The rich genealogical databases in Finland
394 suggest an opportunity to not only pair these analyses at the human-cell and virus level, but also
395 to link observed phenotypes of cellular infection to each patient's clinical, genetic, and/or
396 familial history of herpesvirus disease. Future studies may also seek to link human interaction

397 data, for instance contact networks and/or travel history, to infer how human movement and
398 interactions patterns influence pathogen spread and population dynamics. We anticipate that in
399 the future, this approach would yield fruitful insights not only into direct clinical outcomes of
400 herpesvirus disease, but also generate hypotheses about herpesvirus co-morbidities.

401

402 # Acknowledgements

403 The authors thank professor Päivi Onkamo, and the members of the Hukkanen and Szpara labs
404 for helpful feedback and discussion. We acknowledge Ritva Kajander for preparation of the viral
405 stocks and the viral nucleocapsid DNA, and BSc Kiira Kalke and BSc Mira Hörkkö for help in
406 imaging the viral plaques. This work was supported by research startup funds (MLS) from the
407 Pennsylvania State University, and by a grant from the Pennsylvania Department of Health using
408 CURE funds (MLS). The Department specifically disclaims responsibility for any analyses,
409 interpretations, or conclusions. The funding from the Jane and Aatos Erkko Foundation grant
410 #170046 is acknowledged. We thank Prof. Bernard Roizman, University of Chicago, for the
411 HSV-1 delta305 virus.

412

413 # Figure Legends

414 **Figure 1: Comparison of growth properties of Finnish HSV-1 strains reveals the**
415 **uncoupling of phenotypic and geographic variation.**

416 Histograms comparing virus production for ten Finnish strains of HSV-1, as compared to the
417 HSV-1 reference strain 17+, in different cell types. Panel **(A,** non-human primate**)** shows Vero
418 monkey kidney cells, which are defective in interferon signaling; **(B)** shows HaCaT human
419 keratinocytes; **(C)** shows undifferentiated human SH-SY5Y neuroblastoma cells (neuronal
420 precursor cells); **(D)** shows differentiated human SH-SY5Y neuronal-like cells. Viruses are
421 plotted in descending order based on their viral production. For ease of comparison, the strains
422 are color-coded and divided into the genomic sub-groups used in later figures (see **Figure 3** for
423 illustration). Each histogram bar plots infectious virions in the cell-associated fraction (gray) vs.
424 those released from cells (black). A horizontal line shows the median (solid green) and the
425 average (dashed green) amount of virus production for the ten Finnish HSV-1 strains in each cell
426 type, with a highlighted box (green) on this y-axis added to highlight the multi-log decrease in
427 viral production in **C-D** vs. **A-B**. Each panel has a smaller graph (**inset**) where the extracellular
428 released virus (black) is plotted as a percentage of total virus production (gray). Viruses in each
429 inset are plotted in the same order as the matched panel graph. Bar shows SEM; * = $p=<0.05$, **
430 = $p<0.001$ as compared to the total virus amount (cell associated + extracellular), for each
431 clinical isolate vs HSV-1(17+) reference strain.

432 **Figure 2: RFLP comparison reveals at least two sub-groups of Finnish HSV-1**
433 **strains.**

434 Viral genomic DNA was digested with SalI and separated via electrophoresis to distinguish
435 overall genetic patterns. Classically defined SalI fragment notations (red alphabet letters) are
436 indicated on the left (54). Red crosses (x) indicate variation in the size of SalI fragment G, while
437 diamonds indicate variation in the SalI Q fragment. For ease of comparison, the strains are color-
438 coded the same as in later figures (see **Figure 3** for illustration).

439 **Figure 3. A phylogenetic network of genetic relatedness reveals that Finland HSV-1**
440 **isolates separate into two sub-groups.**

441 A phylogenetic network of genetic relatedness was constructed using SplitsTree4 to demonstrate
442 how the ten new Finland HSV-1 isolates relate to each other (**A**) and to previously described
443 HSV-1 genomes from a diversity of locations (**B**). Geographic clustering of branches and clades
444 resemble those previously described (9, 16–18) and are colored by origin of strains: red for
445 Europe and North America, blue for Asia, and purple for Africa. Finland HSV-1 genomes are
446 indicated in bold, and separate into two sub-groups (**A**), which appear to relate previously
447 recognized Asian and European and North American clades (**B**). Bootstrap values larger than 80
448 supporting clades with three or more strains are shown on the network. The extent of
449 recombination among HSV-1 strains is demonstrated by the parallel edges in the network and a
450 phi-test for recombination. The isolate name, country of origin, GenBank accession, and
451 reference(s) for each previously described HSV-1 genome are listed in **Supplemental Table S1**.

452 **Figure 4. Substantial protein-coding variation exists between Finland HSV-1**
453 **isolates.**

454 Histogram depicts the percent of synonymous nucleotide differences per gene (**A**) and the
455 percent amino acid (AA) variation per protein (**B**) for the ten Finnish HSV-1 strains compared
456 here. Labels on the left in (**A**) indicate the repeat, unique short (US) and unique long (UL)
457 regions that contain each of the sequentially-numbered viral coding regions. Where possible, the
458 common name of each protein is shown to the right of the appropriate histogram bar in (**B**), e.g.
459 UL1 is also known as gL. Nucleotide and AA differences were quantified at the consensus level
460 of each viral genome. See **Supplemental Table S2** for additional data on the number of
461 nucleotide differences and dN/dS ratio for each gene.

462 **Figure 5. Scatter plot illustrates the location and low frequency of minor variants in**
463 **each Finnish HSV-1 genome.**

464 **(A)** Scatter plot demonstrates the position (x-axis) and frequency (y-axis) of minority variants
465 detected in each Finnish HSV-1 genome. Spheres on the graph indicate minor variants in
466 intergenic regions. Squares with gene names indicate MVs located in genes. A color key
467 consistent with that used other figures is included at the lower right. A genome diagram is shown
468 below the x-axis for reference. Areas containing large numbers of tandem repeats, e.g. around
469 UL36 in the U_L region and throughout the RS/RL region, are denoted as gray blocks on the
470 graph. The limit of detection was set at 2% (see Methods for details). **(B)** Histogram depicts the
471 total number of minority variants (y-axis) in each frequency bin (x-axis), summed across all ten
472 Finnish HSV-1 strains. Bins are in 1% increments.

473 **Figure 6. Variations in individual genes provide a better insight into Finnish HSV-1**
474 **strain sub-groups and of specific cellular phenotypes.**

475 Examples are shown here of variation in individual HSV-1 genes that reflect either the Finnish
476 HSV-1 strain sub-groups, or that correlate with and potentially influence specific cellular
477 phenotypes. Diagrams depict three examples of HSV-1 proteins with known functional domains
478 and post-translational modifications (e.g. phosphorylation, glycosylation; see Legend). A subset
479 of each protein is highlighted via an AA alignment of the ten clinical HSV-1 isolates. (**A**) Coding
480 differences in the secreted and virion surface glycoprotein G (gG; US4) correlate with the overall
481 genomic sub-groups (clades) shown in **Figure 3**. (**B**) Variation in copy number of a seven AA
482 tandem repeat block in glycoprotein I (gI; US7) changes the length of a repeating tract of mucin-
483 type O-linked glycosylation sites, as described previously in clinical isolates of HSV-1 (16, 48).
484 (**C**) AA differences in the nuclear egress protein NEC2 (UL34) were observed in the two strains
485 (M-15, F-12g) which showed no detectable extracellular virion release in differentiated neuronal
486 cells (**Figure 1D**). See Discussion for additional insights on the potential phenotypic impacts of
487 the variations shown in B-C.

488

489 # References

490 1. Roizman B, Sears E. 1996. Herpes simplex viruses and their replication, p. 1043–1107. *In*
491 Fields, BN, Knipe, DM, Howley, PM (eds.), Fundamental virology, 3rd ed. Lippincott-
492 Raven, Philadelphia, PA.

