Published in final edited form as:

Infect Genet Evol. 2017 July ; 51: 46–53. doi:10.1016/j.meegid.2017.03.013.

Evolution of the ability to modulate host chemokine networks via gene duplication in Human Cytomegalovirus (HCMV)

Jessica A. Scarborough, John R. Paul, and Juliet V. Spencer[#]

Department of Biology, University of San Francisco, Harney Science Center, 2130 Fulton Street, San Francisco, CA 94117 USA

Abstract

Human cytomegalovirus (HCMV) is a widespread pathogen that is particularly skillful at evading immune detection and defense mechanisms, largely due to extensive co-evolution with its host. One aspect of this co-evolution involves the acquisition of virally encoded G protein-coupled receptors (GPCRs) with homology to the chemokine receptor family. GPCRs are the largest family of cell surface proteins, found in organisms from yeast to humans, and they regulate a variety of cellular processes including development, sensory perception, and immune cell trafficking. The US27 and US28 genes are encoded by human and primate CMVs, but homologs are not found in the genomes of viruses infecting rodents or other species. Phylogenetic analysis was used to investigate the US27 and US28 genes, which are adjacent in the unique short (US) region of the HCMV genome, and their relationship to one another and to human chemokine receptor genes. The results indicate that both US27 and US28 share the same common ancestor with human chemokine receptor CX3CR1, suggesting that a single host gene was captured and a subsequent viral gene duplication event occurred. The US28 gene product (pUS28) has maintained the function of the ancestral gene and has the ability to bind and signal in response to CX3CL1/fractalkine, the natural ligand for CX3CR1. In contrast, pUS27 does not bind to any known chemokine ligand, and the sequence has diverged significantly, highlighted by the fact that pUS27 currently exhibits greater sequence similarity to human CCR1. While the evolutionary advantage of the gene duplication and neofunctionalization event remains unclear, the US27 and US28 genes are highly conserved among different HCMV strains and retained even in laboratory strains that have lost many virulence genes, suggesting that US27 and US28 have each evolved distinct, important functions during virus infection.

1. Introduction

1.1 The Herpesvirus Family

Human cytomegalovirus (HCMV) is a member of the *Herpesviridae* family. These viruses have linear DNA genomes and share two main characteristics, the structure of the virion

[#]Corresponding Author: Juliet V. Spencer, jspencer@usfca.edu, 1-415-422-5470.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

particle and the ability to establish lifelong latent infection of the host. Herpesviruses are highly species-specific, and more than 100 different viruses that infect mammals, birds, reptiles, amphibians, fish, and invertebrates have been identified to date (McGeoch et al., 2008). Some species are infected by multiple, distinct herpesviruses, and there are nine herpesviruses that infect humans: herpes simplex virus-1 (HSV-1), herpes simplex virus-2 (HSV-2), varicella-zoster virus (VZV), Epstein-Barr virus (EBV), human cytomegalovirus (HCMV), human herpesvirus 6A (HHV6A), human herpesvirus 6B (HHV6B), human herpesvirus 7 (HHV7), and human herpesvirus 8 (HHV8, or KSHV, Kaposi's sarcoma associated herpesvirus). Even before the evolution of *Homo sapiens*, an ancient HSV-2 jumped from chimpanzees to *Homo erectus* about 1.6 million years ago (mya), and the ancestral virus that gave rise to HSV-1 infected hominids more than 6 mya (Wertheim et al., 2014). Herpesviruses co-evolve extensively with their specific hosts, frequently through gene capture (Alcendor et al., 2009; McGeoch et al., 2006; Wang et al., 2007). The acquisition of new genes via lateral transfer not only has the potential to increase virus fitness (McGeoch et al., 2006; Wang et al., 2007) but also reduces immune recognition of the pathogen through molecular mimicry (Hughes and Friedman, 2005). Here, we investigated the relationship between two genes that encode G protein-coupled receptors (GPCRs) with immune modulating functions in the genome of HCMV.

1.2 Human Cytomegalovirus (HCMV)

HCMV is a member of the *Betaherpesvirinae* sub-family and is widespread throughout the general population (Cannon et al., 2010). HCMV is adept at evading the host immune system, largely due to having acquired many genes with homology to genes from its human host (McSharry et al., 2012a). HCMV rarely causes disease in a fit host, but more frequently, acts as an opportunistic pathogen. In healthy individuals, HCMV infection produces a strong immune response involving both the innate and adaptive systems. Despite this immune response, even immunocompetent hosts are unable to eliminate the virus or prevent the establishment of latent infection (McSharry et al., 2012b). HCMV's ability to modulate host immune responses through cytokine, chemokine, and chemokine receptor homologs plays a role in this lifelong persistence (McSharry et al., 2012b).

