

RESEARCH ARTICLE

Past and ongoing adaptation of human cytomegalovirus to its host

Alessandra Mozzi¹, Matteo Biolatti², Rachele Cagliani¹, Diego Forni¹,
Valentina Dell'Oste², Chiara Pontremoli¹, Chiara Vantaggiato³, Uberto Pozzoli¹,
Mario Clerici^{4,5}, Santo Landolfo², Manuela Sironi^{1*}

1 Scientific Institute, IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy, **2** Laboratory of Pathogenesis of Viral Infections, Department of Public Health and Pediatric Sciences, University of Turin, Turin, Italy, **3** Scientific Institute, IRCCS E. MEDEA, Laboratory of Molecular Biology, Bosisio Parini, Italy, **4** Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, **5** Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

☯ These authors contributed equally to this work.

* manuela.sironi@BP.LNF.it



OPEN ACCESS

Citation: Mozzi A, Biolatti M, Cagliani R, Forni D, Dell'Oste V, Pontremoli C, et al. (2020) Past and ongoing adaptation of human cytomegalovirus to its host. *PLoS Pathog* 16(5): e1008476. <https://doi.org/10.1371/journal.ppat.1008476>

Editor: Edward Mocarski, Emory University Vaccine Center, UNITED STATES

Received: September 30, 2019

Accepted: March 13, 2020

Published: May 8, 2020

Copyright: © 2020 Mozzi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequences used in this manuscript are publicly accessible through the NCBI database (<http://www.ncbi.nlm.nih.gov/>). The GenBank Accession numbers of all sequences used in this manuscript are listed in the Supporting Information ([S1 Table](#), [S6 Table](#) and [S7 Table](#)).

Funding: This work was supported by the Italian Ministry of Health ("Ricerca Corrente 2019-2020" to MS, "Ricerca Corrente 2018-2020" to DF) and by the Italian Ministry of Education, University and Research ("PRIN 2015", 2015RMNSTA to VDO). The funders had no role in study design, data

Abstract

Cytomegaloviruses (order *Herpesvirales*) display remarkable species-specificity as a result of long-term co-evolution with their mammalian hosts. Human cytomegalovirus (HCMV) is exquisitely adapted to our species and displays high genetic diversity. We leveraged information on inter-species divergence of primate-infecting cytomegaloviruses and intra-species diversity of clinical isolates to provide a genome-wide picture of HCMV adaptation across different time-frames. During adaptation to the human host, core viral genes were commonly targeted by positive selection. Functional characterization of adaptive mutations in the primase gene (*UL70*) indicated that selection favored amino acid replacements that decrease viral replication in human fibroblasts, suggesting evolution towards viral temperance. HCMV intra-species diversity was largely governed by immune system-driven selective pressure, with several adaptive variants located in antigenic domains. A significant excess of positively selected sites was also detected in the signal peptides (SPs) of viral proteins, indicating that, although they are removed from mature proteins, SPs can contribute to viral adaptation. Functional characterization of one of these SPs indicated that adaptive variants modulate the timing of cleavage by the signal peptidase and the dynamics of glycoprotein intracellular trafficking. We thus used evolutionary information to generate experimentally-testable hypotheses on the functional effect of HCMV genetic diversity and we define modulators of viral phenotypes.

Author summary

Human cytomegalovirus (HCMV), which represents the most common infectious cause of birth defects, is perfectly adapted to infect humans. We performed a two-tier analysis of HCMV evolution, by describing selective events that occurred during HCMV adaptation to our species and by identifying more recently emerged adaptive variants in clinical isolates. We show that distinct viral genes were targeted by natural selection over different

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

time frames and we generate a catalog of adaptive variants that represent candidate determinants of viral phenotypic variation. As a proof of concept, we show that adaptive changes in the viral primase modulate viral growth *in vitro* and that selected variants in the UL144 signal peptide affect glycoprotein intracellular trafficking.

Introduction

Cytomegaloviruses (CMVs, family *Herpesviridae*) infect many mammals, including humans and other primates [1]. Human cytomegalovirus (HCMV) infection is very common, with worldwide seroprevalence ranging from 56% to 94% [2]. Whereas HCMV is generally asymptomatic in healthy adults, the virus is an important opportunistic pathogen among immunocompromised individuals such as AIDS patients and transplant recipients [3]. HCMV is also the most common infectious cause of birth defects [4, 5]. Like other herpesviruses, HCMV establishes a persistent infection through latency and recent data suggest that the virus has long-term clinical consequences, especially in the elderly [6].

HCMV possesses the largest genome (~235,000 bp) among human herpesviruses and, more generally, among viruses known to infect humans [7]. Protein-coding genes occupy the great majority of the HCMV genome and are generally divided into core genes, which are shared by all herpesviruses, and non-core or sub-core genes [7–10]. These are specific to herpesvirus genera or even to CMV species and, in contrast to core genes, are often dispensable for viral growth in cell culture [9, 11]. Nonetheless, several non-core proteins play important roles during infection *in vivo*, indicating that most genes contribute to the success of CMVs in natural hosts [1, 9, 12].

CMVs display remarkable species-specificity, which results from long-term co-evolution with and adaptation to their mammalian hosts [13, 14]. Comparison of the HCMV genome with those of non human primate-infecting CMVs indicated that duplication and gene copy number variation of non-core genes were common during CMV evolution and probably contributed to host adaptation [1, 10]. However, coding gene sequence evolution and episodes of positive selection are also known to frequently shape host-pathogen interactions and to contribute to host switches [15].

HCMV genetic diversity is higher than that observed for other human-infecting herpesviruses [16] and mixed infections caused by genetically distant strains are also common [17]. Thus, the within-host HCMV genetic diversity can be very high, suggesting that HCMV populations are the target of strong selective pressures exerted by the host and by competing viral strains. Herein, we leveraged information on inter-species divergence in primate-infecting CMVs and intra-species diversity in HCMV clinical isolates to provide a global picture of HCMV evolution across different time frames.

Results

Evolution of primate CMV coding genes

We first aimed to explore the selective patterns of primate CMV coding genes. Because high sequence diversity can affect evolutionary inference, viruses that infect New World primates were excluded from these analyses. We thus analyzed representative, complete genomes of CMVs that infect great apes, Old World African monkeys, and Old world Asian monkeys (Fig 1A).

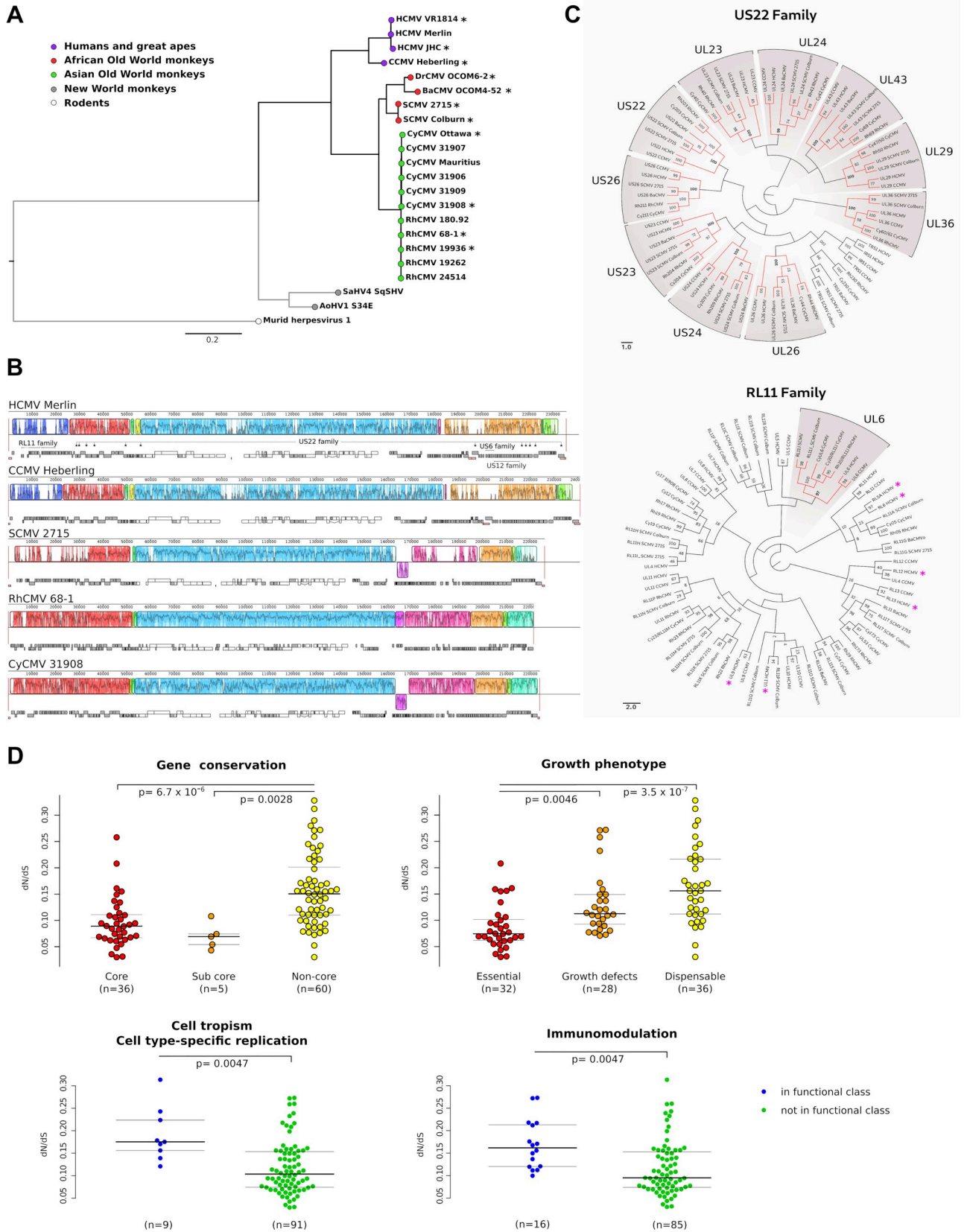


Fig 1. Evolution of catarrhini-infecting CMVs. (A) A maximum-likelihood tree of the full-length amino acid sequence of alkaline nuclease (encoded by the core gene *UL98*) is drawn to exemplify phylogenetic relationships among primate CMVs (CCMV, Chimpanzee cytomegalovirus; DrCMV, Drill monkey cytomegalovirus; BaCMV, Chacma baboon cytomegalovirus; SCMV, Simian cytomegalovirus; CyCMV, Cynomolgus macaque cytomegalovirus; RhCMV, Rhesus macaque cytomegalovirus; SaHV4, Squirrel monkey cytomegalovirus; AoHV1, Owl monkey cytomegalovirus). Murine CMV (murid herpesvirus 1) was used as the outgroup and the tree was constructed using RAXML (version 8.2.12) [19]. Asterisks denote viruses that were included in the analysis of selective patterns of catarrhini-infecting CMVs. (B) Whole-genome alignment of four representative primate CMVs obtained with progressive MAUVE. Each genome is laid out in a horizontal track, with annotated coding regions shown as boxes (white: core genes, gray: non-core genes); repetitive elements are shown as orange boxes. A colored similarity plot generated by progressive MAUVE is also shown: each colored block delimits a genome region that aligns to part of another genome (presumably homologous and free from internal rearrangements) and thus represents a locally collinear block. A similarity profile is plotted within blocks, with its height proportional to the average level of conservation in that region. White areas correspond to regions that could not be aligned. When the similarity plot points downward it indicates an alignment to the reverse strand of the genome. The location of genes belonging to the *US22*, *US12*, *RL11*, and *US6* families is shown. (C) Phylogenetic relationships for large gene families. The protein sequences of family homologs were searched for as described in the Materials and Methods. Phylogenetic trees were constructed using RAXML with 1000 bootstrap replicates (reported at nodes). Orthologous gene groups, shown in red on the tree and denoted by the gray shading, were inferred on the basis of the tree topology and of bootstrap values > 90. Magenta asterisks denote genes that are frequently deleted/mutated in clinical isolates [16]. (D) Analysis of selective patterns. The dN/dS parameter is compared among genes showing different levels of sequence conservation and distinct growth phenotypes (upper panels). Growth phenotypes in human fibroblasts were obtained from a previous work [11] that merged data from two systematic analyses of gene disruption [18, 20]. Statistical significance was assessed by Kruskal-Wallis tests followed by Nemenyi tests as post-hocs (reported in the figure). In the lower panels, genes are grouped based on function. Functional categories were derived from a previous annotation effort that combined multiple information sources [11]. p values derive from Wilcoxon Rank-Sum tests with FDR correction.

<https://doi.org/10.1371/journal.ppat.1008476.g001>

In line with previous observations [1, 9, 10], a whole-genome alignment revealed a large central collinear block, which encompasses the majority of core genes (Fig 1B). Partially due to the presence of gene families, regions flanking core genes are known to be dynamic in terms of gene content [1, 9, 10]. We thus applied a phylogenetic approach to explore gene orthology among members of the largest families (*US22*, *US12*, *RL11*, and *US6*). For the *US22* and *US12* families, one-to-one orthology could be inferred for most genes (Fig 1C and S1 Fig). Conversely, *RL11* and *US6* family members showed murky relationships, most likely due to duplication events that occurred at different time-points during primate CMV evolution (Fig 1C and S1 Fig). Several genes in these families were shown to be dispensable for HCMV growth and to be commonly disrupted in clinical isolates [16, 18] (Fig 1C and S1 Fig).

Analysis of selective patterns was thus performed for all coding genes with reliable one-to-one orthologs in 11 genomes selected to be representative of catarrhini-infecting CMVs (Fig 1A and S1 Table). Gene sequences were rigorously filtered to ensure high quality alignments (see Methods) and genes with short alignments were discarded (S2 Table). Overall, we calculated the average non-synonymous substitution/synonymous substitution rate (dN/dS, also referred to as ω) for 101 genes. All genes had dN/dS values much lower than 1, indicating that purifying selection is the major force acting on CMV coding regions. Comparison among core, sub-core (conserved in *Betaherpesviridae* and *Gammaherpesviridae*), and non-core genes (specific to CMVs) indicated that these latter have significantly higher average dN/dS (Kruskal-Wallis rank sum test, $p = 5.577 \times 10^{-7}$, Nemenyi post-hoc tests are reported in Fig 1D). Consistently, essential genes showed greater evolutionary constraint than dispensable genes or genes that cause growth defects when deleted/mutated, (Kruskal-Wallis rank sum test, $p = 6.6488 \times 10^{-7}$, Nemenyi post-hoc tests are reported in Fig 1D). This observation should however be taken with caution, as growth phenotypes were determined for a cell culture-adapted HCMV strain [11] and do not necessarily correspond to phenotypes that would be observed in clinical isolates or other primate CMVs.

Among functional classes, genes involved in immunomodulation and cell-specific tropism displayed higher dN/dS than genes that have not been reported to participate in these processes (Fig 1D). Significant differences in evolutionary rates were also found among proteins that localize to the cell membrane or in the cell nucleus (S2 Fig). In agreement with this latter finding, genes that participate to DNA replication had lower dN/dS than genes that are not known to be involved in this process, although the difference was not significant (S2 Fig); the

failure to reach statistical significance may be due to the small number of genes included in this functional class. In general, the statistical power to detect differences is low for functional classes/compartments with few annotated genes. Moreover, in these cases, the unbalanced sample sizes complicate the interpretation of the differences.

Adaptive evolution in the HCMV lineage

We next assessed whether adaptation to the human host drove the evolution of specific coding genes in HCMV. We thus considered a phylogeny of primate CMVs that included all fully sequenced catarrhini-infecting CMVs (Fig 1A) and 25 HCMV clinical isolates (S1 Table). We applied a branch-site test [21] to detect positive selection that occurred on the HCMV branch (S2 and S3 Tables). In this test, the branches of the tree are divided a priori into foreground (here, the branch leading to all HCMV isolates) and background lineages (all other branches), and models that allow or disallow positive selection on the foreground lineage(s) are compared. The branch-site test can thus detect lineage-specific selected genes and sites (episodic positive selection).