493 2. Johnston C, Koelle DM, Wald A. 2011. HSV-2 : in pursuit of a vaccine. J Clin Invest
494 121:4600–4609.

495 3. Pebody RG. 2004. The seroepidemiology of herpes simplex virus type 1 and 2 in Europe.
496 Sex Transm Infect 80:185–191.

497 4. Puhakka L, Sarvikivi E, Lappalainen M, Surcel H-M, Saxen H. 2016. Decrease in
498 seroprevalence for herpesviruses among pregnant women in Finland: cross-sectional study
499 of three time points 1992, 2002 and 2012. Infect Dis 48:406–410.

500 5. Looker KJ, Magaret AS, May MT, Turner KME, Vickerman P, Gottlieb SL, Newman LM.
501 2015. Global and regional estimates of prevalent and incident herpes simplex virus type 1
502 infections in 2012. PLoS ONE 10:e0140765.

503 6. Looker KJ, Magaret AS, Turner KME, Vickerman P, Gottlieb SL, Newman LM. 2015.
504 Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012.
505 PloS One 10:e114989.

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

508 8. Kolb AW, Adams M, Cabot EL, Craven M, Brandt CR. 2011. Multiplex sequencing of seven
509 ocular herpes simplex virus type-1 genomes: phylogeny, sequence variability, and SNP
510 distribution. Invest Ophthalmol Vis Sci 52:9061–73.

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

514 10. Parsons LR, Tafuri YR, Shreve JT, Bowen CD, Shipley MM, Enquist LW, Szpara ML.
515 2015. Rapid genome assembly and comparison decode intrastrain variation in human
516 alphaherpesviruses. mBio 6:e02213-14.

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

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

522 13. Greninger AL, Roychoudhury P, Xie H, Casto A, Cent A, Pepper G, Koelle DM, Huang M-
523 L, Wald A, Johnston C, Jerome KR. 2018. Ultrasensitive Capture of Human Herpes
524 Simplex Virus Genomes Directly from Clinical Samples Reveals Extraordinarily Limited
525 Evolution in Cell Culture 3:12.

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

528 15. Roizman B, Knipe DM, Whitley R. 2013. Herpes Simplex Viruses, p. 1823–1897. *In* Fields
529 Virology, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

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

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

535 18. Kolb AW, Ané C, Brandt CR. 2013. Using HSV-1 genome phylogenetics to track past
536 human migrations. PloS One 8:1–9.

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

540 20. Lee K, Kolb AW, Sverchkov Y, Cuellar JA, Craven M, Brandt CR. 2015. Recombination
541 Analysis of Herpes Simplex Virus 1 Reveals a Bias toward GC Content and the Inverted
542 Repeat Regions. J Virol 89:7214–7223.

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

547 22. Lee K, Kolb AW, Larsen I, Craven M, Brandt CR. 2016. Mapping Murine Corneal
548 Neovascularization and Weight Loss Virulence Determinants in the Herpes Simplex Virus
549 1 Genome and the Detection of an Epistatic Interaction between the UL and IRS/US
550 Regions. J Virol 90:8115–8131.

551 23. Pfaff F, Groth M, Sauerbrei A, Zell R. 2016. Genotyping of herpes simplex virus type 1 by
552 whole-genome sequencing. J Gen Virol 97:2732–2741.

553 24. Shipley MM, Renner DW, Ott M, Bloom DC, Koelle DM, Johnston C, Szpara ML. 2018.
554 Genome-wide surveillance of genital herpes simplex virus type 1 from multiple anatomic
555 sites over time. J Infect Dis 218:595–605.

556 25. Greninger AL, Roychoudhury P, Xie H, Casto A, Cent A, Pepper G, Koelle DM, Huang M-
557 L, Wald A, Johnston C, Jerome KR. 2018. Ultrasensitive Capture of Human Herpes
558 Simplex Virus Genomes Directly from Clinical Samples Reveals Extraordinarily Limited
559 Evolution in Cell Culture. mSphere 3:e00283-18.

560 26. Palo JU, Ulmanen I, Lukka M, Ellonen P, Sajantila A. 2009. Genetic markers and population
561 history: Finland revisited. Eur J Hum Genet 17:1336–1346.

562 27. Kerminen S, Havulinna AS, Hellenthal G, Martin AR, Sarin A-P, Perola M, Palotie A,
563 Salomaa V, Daly MJ, Ripatti S, Pirinen M. 2017. Fine-Scale Genetic Structure in Finland.
564 G3amp58 GenesGenomesGenetics 7:3459–3468.

565 28. Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V, Khartanovich V,
566 Balanovsky O, Ongyerth M, Weihmann A, Sajantila A, Kelso J, Pääbo S, Onkamo P, Haak
567 W, Krause J, Schiffels S. 2018. Ancient Fennoscandian genomes reveal origin and spread
568 of Siberian ancestry in Europe. Nat Commun 9.

569 29. Hannelius U, Salmela E, Lappalainen T, Guillot G, Lindgren CM, von Döbeln U, Lahermo
570 P, Kere J. 2008. Population substructure in Finland and Sweden revealed by the use of
571 spatial coordinates and a small number of unlinked autosomal SNPs. BMC Genet 9:54.

572 30. Ilonen J, Reijonen H, Green A, Reunanen A, Knip M, Simell O, Akerblom HK. 2000.
573 Geographical differences within Finland in the frequency of HLA-DQ genotypes associated
574 with type 1 diabetes susceptibility. The Childhood Diabetes in Finland Study Group. Eur J
575 Immunogenetics Off J Br Soc Histocompat Immunogenetics 27:225–230.

576 31. Paavilainen H, Lehtinen J, Romanovskaya A, Nygårdas M, Bamford DH, Poranen MM,
577 Hukkanen V. 2016. Inhibition of clinical pathogenic herpes simplex virus 1 strains with
578 enzymatically created siRNA pools. J Med Virol 88:2196–2205.

579 32. Ziegler T, Waris M, Rautiainen M, Arstila P. 1988. Herpes simplex virus detection by
580 macroscopic reading after overnight incubation and immunoperoxidase staining. J Clin
581 Microbiol 26:2013–2017.

582 33. Hukkanen V, Rehn T, Kajander R, Sjöroos M, Waris M. 2000. Time-resolved fluorometry
583 PCR assay for rapid detection of herpes simplex virus in cerebrospinal fluid. J Clin
584 Microbiol 38:3214–3218.

585 34. Romanovskaya A, Paavilainen H, Nygårdas M, Bamford DH, Hukkanen V, Poranen MM.
586 2012. Enzymatically produced pools of canonical and Dicer-substrate siRNA molecules
587 display comparable gene silencing and antiviral activities against herpes simplex virus.
588 PloS One 7:e51019.

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

591 36. Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE. 1988.
592 Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell
593 line. J Cell Biol 106:761–771.

594 37. Post LE, Mackem S, Roizman B. 1981. Regulation of alpha genes of herpes simplex virus:
595 expression of chimeric genes produced by fusion of thymidine kinase with alpha gene
596 promoters. Cell 24:555–565.

597 38. Tardif KD, Jorgensen S, Langer J, Prichard M, Schlaberg R. 2014. Simultaneous titration
598 and phenotypic antiviral drug susceptibility testing for herpes simplex virus 1 and 2. J Clin
599 Virol 61:382–386.

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

603 40. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak
604 MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S. 2011. Sequence-
605 specific error profile of Illumina sequencers. Nucleic Acids Res 39:1–13.

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

610 42. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM.
611 2012. A program for annotating and predicting the effects of single nucleotide
612 polymorphisms, SnpEff. Fly (Austin) 6:80–92.