Although latent HCMV can reactivate periodically, immune competent hosts typically do not experience clinical symptoms due to strong T-cell responses (McSharry et al., 2012b; Rosa and Diamond, 2012). Immunocompromised patients, however, are much more likely to experience HCMV disease during primary infection or reactivation of latent virus (Wreghitt et al., 2003). HCMV infection is extremely serious in transplant recipients and one of the leading causes of morbidity and mortality in this highly vulnerable population (Ramanan and Razonable, 2013). Additionally, congenital infection can lead to severe birth defects, including hearing loss, mental disability, microcephaly, seizures, and other neuromuscular defects (Manicklal et al., 2013).

Recently, HCMV has been connected to increased malignancy of certain types of cancer. While HCMV is not considered an oncogenic virus, infection can contribute to increased cellular proliferation, inhibition of apoptosis, enhanced chemotaxis, and other characteristics associated with malignancy (Michaelis et al., 2009). The strongest evidence for a role of

HCMV in tumor development has been seen in glioblastomas, where the virus is capable of promoting the malignant phenotype, making the HCMV a potential target in the treatment of malignant gliomas (Dziurzynski et al., 2012).

1.3 G-Protein-Coupled Receptors (GPCRs)

HCMV's ability to establish lifelong latent infection within its host may be in part due to GPCR homologs encoded by the viral genome (Vieira et al., 1998). GPCRs are cell surface receptors, which transmit an intracellular signal upon the binding of an extracellular ligand. Humans have over 800 unique GPCRs, with a wide assortment of functional roles, yet each is similar in physical structure (Chattopadhyay, 2014). These receptors consist of a single polypeptide with seven transmembrane segments, with an N-terminus on the exterior of the cell, and the C-terminus inside the cell, typically associated with heterotrimeric G proteins or other signaling molecules (Fig. 1). Once a ligand binds and activates a GPCR, a conformational change occurs, triggering the activation of a local G-protein and leading to the dispatch of signals within the cell.

HCMV encodes at least four genes that give rise to proteins with predicted seven transmembrane domains: US27, US28, UL33, and UL78 (Chee et al., 1990). In particular, the US27 and US28 gene products (pUS27 and pUS28, respectively) exhibit homology to human chemokine receptors, a subset of the GPCR superfamily that play a crucial role in immune responses. Each chemokine receptor binds to specific chemokines, which are small secreted proteins, and regulates intracellular signaling and immune responses (Murdoch and Finn, 2000). Responses to ligand binding and G protein activation include pro-inflammatory or anti-inflammatory effects, enhanced chemotaxis, and superoxide anion production (Charo and Ransohoff, 2006). Proper functioning of these chemokine receptors is key for mediating damage from an infection or injury. Moreover, dysregulation of chemokine receptors, chemokines, and chemokine attractants is central to many inflammatory and infectious diseases, including psoriasis and allergies (Raman et al., 2011).

In addition to the seven transmembrane domains, the key hallmark features of a chemokine receptor homolog generally include: N-linked glycosylation and multiple negatively charged amino acids in the N-terminus, the capability to form a disulfide bond between the N-terminus and third extracellular loop, multiple positively charged amino acids in the third intracellular loop, conserved amino acids in the transmembrane regions, and many serine and threonine residues in the C-terminus (Beisser et al., 2008). HCMV contains several additional genes that appear to code for seven transmembrane proteins (Das and Pellett, 2007; Lesniewski et al., 2006); however, only pUS27, pUS28, and pUL33 meet the criteria for chemokine receptors, with pUS27 and pUS28 showing the greatest similarity to this receptor family.

1.4 HCMV US27 and US28

pUS27 is a putative chemokine receptor, although it has no known ligands to date (Arnolds et al., 2013). This protein can be found in intracellular vesicles and in the cell membrane of infected cells, and even within the envelope of the virus particle (McSharry et al., 2012b). US27 increases extracellular spreading of HCMV by 10-fold in fibroblasts and endothelial

cells in comparison to a US27 knockout strain (O'Connor and Skenk, 2011). pUS27 may also be responsible for enhancing chemotaxis of virus-infected cells by increasing signaling through CXCR4, a human chemokine receptor (Arnolds et al., 2013). Additionally, pUS27 has been shown to increase cellular proliferation in human embryonic kidney cells (HEK293) by suppressing host negative growth regulators, although the complete underlying mechanism for this effect remains unknown (Lares et al., 2013; Tu and Spencer, 2014).

In contrast to pUS27, pUS28 is a functional chemokine receptor that is known to bind and signal in response to an extensive variety of chemokine ligands including CX₃CL1/fractalkine, CCL2/MCP-1, CCL5/RANTES, and CCL7/MCP-3 (Kuhn et al., 1995). pUS28 appears to play a key role in HCMV infection efficacy and potency, along with increased cell-to-cell spread of the virus (Noriega et al., 2014; Spiess et al., 2015). pUS28 has immune evasion properties due to its ability to internalize host chemokines, thereby decreasing the host immune system's ability to sense and respond to those chemokine signals (Beisser et al., 2002; McSharry et al., 2012b). In addition, US28 is one of a select set of genes expressed during latency, suggesting that this protein may play a role in sensing the host environment (Beisser et al., 2001; Humby and O'Connor, 2015; Mason et al., 2013).