After FDR (false discovery rate) correction, the test detected 34 positively selected genes (S3 Table). Despite their higher constraint during evolution in catarrhini (Fig 1D), core genes were more frequently targeted by positive selection (fraction selected = 41.7%) than non-core plus sub-core genes (fraction selected = 29.2%). Identification of positively selected sites (see Methods) indicated that a similar proportion of codons were called as positively selected in core (0.13%) and non-core/sub-core genes (0.12%).

Positive selection drove the evolution of three capsid proteins and of numerous tegument proteins that participate to an interaction network among themselves and with capsid components (Fig 2A). Four of the selected proteins (UL25, UL45, UL48, and UL69) represent major hubs in this network, which includes several components involved in virion assembly and maturation [22, 23].

Analysis of positively selected sites in *UL85*, *UL86*, and *UL46*, which encode structural constituents of the capsid, indicated that four of them are located at the contact interface with capsid components and/or tegument protein pp150 (UL32) (Fig 2B). In *UL85*, the two positively selected sites localize to the embracing arm, which has an important structural role in capsomer stability. In particular, residue 191 is located at the interface with *UL86* and with the second *UL85* molecule that compose the triplex structure; residue 266 is in close proximity to *UL46* (Fig 2B). Residue 199 in *UL46* and one selected site (residue 741) in *UL86* are placed at the interface with the tegument protein pp150 (Fig 2B).

Only two glycoproteins, US11 and UL55 (glycoprotein B, gB) were among selection targets on the HCMV branch. gB mediates viral-host cell fusion and is a major target antigen [24]. Mapping of positively selected codons relative to known antigenic domains (AD) indicated that two of the three selected sites are located within ADs that elicit neutralizing antibodies [25, 26] (Fig 2C). Finally, positive selection was detected at genes that encode core viral enzymes such as *UL70* (primase) and *UL114* (uracil-DNA glycosylase) (S3 Table).

Positively selected sites in *UL70* modulate viral replication

Natural selection can only act on mutations that determine phenotypic variation. As a proof of concept that the sites we identified do modulate a viral phenotype, we decided to functionally characterize two selected variants in the viral primase (vUL70). In particular, we selected sites G294 and R465 because they are located in regions of high sequence conservation among CMVs (Fig 3A). Thus, point mutations that recapitulate the amino acid state observed in non-human infecting CMVs were introduced in the backbone of an infectious clone (TB40-BAC4)

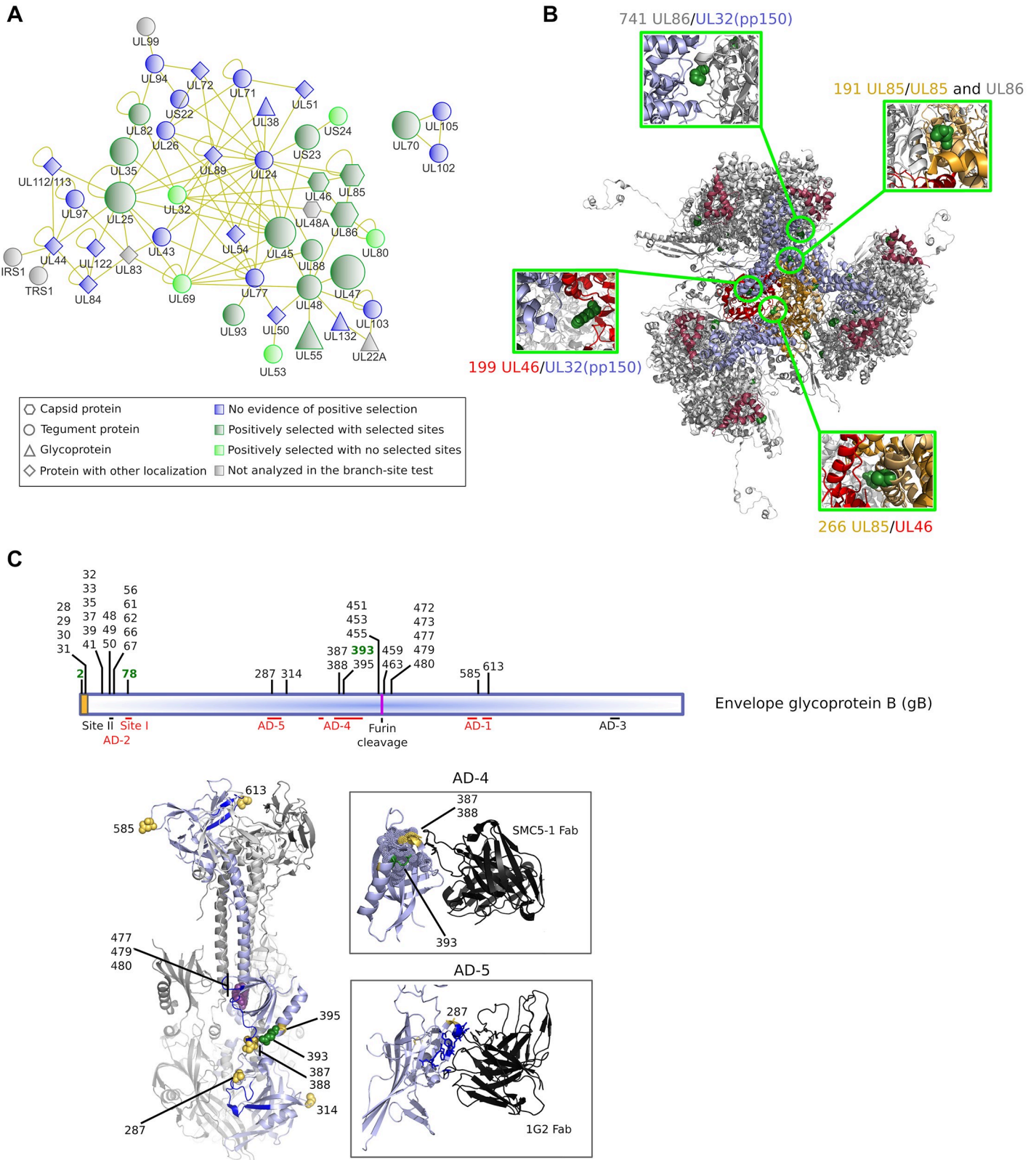


Fig 2. Episodic positive selection on the HCMV branch. (A) Protein-protein interaction network of HCMV proteins. Protein function/localization is coded with shapes, the selective patterns with colors. For positively selected proteins (forest green), node size is proportional to the number of detected positively selected sites.

Protein-protein interaction data were derived from previous analyses [27,28] and from the VirHostNet database [29]. (B) Ribbon representation of the partial capsid structure of HCMV (PDB: 5VKU). Overview of the triplex structure (light orange: UL85 dimer, red:UL46), its surrounding major capsid proteins (UL86, grey), smallest capsid proteins (UL48A, maroon), and tegument proteins pp150 (UL32, light blue). Positively selected sites on the HCMV branch are shown as forest green spheres. Positively selected sites located at the interface with other molecules of the capsomer structure [30] are shown in the inserts. (C) Schematic representation of the gB protein. The signal peptide is in orange. Sites that were positively selected on the HCMV branch are shown in green, those identified by gammaMap are in black. The location of known antigenic determinants is reported [25, 26, 31–33]. Note that AD-2 is composed by two sites: site I (aa 64–84) contains epitopes recognized by neutralizing antibodies, site II (50–54) contains epitopes bound by non-neutralizing antibodies [25]. ADs that elicit neutralizing antibodies are marked in red. Several selected codons also flank the furin cleavage site. The structure of the gB trimer (PDB: 5CXF) is also shown [34]. For clarity, one monomer is represented in light blue, the other two are in gray. Regions encompassing neutralizing epitopes are shown in blue [26, 32, 33]. Selected sites are shown as spheres and color coded (green, selected on the HCMV branch; purple, selected in HCMV strains from at least two compartments; yellow, selected in HCMV strains from urine samples). In the two inserts, AD-4 and AD-5 in complex with neutralizing antibodies are shown. In AD-4 a hydrophobic pocket recognized by SMC5-1 (PDB: 4OT1) [26] is represented with dotted spheres. In AD-5, residues that directly contact 1G2 are marked in blue (PDB: 5C6T) [32].

<https://doi.org/10.1371/journal.ppat.1008476.g002>

of the endotheliotropic HCMV strain TB40/E, either individually or in combination (Fig 3A and S4 Table). After “en passant” mutagenesis, viruses encoding wild-type vUL70 and mutant vUL70 (vUL70 G294L, vUL70 R465K, and vUL70 dm, for double mutant) were obtained (Fig 3A). As a control, mutant viruses carrying alanine at positions 294 and 465 were also generated (vUL70 G294A and vUL70 R465A) (Fig 3A).

To compare the cell-to-cell spreading efficiency of the mutant viruses to that of the wild-type parental strain, a focus expansion assay (FEA) was performed on HFFs (human foreskin fibroblasts, where HCMV is typically propagated) and ARPE-19 cells (an *in vitro* epithelial cell model). Infections were performed with serial dilutions and cells were stained for HCMV IEA protein expression. The numbers of IEA positive foci in each cell dilution were comparable between the wild-type and the mutant viruses, in both HFFs and in ARPE-19 (S5 Table), suggesting that the overall replication dynamics of the viral strains was not affected by the UL70 mutations.

In agreement with previous works [35, 36], when HFFs were infected, all HCMV strains formed enlarged and diffused plaques throughout the culture (Fig 3B). In contrast, plaques appeared smaller and confined in ARPE-19 cells (Fig 3B). By measuring the plaque area and the number of IEA positive cells in each focus, we observed significant differences in the spreading capability in HFF among the mutant viruses. In fact, the vUL70 G294L, vUL70 R465K, and vUL70 dm mutants displayed a higher number of IEA positive HFFs in infected foci compared to the wild-type (Fig 3C and 3D). Conversely, none of the alanine-substitution mutants significantly affected FEA results compared to the wild-type virus. Consistently, the relative plaque area of the infected foci mirrored the IEA foci count, although a significant area increase was obtained for the double mutant only (Fig 3D). No significant difference among viruses was observed in ARPE-19 cells. The different phenotypes between the two cell types is most likely due to the fact that, compared to HFFs, ARPE-19 cells are less permissive to HCMV infection and support lower levels of virus release [37].

Given the results above, we assessed whether the three mutant viruses with amino acid substitutions observed in non-human infecting CMVs (vUL70 G294L, vUL70 R465K, and vUL70 dm) had higher fitness than the wild-type virus. To this aim, we performed head-to-head competition assays in HFFs, by infecting the cells with a 1:1 ratio of wild-type and each mutant virus with a multiplicity of infection (MOI) of 0.0005 PFU/ml. Viral growth was maintained for 28 days and the relative ratio of wild-type and mutant virus was assessed at different time points through deep sequencing. Results indicated that all mutants out-competed the wild-type, with their relative proportions significantly increasing over time (all *p* values < 0.01, Cochran Armitage tests for trend). The effect was evident since the earlier time point (five days post infection), especially for the vUL70 R465K and vUL70 dm viruses (Fig 3E). At late time points, the relative proportions of all viruses tended to stabilize, most likely due to limiting conditions in cell culture. In general, the strongest competition over the wild-type was

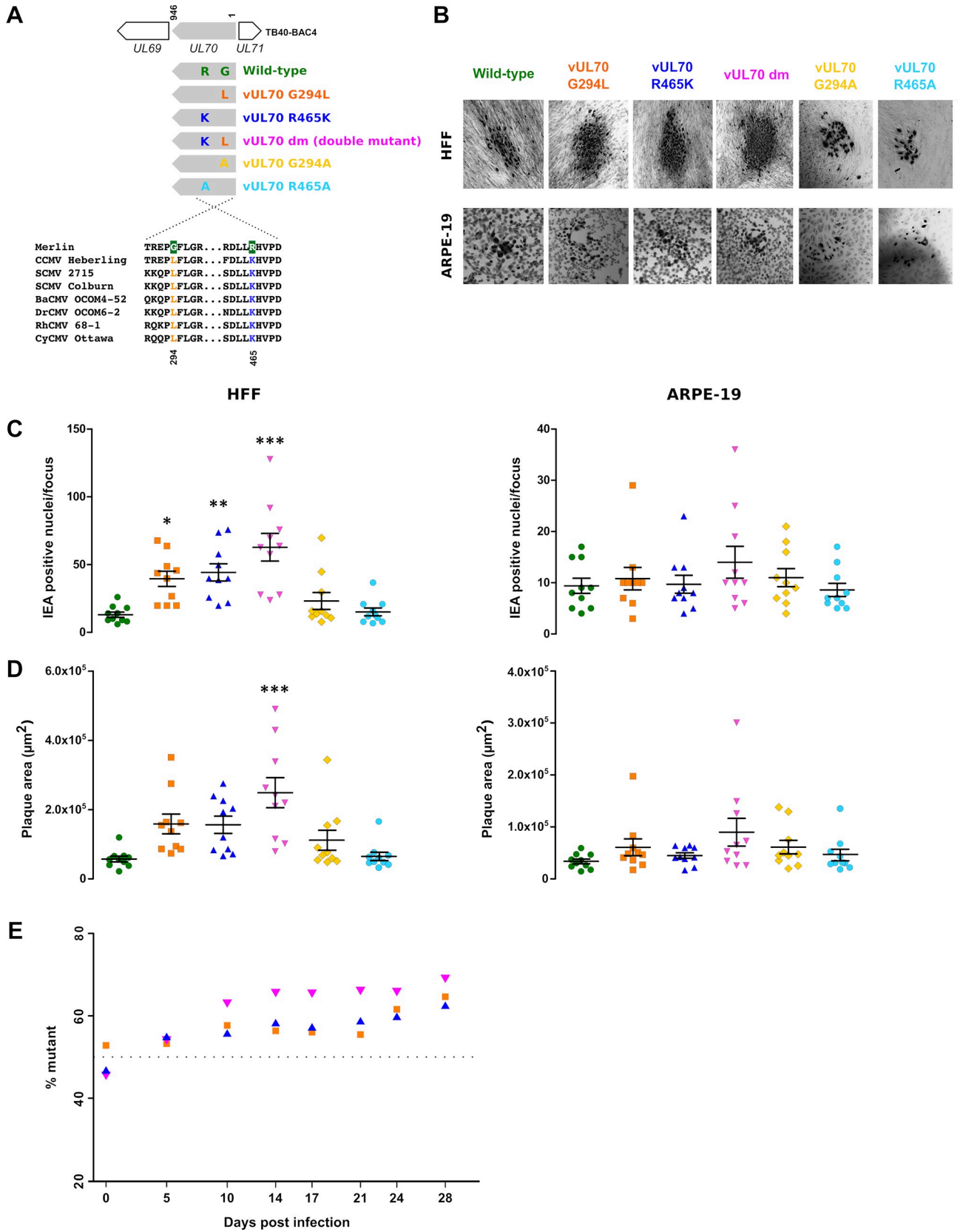


Fig 3. Phenotypic effect of positively selected sites. (A) Generation of vUL70 mutant strains. The HCMV *UL70* gene region is schematically represented together with the induced amino acid replacements at positively selected sites (G294 and R465). The alignment of a few catarrhini-infecting CMVs are shown. Wild-type and mutant viral strains are color-coded. (B) Representative light microscopy images of infectious foci are shown as seen in HFFs or ARPE-19 cells (5 days post infection). Original magnification: 20X. (C) IEA positive nuclei/focus. Each data point of the graph represents the number of IEA-positive nuclei of one focus at day 5 post infection. Bold horizontal lines depict the median number of IE antigen-positive cells/focus for each virus. One-way ANOVA followed by Tukey's post-hoc test was used for comparison of mutant viruses *versus* the wild-type (* $P < 0.05$; ** $P < 0.01$; ***, $P < 0.001$). (D) Plaque areas were calculated using ImageJ software. Horizontal lines represent mean values \pm standard deviations. One-way ANOVA followed by Tukey's post-hoc test was used for comparison of mutant viruses *versus* the wild-type (***, $P < 0.001$). (E) Head-to-head competition assays of three mutant viruses against the wild-type. HFFs were infected with a 1:1 ratio of wild-type and each mutant virus (MOI: 0.0005 PFU/ml). At different time points, total genomic DNA was extracted and subjected to deep sequencing. The relative proportion of mutant viral DNA is plotted. Each point represents the mean of two replicates.