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

616 44. Igarashi K, Fawl R, Roller RJ, Roizman B. 1993. Construction and properties of a
617 recombinant herpes simplex virus 1 lacking both S-component origins of DNA synthesis. J
618 Virol 67:2123–2132.

619 45. Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M,
620 Goldman D, Long JC. 1998. Dual Origins of Finns Revealed by Y Chromosome Haplotype
621 Variation. Am J Hum Genet 62:1171–1179.

622 46. Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L. 2003. The
623 interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in
624 chromosomes of Finnish populations with different histories. Hum Mol Genet 12:51–59.

625 47. Szpara ML, Tafuri YR, Parsons L, Shamim SR, Verstrepen KJ, Legendre M, Enquist LW.
626 2011. A wide extent of inter-strain diversity in virulent and vaccine strains of
627 alphaherpesviruses. PLoS Pathog 7:1–23.

628 48. Norberg P, Olofsson S, Tarp MA, Clausen H, Bergstrom T, Liljeqvist JA. 2007.
629 Glycoprotein I of herpes simplex virus type 1 contains a unique polymorphic tandem-
630 repeated mucin region. J Gen Virol 88:1683–1688.

631 49. Reynolds AE, Wills EG, Roller RJ, Ryckman BJ, Baines JD. 2002. Ultrastructural
632 localization of the herpes simplex virus type 1 UL31, UL34, and US3 proteins suggests
633 specific roles in primary envelopment and egress of nucleocapsids. J Virol 76:8939–52.

634 50. Roller RJ, Bjerke SL, Haugo AC, Hanson S. 2010. Analysis of a Charge Cluster Mutation of
635 Herpes Simplex Virus Type 1 UL34 and Its Extragenic Suppressor Suggests a Novel

636 Interaction between pUL34 and pUL31 That Is Necessary for Membrane Curvature around
637 Capsids. J Virol 84:3921–3934.

638 51. Davison AJ, Wilkie NM. 1981. Nucleotide sequences of the joint between the L and S
639 segments of herpes simplex virus types 1 and 2. J Gen Virol 55:315–31.

640 52. Wathen MW, Pirtle EC. 1984. Stability of the pseudorabies virus genome after in vivo serial
641 passage. J Gen Virol 65:1401–4.

642 53. Umene K, Kawana T. 2003. Divergence of reiterated sequences in a series of genital isolates
643 of herpes simplex virus type 1 from individual patients. J Gen Virol 84:917–923.

644 54. Buchman TG, Simpson T, Nosal C, Roizman B. 1980. The Stucture of Herpes Simplex Virus
645 DNA and its Application to Molecular Epidemiology. Ann N Y Acad Sci 354:279–290.

646 55. Rastrojo A, López-Muñoz AD, Alcamí A. 2017. Genome Sequence of Herpes Simplex Virus
647 1 Strain SC16. Genome Announc 5:e01392-16.

648 56. Skoblov MY, Lavrov AV, Bragin AG, Zubtsov DA, Andronova VL, Galegov GA, Skoblov
649 YS. 2017. The genome nucleotide sequence of herpes simplex virus 1 strain L2. Russ J
650 Bioorganic Chem 43:140–142.

651 57. Petro CD, Weinrick B, Khajoueinejad N, Burn C, Sellers R, Jacobs WR, Herold BC. 2016.
652 HSV-2 ΔgD elicits FcγR-effector antibodies that protect against clinical isolates. JCI
653 Insight 1.

654 58. Bondre VP, Sankararaman V, Andhare V, Tupekar M, Sapkal GN. 2016. Genetic
655 characterization of human herpesvirus type 1: Full-length genome sequence of strain
656 obtained from an encephalitis case from India. Indian J Med Res 144:750–760.

657 59. Parsons LR, Tafuri YR, Shreve JT, Bowen CD, Shipley MM, Enquist LW, Szpara ML.
658 2015. Rapid Genome Assembly and Comparison Decode Intrastrain Variation in Human
659 Alphaherpesviruses. mBio 6:e02213-14.

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

663

664 # Tables

665 **Table 1: The Finnish HSV-1 strains used for genome sequence comparisons.**

Sample	Gender	Code	Age	Lesion or diagnosis	Acyclovir sensitivity, IC50 value (µg/ml)
H1211	female	F-11	21	blister	0.25
H1215	male	M-15	5	blister	0.03
H12113	female	F-13	29	blister	0.11
H12114	female	F-14g	23	genital blister	0.50
H12117	female	F-17	57	blister	0.15
H12118	female	F-18g	19	genital blister	0.13
H1311	female	F-11*l*	59	blister (lip)	0.31
H1312	male	M-12	39	blister	0.13
H1412	female	F-12g	28	genital lesion	0.09
H15119	male	M-19	36	blister	0.15

666 *The IC50 value for HSV-1 delta305 (a TK-negative, ACV-resistant control) was 8.08 and for
667 the reference strain HSV-1(17+) it was 0.13. The limit of resistance was ≥2.0 µg/ml.

668

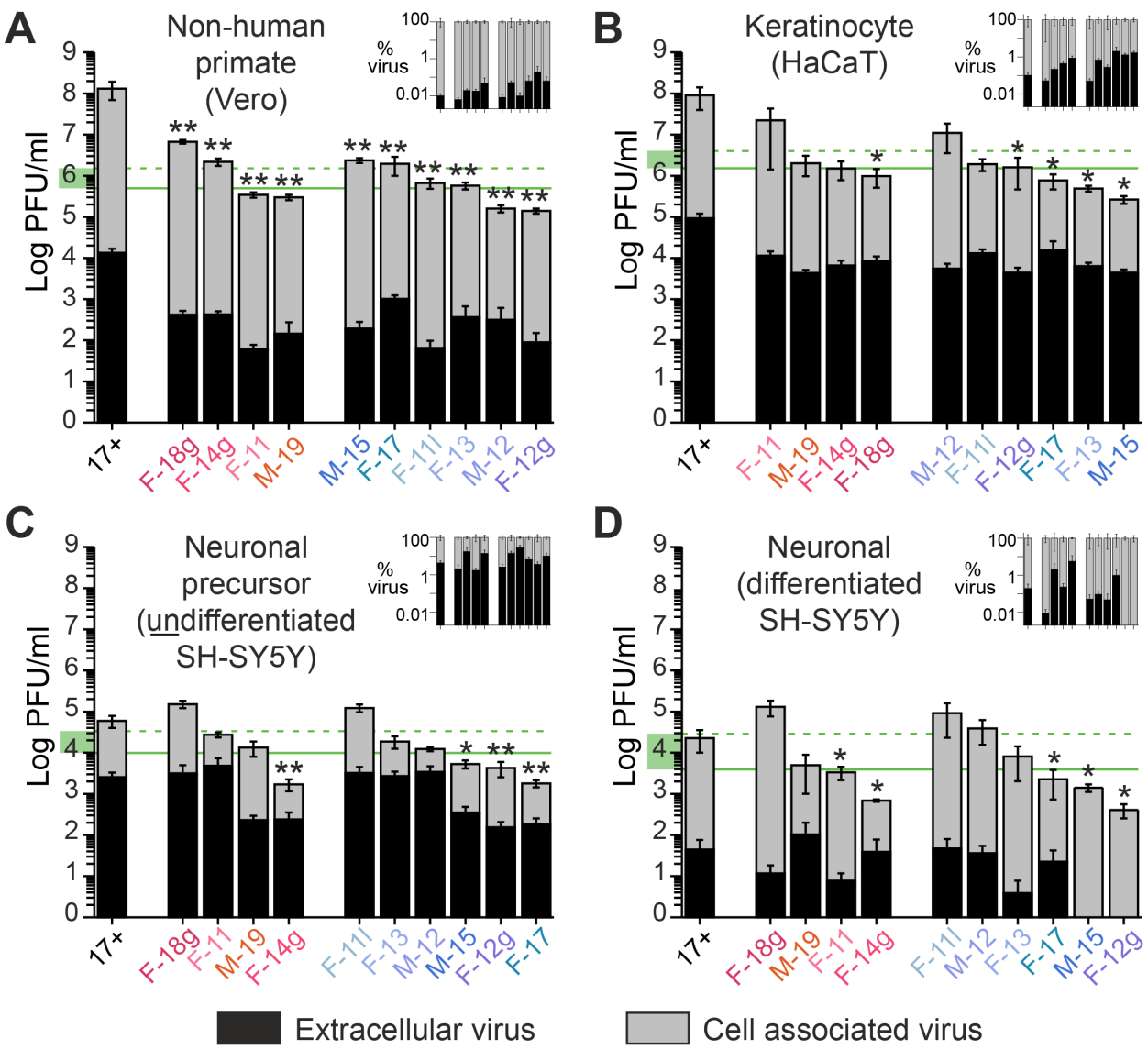
669 **Table 2. Sequencing statistics for Finnish HSV-1 strains.**