This study examines the evolutionary history of US27 and US28, which are hypothesized to have arisen via gene duplication. Here, we performed a phylogenetic analysis confirming that US27 and US28 are paralogs that arose via a gene duplication event that occurred in an ancestral viral genome. We found that US27 and US28 likely share common ancestry with the human gene CX3CR1.

2. Materials and Methods

2.1 Sequences

The pUS27 and pUS28 sequences from six different HCMV strains were analyzed: AD169, Towne, Toledo, Merlin, TR, and TB40/E. Genbank Accession numbers for the protein sequences examined in this study are as follows: US27 AD169 (ABG73076.1), US27 Towne (ABG73086.1), US27 Toledo (AAS49024.1), US27 Merlin (YP_081611.1), US27 TR (AGL96752.1), US27 TB40/E (ABV71517.1), US28 AD169 (ACL51230.1), US28 Towne (ACM48113.1), US28 Toledo (AAS49025.1), US28 Merlin (YP_081612.1), US28 TR (AGL96753.1), and US28 TB40/E (ABV71518.1).

The pUS27 and pUS28 amino acid sequences were independently searched through GenBank using the NCBI's protein Basic Local Alignment Search Tool (BLAST). This search compared the HCMV protein sequences to the entire human genome, displaying the top 100 matches by E-value (Madden, 2002). All unique chemokine receptors from this list of the top 100 matches were compiled and used for further analysis. This BLAST tool was also utilized to provide E-values between US27, US28 and their closest human chemokine receptor matches. Genbank Accession numbers for the human chemokine receptors are: CCR1 (NP_001286.1), CCR2 (AGC02843.1), CCR3 (NP_001828.1), CCR4 (NP_005499.1), CCR5 (NP_000570.1), CCR6 (NP_004358.2), CCR8 (NP_005192.1), CCR9 (AAH95516.1), CX3CR1 (NP_001328.1), CXCR3 (NP_001495.1), and CXCR4

(CAA12166.1). The outgroup used in phylogenetic analysis is a human G-protein coupled receptor that is not part of the chemokine receptor family, the Oxoeicosanoid Receptor (12NP_062545.1).

2.2 Sequence alignments

Protein alignments between HCMV strains were performed using the MUSCLE multiple sequence alignment tool, version 3.8 (Edgar, 2004), and alignment matrices were created by Clustal2.1 (Goujon et al., 2010; Larkin et al., 2007) within Geneious v. 9.1, a bioinformatics phylogenetics software, with the default settings used (Kearse et al., 2012). The alignments were then further edited by hand to replace blanks at the beginning and end of the alignment with question marks to represent missing data.

2.3 Phylogenetic Analysis

Using Geneious v. 9.1, phylogenetic trees using Bayesian and maximum likelihood method were inferred to analyze the evolutionary history of the pUS27 and pUS28 proteins in relation to selected human chemokine receptors. The maximum likelihood analysis was inferred using PhyML v. 3.0 plugin (Guindon et al., 2010; Guindon and Gascuel, 2003) on the Geneious v. 9.1 software. Node values display bootstrap values (with 100 replicates) while all other settings utilized the default Geneious specifications. The Bayesian phylogenetic analysis was run utilizing the MrBayes plugin on Geneious software. A chain length of 10,100,000, subsampling frequency of 2,000, and burn-in length of 1,100,00 were used, with all other parameters set to default Geneious specifications (Huelsenbeck and Ronquist, 2001). Node values represent posterior probabilities of the clades.

3. Results

3.1 Protein Alignments

Because there are multiple HCMV strains that have been sequenced and are used for *in vitro* experiments, variability in the pUS27 and pUS28 sequences among six different strains was first examined. These strains were chosen because they are prominent laboratory strains (AD169, Towne) and clinical isolates (TB40/E, Toledo, TR, Merlin) (Hosogai et al., 2015; Wilkinson et al., 2015). Sequence alignments were performed and percent identity among strains evaluated (Table 1, 2). The results indicate that the six HCMV strains examined in this study show minimal variation within the pUS27 and pUS28 primary protein sequence. For pUS28, the lowest percent identity was between the strains TB40/E and TR (both clinical strains) with 97.46% identity match (Table 1). The closest identity was seen between AD169 and Towne, both laboratory strains that share 100% identity match within the pUS28 protein (Table 1). For pUS27, there was slightly more variation between strains (Table 2). The greatest variation was between TB40/E and Merlin, both clinical strains, with 94.74% identity match within the pUS27 protein. The closest similarity was between Towne and Toledo with a 99.45% identity match within the pUS27 protein. Thus, both genes are highly conserved among different virus strains, indicating their functions are likely beneficial for the virus.