<https://doi.org/10.1371/journal.ppat.1008476.g003>

observed for the double mutant (Fig 3E). Calculation of the relative fitness gain (see [methods](#)) of the three mutant viruses yielded values of 2.5% (vUL70 G294L), 3.3% (vUL70 R465K), and 5.1% (vUL70 dm), in good agreement with the IEA foci count results.

Overall, these data indicate that the introduction of amino acid residues observed in non-human primate-infecting CMVs favors viral replication in human fibroblasts.

Selective patterns in HCMV clinical isolates

Genome-wide evolutionary analyses of coding genes in HCMV isolates were previously reported [16, 38, 39]. These studies, however, applied methodologies that are best-suited to study inter-species diversity [16] (i.e., long-term evolutionary processes [40]), or used population genetics approaches to focus on within-host selection [39]. Alternatively, SNP enrichment was calculated to obtain an estimate of preferential selection in specific body compartments [41].

Herein, we applied a population genetics-phylogenetics approach to study the evolution of coding genes in HCMV strains from different compartments. Specifically, we used the gammaMap program [32], that jointly uses intra-species variation and inter-species diversity, to estimate the distribution of fitness effects (i.e. selection coefficients, γ , expressed as discrete categories from -500 to 100) along coding regions (see [Materials and Methods](#) and [S6 Table](#)). GammaMap is relatively insensitive to demography and recombination [42]. Analyses were performed for all coding genes in clinical isolates deriving from amniotic fluid, urine, and blood/plasma ([S7 Table](#)).

Overall, we observed a strong preponderance of codons evolving under negative selection ($-500 \leq \gamma \leq -1$) and a non-negligible fraction of positive selection ($\gamma > 1$) signals. No substantial differences in the distribution of selection coefficients was observed across the three body compartments (Fig 4A). The selective patterns were also similar between genes that were or were not called as positively selected in the branch-site test, suggesting that the evolutionary processes that drive host adaptation do not necessarily parallel those that occur at the intra-species level (S3 Fig). Core genes, however, showed a larger proportion of strongly constrained codons compared to non-core/sub-core genes, whereas the immunomodulation and glycoprotein gene functions showed comparatively fewer sites under negative selection and more with positive selection signals (Fig 4A). Despite these differences, most genes had a proportion of codons evolving under strong to moderate negative selection ($\gamma < -10$) (S4 Fig).

To define the signals of positive selection, we estimated codon-wise posterior probabilities for each selection coefficient. We called a codon as positively selected if its cumulative posterior probability of $\gamma \geq 1$ was > 0.80 . A total of 32 genes (5 core, 27 non-core) were found to be positively selected (i.e., to have at least one selected codon) in at least one compartment (Fig 4B, S8 Table). Although there was a large overlap among compartments, isolates from urine and blood/plasma had more abundant and more divergent signals of positive selection

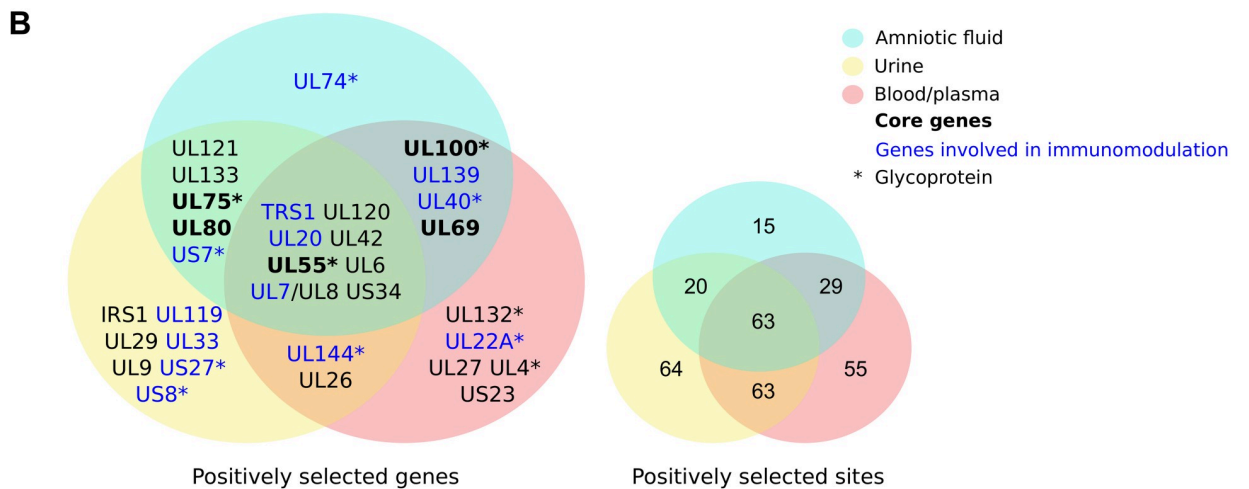
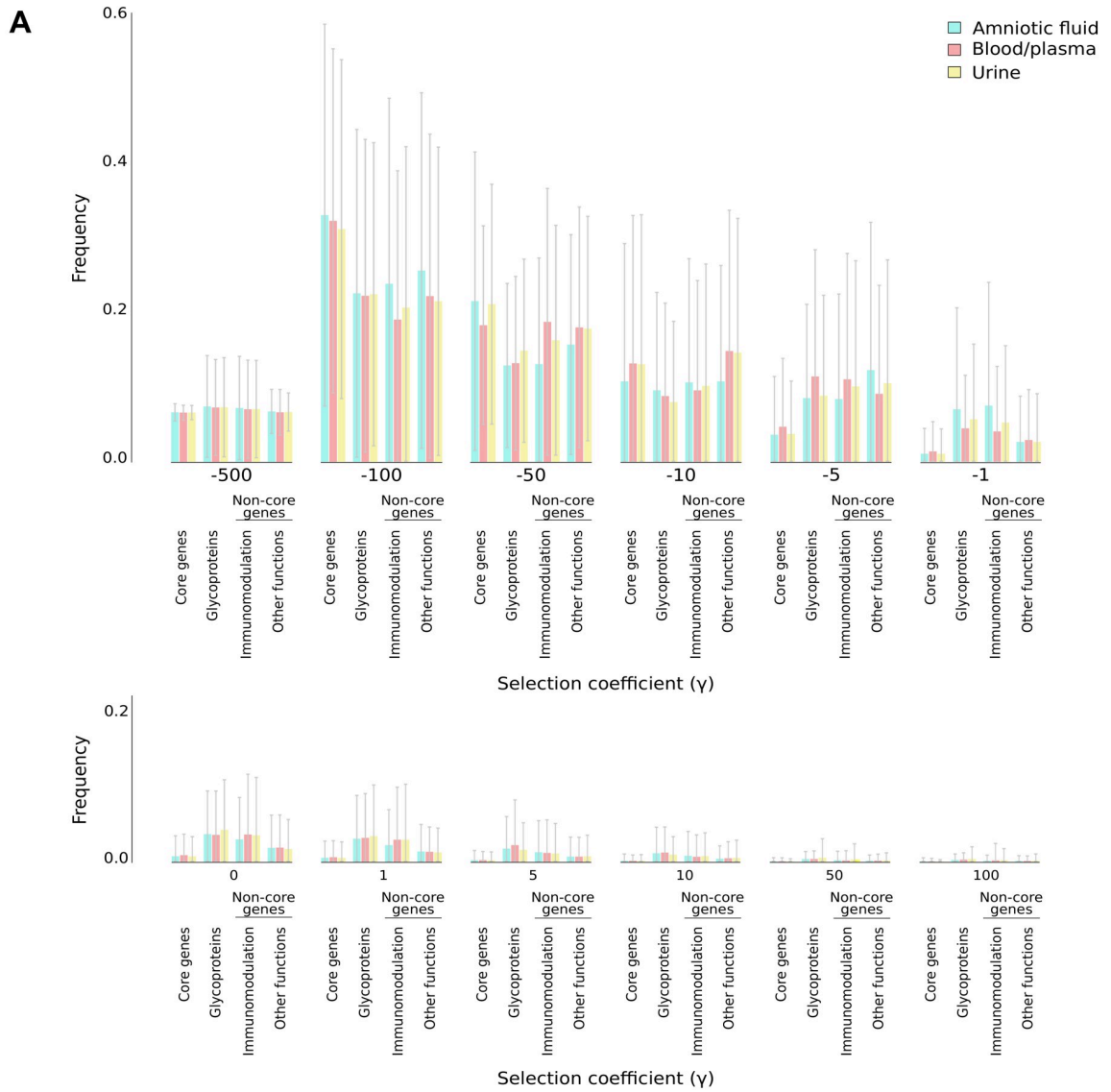


Fig 4. Selective patterns in HCMV clinical isolates from different compartments. (A) Distribution of selection coefficients (γ) for all coding genes of HCMV clinical isolates sampled from amniotic fluid, urine, and blood/plasma. Selection coefficients were calculated for all codons and genes were grouped on the basis of their conservation (core, non-core/sub-core) and of the functional classification of the encoded protein (glycoprotein, both core and non-core; immunomodulation, non-core only). Large error bars are caused by the difficulty in determining the relative frequency of similar selection coefficients. (B) Venn diagrams of positively selected genes and sites in HCMV isolates.

<https://doi.org/10.1371/journal.ppat.1008476.g004>

compared to those from amniotic fluid, in which the only compartment-specific signal was at *UL74* (glycoprotein O, gO) (Fig 4B).

Among positively selected genes, glycoproteins were significantly enriched (Fisher exact test with FDR correction, $p = 0.0195$). A tendency for enrichment was also observed for genes with immunomodulatory function (Fisher exact test with FDR correction, $p = 0.0514$), but not for genes involved in cell tropism (Fisher exact test with FDR correction, $p = 1$).

We detected positively selected sites in nine genes (*UL33*, *UL4*, *UL55*, *UL74*, *UL75*, *UL100*, *UL119*, *UL132*, and *US27*) coding for viral envelope proteins. The majority of such sites (88.3%) are located in protein regions exposed on the virion surface (S5 Fig and S8 Table). Although exposed regions account for a large portion of the protein sequences, the number of selected sites was significantly higher than expected (Binomial test, p value = 0.00925). Likewise, eight genes (*UL6*, *UL7/8*, *UL9*, *UL42*, *UL120*, *UL121*, *UL139*, and *UL144*) that encode proteins expressed at the host cell membrane were targeted by positive selection and most sites (90.8%) were in the extracellular domains (Binomial test, p value = 0.001955) (S5 Fig and S8 Table).

Overall, these findings suggest that a major selective pressure on these genes is exerted by the host immune system. For instance, several selected sites in gB map within ADs (Fig 2C), and in gH (*UL75*) positively selected sites localize to a structural epitope within antigenic site 7 [43] (Fig 5A). Additional details on the location of positively selected sites, sequence polymorphism, and ADs for gB, gH, gM and gO are reported in S6 Fig. As for *UL144*, positively selected sites are directly involved in BTLA (B and T lymphocyte attenuator) binding (Fig 5B) [44].

In *UL144*, nine sites were also located in the short signal peptide (SP) (Fig 5C). More generally, we found evidence of positive selection in the SPs of several proteins (*UL9*, *UL20*, gB, gH, *UL132*, *US7*, in addition to *UL144*) (S5 Fig and S8 Table). Overall, 9% of the positively selected sites we detected were located in SPs, a proportion significantly higher than expected (Binomial test, p value = 0.001154).

Positively selected sites in the *UL144* signal peptide modulate intracellular trafficking and the timing of cleavage by the signal peptidase

Given the results above, we decided to functionally characterize the effects of selected variants within the *UL144* SP. We focused on *UL144* because its SP was strongly targeted by selection and because *UL144* is a membrane glycoprotein with important immunomodulatory functions [52, 53]. The nine sites in the SP were called as positively selected in blood/plasma and urine samples; their amino acid states differ among previously described *UL144* genotypes [54].

We analyzed the intracellular trafficking of two *UL144* sequences: one that is identical to the protein encoded by the Merlin strain and corresponds to genotype A (gtA) and one that carries amino acid replacements at the 9 positively selected sites and corresponds to the SP of genotype B (gtB). These two variants have similar representation among clinical isolates (Fig 5C). The prediction of protein secondary structure indicated a shorter helix domain for the gtA SP compared to gtB (Fig 5C). This feature was reported to facilitate the cleavage by the

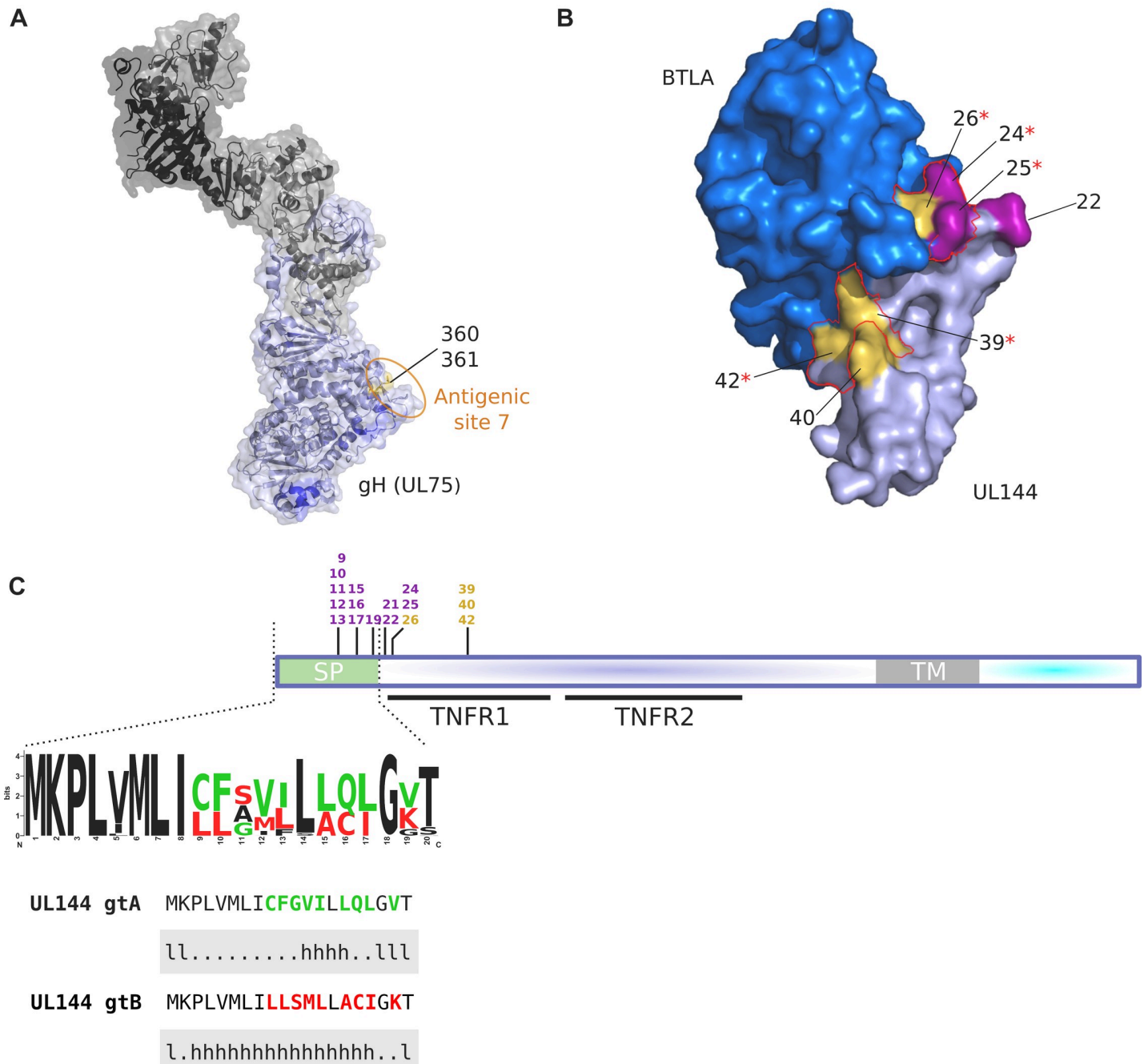


Fig 5. Positively selected sites in clinical isolates. (A) Structure of the HCMV pentamer (PDB: 5VOC) [45]. gH is shown in light blue, gL, UL128, UL130, and UL131 in shades of grey. The location of known antigenic sites is shown in blue [46–48]. Positively selected sites in urine-derived samples are shown in yellow and are surface-exposed. These sites localize to antigenic site 7 [43]. (B) 3D structure of HCMV UL144 (light purple) in complex with BTLA (blue) (PDB: 6NYP). Positively selected sites in urine-derived samples and in at least two different compartments are shown in yellow and purple, respectively. Residues critical for BTLA binding are highlighted with a red outline. Red asterisks denote positively selected sites directly involved in BTLA binding [49]. (C) Schematic representation of the UL144 protein. Positively selected sites are color coded as in (B). The location of tumor necrosis factor receptor/nerve growth factor receptor repeats (TNFR1 and TNFR2) is reported. The signal peptide (SP) is in green. The sequence logos recapitulate the amino acidic variation observed in SPs of HCMV clinical samples. Information content (bits) is plotted as a function of amino acid position. Sequence logos were generated with WebLogo [50]. The amino acid state of positively selected sites that define the two SPs are reported (gtA, green; gtB, red). The secondary structure prediction (l, loop; h, helix), generated with PredictProtein [51], is also shown below both sequences.