Sample	Code	% Depth ≥100	Average coverage	Raw Sequence Reads	Used for Assembly	% Viral	GenBank Accession^
H1211	F-11	96%	1,027x	4.2 million	2.8 million	66%	MH999843
H1215	M-15	99%	2,808x	5.5 million	3.8 million	70%	MH999846
H12113	F-13	98%	2,354x	3.8 million	2.4 million	64%	MH999842
H12114	F-14g	98%	2,078x	5.9 million	3.8 million	66%	MH999844
H12117	F-17	98%	1,976x	4.1 million	2.7 million	66%	MH999845
H12118	F-18g	98%	1,678x	4.2 million	2.9 million	69%	MH999847
H1311	F-11*l*	99%	2,541x	4.6 million	3.0 million	65%	MH999848
H1312	M-12	98%	2,415x	6.0 million	3.7 million	61%	MH999849
H1412	F-12g	98%	1,524x	4.5 million	3.1 million	69%	MH999851
H15119	M-19	99%	1,536x	4.8 million	3.0 million	64%	MH999850

670 ^See GenBank flatfile attached for review purposes.

671

672 # Figures & Figure Legends



673

674 **Figure 1: Comparison of growth properties of Finnish HSV-1 strains reveals the**

675 **uncoupling of phenotypic and geographic variation.**

676 Histograms comparing virus production for ten Finnish strains of HSV-1, as compared to the

677 HSV-1 reference strain 17+, in different cell types. Panel **(A,** non-human primate**)** shows Vero

678 monkey kidney cells, which are defective in interferon signaling; **(B)** shows HaCaT human
679 keratinocytes; **(C)** shows undifferentiated human SH-SY5Y neuroblastoma cells (neuronal
680 precursor cells); **(D)** shows differentiated human SH-SY5Y neuronal-like cells. Viruses are
681 plotted in descending order based on their viral production. For ease of comparison, the strains
682 are color-coded and divided into the genomic sub-groups used in later figures (see **Figure 3** for
683 illustration). Each histogram bar plots infectious virions in the cell-associated fraction (gray) vs.
684 those released from cells (black). A horizontal line shows the median (solid green) and the
685 average (dashed green) amount of virus production for the ten Finnish HSV-1 strains in each cell
686 type, with a highlighted box (green) on this y-axis added to highlight the multi-log decrease in
687 viral production in **C-D** vs. **A-B**. Each panel has a smaller graph (**inset**) where the extracellular
688 released virus (black) is plotted as a percentage of total virus production (gray). Viruses in each
689 inset are plotted in the same order as the matched panel graph. Bar shows SEM; * = p=<0.05, **
690 = p<0.001 as compared to the total virus amount (cell associated + extracellular), for each
691 clinical isolate vs HSV-1(17+) reference strain.

692

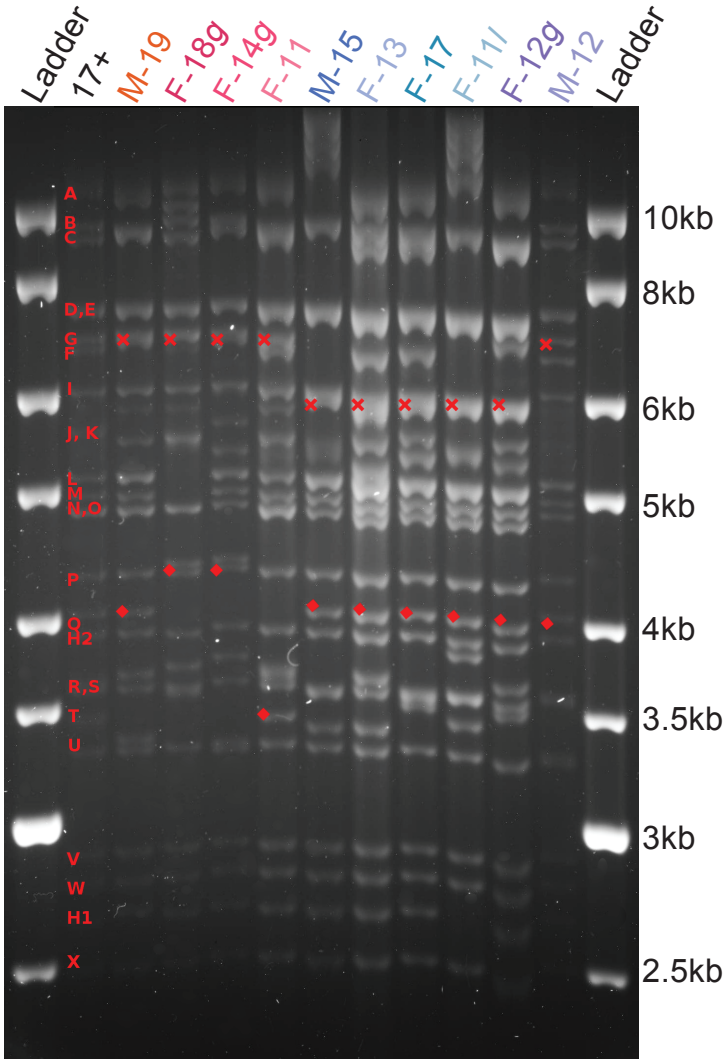
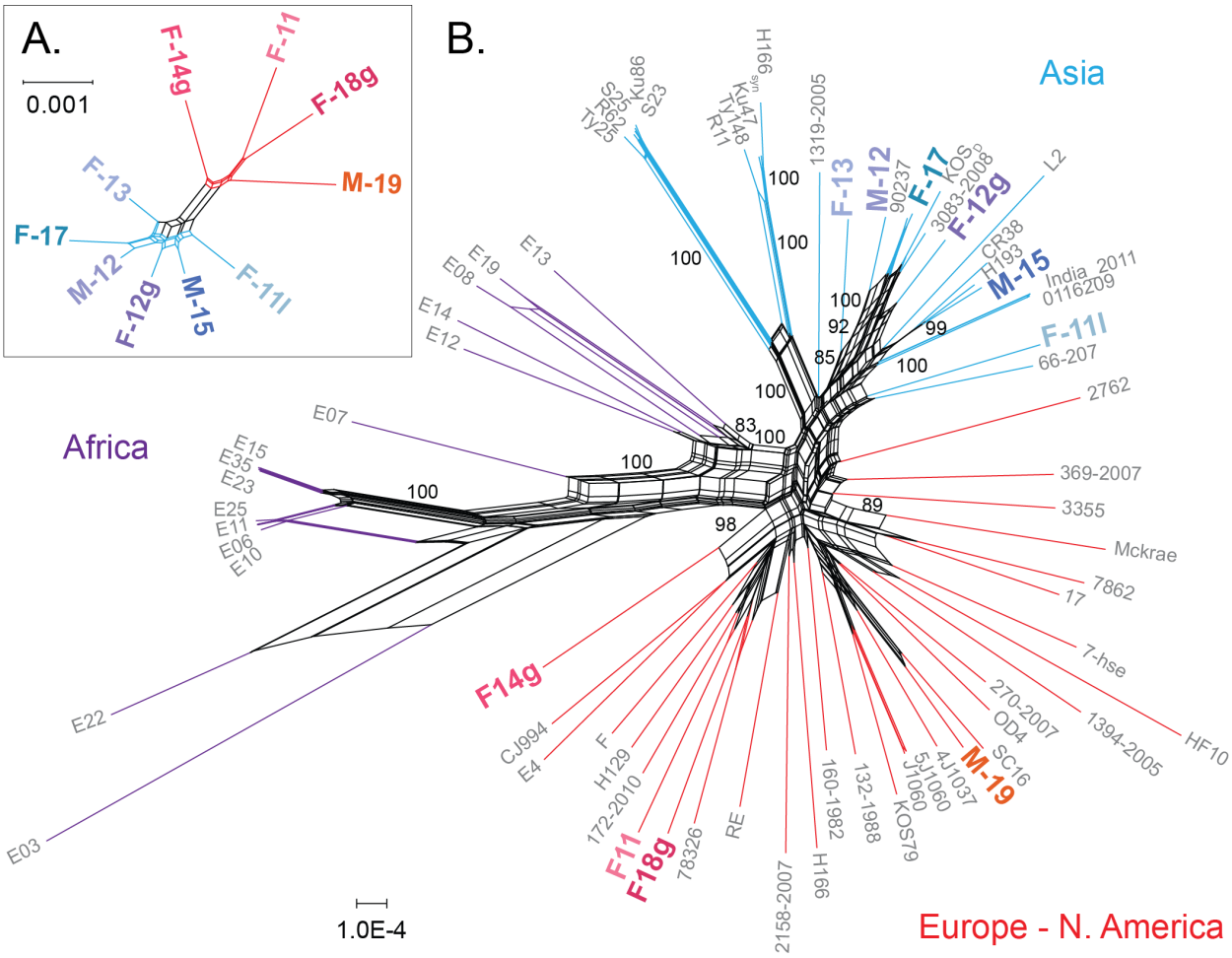