Since the variability among strains was minimal, gross alignments used sequences from HCMV strain AD169, the first HCMV strain to be sequenced and annotated, which has been extensively used during laboratory research (Wilkinson et al., 2015). A BLAST search for the most similar human chemokine receptors to pUS27 and pUS28 was performed, and alignments between complete primary protein sequences were evaluated by Expect (E) value. The closer the E-value is to zero, the more significant the match. The E-values for pUS27 and pUS28 alignments were quite low, and therefore, the results are represented as reciprocal E-value for more clear visualization of closest alignment matches. As shown in Fig 2A, pUS27 shares the highest similarity with CCR1, although the difference between E-values for alignments of pUS27 with CCR1 and pUS27 with CCR4 (the second most closely matched to pUS27) are not significantly different. On the other hand, pUS28 shows a very strong sequence similarity to CX3CR1 in particular (Fig 2B). The difference in alignments between pUS28 and CX3CR1 and CCR4 (the second most closely matched to both pUS27 and pUS28) is very substantial. Alignment of pUS27 and pUS28 results in an E-value of $1.0E^{-49}$ (Fig 3). These results demonstrate that pUS27 is more similar to pUS28 than to any human chemokine receptor; however, pUS28 is more similar to CX3CR1 than to pUS27.

3.2 Phylogenetic Analysis

Phylogenetic analysis was used to investigate the evolutionary history of the US27 and US28 genes. A common ancestor would indicate a shared evolutionary history with the two genes evolving via a gene duplication event. If US27 and US28 had separate ancestral genes, then it is more likely that the two genes evolved via two separate gene captures between HCMV and the host DNA. The inferred maximum likelihood phylogenetic tree supports the hypothesis that US28 and US27 share a common ancestor (Fig 4). Additionally, the common ancestor of these two viral genes shares a common ancestor with human CX3CR1. The bootstrap support values show strong support for a common ancestor between pUS27 and pUS28. There is very weak support, however, for pCX3CR1 being the closest human protein to pUS27 and pUS28's common ancestor. Bayesian phylogenetic analysis is concordant with the maximum likelihood phylogeny, showing a common ancestral gene between US28 and US27 with that viral ancestral gene sharing a common ancestral gene with human CX3CR1 (Fig 5). The posterior probabilities are highly supportive of both US27 and US28's common ancestor and this common ancestor's relationship to CX3CR1. These results suggest that a gene capture event led to the common ancestor of US28 and US27 within the viral genome. Then, within the viral DNA, there was a subsequent gene duplication event that ultimately led to US28 and US27 evolving from one common ancestral viral gene. Both Bayesian and maximum likelihood phylogenetic analysis support the notion that US27 and US28 most likely arose from a gene duplication event, and they likely share a common ancestor with human CX3CR1.

4. Discussion

Although the evolutionary history of viruses can be difficult to determine due to significant variation between strains and high rates of mutation, the evidence presented here indicates that gene duplication led to the presence of the US27 and US28 genes in the HCMV genome. Beisser et al. previously showed that US27 and US28, along with UL78, cluster

with CX3CR1 in a phylogenetic tree, but the study did not evaluate the evolutionary history of the viral genes (Beisser et al., 2002). Our analyses suggest it is likely that an ancestral cytomegalovirus captured a gene from its host, then integrated this gene into its own genome. This newly integrated gene, which shares ancestry with human CX3CR1, would have then undergone a gene duplication event within the viral genome, which led to the creation US27 and US28, paralogs in the HCMV genome. The fate of duplicated genes varies, and gene loss, subfunctionalization, and neofunctionalization are some possible outcomes.

Neofunctionalization occurs when one gene from a gene duplication is constrained by selective pressure, while the other gene is not (Gibson and Goldberg, 2009). In the neofunctionalization scenario, one protein has a low amino acid substitution rate due to purifying selection pressure, causing it to remain more similar to its parent gene. The second gene may take on a new function because of the lack of selective pressure enables evolution of new functionalities that are beneficial to the virus (Zhang, 2003). On the other hand, subfunctionalization occurs when there is no evolutionary advantage to the extra gene copy (Zhang, 2003). With a lack of selective pressure, both genes can mutate, while keeping complementary functions of the original genes. This can be an important form of evolution because it allows enhanced evolution of specific aspects of a gene without trading off fitness for other aspects of a gene.