<https://doi.org/10.1371/journal.ppat.1008476.g005>

signal peptidase in *E. coli* [55] and changes in the helix structure of the HIV-1 envelope glycoprotein (Env) SP modulate the timing of its cleavage [56].

The two UL144 variants with a DDK tag (hereafter referred to as gtA and gtB) were transiently expressed in HeLa cells and detected with an anti-DDK antibody (green). Twenty-four hours after transfection, both gtA and gtB reached the cell membrane as expected (Fig 6A), but were also differentially detected in intracellular vesicles (Fig 6B). Indeed, while gtB mainly co-localized with the early endosomal marker EEA1 (blue, co-localization light-blue), the gtA variant showed predominant co-localization with the lysosomal marker LAMP1 (red, co-localization yellow) (Fig 6B). Endosomes represent the first compartment of the endocytic pathway for the sorting of internalized protein from the plasma membrane, that terminates in the lysosomes. The differences in the localization of the two UL144 variants in these compartments (Fig 6B) therefore suggest a sorting delay of gtB compared to gtA.

To monitor the differences in the cellular trafficking of the gtA and gtB variants, we analyzed their localization at different times after transfection. Membrane proteins are commonly targeted to the endoplasmic reticulum (ER) by the signal peptide through the Sec61 translocon [57]. We therefore evaluated the co-localization of gtA and gtB with the translocon complex (Sec61a) and with calreticulin, an ER luminal resident protein, at 3, 4 and 5 hours after transfection. Three hours after transfection, the two variants showed a similar co-localization with Sec61a (red) (Fig 7A), indicating that the initial contact with the translocon occurs with the same efficiency. Conversely at 4 hours, while the gtA variant moved to other ER compartments mainly co-localizing with calreticulin (red), gtB was still associated with Sec61a and its co-localization with calreticulin was delayed compared to gtA (Fig 7A and 7B), indicating that gtB is retained for a longer time by the translocon complex. Differences in the sorting dynamics of the two variants were also evident at 6 and 9 hours post transfection, when we analyzed the co-localization of the UL144 variants with the early endosomal marker EEA1 (blue) and with the lysosomal marker LAMP1 (red). At 6 hours post transfection, gtA was mainly found in early endosomes, whereas most of gtB was still located within the ER and only a minor co-localization with EEA1 was detected (Fig 7C and 7D). None of the two variants localized to the lysosomes at this time point (Fig 7D). The delayed processing of gtB remains unchanged at 9 hours post transfection (Fig 7C and 7D) and eventually explains the differential localization observed at 24 hours.

These results, also confirmed in HEK-293 cells (S7 Fig), indicate that the selected sites in the UL144 SP are sufficient to alter the intracellular trafficking of the glycoprotein.

To determine whether the delayed trafficking of the UL144 gtB variant is due to late cleavage of the SP, we generated two expression vectors carrying the DDK tag immediately downstream the gtA and gtB SPs (Fig 8A). Immunofluorescence analysis was then performed 3, 4 and 5 hours after transfection using two antibodies: the anti-M1-DDK antibody, that specifically recognizes the free N-terminal end of the DDK epitope and thus stains the protein only after SP cleavage [58], and the anti-DDK antibody used above, which binds the epitope irrespective of its sequence context (Fig 8A). At 3 hours after transfection, while almost all the cells transfected with gtA were positive for both M1-DDK (green) and DDK (red) staining, indicating SP removal, the 35% of the cells transfected with gtB still presented the SP and were therefore stained by the anti-DDK antibody only (Fig 8B and 8C). This delay in the removal of the SP from gtB was also maintained at 4 hours after transfection. It was not until 5 hours post transfection that the SP was cleaved from the gtB variant in all the analyzed cells.

Discussion

CMVs display a remarkable host-specificity even *in vitro*—i.e., in the absence of the host adaptive immune response. Such specificity is mainly driven by post-entry events [59–63]. In the

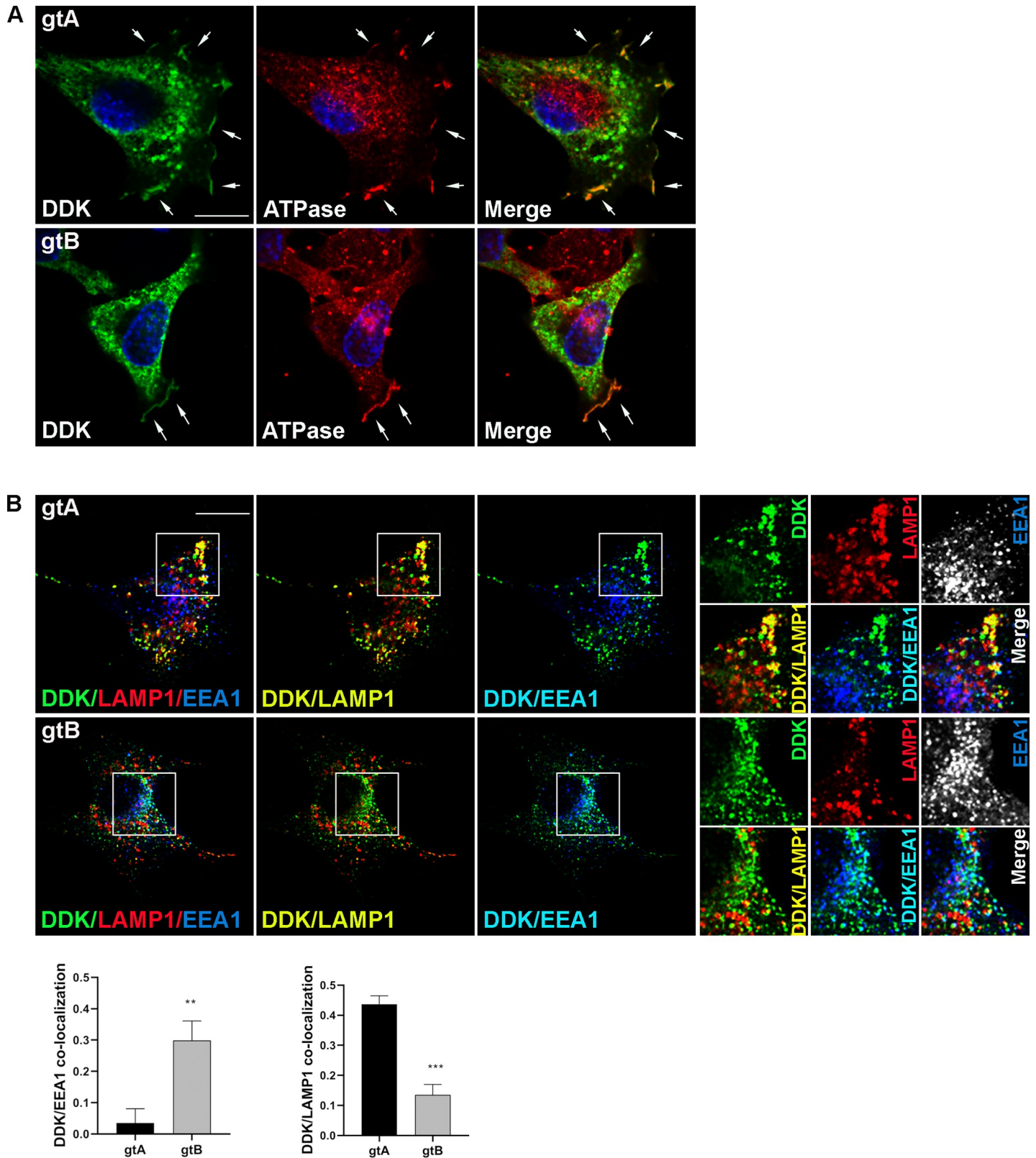


Fig 6. Functional analysis of positively selected sites in the UL144 SP. HeLa cells were transfected with pCMV6-UL144 gtA and pCMV6-UL144 gtB. Twenty-four hours later, cells were fixed and immunostained. (A) Localization at the plasma membrane. Cells were stained with antibodies against the DDK tag (green) and the plasma membrane protein sodium potassium ATPase (red). Nuclei were counterstained with DAPI. Arrows indicates localization at the plasma membrane. Scale bar:

10 μm . (B) Localization in the endo-lysosomal compartment. Cells were immunostained with antibodies against the DDK tag (green), the lysosomal marker LAMP1 (red) and the early endosomal marker EEA1 (blue). Co-localization of DDK with LAMP1 (yellow) or EEA1 (light blue) is shown in the merge images. The small panels show an higher magnification of the area indicated in the squares. Scale bar: 10 μm . Pearson's correlation coefficients for DDK/LAMP1 and DDK/EEA1 co-localization are reported in the graphs as mean \pm SEM (t test; $n > 30$) (**, $P < 0.01$; ***, $P < 0.001$).

<https://doi.org/10.1371/journal.ppat.1008476.g006>

case of HCMV, chimpanzee primary fibroblasts allow low-level viral replication [64], but other non-human primate cells do not support the production of infectious virions [59–61]. This clearly implies that HCMV must have adapted to efficiently complete its infectious cycle in human cells. Our data indicate that, although core genes were generally more constrained during primate CMV evolution, a number of them were targeted by positive selection during HCMV speciation. We found evidence of selection at capsid components and at numerous tegument proteins that drive important steps in virion maturation and assembly. Such steps are orchestrated by an interplay of interactions among viral- and host-encoded proteins [23], suggesting that the selection signals we identified represent polygenic adaptation of HCMV to replication in human cells.

Evidence of selection was also detected at two core viral enzymes, including the primase, which we decided to characterize experimentally. We found that UL70 mutations that recapitulate the amino acid state observed in non-human infecting CMVs display increase cell-to-cell spreading efficiency in human fibroblasts. Competition experiments also indicated that mutant viruses have higher fitness than a virus carrying the same amino acid states as circulating HCMV strains. Although this finding may seem counter-intuitive, it is known that HCMV has evolved strategies to reduce or control its replication rate and, most likely, its virulence. Such strategies were referred to as “temperance” by Dunn and coworkers, who performed a large-scale analysis of the effect of HCMV gene mutations on viral growth [18]. They found that deletion of some viral genes (e.g., *UL10*, *UL23*, and *US16*) increased rather than decreased replication in specific cell types and suggested that this temperance effect is necessary to promote long-term viral co-existence with its host [18]. In analogy, we found that the cell-to-cell spreading effect of UL70 mutations is cell-type specific, as only minor, non significant differences were observed in an epithelial cell type. Nonetheless, additional experiments will be required to assess whether adaptive mutations in the *UL70* gene arose during HCMV evolution as temperance factors.

Analysis of the distribution of fitness effects in HCMV clinical isolates clearly indicated that the selective events that occurred during HCMV speciation (i.e., most likely as a result of human adaptation) differ considerably from the selective patterns that govern intra-species variation (i.e., ongoing or recent adaptive events). The selective pressure exerted by the host immune system has likely played a major role in the shaping of genetic diversity among circulating strains. In line with previous findings, we found that glycoproteins and proteins involved in immunomodulation tend to be less constrained and to display the strongest signals of positive selection [16, 38, 41]. This effect was observed in all compartments, that also showed a similar overall distribution of selection coefficients. This is in contrast with previous suggestions that isolates from vascular compartments are more selectively constrained than those from urine [41]. Samples from blood/plasma and urine had more abundant and more diverse selection signals compared to those isolated from amniotic fluid, a finding that is not necessarily due to the specific compartments, but may depend on the number of available isolates and/or on their diversity. In fact, several positively selected genes and sites were shared among compartments, revealing limited tissue-specific adaptive events.

Some of the positively selected genes we identified using the population genetics-phylogenetic approach were previously found to represent positive selection targets during intra-host

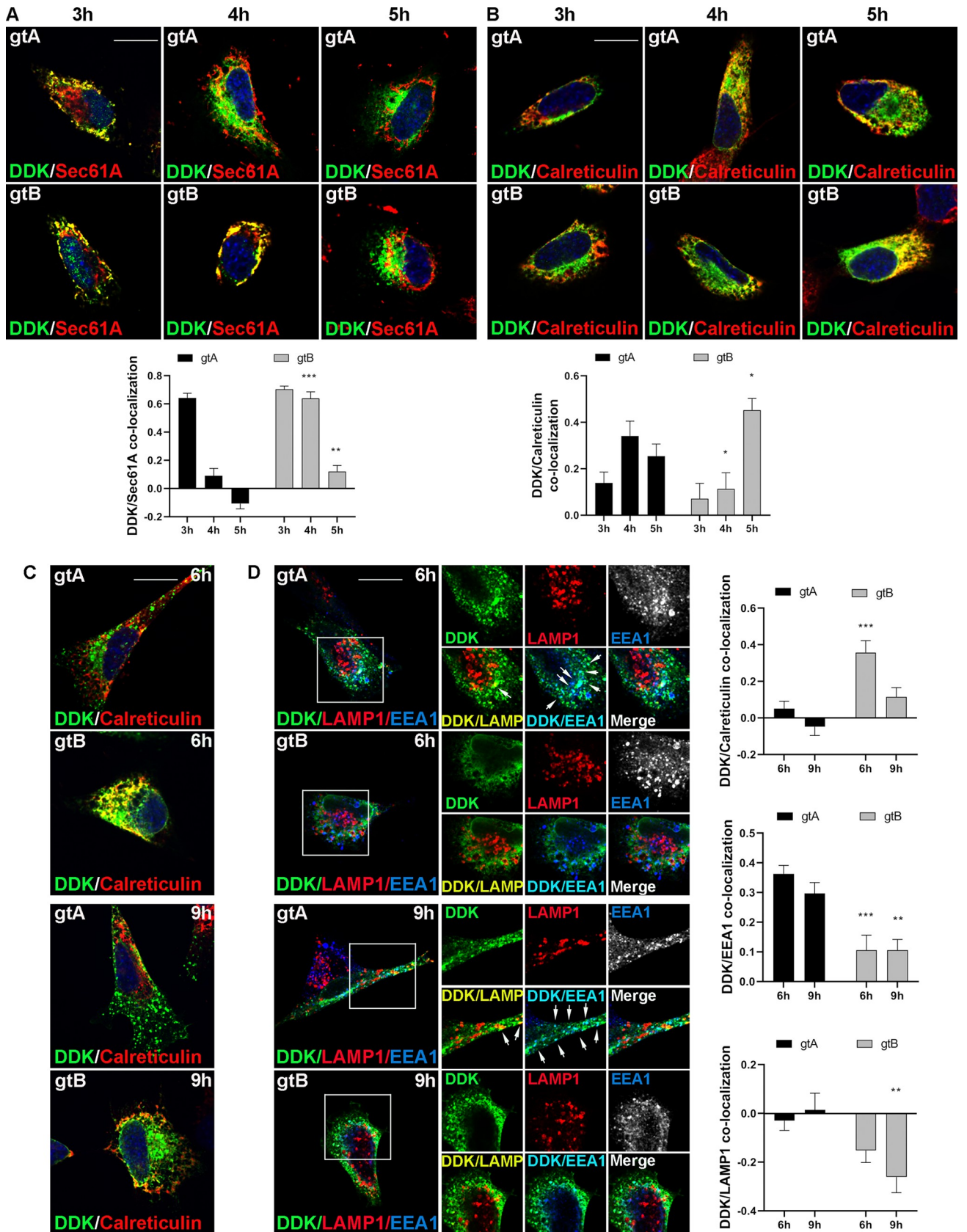


Fig 7. Time-course analysis of UL144 gtA and gtB intracellular trafficking. HeLa cells were transfected with pCMV6-UL144 gtA and pCMV6-UL144 gtB and fixed at different time points. Three, four and five hours post transfection cells were immunostained with anti-DDK (green) and anti-Sec61A (red) Abs (A) or with anti-DDK (green) and anti-Calreticulin (red) Abs (B). Nuclei were counterstained with DAPI. Yellow indicates co-localization. Scale bar: 10 μ m. Pearson's correlation coefficients for DDK/Sec61A and DDK/Calreticulin co-localization are reported in the graphs as mean \pm SEM (two way ANOVA; $n > 25$) (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (C) Six and nine hours post transfection cells were immunostained with anti-DDK (green) and anti-Calreticulin (red) antibodies. Nuclei were counterstained with DAPI. Yellow indicates co-localization. Scale bar: 10 μ m. (D) Six and nine hours post transfection cells were immunostained with anti-DDK (green), anti-LAMP1 (red) and anti-EEA1 (blue) antibodies. The small panels show an higher magnification of the area indicated in the squares. The arrows indicate co-localizing vesicles. Scale bar: 10 μ m. Pearson's correlation coefficients for DDK/Calreticulin, DDK/EEA1 and DDK/LAMP1 co-localization are reported in the graphs as mean \pm SEM (two way ANOVA; $n > 25$) (**, $P < 0.01$; ***, $P < 0.001$).