Figure 2: RFLP comparison reveals at least two sub-groups of Finnish HSV-1 strains.

Viral genomic DNA was digested with SalI and separated via electrophoresis to distinguish overall genetic patterns. Classically defined SalI fragment notations (red alphabet letters) are indicated on the left (54). Red crosses (x) indicate variation in the size of SalI fragment G, while diamonds indicate variation in the SalI Q fragment. For ease of comparison, the strains are color-coded the same as in later figures (see **Figure 3** for illustration).



701

702 **Figure 3. A phylogenetic network of genetic relatedness reveals that Finland HSV-1**

703 **isolates separate into two sub-groups.**

704 A phylogenetic network of genetic relatedness was constructed using SplitsTree4 to demonstrate

705 how the ten new Finland HSV-1 isolates relate to each other (**A**) and to previously described

706 HSV-1 genomes from a diversity of locations (**B**). Geographic clustering of branches and clades

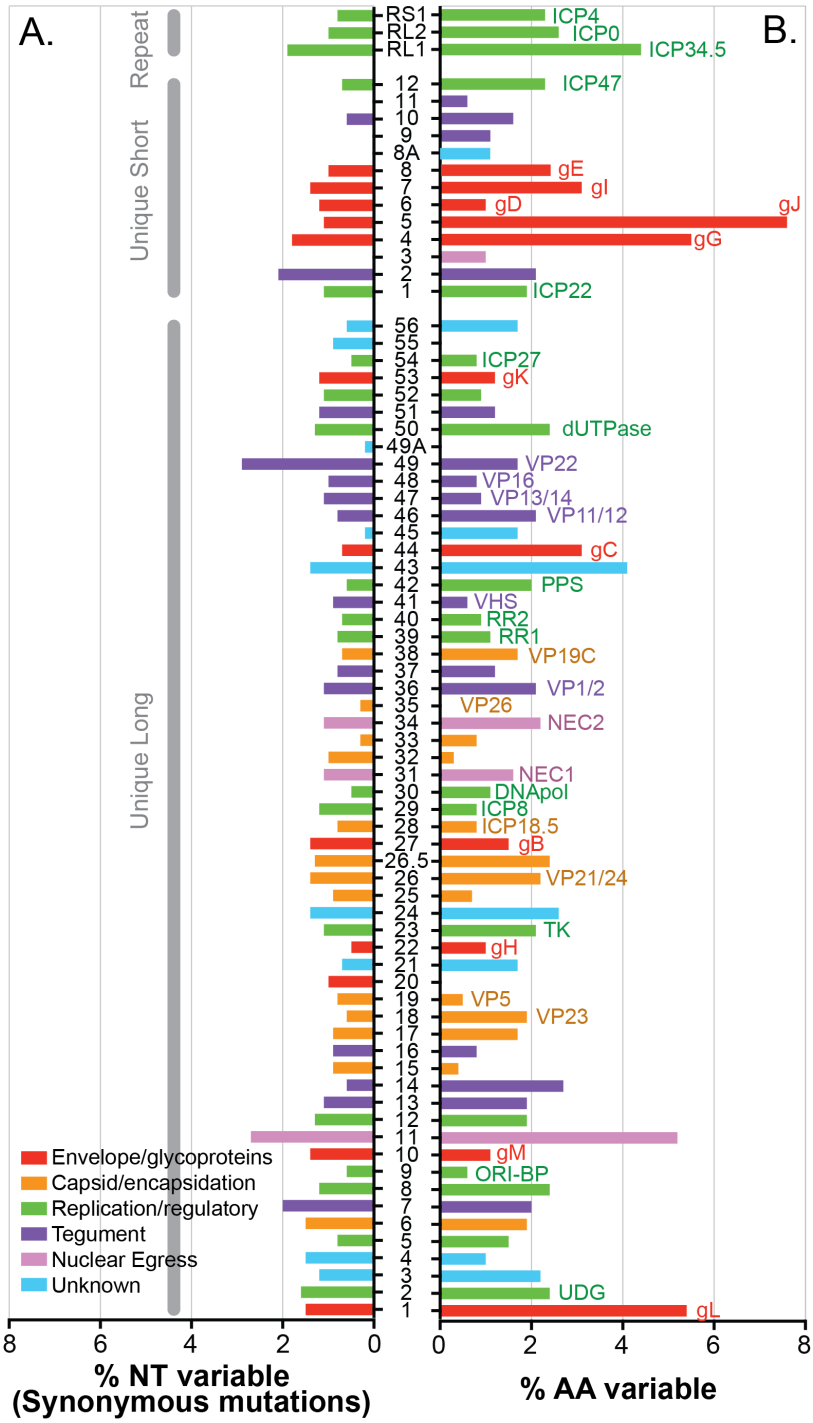
707 resemble those previously described (9, 16–18) and are colored by origin of strains: red for

708 Europe and North America, blue for Asia, and purple for Africa. Finland HSV-1 genomes are

709 indicated in bold, and separate into two sub-groups (**A**), which appear to relate previously

710 recognized Asian and European and North American clades (**B**). Bootstrap values larger than 80

711 supporting clades with three or more strains are shown on the network. The extent of
712 recombination among HSV-1 strains is demonstrated by the parallel edges in the network and a
713 phi-test for recombination. The isolate name, country of origin, GenBank accession, and
714 reference(s) for each previously described HSV-1 genome are listed in **Supplemental Table S1**.