Because pUS28 exhibits far higher similarity to the human CX3CR1 protein than pUS27, it is likely that US27 underwent neofunctionalization after divergence. This model would propose that after gene duplication occurred, the virus could keep the functionality of pUS28, while pUS27 was less constrained and eventually gained a new function. It is notable that pUS27 does not bind any known human chemokines (Arnolds et al., 2013), but is present in the viral envelope (Margulies and Gibson, 2007) and has the ability to facilitate extracellular virus spread, as demonstrated by viral mutants lacking US27 that display reduced virus spreading in culture (O'Connor and Shenk, 2011). In addition, pUS27 also modulates host chemokine networks by enhancing the signaling activity of human CXCR4 (Arnolds et al., 2013), suggesting that CXCR4 may activate cellular pathways that are beneficial to HCMV. However, other possible scenarios, such as subfunctionalization, cannot be ruled out. This model would predict that the ancestral host gene had a range of functions, and after the gene duplication event, US27 and US28 each retained a different subset of the original gene's functions. This would give both US27 and US28 important but non-redundant functions, and ensure that both genes were retained in the genome. While US27 has not been evaluated in patient samples, only minimal variation in US28 sequences was found in samples from HIV-positive patients (Goffard et al., 2006) or from patients with congenital HCMV infection (Arav-Boger et al., 2002). We found that both proteins were highly conserved among a variety of laboratory and clinical strains of HCMV (Table 1, 2), which could support the subfunctionalization model.

Other possible outcomes for duplicated genes include gene loss or regulatory differentiation (Louis, 2007). Since regulatory sequences are not transcribed into mRNA, our analyses of amino acid sequences are not sufficient for examining the differences between US27 and US28 gene regulatory patterns (Zhang, 2003). Gene loss, on the other hand, is an intriguing

possibility. The genomes of green monkey cytomegalovirus (GMCMV) and Rhesus cytomegalovirus (RhCMV) each harbor five adjacent genes with homology to HCMV US28 (Alcendor et al., 2009; Hansen et al., 2003; Martin, 2000). It is conceivable that the ancestor of HCMV also had five genes and subsequently lost all but US27 and US28, which together retain all the most desirable functions of the ancestral gene. Chimpanzee cytomegalovirus also has only two GCPR genes, considered orthologs of HCMV US27 and US28 (Davison et al., 2003). Penfold et al. found that only one of the RhCMV US28 homologs genes had a ligand binding profile similar to HCMV US28 (Penfold et al., 2003), but little is known about the functions of the other genes in this cluster. Further research is necessary to understand the evolutionary advantage of the five tandem copies in GMCCMV and RhCMV, which could potentially shed light on the functions of US27 during HCMV infection.

Our analyses and published experimental evidence (Arnolds et al., 2013; Margulies and Gibson, 2007; O'Connor and Shenk, 2011) support the notion that pUS27 likely underwent neofunctionalization after a gene duplication event while US28 had greater selective pressure to retain the original gene function, keeping the amino acid substitution rate lower. Intriguingly, HCMV also encodes two adjacent chemokine genes, UL146 and UL147, which are found in clinical strains but not in laboratory strains AD169 or Towne (Cha et al., 1996; Prichard et al., 2001). UL146 is a functional chemokine that induces chemotaxis of neutrophils expressing CXCR2, but UL147 does not appear to have chemotactic ability (Penfold et al., 1999). The two genes are believed to be the result of a relatively recent gene duplication event, and evidence indicates that UL146 is the more rapidly evolving paralog (Arav-Boger et al., 2005). The UL146 gene is among the most highly variable in the HCMV genome among sequenced clinical isolates (Arav-Boger et al., 2006; Bradley et al., 2008), but no specific polymorphisms have been linked to disease outcome or severity (Arav-Boger et al., 2006; Heo et al., 2008). The loss of these genes from laboratory strains suggests they are not required for virus replication in culture but may be important during natural infection of the host, although specific roles in virulence or virus dissemination have been difficult to confirm experimentally. Like US27 and US28, the UL146 and UL147 genes are absent from the genomes of rodent CMVs commonly used for *in vivo* infection models. Additional work is needed to fully understand the mechanisms by which HCMV controls chemokine networks and cell migration patterns to promote virus persistence.

Both maximum likelihood and Bayesian phylogenetic analyses indicate a common ancestor between US27 and US28, suggesting that a gene duplication event within the viral genome led to their existence after capture of a host gene, and that both genes share a common ancestor with human CX3CR1. The relationships inferred for the human chemokine receptors by these analyses also make sense based on current knowledge. For example, the clades containing CCR1, CCR2, CCR3, CCR4, CCR5, and CCR8 (Fig 4, 5) are consistent with the observation that these genes are all clustered together on human Chromosome 3, suggesting that gene duplications have played a role in the development of multiple, diverse human chemokine receptors in the human genome. The fact that these proteins formed clades in both the maximum likelihood and Bayesian phylogenies corroborates that these methods are successful at grouping together proteins that share evolutionary history.