<https://doi.org/10.1371/journal.ppat.1008476.g007>

evolution (*UL7*, *US7*, *UL74*, *US27*, *UL20*, *UL132*, *UL80*, *UL55*) [39] or to display an excess of nonsynonymous polymorphism in plasma samples (*UL55*, *UL74*, *UL75*) [41]. In gB (*UL55*) and gH (*UL75*), most positively selected sites were located within ADs and, in the case of *UL144*, several sites were within the interaction surface with BTLA, again suggesting immune system-driven selective pressure. gB and gH are highly immunogenic and represent major targets of neutralizing antibodies. Extensive analysis of antibody responses indicated that different ADs elicit neutralizing antibodies in a variable proportion of naturally infected or immunized individuals [65–67]. In some instances, such as the AP86 epitope in the N-terminus of gH, where no positively selected sites were detected, variability among HCMV strains is known to modulate the antibody response [68]. In the case of gB ADs, it is presently unclear whether viral diversity affects neutralization capacity. Recently, a correlation was observed between anti-AD-2 serum titers and protection from viremia in a vaccination trial [69], although only a proportion of infected individuals developed AD-2 neutralizing responses. The majority of positively selected sites we found in AD-2 map to a site which is known to elicit non-neutralizing antibodies, raising questions about their adaptive significance in terms of immune evasion. A possibility is that, as previously shown for AD-1, a competition between neutralizing and non-neutralizing antibodies at site I and site II represent a mechanism to evade humoral responses [70]. Thus, although we found a very good correspondence between the location of ADs and selected sites, the relevance of the latter for immunoevasion and, ultimately, vaccine design, remains to be evaluated.

In gB, we also detected several selected sites across the furin cleavage site. Using a different approach, variants in this region were previously shown to be targeted by positive selection and to modulate the kinetic of furin cleavage [71]. As for other herpesviruses with furin cleavable gBs, proteolytic processing by furin is dispensable for HCMV growth in cell culture [72–75]. However, because the furin cleavage site is evolutionary conserved in several members of the *Herpesviridae* family, processing may confer some advantages under specific circumstances or in specific cell types. For instance, in other herpesviruses, loss of gB cleavage decreases viral cell-to-cell spread, sometimes in a cell-type dependent manner [73, 74, 76, 77]. Cell-to-cell spread is a typical feature of HCMV clinical isolates [78], suggesting that variation at the selected sites might modulate furin cleavage and viral dissemination in specific cell types or tissues.

An unexpected finding was the identification of SPs as major targets of positive selection in HCMV isolates. Although they were considered for a long time as simple and interchangeable targeting signals, emerging evidence indicate that SPs can regulate important steps in protein processing, including translocon interaction efficiency, glycosylation patterns, and folding [79]. However, few naturally occurring variants in the SPs of viral proteins were characterized. Herein we show that positively selected sites within the *UL144* SP modulate the dynamics of SP processing and of protein sorting. In fact, we show that the delayed trafficking of *UL144* gtB compared to gtA is due to late cleavage by of the SP. This is in line with experiment on

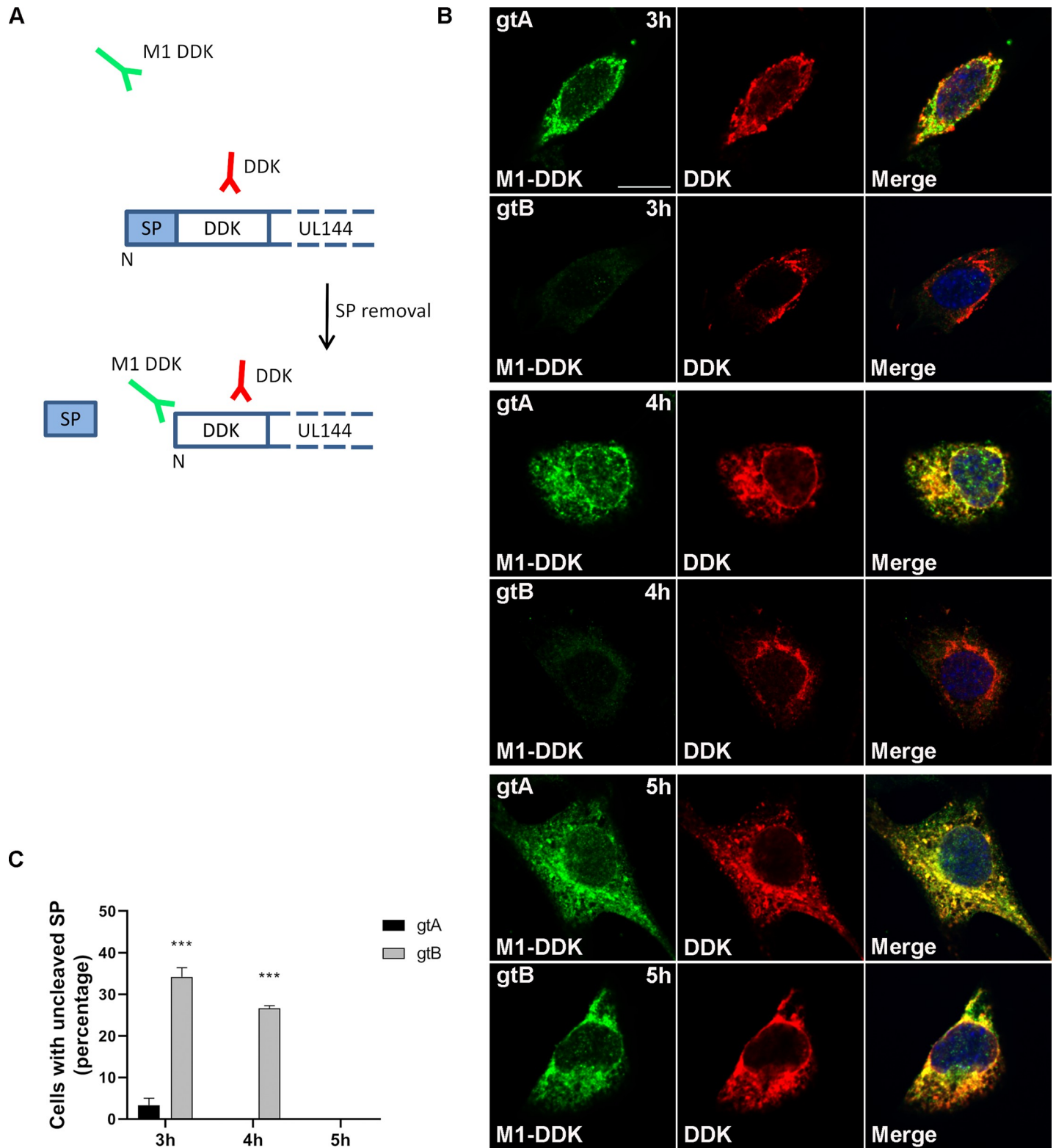


Fig 8. Time-course analysis of signal peptide cleavage. (A) Schematic representation of the vectors used in the experiment. To analyze the removal of the signal peptide from the gtA and gtB variants, a DDK tag was cloned immediately downstream the SP. The DDK tag was detected by using two antibodies: an anti-M1-DDK Ab (green signal) that recognizes only the free N-terminal end of the DDK tag (generated by the cleavage of the SP), and an anti-DDK Ab (red signal) that recognizes the tag independently of its position in the protein and stains all the cells transfected with UL144 variants. (B) HeLa cells were transfected and fixed at different time points.

Three, four, and five hours post transfection cells were immunostained with anti-M1-DDK and anti-DDK Abs. Nuclei were counterstained with DAPI. Yellow in the merge images indicates co-localization. Scale bar: 10 μm . (C) Cells only stained with the anti-DDK Ab (red only, indicating an uncleaved SP) were counted, normalized to the total number of analyzed cells and reported as percentage (two way ANOVA; $n > 60$) (***, $P < 0.001$).

<https://doi.org/10.1371/journal.ppat.1008476.g008>

engineered and mammalian proteins showing that the SP sequence can determine the timing of its cleavage [80]. Following translocation through the Sec61 channel, the SP is inserted into the ER membrane and, for most proteins, co-translationally cleaved [81]. However, some SPs are cleaved post-translationally. This is the case of another HCMV protein, US11, an ER-resident protein: the sequence of the US11 SP determines its late removal, although the functional significance of this unusual processing is still unknown [82]. The best-studied case of late SP removal, however, is that of the HIV-1 Env protein. The Env SP acts as a membrane tether for at least 15 min after protein synthesis and its cleavage occurs only after initial folding of the ectodomain [83]. The SP sequence is sufficient to dictate late cleavage [56], which, in turn, determines correct protein folding and eventually modulates Env antigenic properties and glycosylation patterns, as well as Env incorporation into virions and HIV-1 infectivity [56, 84, 85]. Because the Env SP is highly variable, it was suggested that, although not present in the mature protein, it is subject to antibody-mediated immune pressure [84]. Notably, we also detected positively selected sites in the SPs of gB and gH, which represent major targets of neutralizing antibodies and promising candidates for vaccine development. An alternative possibility is that the changes in sorting dynamics and SP cleavage timing affect other protein functions, such as the efficiency of UL144 binding to its cellular partner or the amount of mature protein that is delivered to functionally relevant location. Other experiments will be needed to disentangle these possibilities.

Molecular evolution analyses can provide information on the location and nature of adaptive changes in genomic regions, thus highlighting the presence of functional variation either at the inter- or intra-species level. We performed a two-tier analysis of HCMV evolution, by describing selective events that occurred during HCMV speciation and by identifying more recently emerged adaptive variants in clinical isolates. We provide proof-of-concept validation of the functional effect of positively selected sites in two viral genes. Thus, we used evolutionary information to generate experimentally-testable hypotheses concerning the functional effect of HCMV diversity, eventually generating a catalog of candidate modulators of viral phenotypes.

Materials and methods

Sequences and alignments

Viral genome sequences were retrieved from the National Center for Biotechnology Information database (NCBI, <http://www.ncbi.nlm.nih.gov/>). Only complete genome sequences were included in this study (S1 and S7 Tables). For each genome, we retrieved coding sequences corresponding to the 169 ORFs annotated for the Merlin strain (NC_006273). Orthology was inferred according to previously reported analyses and genome annotations, as well as by applying a phylogenetic approach to the largest gene families (see below) [14, 86–89]. Gene alignments were generated using MAFFT (version 7.392) [90], setting sequence type as codons and using default parameters; unreliably aligned codons were filtered using GUIDANCE2 [91] with a score of 0.90 [92]. The resulting alignments were manually inspected.

Whole genome alignments were obtained using Progressive MAUVE 2.3.1 [93, 94], a program designed to construct multiple genome alignments in the presence of large-scale changes such as deletions, rearrangements and inversions. MAUVE identifies and aligns regions of

local collinearity (locally collinear blocks, LCBs). Each LCB is a region of sequence homology shared by two or more of the genomes being aligned. MAUVE was run using default parameters.

Coding gene annotations and protein interaction network

Classification based on gene conservation was obtained from previous works [7, 10]. Functional classification was obtained from a survey of HCMV proteins [11]. For protein localization, information was obtained from gene ontology classifications (component category) via the QuickGO server (<https://www.ebi.ac.uk/QuickGO/>). Because the same gene can be associated with different functional classification and/or GO component terms (i.e., terms are not independent), we compared genes associated with a given term with those that are not associated to it. Terms were included only if they had more than five contributing genes.

Growth phenotypes were obtained from Van Damme et al. [11]. Phenotypes were made independent by considering as essential (E) only genes that were unequivocally described as such—e.g., genes annotated as essential/growth defective (GD) were assigned to the GD category- and by ascribing to the GD category any gene for which growth defect have been described at least once. Finally, the enhanced growth phenotype (2 genes) was collated to the dispensable (D) category.

Data on protein-protein interactions were derived from two yeast two-hybrid analyses that tested binary interaction between more than 30 proteins each [27, 28]. Additional protein-protein interaction data were derived from the VirHostNet database (<http://virhostnet.prabi.fr>) [29]. All interactions (physical associations) between HCMV-encoded proteins were retrieved from VirHostNet, irrespective of the viral strain to which they referred. Data from the two studies and from VirHostNet were merged and a network was visualized using Cytoscape version 3.6.0 [95].

Orthology assessment for large gene families

In order to make a reliable orthology analysis of the four largest gene families (*RL11*, *US22*, *US6*, and *US12* families), we used HCMV proteins belonging to these families to query the genomes of primate CMVs using different bioinformatic tools. Specifically we used: 1) protein BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>); 2) the Conserved Domain Search service (CD-Search), a web-based tool for the detection of structural and functional domains in protein sequences from the NCBI's Conserved Domain Database (CDD) server [96]; 3) the HMM-HMM-based lightning-fast iterative sequence search (HHblits), which uses a profile-profile alignment prefilter enabling fast, iterative and accurate sequence searches [97]. All identified homologs were aligned using MAFFT (version 7.392) using default parameters. Phylogenetic trees were constructed using RAxML (version 8.2.12, PROTGAMMAJTT matrix) with 1000 bootstrap replicates [98]. Orthologous genes were inferred on the basis of the tree topology (sequences clustering by viral species) and of bootstrap values (support higher than 90).

Selective patterns in catarrhini-infecting CMVs

To obtain an overview of the selective patterns of primate CMV coding genes, we analyzed a viral phylogeny composed of 11 strains including 3 CMVs that infect great apes and 8 that infect either African or Asian Old World monkeys (S1 Table). Only one-to-one orthologs were included in the analysis. Alignments that, after GUIDANCE filtering, had less than 250 aligned nucleotides were discarded (S2 Table). The average dN/dS parameter was calculated using the single-likelihood ancestor counting (SLAC) method [99]. Inputs were the multiple sequence alignments and trees generated with the phyML program (version 3.1). For the latter, we

applied a General Time Reversible (GTR) model plus gamma-distributed rates and 4 substitution rate categories, a fixed proportion of invariable sites, and a BioNJ starting tree [100].