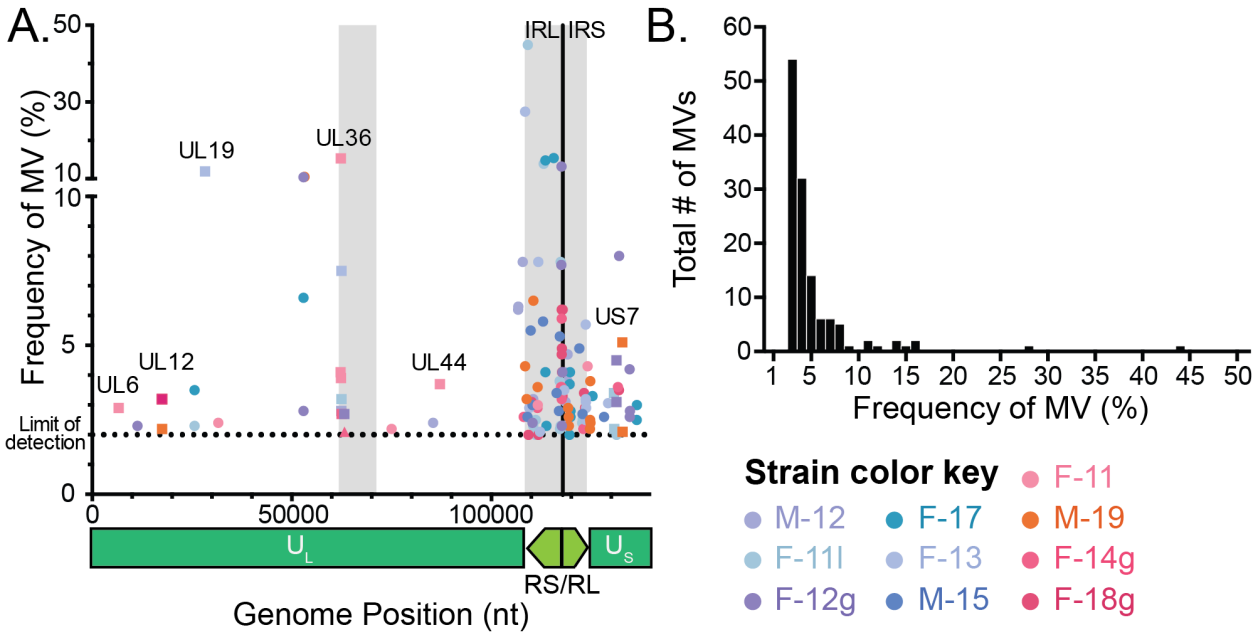
715

716 **Figure 4. Substantial protein-coding variation exists between Finland HSV-1**

717 **isolates.**

718 Histogram depicts the percent of synonymous nucleotide differences per gene (**A**) and the

719 percent amino acid (AA) variation per protein (**B**) for the ten Finnish HSV-1 strains compared
720 here. Labels on the left in (**A**) indicate the repeat, unique short (US) and unique long (UL)
721 regions that contain each of the sequentially-numbered viral coding regions. Where possible, the
722 common name of each protein is shown to the right of the appropriate histogram bar in (**B**), e.g.
723 UL1 is also known as gL. Nucleotide and AA differences were quantified at the consensus level
724 of each viral genome. See **Supplemental Table S2** for additional data on the number of
725 nucleotide differences and dN/dS ratio for each gene.



726

727 **Figure 5. Scatter plot illustrates the location and low frequency of minor variants in**

728 **each Finnish HSV-1 genome.**

729 **(A)** Scatter plot demonstrates the position (x-axis) and frequency (y-axis) of minority variants

730 detected in each Finnish HSV-1 genome. Spheres on the graph indicate minor variants in

731 intergenic regions. Squares with gene names indicate MVs located in genes. A color key

732 consistent with that used other figures is included at the lower right. A genome diagram is shown

733 below the x-axis for reference. Areas containing large numbers of tandem repeats, e.g. around

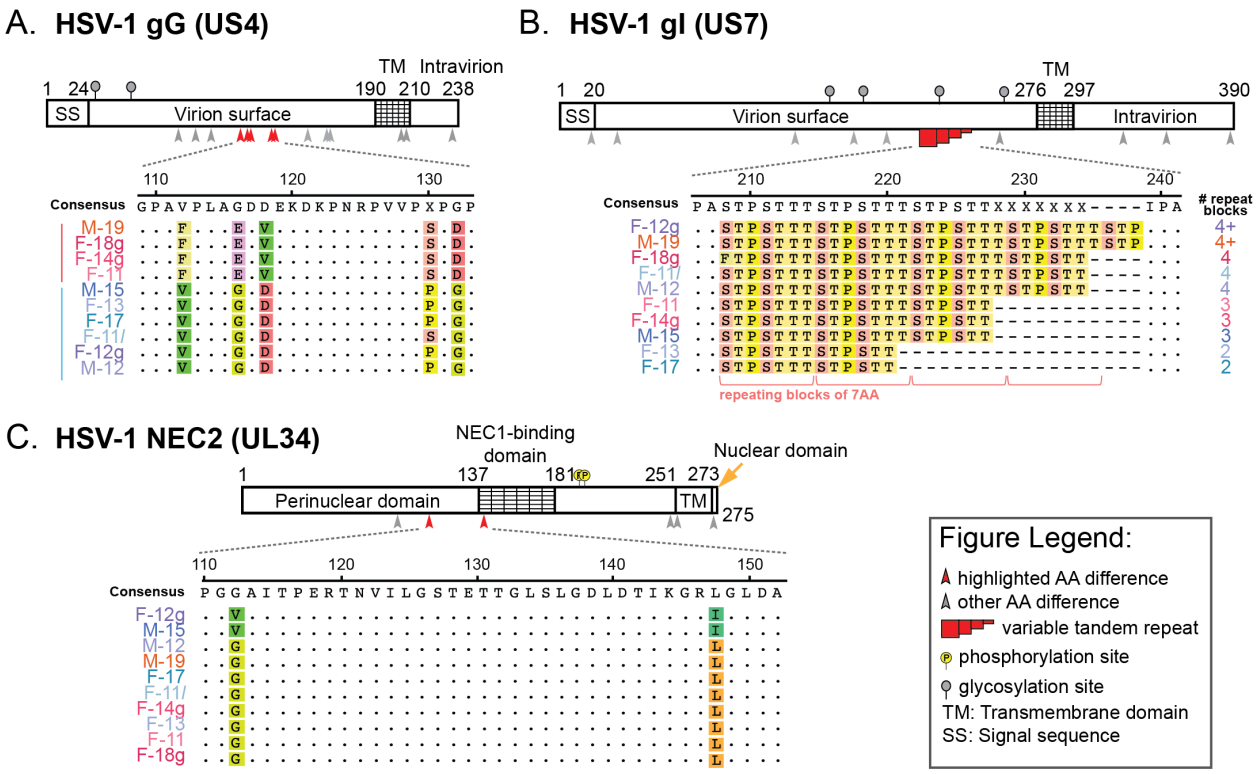
734 UL36 in the U_L region and throughout the RS/RL region, are denoted as gray blocks on the

735 graph. The limit of detection was set at 2% (see Methods for details). **(B)** Histogram depicts the

736 total number of minority variants (y-axis) in each frequency bin (x-axis), summed across all ten

737 Finnish HSV-1 strains. Bins are in 1% increments.

738



739

740 **Figure 6. Variations in individual genes provide a better insight into Finnish HSV-1**

741 **strain sub-groups and of specific cellular phenotypes.**

742 Examples are shown here of variation in individual HSV-1 genes that reflect either the Finnish

743 HSV-1 strain sub-groups, or that correlate with and potentially influence specific cellular

744 phenotypes. Diagrams depict three examples of HSV-1 proteins with known functional domains

745 and post-translational modifications (e.g. phosphorylation, glycosylation; see Legend). A subset

746 of each protein is highlighted via an AA alignment of the ten clinical HSV-1 isolates. (**A**) Coding

747 differences in the secreted and virion surface glycoprotein G (gG; US4) correlate with the overall

748 genomic sub-groups (clades) shown in **Figure 3**. (**B**) Variation in copy number of a seven AA

749 tandem repeat block in glycoprotein I (gI; US7) changes the length of a repeating tract of mucin-

750 type O-linked glycosylation sites, as described previously in clinical isolates of HSV-1 (16, 48).