The CX3CR1 gene, also located on Chromosome 3, codes for a chemokine receptor that binds CX3C Ligand 1 (CX3CL1), also known as fractalkine. Activation of CX3CR1 regulates cell adhesion and directed movement (Imai et al., 1997), and the binding of CX3CL1/fractalkine leads to the attraction of intraepithelial lymphocytes (Isse et al., 2005). This receptor is largely associated with expression on lymphocytes, but CX3CR1 expression has also been observed in microglial cells and macrophages, where interactions with CX3CL1/fractalkine facilitate communications with neurons (Limotala and Ransohoff, 2014). Additionally, CX3CL1/fractalkine contributes to immune function by stimulating the migration of leukocytes during both physiological and pathological conditions. This ligand can exist as either a membrane bound or isolated soluble protein (Haskell et al., 1999). When it is in a soluble state, CX3CL1/fractalkine serves an important role as chemoattractant for the CD8+ and CD4+ T cells, natural killer cells, and monocytes that express the CX3CR1 receptor. In a membrane-bound state, CX3CL1/fractalkine and CX3CR1 form strong binding interactions (Haskell et al., 1999).

The human gene CX3CR1 contains 4 exons and 3 introns (Garin et al., 2002), yet neither the US27 and US28 genes nor the GMCMV, RhCMV, and CCMV orthologs contain any introns (Rawlingson and Barrell, 1993). The exact mechanism of gene capture has not been determined; however, since US27 and US28 lack introns, one possibility is that an ancestral cytomegalovirus acquired the host gene ancestor to US27 and US28 through some retroviral intermediate, perhaps during a mixed infection (McGeoch et al., 2008). The US27 transcript was reported to be spliced within the 5' untranslated region (Beisser et al., 2002), but the significance of this observation and any effects on US27 gene expression remain unknown.

CX3CR1 can serve as a co-receptor for entry of human immunodeficiency virus (HIV), and specific polymorphisms of CX3CR1 have been linked to an increase in susceptibility to HIV infection and more rapid progression to acquired immunodeficiency syndrome, or AIDS (Parczewski et al., 2014). There are also increased levels of CX3CR1 expression within HIV-positive monocytes (Krishnan et al., 2014). Moreover, active HCMV infection in HIV-infected patients has been connected with perpetually low CD4+ T cell counts (Munawwar and Singh, 2016). It's possible that one reason HCMV co-infection is an increased risk factor for the progression of HIV to AIDS (Munawwar and Singh, 2016; Pleskoff et al., 1997b) is that pUS28 has similar binding properties to CX3CR1 and facilitates HIV infection. In fact, pUS28 has been shown to be able to function as a co-receptor to mediate HIV entry (Pleskoff et al., 1997a). Additional investigation is needed to better understand the role of HCMV in the progression of HIV-infection to AIDS and whether cooperation between pathogens may have played a role in the acquisition of the ancestral host gene by ancient CMVs.

Manipulation of host immune responses is a critical factor for the establishment of persistent HCMV infection. pUS28 is a bona fide viral chemokine receptor that contributes to immune evasion in multiple ways, while the binding and signaling mechanisms of pUS27 remain unclear. However, there is no doubt that the US27 and US28 genes arose from a common ancestor and their protein products have developed distinct roles in the infection, spread, and persistence of HCMV.

Acknowledgments

The authors thank Dr. James Sikes for helpful discussion and critical reading of the manuscript and four anonymous reviewers for their insightful comments. This work was supported by NIH grant AI11232 (to JVS) and USF Faculty Development Funds.