Differences in dN/dS among catarrhini-infecting CMVs genes grouped on the basis of gene conservation and gene dispensability were evaluated using Kruskal-Wallis tests. Nemenyi tests with χ^2 distribution to account for ties were used as post-hocs. Differences in dN/dS among genes grouped on the basis of functional classification and protein localization were evaluated by Wilcoxon Rank-Sum tests. FDR correction was applied to account for multiple tests. These calculations were performed in the R environment using the PMCMR package [101].

Detection of positive selection in the HCMV lineage

Analyses were performed on a phylogeny that included all fully-sequenced catarrhini-infecting CMVs ($n = 16$) and 25 HCMV clinical isolates with no history of passaging in cell culture. Strains were selected to derive from different tissues and from different countries (S1 Table).

Phylogenetic trees were reconstructed using phyML (version 3.1). Each alignment was screened for the presence of recombination using GARD [102], a genetic algorithm implemented in the HYPHY suite (version 2.2.4). This method uses phylogenetic incongruence among segments in the alignment to detect the best-fit number and location of recombination breakpoints. The statistical significance of putative breakpoints is determined through Kishino-Hasegawa (HK) tests. When evidence of recombination was detected ($p\text{-value} < 0.01$), the coding alignment was split on the basis of the recombination breakpoints and sub-regions were used as the input for molecular evolution analyses. Recombination breakpoints were identified in 12 genes: *UL29*, *UL44*, *UL48*, *UL54*, *UL55*, *UL56*, *UL57*, *UL74*, *UL75*, *UL123*, *US23* and *US28*. Only resulting alignments that, after GUIDANCE filtering had a length ≥ 250 nt were considered for subsequent analyses.

To search for episodic positive selection on the HCMV branch, we applied the branch-site likelihood ratio tests from codeml (“test 2”) [21]. In this test, branches are divided *a priori* into foreground (those to be analyzed for positive selection, in this case the branch leading to the 25 HCMV isolates) and background lineages (in this case all other branches), and a likelihood ratio test is applied to compare a model (MA) that allows positive selection on the foreground lineages with a model (MA1) that does not allow such positive selection. Twice the difference of likelihood for the two models ($\Delta \ln L$) is then compared to a χ^2 distribution with one degree of freedom [21]. The analyses were performed using an F3X4 codon frequency models. An FDR correction was applied to account for multiple tests.

To identify sites evolving under positive selection on HCMV lineage we used the BEB analysis from MA (with a cutoff of 0.90) and the Mixed Effects Model of Evolution (MEME) (with the default cutoff of 0.1) [103]. MEME allows the distribution of nonsynonymous substitution/synonymous substitution rate (dN/dS, also referred to as ω) to vary from site to site and from branch to branch at a site. To limit false positives, only sites confirmed by both methods were considered as positively selected.

Analysis of selective patterns in HCMV isolates

HCMV genomes were derived from the NCBI database (NCBI, <http://www.ncbi.nlm.nih.gov>). Based on the number of available sequences, we analyzed HCMV isolates derived from amniotic fluid, urine and blood/plasma. Most of these latter derive from transplant recipients, whereas little information on the clinical history of the infected subjects was available for the other isolates (S7 Table). Only isolates that were directly sequenced with no *in vitro* passages were included. When stop codon/frameshift mutations were identified (e.g., in *UL9*), the gene of the isolate was removed from analysis.

For all genes, a HCMV outgroup sequence was reconstructed by maximum-likelihood from a balanced subset of no/low passaged HCMV strains derived from different compartments (S6 Table). Maximum-likelihood inference was obtained with FastML [104]. The distribution of fitness effects (DFE) along the 169 ORFs annotated for the Merlin strain was estimated using GammaMap [42]. We assumed θ (neutral mutation rate per site), k (transitions/transversions ratio), and T (branch length) to vary among genes following log-normal distributions, whereas p (probability of adjacent codons to share the same selection coefficient) following a log-uniform distribution. For each gene we set the neutral frequencies of non-STOP codons (1/61). For selection coefficients we considered a uniform Dirichlet distribution with the same prior weight for each selection class. For each gene we run 100,000 iterations with thinning interval of 10 iterations.

3D structures

All crystallographic structures were derived from the Protein Data Bank (PDB) (see IDs in Figure legends). Images were rendered using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.2 Schrödinger, LLC).

Cell lines and HCMV strain

Primary human foreskin fibroblasts (HFFs, ATCC SCRC-1041, male), human retinal pigment epithelial cells (ARPE-19, ATCC CRL-2302, male), human epithelial adenocarcinoma cells (HeLa, ATCC CCL-2, female) and human embryonic kidney cells (HEK-293, ATCC CRL-1573, female) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS, Euroclone, Milano, Italy), 2 mM L-glutamine and 100 U/ml penicillin/streptomycin (Invitrogen, Carlsbad, CA, USA, Thermo Fisher Scientific, Waltham, MA, USA). All cell lines were grown at 37°C in a humidified 5% CO₂ incubator.

HCMV strain TB40-BAC4 was propagated and titrated on HFFs by standard plaque assay [105, 106].

Plasmids and Antibodies

pEPkan-S template plasmid (Addgene plasmid # 41017) [107] was used for the mutagenesis of vUL70. pSGIE86 IE1-encoding plasmid was used in quantitative nucleic acid analysis [108].

pCMV6-UL144 (gtA) (VC101305) was purchased from Origene. The sequence of UL144 (gtB) was synthesized and cloned in pCMV6-Entry vector by Origene custom service.

To study the timing of SP removal, the 24 bp DDK tag sequence (also referred to as FLAG) was inserted downstream the SP by a PCR site-direct mutagenesis approach. Briefly, for each construct (gtA and gtB), a pair of complementary primers were designed between the SP and the next sequence portion, and the DDK tag sequence was inserted right in the middle of both complementary primers at the end of SP sequence (S4 Table and Fig 8A). PCR was carried out in a 50 μ l reaction containing 10 ng of specific plasmid template, 10 μ M of each of the two primer pairs, 200 μ M dNTPs, and 3 units of Pfu DNA polymerase (Promega). Reactions were performed under the following cycling conditions: initial denaturation at 95°C for 2 min; 35 cycles at 95°C for 30 s, 70°C for 30 s, and 72°C for 10 min; and a final extension at 72°C for 15 min. The PCR products were treated with 5 units of *DpnI* (New England BioLabs) at 37°C for 2 hours and 5 μ l aliquot of each digested PCR products was transformed into *E. coli* DH5 α competent cells (ThermoFisher Scientific) by heat shock. The transformed cells were spread on a Luria-Bertani (LB) plate containing kanamycin antibiotic and incubated at 37°C overnight. Colonies from each plate were grown and the plasmid DNA was isolated using Qiagen

miniprep kits (Qiagen). To verify the correctness of the constructs, clones were sequenced using specific primers on pCMV6 backbone (S4 Table).

Mouse monoclonal anti-DDK—Clone 4C5 (TA50011-100, Origene) was used to detect DDK-tagged UL144. Rabbit antibodies against Sodium Potassium ATPase (ab76020), LAMP1 (ab24170), and SEC61a (ab221288) were purchased from Abcam. Goat polyclonal anti-EEA1 (N-19, sc-6415) was purchased from Santa Cruz Biotechnology. Rabbit polyclonal anti-calreticulin (PA3-900) was purchased from Thermo Fisher Scientific.

For the evaluation of the timing of SPs removal, rabbit anti-DDK polyclonal antibody (TA100023, Origene) and mouse monoclonal ANTI-FLAG M1 antibody (F3040, Merck) were used to detect the DDK-tag (irrespective of context) and the free N-terminus DDK-tag, respectively.

Alexa Fluor secondary antibodies Goat anti-Mouse IgG (H+L) - 488 (A-11001), Goat anti-Rabbit IgG (H+L) - 546 (A-11010), Donkey anti-Mouse IgG (H+L) - 488 (A-21202), Donkey anti-Rabbit IgG (H+L) - 546 (A-10040), and Donkey anti-Goat IgG (H+L) - 647 (A-21447) were purchased from Thermo Fisher Scientific.

HCMV mutagenesis

Viruses used in this study were all bacterial artificial chromosome (BAC) clones. The clones of the endotheliotropic HCMV strain TB40-BAC4 (wild-type) viruses were generated using a markerless two-step RED-GAM recombination protocol [107, 109]. RED recombination was used to modify the TB40-BAC4 BAC that is derived from the endotheliotropic HCMV strain TB40/E [110]. The BAC of mutant virus vUL70 was generated with specific primers listed in S4 Table. Primers contained sequences of homology upstream and downstream the sites to be mutated, the mutation (lowercase), and sequences homologous to the pEPkan-S template plasmid (underlined) (Addgene plasmid # 41017; <http://n2t.net/addgene:41017>; RRID: Addgene_41017) [107]. All generated recombinant BAC DNAs were controlled for integrity and correctness by restriction length polymorphism and sequencing of the mutated region. HFF cells were used for reconstitution of recombinant viruses and virus stock production. The reconstitution from BAC DNA was performed as previously described [111].

Focus Expansion Assay (FEA)

To analyze BAC clone transmission and spreading on different cell types, focus expansion assay (FEA) was performed as previously described [36]. Serial dilutions of HFFs (104 to 1) infected by the indicated viruses were co-cultured with an excess of uninfected HFFs or ARPE-19 cells in a 96 well plate. After 5 days of cocultivation, cells were fixed with cold methanol and HCMV immediate early antigen (CMV IE1/2 Monoclonal Antibody—CH160, Vyrusis Corporation) was detected by indirect immunoperoxidase staining (VECTASTAIN Universal Quick HRP Kit, R.T.U., Vector Laboratories). Stained slides were read with a Leica ICC50 HD microscope (Leica Microsystems). Infectious foci were defined as clusters of three or more antigen-positive cells. The number of infected cells in ten representative foci were counted. Plaque area and IEA positive nuclei were calculated using ImageJ software.

Statistical tests were performed using GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego California USA). One-way ANOVA followed by Tukey's post-hoc test was used for comparison of mutant viruses versus the wild-type.

Growth competition assay

HFFs were plated in a p24 well and infected with 0.0005 PFU/ml of wild-type and vUL70 G294L each, wild-type and vUL70 R465K each, wild-type and vUL70 dm each. Virus growth

was maintained for 28 days, while cells were split and fresh culture media replenished twice a week. At different time points, cells were collected and total genomic DNA was extracted using TRI Reagent solution (Sigma-Aldrich), according to the manufacturer's protocol, and eluted into 30 μ L of water. Each head-to-head competition consisted of 2 biological replicates. The relative proportion of wild-type and mutant viruses was assessed using deep sequencing.

Library preparation for Illumina sequencing on NextSeq 550 platform was performed in accordance with Illumina tech note 15044223 Rev-B, using PCRIBIO HS VeriFi Mix (PCR Biosystems), with two primer sets built to amplify PCR fragments around the 294 and 465 variants (S4 Table).

After purification with Ampure XP (Beckman Genomics), samples were quantified using Qubit high sensitivity reagents (ThermoFisher Scientific) and then pooled. Final pools of up to 96 libraries were quantified using high sensitivity Bioanalyzer reagents (Agilent) prior to sequencing 2X150 bp long paired end reads of the amplicons with NextSeq. Samples were supplemented with 1% phiX control library (Illumina). All reagents were used according to manufacturer's recommendations.

The generated reads were processed to remove adaptors and low quality bases ($q < 30$). The bwa aligner [112] was used to map the selected reads to the *UL70* gene. The aligned sequences were analyzed with samtools [113] and homemade R-scripts to count the number of reads deriving from wild-type or mutant viruses. The statistical significance of an increasing trend in mutant read counts across time points were assessed with Cochran Armitage tests for trend. These were performed by considering the counts of sequence reads corresponding to the wild-type and the mutant virus as the response variable and time points as the explanatory variable with ordered levels.

Relative fitness (f) was calculated by applying the following formula:

$$f = \frac{1}{t} \cdot \ln \frac{p(t) \cdot q(0)}{q(t) \cdot p(0)}$$

where t is time (expressed in days), p and q are the proportion of reads deriving from mutants and wild-type viruses, respectively [114, 115].

Immunofluorescence and confocal microscopy

HeLa/HEK-293 cells were seeded (0.3×10^5 cells/well) onto coverslips treated with 0.1 μ g/mL poly-L-lysine. Transient transfections were performed using Lipofectamine 2000 (Thermo Fisher Scientific, 11,668,027) with 2.5 μ g of plasmid DNA (pCMV6-UL144 gtA and pCMV6-UL144 gtB). At 3, 4, 5, 6, 9 and 24 hours after transfection, cells were fixed with 4% paraformaldehyde (Santa Cruz Biotechnology, sc-281692) and permeabilized with phosphate-buffered saline (PBS; Euroclone, ECB4053L) containing 0.1% saponin (Sigma-Aldrich, S4521) and 1% bovine serum albumin (Sigma-Aldrich, A9647).

Cells were then incubated in the same buffer for 2h with primary Abs (1:50) and revealed using the secondary Abs Alexa Fluor 488, 546 and 647 (1:500).

Confocal microscopy was performed with a Yokogawa CSU-X1 spinning disk confocal on a Nikon Ti-E inverted microscope equipped with a Nikon 60x/1.40 oil Plan Apochromat objective and were acquired with an Andor Technology iXon3 DU-897-BV EMCCD camera (Nikon Instruments S.p.A., Firenze, Italy). The investigator was blinded as to the nature of the sample analyzed. Pearson correlation coefficients for protein co-localization were determined with Fiji ImageJ software using the COLOC2 plugin.

Statistical analyses were performed using GraphPad Prism version 8.1.1 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com).

Significance was calculated with two-way ANOVA followed by Sidak's multiple comparisons test or with Student's *t* test for unpaired variables (two-tailed), as detailed in the legend of the figures (Figs 6, 7 and 8). Pearson's correlation coefficients are reported as mean \pm standard error of the mean (SEM); *n* represents individual data. For the statistical analysis, a Fisher transformation in z-scores of Pearson's correlation coefficients was performed. *p* values of less than 0.05 were considered significant.