751 (**C**) AA differences in the nuclear egress protein NEC2 (UL34) were observed in the two strains

752 (M-15, F-12g) which showed no detectable extracellular virion release in differentiated neuronal
753 cells (**Figure 1D**). See Discussion for additional insights on the potential phenotypic impacts of
754 the variations shown in B-C.

755

Supplemental Tables

Supplemental Table S1: List of previously published HSV-1 genomes used for phylogenetic analyses

Virus Isolate	Country (with location detail, if available)	GenBank Accession #	References
H1211 / F-11*	Finland	MH999843	(31), present
H1215 / M-15*	Finland	MH999846	(31), present
H12113 / F-13*	Finland	MH999842	present
H12114 / F-14g*	Finland	MH999844	(31), present
H12117 / F-17*	Finland	MH999845	(31), present
H12118 / F-18g*	Finland	MH999847	(31), present
H1311 / F11*l*	Finland	MH999848	present
H1312 / M-12*	Finland	MH999849	present
H1412 / F-12g*	Finland	MH999851	present
H15119 / M-19*	Finland	MH999850	present
SC16	Spain (Madrid)	KX946970	(55)
172/2010	Germany	LT594105	(12)
2158/2007	Germany	LT594106	(12)
3083/2008	Germany	LT594107	(12)
1319/2005	Germany	LT594108	(12)
270/2007	Germany	LT594109	(12)
66/2007	Germany	LT594110	(12)
1394/2005	Germany	LT594111	(12)
369/2007	Germany	LT594112	(12)
160/1982	Germany	LT594192	(12)
132/1998	Germany	LT594457	(12)
L2	Russia (Moscow)	KT780616	(56)
B^3x1.1	U.S.A. (Bronx, NY)	KU310657	(57)
B^3x1.2	U.S.A. (Bronx, NY)	KU310658	(57)
B^3x1.3	U.S.A. (Bronx, NY)	KU310659	(57)
B^3x1.4	U.S.A. (Bronx, NY)	KU310660	(57)
B^3x1.5	U.S.A. (Bronx, NY)	KU310661	(57)
H193	U.S.A.	KT425108	n/a
KOS79	U.S.A. (Madison, WI)	KT425109	(11)
CJ994	U.S.A. (Madison, WI)	KR011283	(20)
HSV-1/0116209/India/2011	India	KJ847330	(58)
H166	U.S.A.	KM222726	(59)
H166syn	U.S.A.	KM222727	(59)
RE	U.S.A.	KF498959	n/a
OD4	U.S.A. (Madison, WI)	JN420342	(8)
17	U.K. (Glasgow)	JN555585	(9)
CR38	China (Shenyang)	HM585508	(9)

Virus Isolate	Country (with location detail, if available)	GenBank Accession #	References
E03	Kenya (Nairobi)	HM585509	(9)
E06	Kenya (Nairobi)	HM585496	(9)
E07	Kenya (Nairobi)	HM585497	(9)
E08	Kenya (Nairobi)	HM585498	(9)
E10	Kenya (Nairobi)	HM585499	(9)
E11	Kenya (Nairobi)	HM585500	(9)
E12	Kenya (Nairobi)	HM585501	(9)
E13	Kenya (Nairobi)	HM585502	(9)
E14	Kenya (Nairobi)	HM585510	(9)
E15	Kenya (Nairobi)	HM585503	(9)
E19	Kenya (Nairobi)	HM585511	(9)
E22	Kenya (Nairobi)	HM585504	(9)
E23	Kenya (Nairobi)	HM585505	(9)
E25	Kenya (Nairobi)	HM585506	(9)
E35	Kenya (Nairobi)	HM585507	(9)
R11	South Korea (Seoul)	HM585514	(9)
R62	South Korea (Seoul)	HM585515	(9)
S23	Japan (Sapporo)	HM585512	(9)
S25	Japan (Sapporo)	HM585513	(9)
F	U.S.A. (Chicago, IL)	GU734771	(9)
H129	U.S.A. (San Francisco, CA)	GU734772	(9)
McKrae	U.S.A. (Gainesville, FL)	JQ730035, JX142173	(9)
KOS	U.S.A. (Houston, TX)	JQ673480, JQ780693	(9)
HF10	U.S.A. (New York, NY)	DQ889502	(9)
Ty_25	Japan	MH999840	n/a
Ty_148	Japan	MH999841	n/a
K_86	Japan	MH999839	n/a
K_47	Japan	MH999838	n/a
7-hse [†]	Sweden or U.S.A.	SRX056767	(60)
7862 [†]	Sweden or U.S.A.	SRX056770	(60)
3355 [†]	Sweden or U.S.A.	SRX056769	(60)
2762 [†]	Sweden or U.S.A.	SRX056768	(60)
90237 [†]	Sweden or U.S.A.	SRX056772	(60)
E4 [†]	Sweden or U.S.A.	SRX056773	(60)
78326 [†]	Sweden or U.S.A.	SRX056771	(60)
4-J1037 [†]	Sweden or U.S.A.	SRX056760	(60)
J1061 [†]	Sweden or U.S.A.	SRX056774	(60)
5-J1061	Sweden or U.S.A.	n/a	(60)

759 [*] newly sequenced strains from Finland in this study.

760 [†] Sequence data deposited in GenBank short read archive

761

762 **Supplemental Table S2: Number of nucleotide and amino acid (AA) differences observed in each set of Finnish or**

763 **worldwide published HSV1 genomes.**

Gene*	ORF length in NT^	protein length (#AA)	10 Finnish # NT diff.^	10 Finnish # AA changes	10 Finnish %AA diff.	10 Finnish dN/dS	All HSV1 # NT diff.	All HSV1 # AA changes	All HSV1 %AA diff.	All HSV1 dN/dS ratio
UL1	675	224	22	12	5.4	1.20	58	31	13.8	1.15
UL2	1005	334	24	8	2.4	0.50	62	28	8.4	0.82
UL3	675	224	13	5	2.2	0.63	43	19	8.5	0.79
UL4	600	199	11	2	1.0	0.22	43	18	9.0	0.72
UL5	2649	882	35	13	1.5	0.59	106	38	4.3	0.56
UL6	2031	676	43	13	1.9	0.43	107	30	4.4	0.39
UL7	891	296	24	6	2.0	0.33	47	17	5.7	0.57
UL8	2253	750	44	18	2.4	0.69	124	62	8.3	1.00
UL9	2556	851	21	5	0.6	0.31	110	32	3.8	0.41
UL10	1422	473	25	5	1.1	0.25	81	30	6.3	0.59
UL11	291	96	13	5	5.2	0.63	26	17	17.7	1.89
UL12	1881	626	36	12	1.9	0.50	87	30	4.8	0.53
UL13	1557	518	27	10	1.9	0.59	72	33	6.4	0.85
UL14	660	219	10	6	2.7	1.50	27	15	6.8	1.25
UL15	2208	735	22	3	0.4	0.16	80	16	2.2	0.25
UL16	1122	373	13	3	0.8	0.30	55	19	5.1	0.53
UL17	2112	703	30	12	1.7	0.67	108	39	5.5	0.57
UL18	957	318	12	6	1.9	1.00	41	15	4.7	0.58
UL19	4125	1374	39	7	0.5	0.22	139	40	2.9	0.40
UL20	669	222	7	0	0.0	0.00	29	9	4.1	0.45
UL21	1608	535	20	9	1.7	0.82	70	25	4.7	0.56
UL22	2517	838	21	8	1.0	0.62	120	50	6.0	0.71
UL23	1131	376	21	8	2.1	0.62	63	26	6.9	0.70
UL24	810	269	18	7	2.6	0.64	60	28	10.4	0.88