References

- Alcendor DJ, Zong J, Dolan A, Gatherer D, Davison AJ, Hayward GS. Patterns of divergence in the vCXCL and vGPCR gene clusters in primate cytomegalovirus genomes. *Virology*. 2009; 395:21–32. [PubMed: 19818982]
- Arav-Boger R, Foster CB, Zong JC, Pass RF. Human cytomegalovirus-encoded alpha -chemokines exhibit high sequence variability in congenitally infected newborns. *J Infect Dis*. 2006; 193:788–791. [PubMed: 16479512]
- Arav-Boger R, Willoughby RE, Pass RF, Zong JC, Jang WJ, Alcendor D, Hayward GS. Polymorphisms of the cytomegalovirus (CMV)-encoded tumor necrosis factor-alpha and beta-chemokine receptors in congenital CMV disease. *J Infect Dis*. 2002; 186:1057–1064. [PubMed: 12355354]
- Arav-Boger R, Zong JC, Foster CB. Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes*. 2005; 31:65–72. [PubMed: 15965610]
- Arnolds KL, Lares AP, Spencer JV. The US27 gene product of human cytomegalovirus enhances signaling of host chemokine receptor CXCR4. *Virology*. 2013; 439:122–131. [PubMed: 23490053]
- Beisser, PS., Goh, C-S., Cohen, FE., Michelson, S. *Viral Chemokine Receptors and Chemokines in Human Cytomegalovirus Trafficking and Interaction with the Immune System, Viral Proteins Counteracting Host Defenses*. Springer; Berlin Heidelberg; 2002. p. 203-234.
- Beisser PS, Laurent L, Virelizier JL, Michelson S. Human cytomegalovirus chemokine receptor gene US28 is transcribed in latently infected THP-1 monocytes. *J Virol*. 2001; 75:5949–5957. [PubMed: 11390596]
- Beisser, PS., Lavreysen, H., Bruggeman, CA., Vink, C. Springer Science + Business Media. 2008. Chemokines and Chemokine Receptors Encoded by Cytomegaloviruses, *Current Topics in Microbiology and Immunology*; p. 221-242.
- Bradley AJ, Kovacs IJ, Gatherer D, Dargan DJ, Alkharsah KR, Chan PK, Carman WF, Dedicoat M, Emery VC, Geddes CC, Gerna G, Ben-Ismael B, Kaye S, McGregor A, Moss PA, Puzsai R, Rawlinson WD, Scott GM, Wilkinson GW, Schulz TF, Davison AJ. Genotypic analysis of two hypervariable human cytomegalovirus genes. *J Med Virol*. 2008; 80:1615–1623. [PubMed: 18649324]
- Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol*. 2010; 20:202–213. [PubMed: 20564615]
- Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol*. 1996; 70:78–83. [PubMed: 8523595]
- Charo IF, Ransohoff RM. The many roles of chemokines and chemokine receptors in inflammation. *N Engl J Med*. 2006; 354:610–621. [PubMed: 16467548]
- Chattopadhyay A. GPCRs: Lipid-Dependent Membrane Receptors That Act as Drug Targets. *Advances in Biology*. 2014; 2014:1–12.
- Chee MS, Satchwell SC, Preddie E, Weston KM, Barrell BG. Human cytomegalovirus encodes three G protein-coupled receptor homologues. *Nature*. 1990; 344:774–777. [PubMed: 2158627]
- Das S, Pellett PE. Members of the HCMV US12 family of predicted heptaspanning membrane proteins have unique intracellular distributions, including association with the cytoplasmic virion assembly complex. *Virology*. 2007; 361:263–273. [PubMed: 17188320]
- Davison AJ, Dolan A, Akter P, Addison C, Dargan DJ, Alcendor DJ, McGeoch DJ, Hayward GS. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol*. 2003; 84:17–28. [PubMed: 12533697]

- Dziurzynski K, Chang SM, Heimberger AB, Kalejta RF, McGregor Dallas SR, Smit M, Soroceanu L, Cobbs CS. Consensus on the role of human cytomegalovirus in glioblastoma. *Neuro-Oncology*. 2012; 14:246–255. [PubMed: 22319219]
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004; 32:1792–1797. [PubMed: 15034147]
- Garin A, Pellet P, Deterre P, Debré P, Combadière C. Cloning and functional characterization of the human fractalkine receptor promoter regions. *Biochemical Journal*. 2002; 368:753–760. [PubMed: 12234253]
- Gibson TA, Goldberg DS. Questioning the Ubiquity of Neofunctionalization. *PLoS Computational Biology*. 2009;5.
- Goffard A, Gault E, Rozenberg F, Moret N, Hober D, Deny P. Comparative sequence analysis of US28 gene of human cytomegalovirus strains isolated from HIV-positive patients. *Virus Genes*. 2006; 33:175–181. [PubMed: 16972032]
- Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*. 2010; 38:W695–W699. [PubMed: 20439314]
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. 2010; 59:307–321. [PubMed: 20525638]
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52:696–704. [PubMed: 14530136]
- Hansen SG, Strelow LI, Franchi DC, Anders DG, Wong SW. Complete Sequence and Genomic Analysis of Rhesus Cytomegalovirus. *Journal of Virology*. 2003; 77:6620–6636. [PubMed: 12767982]
- Haskell CA, Cleary MD, Charo IF. Molecular uncoupling of fractalkine-mediated cell adhesion and signal transduction. Rapid flow arrest of CX3CR1-expressing cells is independent of G-protein activation. *Journal of Biological Chemistry*. 1999; 274:10053–10058. [PubMed: 10187784]
- Heo J, Petheram S, Demmler G, Murph JR, Adler SP, Bale J, Sparer TE. Polymorphisms within human cytomegalovirus chemokine (UL146/UL147) and cytokine receptor genes (UL144) are not predictive of sequelae in congenitally infected children. *Virology*. 2008; 378:86–96. [PubMed: 18556037]
- Hosogai M, Shima N, Nakatani Y, Inoue T, Iso T, Yokoo H, Yorifuji H, Akiyama H, Kishi S, Isomura H. Analysis of human cytomegalovirus replication in primary cultured human corneal endothelial cells. *British Journal of Ophthalmology*. 2015; 99:1583–1590. [PubMed: 26261231]
- Huelsenbeck J, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17:754–755. [PubMed: 11524383]
- Hughes AL, Friedman R. Poxvirus genome evolution by gene gain and loss. *Molecular Phylogenetics and Evolution*. 2005; 35:186–195. [PubMed: 15737590]
- Humby MS, O'Connor CM. Human Cytomegalovirus US28 Is Important for Latent Infection of Hematopoietic Progenitor Cells. *J Virol*. 2015; 90:2959–2970. [PubMed: 26719258]
- Imai T, Hieshima K, Haskell C, Baba M, Magira M, Nishimura M, Kakizaki M, Takagi S, Nomiyama H, Schall TJ, Yoshie O. Identification and molecular characterization of fractalkine receptor CX3CR1, which mediates both leukocyte migration and adhesion. *Archivum Immunologiae et Therapia Experimentalis*. 1997; 91:521–530.
- Isse K, Harada K, Zen Y, Kamihira T, Shimoda S, Harada M, Nakanuma Y. Fractalkine and CX3CR1 Are Involved in the Recruitment of Intraepithelial Lymphocytes of Intrahepatic Bile Ducts. *Hepatology*. 2005; 41:506–516. [PubMed: 15726664]
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentijies P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28:1647–1649. [PubMed: 22543367]
- Krishnan S, Wilson EMP, Sheikh V, Rupert A, Mendoza D, Yang J, Lempicki R, Migueles SA, Sereti I. Evidence for Innate Immune System Activation in HIV Type 1–Infected Elite Controllers. *Journal of Infectious Diseases*. 2014; 209:931–939.