Supporting information

S1 Fig. Phylogenetic relationships for the *US6* and *US12* gene families. Protein sequences of family homologs were searched for as described in the Materials and Methods. Phylogenetic trees were constructed using RAxML with 1000 bootstrap replicates (posterior probabilities are reported at nodes). Orthologous gene groups, shown in red on the tree and denoted by the gray shading, were inferred on the basis of the tree topology and of bootstrap values > 70 . Magenta asterisks denote genes that are frequently deleted/mutated in clinical isolates [16].
(PDF)

S2 Fig. Analysis of selective patterns for catarrhini-infecting CMVs. The dN/dS parameter is compared among genes showing different function (A) or by the location of the encoded protein (B). *p* values derive from Wilcoxon Rank-Sum tests with FDR correction. Star indicates significant *p* value (< 0.05).
(PDF)

S3 Fig. Selective patterns in HCMV clinical isolates from different compartments. Distribution of selection coefficients (γ) for all coding genes of HCMV clinical isolates sampled from amniotic fluid, urine, and blood/plasma. Selection coefficients were calculated for all codons and genes were grouped on the basis of the results of the branch-site analysis.
(PDF)

S4 Fig. Population genetics-phylogenetics analysis of HCMV genes. Violin plots (median, white dot; interquartile range, black bar) of selection coefficients for genes of HCMV isolates deriving from the blood/plasma (red), urine (yellow), and amniotic fluid (light blue). Selection coefficients (γ) are classified as strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (-1), moderately deleterious (-5 , -10), strongly deleterious (-50 , -100), and inviable (-500). The gray shading denotes different degrees of constraint based on selection coefficients.
(PDF)

S5 Fig. Positively selected sites detected by gammaMap analysis. Positively selected sites (black arrows) were mapped onto the topological domains of HCMV proteins. Protein domain information was obtained from the Uniprot and SMART databases. Positions refer to proteins of the Merlin strain (NC_006273) (see also S8 Table).
(PDF)

S6 Fig. Positive selection, sequence diversity, and ADs in envelope glycoproteins. Protein domain information and positively selected sites of gB, gH, gM and gO are as in S5 Fig. Plots below the schematic representations of protein domains report the number of amino acidic substitutions per site, as inferred from the sequences used for gammaMap analysis (see S7 Table) and provided as a measure of polymorphism. Black and blue triangles denote sites identified with gammaMap or with the branch-site test, respectively. Grey boxes indicated linear epitopes mapped onto ADs. The name of major neutralizing antibodies targeting these epitopes are reported. gB and gH present some region with high sequence polymorphism. They

are also highly immunogenic and represent major targets of neutralizing antibodies (see text). gM displays very low levels of sequence diversity among strains, whereas high sequence divergence was observed in gO, especially at the N-terminus, coincident with the signals of positive selection. For gO, no AD/epitope has been identified yet.
(PDF)

S7 Fig. Analysis of UL144 gtA and gtB localization in HEK-293 cells. HEK-293 cells were transfected with pCMV6-UL144 gtA and pCMV6-UL144 gtB. **(A)** Twenty-four hours after transfection cells were fixed and immunostained with antibodies against the DDK tag (green) and the plasma membrane protein sodium potassium ATPase (red). Nuclei were counterstained with DAPI. Arrows indicates co-localization at the plasma membrane. Scale bar: 10 μ m. **(B)** Twenty-four hours after transfection cells were fixed and immunostained with antibodies against the DDK tag (green), the lysosomal marker LAMP1 (red), and the early endosomal marker EEA1 (blue). Co-localization of DDK with LAMP1 (yellow) or EEA1 (light blue) is showed in the merge images. The small panels show a higher magnification of the area indicated in the squares. Scale bar: 10 μ m. Pearson's correlation coefficients for DDK/LAMP1 or DDK/EEA1 co-localization were reported in the graphs as mean \pm SEM (*t* test; *n* > 30). Scale bar: 10 μ m. **(C)** Four hours after transfection cells were fixed and immunostained with antibodies against the DDK tag (green) and Sec61A (red). Nuclei were counterstained with DAPI. Yellow in the merge images indicates co-localization. Pearson's correlation coefficients for DDK/Sec61A co-localization were reported in the graphs as mean \pm SEM (*t* test; *n* > 25). Scale bar: 10 μ m. **(D)** Six hours after transfection cells were fixed and immunostained with antibodies against the DDK tag (green) and calreticulin (red). Nuclei were counterstained with DAPI. Yellow in the merge images indicates co-localization. Pearson's correlation coefficients for DDK/Calreticulin co-localization were reported in the graphs as mean \pm SEM (*t* test; *n* > 25). Scale bar: 10 μ m.
(PDF)

S1 Table. List of sequences used for the branch-site test.
(PDF)

S2 Table. CMV genes excluded from the branch-site test.
(PDF)

S3 Table. Likelihood ratio test (LRT) statistics for models of variable selective pressure on the HCMV branch.
(PDF)

S4 Table. List of primers.
(PDF)

S5 Table. Focus expansion assay (FEA).
(PDF)

S6 Table. List of HCMV strains used for HCMV ancestral outgroup reconstruction.
(PDF)

S7 Table. List of HCMV strains used for gammaMap analyses.
(PDF)

S8 Table. List of positively selected sites detected with gammaMap.
(XLSX)

Acknowledgments

We thank Andrea Citterio for providing technical support in NGS analysis.

Author Contributions

Conceptualization: Santo Landolfo, Manuela Sironi.

Data curation: Rachele Cagliani, Valentina Dell'Oste.

Formal analysis: Alessandra Mozzi, Matteo Biolatti, Rachele Cagliani, Diego Forni, Valentina Dell'Oste, Uberto Pozzoli, Manuela Sironi.

Funding acquisition: Diego Forni, Valentina Dell'Oste, Manuela Sironi.

Investigation: Alessandra Mozzi, Rachele Cagliani, Diego Forni, Valentina Dell'Oste, Chiara Pontremoli, Chiara Vantaggiato, Mario Clerici.

Methodology: Alessandra Mozzi, Matteo Biolatti, Rachele Cagliani, Diego Forni, Valentina Dell'Oste, Chiara Pontremoli, Chiara Vantaggiato.

Software: Diego Forni, Uberto Pozzoli.

Supervision: Mario Clerici, Manuela Sironi.

Visualization: Alessandra Mozzi, Matteo Biolatti, Diego Forni.

Writing – original draft: Matteo Biolatti, Santo Landolfo, Manuela Sironi.

Writing – review & editing: Mario Clerici, Manuela Sironi.

References

1. Barry PA, William Chang W. (2007) Primate betaherpesviruses. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, et al, editors. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge: Cambridge University Press; 2007. pp. 1051–1075.
2. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, et al. (2019) Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev Med Virol* 29(3): e2034. <https://doi.org/10.1002/rmv.2034> PMID: 30706584
3. Griffiths P, Baraniak I, Reeves M. (2015) The pathogenesis of human cytomegalovirus. *J Pathol* 235(2): 288–297. <https://doi.org/10.1002/path.4437> PMID: 25205255
4. Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. (2013) The "silent" global burden of congenital cytomegalovirus. *Clin Microbiol Rev* 26(1): 86–102. <https://doi.org/10.1128/CMR.00062-12> PMID: 23297260
5. Rawlinson WD, Boppana SB, Fowler KB, Kimberlin DW, Lazzarotto T, et al. (2017) Congenital cytomegalovirus infection in pregnancy and the neonate: Consensus recommendations for prevention, diagnosis, and therapy. *Lancet Infect Dis* 17(6): e177–e188. [https://doi.org/10.1016/S1473-3099\(17\)30143-3](https://doi.org/10.1016/S1473-3099(17)30143-3) PMID: 28291720
6. Wang GC, Kao WH, Murakami P, Xue QL, Chiou RB, et al. (2010) Cytomegalovirus infection and the risk of mortality and frailty in older women: A prospective observational cohort study. *Am J Epidemiol* 171(10): 1144–1152. <https://doi.org/10.1093/aje/kwq062> PMID: 20400465
7. Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, Lee L, et al. (2004) Genetic content of wild-type human cytomegalovirus. *J Gen Virol* 85(Pt 5): 1301–1312. <https://doi.org/10.1099/vir.0.79888-0> PMID: 15105547
8. Murphy E, Yu D, Grimwood J, Schmutz J, Dickson M, et al. (2003) Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* 100(25): 14976–14981. <https://doi.org/10.1073/pnas.2136652100> PMID: 14657367
9. McGeoch DJ, Rixon FJ, Davison AJ. (2006) Topics in herpesvirus genomics and evolution. *Virus Res* 117(1): 90–104. <https://doi.org/10.1016/j.virusres.2006.01.002> PMID: 16490275
10. Davison AJ. (2007) Comparative analysis of the genomes. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, et al, editors. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge: Cambridge University Press; 2007. pp. 10–26.

11. Van Damme E, Van Loock M. (2014) Functional annotation of human cytomegalovirus gene products: An update. *Front Microbiol* 5: 218. <https://doi.org/10.3389/fmicb.2014.00218> PMID: 24904534
12. Mocarski E Jr. (2007) Betaherpes viral genes and their functions. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, et al, editors. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge: Cambridge University Press; 2007. pp. 204–230.
13. McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EA. (1995) Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J Mol Biol* 247(3): 443–458. <https://doi.org/10.1006/jmbi.1995.0152> PMID: 7714900
14. Russell JN, Marsh AK, Willer DO, Ambagala AP, Dzamba M, et al. (2016) A novel strain of cynomolgus macaque cytomegalovirus: Implications for host-virus co-evolution. *BMC Genomics* 17: 277-016–2588-3.
15. Sironi M, Cagliani R, Forni D, Clerici M. (2015) Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet* 16(4): 224–236. <https://doi.org/10.1038/nrg3905> PMID: 25783448
16. Sijmons S, Thys K, Mbong Ngwese M, Van Damme E, Dvorak J, et al. (2015) High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J Virol* 89(15):7673–7695. <https://doi.org/10.1128/JVI.00578-15> PMID: 25972543
17. Cudini J, Roy S, Houldcroft CJ, Bryant JM, Depledge DP, et al. (2019) Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. *Proc Natl Acad Sci U S A* 116(12): 5693–5698. <https://doi.org/10.1073/pnas.1818130116> PMID: 30819890
18. Dunn W, Chou C, Li H, Hai R, Patterson D, et al. (2003) Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A* 100(24): 14223–14228. <https://doi.org/10.1073/pnas.2334032100> PMID: 14623981
19. Stamatakis A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
20. Yu D, Silva MC, Shenk T. (2003) Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A* 100(21): 12396–12401. <https://doi.org/10.1073/pnas.1635160100> PMID: 14519856
21. Zhang J, Nielsen R, Yang Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22(12): 2472–2479. <https://doi.org/10.1093/molbev/msi237> PMID: 16107592
22. Smith RM, Kosuri S, Kerry JA. (2014) Role of human cytomegalovirus tegument proteins in virion assembly. *Viruses* 6(2): 582–605. <https://doi.org/10.3390/v6020582> PMID: 24509811
23. Tandon R, Mocarski ES. (2012) Viral and host control of cytomegalovirus maturation. *Trends Microbiol* 20(8): 392–401. <https://doi.org/10.1016/j.tim.2012.04.008> PMID: 22633075
24. Malito E, Chandramouli S, Carfi A. (2018) From recognition to execution—the HCMV pentamer from receptor binding to fusion triggering. *Curr Opin Virol* 31: 43–51. <https://doi.org/10.1016/j.coviro.2018.05.004> PMID: 29866439
25. Meyer H, Sundqvist VA, Pereira L, Mach M. (1992) Glycoprotein gp116 of human cytomegalovirus contains epitopes for strain-common and strain-specific antibodies. *J Gen Virol* 73 (Pt 9)(Pt 9): 2375–2383.
26. Spindler N, Diestel U, Stump JD, Wieggers AK, Winkler TH, et al. (2014) Structural basis for the recognition of human cytomegalovirus glycoprotein B by a neutralizing human antibody. *PLoS Pathog* 10 (10): e1004377. <https://doi.org/10.1371/journal.ppat.1004377> PMID: 25299639
27. Phillips SL, Bresnahan WA. (2011) Identification of binary interactions between human cytomegalovirus virion proteins. *J Virol* 85(1): 440–447. <https://doi.org/10.1128/JVI.01551-10> PMID: 20962080
28. To A, Bai Y, Shen A, Gong H, Umamoto S, et al. (2011) Yeast two hybrid analyses reveal novel binary interactions between human cytomegalovirus-encoded virion proteins. *PLoS One* 6(4): e17796. <https://doi.org/10.1371/journal.pone.0017796> PMID: 21483780
29. Guirimand T, Delmotte S, Navratil V. (2015) VirHostNet 2.0: Surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* 43(Database issue): D583–7. <https://doi.org/10.1093/nar/gku1121> PMID: 25392406
30. Yu X, Jih J, Jiang J, Zhou ZH. (2017) Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. *Science* 356(6345): <https://doi.org/10.1126/science.aam6892> PMID: 28663444

31. Silvestri M, Sundqvist VA, Ruden U, Wahren B. (1991) Characterization of a major antigenic region on gp55 of human cytomegalovirus. *J Gen Virol* 72 (Pt 12)(Pt 12): 3017–3023.
32. Chandramouli S, Ciferri C, Nikitin PA, Calo S, Gerrein R, et al. (2015) Structure of HCMV glycoprotein B in the postfusion conformation bound to a neutralizing human antibody. *Nat Commun* 6: 8176. <https://doi.org/10.1038/ncomms9176> PMID: 26365435
33. Ohlin M, Sundqvist VA, Mach M, Wahren B, Borrebaeck CA. (1993) Fine specificity of the human immune response to the major neutralization epitopes expressed on cytomegalovirus gp58/116 (gB), as determined with human monoclonal antibodies. *J Virol* 67(2): 703–710. PMID: 7678304
34. Burke HG, Heldwein EE. (2015) Crystal structure of the human cytomegalovirus glycoprotein B. *PLoS Pathog* 11(10): e1005227. <https://doi.org/10.1371/journal.ppat.1005227> PMID: 26484870
35. Scrivano L, Sinzger C, Nitschko H, Koszinowski UH, Adler B. (2011) HCMV spread and cell tropism are determined by distinct virus populations. *PLoS Pathog* 7(1): e1001256. <https://doi.org/10.1371/journal.ppat.1001256> PMID: 21249233
36. Sinzger C, Knapp J, Plachter B, Schmidt K, Jahn G. (1997) Quantification of replication of clinical cytomegalovirus isolates in cultured endothelial cells and fibroblasts by a focus expansion assay. *J Virol Methods* 63(1–2): 103–112. [https://doi.org/10.1016/s0166-0934\(97\)02082-x](https://doi.org/10.1016/s0166-0934(97)02082-x) PMID: 9015280
37. Mariame B, Kappler-Gratias S, Kappler M, Balor S, Gallardo F, et al. (2018) Real-time visualization and quantification of human cytomegalovirus replication in living cells using the ANCHOR DNA labeling technology. *J Virol* 92(18): <https://doi.org/10.1128/JVI.00571-18> Print 2018 Sep 15. PMID: 29950406
38. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. (2011) Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog* 7(5): e1001344. <https://doi.org/10.1371/journal.ppat.1001344> PMID: 21625576
39. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, et al. (2013) Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLoS Genet* 9(9): e1003735. <https://doi.org/10.1371/journal.pgen.1003735> PMID: 24086142
40. Kryazhimskiy S, Plotkin JB. (2008) The population genetics of dN/dS. *PLoS Genet* 4(12): e1000304. <https://doi.org/10.1371/journal.pgen.1000304> PMID: 19081788
41. Renzette N, Pokalyuk C, Gibson L, Bhattacharjee B, Schleiss MR, et al. (2015) Limits and patterns of cytomegalovirus genomic diversity in humans. *Proc Natl Acad Sci U S A* 112(30): E4120–8. <https://doi.org/10.1073/pnas.1501880112> PMID: 26150505
42. Wilson DJ, Hernandez RD, Andolfatto P, Przeworski M. (2011) A population genetics-phylogenetics approach to inferring natural selection in coding sequences. *PLoS Genet* 7(12): e1002395. <https://doi.org/10.1371/journal.pgen.1002395> PMID: 22144911
43. Ha S, Li F, Troutman MC, Freed DC, Tang A, et al. (2017) Neutralization of diverse human cytomegalovirus strains conferred by antibodies targeting viral gH/gL/pUL128-131 pentameric complex. *J Virol* 91(7): <https://doi.org/10.1128/JVI.02033-16> Print 2017 Apr 1. PMID: 28077654
44. Sedy JR, Balmert MO, Ware BC, Smith W, Nemcovicova I, et al. (2017) A herpesvirus entry mediator mutein with selective agonist action for the inhibitory receptor B and T lymphocyte attenuator. *J Biol Chem* 292(51): 21060–21070. <https://doi.org/10.1074/jbc.M117.813295> PMID: 29061848
45. Chandramouli S, Malito E, Nguyen T, Luisi K, Donnarumma D, et al. (2017) Structural basis for potent antibody-mediated neutralization of human cytomegalovirus. *Sci Immunol* 2(12): <https://doi.org/10.1126/sciimmunol.aan1457> PMID: 28783665
46. Ciferri C, Chandramouli S, Leitner A, Donnarumma D, Cianfrocco MA, et al. (2015) Antigenic characterization of the HCMV gH/gL/gO and pentamer cell entry complexes reveals binding sites for potently neutralizing human antibodies. *PLoS Pathog* 11(10): e1005230. <https://doi.org/10.1371/journal.ppat.1005230> PMID: 26485028
47. Fouts AE, Comps-Agrar L, Stengel KF, Ellerman D, Schoeffler AJ, et al. (2014) Mechanism for neutralizing activity by the anti-CMV gH/gL monoclonal antibody MSL-109. *Proc Natl Acad Sci U S A* 111(22): 8209–8214. <https://doi.org/10.1073/pnas.1404653111> PMID: 24843144
48. Gardner TJ, Stein KR, Duty JA, Schwarz TM, Noriega VM, et al. (2016) Functional screening for anti-CMV biologics identifies a broadly neutralizing epitope of an essential envelope protein. *Nat Commun* 7: 13627. <https://doi.org/10.1038/ncomms13627> PMID: 27966523
49. Bitra A, Nemcovicova I, Picarda G, Doukov T, Wang J, et al. (2019) Structure of human cytomegalovirus UL144, an HVEM orthologue, bound to the B and T cell lymphocyte attenuator. *J Biol Chem* 294(27): 10519–10529. <https://doi.org/10.1074/jbc.RA119.009199> PMID: 31126984
50. Crooks GE, Hon G, Chandonia JM, Brenner SE. (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6): 1188–1190. <https://doi.org/10.1101/gr.849004> PMID: 15173120