Gene*	ORF length in NT^	protein length (#AA)	10 Finnish # NT diff.^	10 Finnish # AA changes	10 Finnish %AA diff.	10 Finnish dN/dS	All HSV1 # NT diff.	All HSV1 # AA changes	All HSV1 %AA diff.	All HSV1 dN/dS ratio
UL25	1743	580	19	4	0.7	0.27	74	25	4.3	0.51
UL26	1908	635	40	14	2.2	0.54	92	34	5.4	0.59
UL26.5	990	329	21	8	2.4	0.62	45	17	5.2	0.61
UL27	2715	904	51	14	1.5	0.38	123	40	4.4	0.48
UL28	2358	785	26	6	0.8	0.30	87	23	2.9	0.36
UL29	3591	1196	54	10	0.8	0.23	154	33	2.8	0.27
UL30	3708	1235	32	13	1.1	0.68	154	49	4.0	0.47
UL31	921	306	15	5	1.6	0.50	34	13	4.2	0.62
UL32	1791	596	20	2	0.3	0.11	85	28	4.7	0.49
UL33	393	130	2	1	0.8	1.00	14	4	3.1	0.40
UL34	828	275	15	6	2.2	0.67	37	14	5.1	0.61
UL35	339	112	1	0	0.0	0.00	8	1	0.9	0.14
UL36	9420	3139	166	66	2.1	0.66	1271	440	14.0	0.53
UL37	3372	1123	41	13	1.2	0.46	140	60	5.3	0.75
UL38	1398	465	18	8	1.7	0.80	70	30	6.5	0.75
UL39	3414	1137	42	13	1.1	0.45	157	61	5.4	0.64
UL40	1023	340	10	3	0.9	0.43	38	12	3.5	0.46
UL41	1470	489	16	3	0.6	0.23	52	22	4.5	0.73
UL42	1467	488	19	10	2.0	1.11	76	39	8.0	1.05
UL43	1254	417	35	17	4.1	0.94	90	50	12.0	1.25
UL44	1536	511	26	16	3.1	1.60	97	55	10.8	1.31
UL45	519	172	4	3	1.7	3.00	22	8	4.7	0.57
UL46	2157	718	33	15	2.1	0.83	111	65	9.1	1.41
UL47	2082	693	29	6	0.9	0.26	81	29	4.2	0.56
UL48	1473	490	18	4	0.8	0.29	69	25	5.1	0.57
UL49	276	301	13	5	1.7	0.63	42	18	6.0	0.75
UL49A	906	91	2	0	0.0	0.00	8	3	3.3	0.60
UL50	1116	371	23	9	2.4	0.64	65	26	7.0	0.67

Gene*	ORF length in NT^	protein length (#AA)	10 Finnish # NT diff.^	10 Finnish # AA changes	10 Finnish %AA diff.	10 Finnish dN/dS	All HSV1 # NT diff.	All HSV1 # AA changes	All HSV1 %AA diff.	All HSV1 dN/dS ratio
UL51	735	244	12	3	1.2	0.33	33	13	5.3	0.65
UL52	3177	1058	43	9	0.9	0.26	129	41	3.9	0.47
UL53	1017	338	16	4	1.2	0.33	36	12	3.6	0.50
UL54	1539	512	11	4	0.8	0.57	73	34	6.6	0.87
UL55	561	186	5	0	0.0	0.00	25	9	4.8	0.56
UL56	705	234	8	4	1.7	1.00	34	19	8.1	1.27
US1	1263	420	22	8	1.9	0.57	75	39	9.3	1.08
US2	876	291	24	6	2.1	0.33	44	14	4.8	0.47
US3	1446	481	14	5	1.0	0.56	57	21	4.4	0.58
US4	717	238	26	13	5.5	1.00	68	36	15.1	1.13
US5	279	92	10	7	7.6	2.33	25	17	18.5	2.13
US6	1185	394	18	4	1.0	0.29	67	18	4.6	0.37
US7	1173	390	28	12	3.1	0.75	119	57	14.6	0.92
US8	1653	550	30	13	2.4	0.76	79	38	23.9	0.93
US8A	480	159	3	3	1.9	1.00	49	15	2.7	0.44
US9	273	90	1	1	1.1	1.00	12	4	4.4	0.50
US10	939	312	11	5	1.6	0.83	65	29	9.3	0.81
US11	486	161	1	1	0.6	1.00	42	19	11.8	0.83
US12	267	88	4	2	2.3	1.00	17	10	11.4	1.43
RL1	747	248	25	11	4.4	0.79	422	90	36.3	0.27
RL2	2328	775	44	20	2.6	0.83	180	91	11.7	1.02
RS1	3897	1298	62	30	2.3	0.94	345	207	15.9	1.50

764 * See **Supplemental Table S1** for list of strain names and genome accessions for comparisons of 10 Finnish and all published HSV1 genomes.
765 See **Supplemental Table S3** for list of strains excluded from selected gene alignments due to missing data in GenBank.
766 ^ Abbreviations: ORF, open reading frame; NT, nucleotide; AA, amino acid; diff., difference; %, percent; dN/dS, ratio of nonsynonymous to
767 synonymous nucleotide changes

768 **Supplemental Table S3: HSV-1 strains used to calculate nucleotide and amino acid**

769 **(AA) differences in Supplemental Table S2 and Figure 4.**

Gene	# of strains in nucleotide alignment	# of strains in AA alignment	Strains not used in nucleotide and AA alignments*
UL7	60	60	B^3x1.1, B^3x1.2, B^3x1.3, B^3x1.4, B^3x1.5
UL12	61	61	B^3x1.3, B^3x1.4, B^3x1.5, RE
UL13	58	58	E06, E25, E11, OD4, B^3x1.3, B^3x1.4, B^3x1.5
UL15	60	60	B^3x1.1, B^3x1.2, B^3x1.3, B^3x1.4, B^3x1.5
UL19	61	61	B^3x1.3, B^3x1.4, B^3x1.5, RE
UL26.5	61	61	B^3x1.3, B^3x1.4, B^3x1.5, HF10
UL41	61	61	B^3x1.3, B^3x1.4, B^3x1.5, OD4
UL43	61	61	B^3x1.3, B^3x1.4, B^3x1.5, HF10
UL48	60	60	B^3x1.1, B^3x1.2, B^3x1.3, B^3x1.4, B^3x1.5
UL49A	51	51	B^3x1.3, B^3x1.4, B^3x1.5, HF10, 66_2007, 369_2007, 160_1982, 1394_2005, 132_1998, 3083_2008, 2158_2007, 172_2010, 270_2007, 1319_2005
UL55	59	59	E35, E13, KUTy25, B^3x1.3, B^3x1.4, B^3x1.5
UL56	60	60	B^3x1.3, B^3x1.4, B^3x1.5, E35, HF10
US1	60	60	B^3x1.1, B^3x1.2, B^3x1.3, B^3x1.4, B^3x1.5
US8A	61	61	HF10, B^3x1.3, B^3x1.4, B^3x1.5
RL1	58	58	B^3x1.3, B^3x1.4, B^3x1.5, OD4, 1394_2005, CJ994, L2
RL2	61	61	B^3x1.3, B^3x1.4, B^3x1.5, B^3x1.2
RS1	59	59	B^3x1.3, B^3x1.4, B^3x1.5, S23, CJ994, L2
All other genes	62	62	B^3x1.3, B^3x1.4, B^3x1.5

770 * Strains were excluded due to missing data (e.g. due to sequencing gaps) or annotations in
771 GenBank.
772