51. Rost B, Yachdav G, Liu J. (2004) The PredictProtein server. *Nucleic Acids Res* 32(Web Server issue): W321–6. <https://doi.org/10.1093/nar/gkh377> PMID: 15215403
52. Cheung TC, Humphreys IR, Potter KG, Norris PS, Shumway HM, et al. (2005) Evolutionarily divergent herpesviruses modulate T cell activation by targeting the herpesvirus entry mediator cosignaling pathway. *Proc Natl Acad Sci U S A* 102(37): 13218–13223. <https://doi.org/10.1073/pnas.0506172102> PMID: 16131544
53. Poole E, King CA, Sinclair JH, Alcami A. (2006) The UL144 gene product of human cytomegalovirus activates NFkappaB via a TRAF6-dependent mechanism. *Embo j* 25(18): 4390–4399. <https://doi.org/10.1038/sj.emboj.7601287> PMID: 16932746
54. Paradowska E, Studzinska M, Suski P, Kasztelewicz B, Wisniewska-Ligier M, et al. (2015) Human cytomegalovirus UL55, UL144, and US28 genotype distribution in infants infected congenitally or postnatally. *J Med Virol* 87(10): 1737–1748. <https://doi.org/10.1002/jmv.24222> PMID: 25926093
55. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. (2018) A comprehensive review of signal peptides: Structure, roles, and applications. *Eur J Cell Biol* 97(6): 422–441. <https://doi.org/10.1016/j.ejcb.2018.06.003> PMID: 29958716
56. Snapp EL, McCaul N, Quandt M, Cabartova Z, Bontjer I, et al. (2017) Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. *Elife* 6: <https://doi.org/10.7554/eLife.26067> PMID: 28753126
57. Shao S, Hegde RS. (2011) Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol* 27: 25–56. <https://doi.org/10.1146/annurev-cellbio-092910-154125> PMID: 21801011
58. Prickett KS, Amberg DC, Hopp TP. (1989) A calcium-dependent antibody for identification and purification of recombinant proteins. *Biotechniques* 7(6): 580–589. PMID: 2698650
59. Fioretti A, Furukawa T, Santoli D, Plotkin SA. (1973) Nonproductive infection of guinea pig cells with human cytomegalovirus. *J Virol* 11(6): 998–1003. PMID: 4351465
60. Lafemina RL, Hayward GS. (1988) Differences in cell-type-specific blocks to immediate early gene expression and DNA replication of human, simian and murine cytomegalovirus. *J Gen Virol* 69 (Pt 2): 355–374.
61. Ellsmore V, Reid GG, Stow ND. (2003) Detection of human cytomegalovirus DNA replication in non-permissive vero and 293 cells. *J Gen Virol* 84(Pt 3): 639–645. <https://doi.org/10.1099/vir.0.18812-0> PMID: 12604816
62. Jurak I, Brune W. (2006) Induction of apoptosis limits cytomegalovirus cross-species infection. *Embo j* 25(11): 2634–2642. <https://doi.org/10.1038/sj.emboj.7601133> PMID: 16688216
63. Schumacher U, Handke W, Jurak I, Brune W. (2010) Mutations in the M112/M113-coding region facilitate murine cytomegalovirus replication in human cells. *J Virol* 84(16): 7994–8006. <https://doi.org/10.1128/JVI.02624-09> PMID: 20519391
64. Perot K, Walker CM, Spaete RR. (1992) Primary chimpanzee skin fibroblast cells are fully permissive for human cytomegalovirus replication. *J Gen Virol* 73 (Pt 12)(Pt 12): 3281–3284.
65. Potzsch S, Spindler N, Wieggers AK, Fisch T, Rucker P, et al. (2011) B cell repertoire analysis identifies new antigenic domains on glycoprotein B of human cytomegalovirus which are target of neutralizing antibodies. *PLoS Pathog* 7(8): e1002172. <https://doi.org/10.1371/journal.ppat.1002172> PMID: 21852946
66. Nelson CS, Huffman T, Jenks JA, Cisneros de la Rosa E, Xie G, et al. (2018) HCMV glycoprotein B subunit vaccine efficacy mediated by nonneutralizing antibody effector functions. *Proc Natl Acad Sci U S A* 115(24): 6267–6272. <https://doi.org/10.1073/pnas.1800177115> PMID: 29712861
67. Xia L, Tang A, Meng W, Freed DC, He L, et al. (2017) Active evolution of memory B-cells specific to viral gH/gL/pUL128/130/131 pentameric complex in healthy subjects with silent human cytomegalovirus infection. *Oncotarget* 8(43): 73654–73669. <https://doi.org/10.18632/oncotarget.18359> PMID: 29088734
68. Urban M, Britt W, Mach M. (1992) The dominant linear neutralizing antibody-binding site of glycoprotein gp86 of human cytomegalovirus is strain specific. *J Virol* 66(3): 1303–1311. PMID: 1371164
69. Baraniak I, Kropff B, McLean GR, Pichon S, Piras-Douce F, et al. (2018) Epitope-specific humoral responses to human cytomegalovirus glycoprotein-B vaccine with MF59: Anti-AD2 levels correlate with protection from viremia. *J Infect Dis* 217(12): 1907–1917. <https://doi.org/10.1093/infdis/jiy102> PMID: 29528415
70. Speckner A, Glykofrydes D, Ohlin M, Mach M. (1999) Antigenic domain 1 of human cytomegalovirus glycoprotein B induces a multitude of different antibodies which, when combined, results in incomplete virus neutralization. *J Gen Virol* 80 (Pt 8)(Pt 8): 2183–2191.

71. Stangherlin LM, de Paula FN, Icimoto MY, Ruiz LGP, Nogueira ML, et al. (2017) Positively selected sites at HCMV gB furin processing region and their effects in cleavage efficiency. *Front Microbiol* 8: 934. <https://doi.org/10.3389/fmicb.2017.00934> PMID: 28588572
72. Oliver SL, Sommer M, Zerboni L, Rajamani J, Grose C, et al. (2009) Mutagenesis of varicella-zoster virus glycoprotein B: Putative fusion loop residues are essential for viral replication, and the furin cleavage motif contributes to pathogenesis in skin tissue in vivo. *J Virol* 83(15): 7495–7506. <https://doi.org/10.1128/JVI.00400-09> PMID: 19474103
73. Vallbracht M, Brun D, Tassinari M, Vaney MC, Pehau-Arnaudet G, et al. (2017) Structure-function dissection of pseudorabies virus glycoprotein B fusion loops. *J Virol* 92(1): <https://doi.org/10.1128/JVI.01203-17> Print 2018 Jan 1. PMID: 29046441
74. Kopp A, Blewett E, Misra V, Mettenleiter TC. (1994) Proteolytic cleavage of bovine herpesvirus 1 (BHV-1) glycoprotein gB is not necessary for its function in BHV-1 or pseudorabies virus. *J Virol* 68(3): 1667–1674. PMID: 8107227
75. Strive T, Borst E, Messerle M, Radsak K. (2002) Proteolytic processing of human cytomegalovirus glycoprotein B is dispensable for viral growth in culture. *J Virol* 76(3): 1252–1264. <https://doi.org/10.1128/JVI.76.3.1252-1264.2002> PMID: 11773401
76. Okazaki K. (2007) Proteolytic cleavage of glycoprotein B is dispensable for in vitro replication, but required for syncytium formation of pseudorabies virus. *J Gen Virol* 88(Pt 7): 1859–1865. <https://doi.org/10.1099/vir.0.82610-0> PMID: 17554016
77. Sorem J, Longnecker R. (2009) Cleavage of epstein-barr virus glycoprotein B is required for full function in cell-cell fusion with both epithelial and B cells. *J Gen Virol* 90(Pt 3): 591–595. <https://doi.org/10.1099/vir.0.007237-0> PMID: 19218203
78. Jackson JW, Sparer T. (2018) There is always another way! cytomegalovirus' multifaceted dissemination schemes. *Viruses* 10(7): <https://doi.org/10.3390/v10070383> PMID: 30037007
79. Hegde RS, Bernstein HD. (2006) The surprising complexity of signal sequences. *Trends Biochem Sci* 31(10): 563–571. <https://doi.org/10.1016/j.tibs.2006.08.004> PMID: 16919958
80. Rutkowski DT, Ott CM, Polansky JR, Lingappa VR. (2003) Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen. *J Biol Chem* 278(32): 30365–30372. <https://doi.org/10.1074/jbc.M302117200> PMID: 12771148
81. Martoglio B, Dobberstein B. (1998) Signal sequences: More than just greasy peptides. *Trends Cell Biol* 8(10): 410–415. [https://doi.org/10.1016/s0962-8924\(98\)01360-9](https://doi.org/10.1016/s0962-8924(98)01360-9) PMID: 9789330
82. Rehm A, Stern P, Ploegh HL, Tortorella D. (2001) Signal peptide cleavage of a type I membrane protein, HCMV US11, is dependent on its membrane anchor. *Embo j* 20(7): 1573–1582. <https://doi.org/10.1093/emboj/20.7.1573> PMID: 11285222
83. Land A, Zonneveld D, Braakman I. (2003) Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *Faseb j* 17(9): 1058–1067. <https://doi.org/10.1096/fj.02-0811com> PMID: 12773488
84. Yoltz J, Schwing C, Chang J, Van Ryk D, Nawaz F, et al. (2018) Signal peptide of HIV envelope protein impacts glycosylation and antigenicity of gp120. *Proc Natl Acad Sci U S A* 115(10): 2443–2448. <https://doi.org/10.1073/pnas.1722627115> PMID: 29463753
85. Upadhyay C, Feyznehzhad R, Yang W, Zhang H, Zolla-Pazner S, et al. (2018) Alterations of HIV-1 envelope phenotype and antibody-mediated neutralization by signal peptide mutations. *PLoS Pathog* 14(1): e1006812. <https://doi.org/10.1371/journal.ppat.1006812> PMID: 29370305
86. Davison AJ, Dolan A, Akter P, Addison C, Dargan DJ, et al. (2003) The human cytomegalovirus genome revisited: Comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* 84(Pt 1): 17–28. <https://doi.org/10.1099/vir.0.18606-0> PMID: 12533697
87. Hansen SG, Strelow LI, Franchi DC, Anders DG, Wong SW. (2003) Complete sequence and genomic analysis of rhesus cytomegalovirus. *J Virol* 77(12): 6620–6636. <https://doi.org/10.1128/JVI.77.12.6620-6636.2003> PMID: 12767982
88. Rivaiier P, Kaur A, Johnson RP, Wang F. (2006) Genomic sequence of rhesus cytomegalovirus 180.92: Insights into the coding potential of rhesus cytomegalovirus. *J Virol* 80(8): 4179–4182. <https://doi.org/10.1128/JVI.80.8.4179-4182.2006> PMID: 16571834
89. Blewett EL, Sherrod CJ, Texier JR, Conrad TM, Dittmer DP. (2015) Complete genome sequences of mandrillus leucophaeus and papio ursinus cytomegaloviruses. *Genome Announc* 3(4): <https://doi.org/10.1128/genomeA.00781-15> PMID: 26251484
90. Katoh K, Standley DM. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4): 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690

91. Sela I, Ashkenazy H, Katoh K, Pupko T. (2015) GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res* 43(W1): W7–14. <https://doi.org/10.1093/nar/gkv318> PMID: 25883146
92. Privman E, Penn O, Pupko T. (2012) Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol* 29(1): 1–5. <https://doi.org/10.1093/molbev/msr177> PMID: 21772063
93. Darling AC, Mau B, Blattner FR, Perna NT. (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7): 1394–1403. <https://doi.org/10.1101/gr.2289704> PMID: 15231754
94. Darling AE, Mau B, Perna NT. (2010) progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6): e11147. <https://doi.org/10.1371/journal.pone.0011147> PMID: 20593022
95. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. (2007) Integration of biological networks and gene expression data using cytoscape. *Nat Protoc* 2(10): 2366–2382. <https://doi.org/10.1038/nprot.2007.324> PMID: 17947979
96. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue): D222–6. <https://doi.org/10.1093/nar/gku1221> PMID: 25414356
97. Remmert M, Biegert A, Hauser A, Soding J. (2011) HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9(2): 173–175. <https://doi.org/10.1038/nmeth.1818> PMID: 22198341
98. Stamatakis A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
99. Kosakovsky Pond SL, Frost SD. (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22(5): 1208–1222. <https://doi.org/10.1093/molbev/msi105> PMID: 15703242
100. Guindon S, Delsuc F, Dufayard JF, Gascuel O. (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol* 537: 113–137. https://doi.org/10.1007/978-1-59745-251-9_6 PMID: 19378142
101. Pohlert T. (2014) The pairwise multiple comparison of mean ranks package (PMCMR). R Package: 2004–2006.
102. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23(10): 1891–1901. <https://doi.org/10.1093/molbev/msl051> PMID: 16818476
103. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8(7): e1002764. <https://doi.org/10.1371/journal.pgen.1002764> PMID: 22807683
104. Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, et al. (2012) FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res* 40(Web Server issue): W580–4. <https://doi.org/10.1093/nar/gks498> PMID: 22661579
105. Gariano GR, Dell'Oste V, Bronzini M, Gatti D, Lukanini A, et al. (2012) The intracellular DNA sensor IFI16 gene acts as restriction factor for human cytomegalovirus replication. *PLoS Pathog* 8(1): e1002498. <https://doi.org/10.1371/journal.ppat.1002498> PMID: 22291595
106. Biolatti M, Dell'Oste V, Pautasso S, von Einem J, Marschall M, et al. (2016) Regulatory interaction between the cellular restriction factor IFI16 and viral pp65 (pUL83) modulates viral gene expression and IFI16 protein stability. *J Virol* 90(18): 8238–8250. <https://doi.org/10.1128/JVI.00923-16> PMID: 27384655
107. Tischer BK, von Einem J, Kaufer B, Osterrieder N. (2006) Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. *Biotechniques* 40(2): 191–197. <https://doi.org/10.2144/000112096> PMID: 16526409
108. Klucher KM, Sommer M, Kadonaga JT, Spector DH. (1993) In vivo and in vitro analysis of transcriptional activation mediated by the human cytomegalovirus major immediate-early proteins. *Mol Cell Biol* 13(2): 1238–1250. <https://doi.org/10.1128/mcb.13.2.1238> PMID: 8423789
109. Tischer BK, Smith GA, Osterrieder N. (2010) En passant mutagenesis: A two step markerless red recombination system. *Methods Mol Biol* 634: 421–430. https://doi.org/10.1007/978-1-60761-652-8_30 PMID: 20677001

110. Sinzger C, Hahn G, Digel M, Katona R, Sampaio KL, et al. (2008) Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J Gen Virol* 89 (Pt 2): 359–368. <https://doi.org/10.1099/vir.0.83286-0> PMID: 18198366
111. Bronzini M, Luginini A, Dell'Oste V, De Andrea M, Landolfo S, et al. (2012) The US16 gene of human cytomegalovirus is required for efficient viral infection of endothelial and epithelial cells. *J Virol* 86(12): 6875–6888. <https://doi.org/10.1128/JVI.06310-11> PMID: 22496217
112. Li H, Durbin R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26(5): 589–595. <https://doi.org/10.1093/bioinformatics/btp698> PMID: 20080505
113. Li H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21): 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
114. Emery VC, Cope AV, Bowen EF, Gor D, Griffiths PD. (1999) The dynamics of human cytomegalovirus replication in vivo. *J Exp Med* 190(2): 177–182. <https://doi.org/10.1084/jem.190.2.177> PMID: 10432281
115. Hartl DL, Clark AG. (1997) Principles of population genetics. Sinauer associates Sunderland, MA.