- Kuhn DE, Beall CJ, Kolattukudy PE. The Cytomegalovirus US28 Protein Binds Multiple CC Chemokines with High Affinity. *Biochemical and Biophysical Research Communications*. 1995; 211:325–330. [PubMed: 7540006]
- Lares AP, Tu CC, Spencer JV. The human cytomegalovirus US27 gene product enhances cell proliferation and alters cellular gene expression. *Virus Res*. 2013; 176:312–320. [PubMed: 23850869]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
- Lesniewski M, Das S, Skomorowska-Prokvolit Y, Wang FZ, Pellett PE. Primate cytomegalovirus US12 gene family: a distinct and diverse clade of seven-transmembrane proteins. *Virology*. 2006; 354:286–298. [PubMed: 16904149]
- Limotala C, Ransohoff RM. Modulating neurotoxicity through CX3CL1/CX3CR1 signaling. *Frontiers in Cellular Neuroscience*. 2014;8. [PubMed: 24478631]
- Louis EJ. Evolutionary genetics: Making the most of redundancy. *Nature*. 2007; 449:673–674. [PubMed: 17928851]
- Madden, T. The BLAST Sequence Analysis Tool. In: McEntyre, J., Ostell, J., editors. *The NCBI Handbook*. National Center for Biotechnology information; Bethesda, MD: 2002.
- Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The “silent” global burden of congenital cytomegalovirus. *Clin Microbiol Rev*. 2013; 26:86–102. [PubMed: 23297260]
- Margulies BJ, Gibson W. The chemokine receptor homologue encoded by US27 of human cytomegalovirus is heavily glycosylated and is present in infected human foreskin fibroblasts and enveloped virus particles. *Virus Res*. 2007; 123:57–71. [PubMed: 16963142]
- Martin WJ. Chemokine receptor-related genetic sequences in an african green monkey simian cytomegalovirus-derived stealth virus. *Exp Mol Pathol*. 2000; 69:10–16. [PubMed: 10891288]
- Mason GM, Jackson S, Okecha G, Poole E, Sissons JG, Sinclair J, Wills MR. Human cytomegalovirus latency-associated proteins elicit immune-suppressive IL-10 producing CD4(+) T cells. *PLoS Pathog*. 2013; 9:e1003635. [PubMed: 24130479]
- McGeoch, DJ., Davison, AJ., Dolan, A., Gatherer, D., Sevilla-Reyes, EE. Molecular Evolution of the Herpesvirales. In: Domingo, E.Parrish, CR., Holland, JJ., editors. *Origin and Evolution of Viruses*. Elsevier; 2008. p. 447-475.
- McGeoch DJ, Rixon FJ, Davison AJ. Topics in herpesvirus genomics and evolution. *Virus Research*. 2006; 117:90–104. [PubMed: 16490275]
- McSharry B, Avdic S, Slobedman B. Human Cytomegalovirus Encoded Homologs of Cytokines, Chemokines and their Receptors: Roles in Immunomodulation. *Viruses*. 2012a; 4:2448–2470. [PubMed: 23202490]
- McSharry BP, Avdic S, Slobedman B. Human Cytomegalovirus Encoded Homologs of Cytokines, Chemokines and their Receptors: Roles in Immunomodulation. *Viruses*. 2012b; 4:2448–2470. [PubMed: 23202490]
- Michaelis M, Doerr HW, Cinatl J. The Story of Human Cytomegalovirus and Cancer: Increasing Evidence and Open Questions. *Neoplasia*. 2009; 11:1–9. [PubMed: 19107226]
- Munawwar A, Singh S. Human Herpesviruses as Copathogens of HIV Infection, Their Role in HIV Transmission, and Disease Progression. *Journal of Laboratory Physicians*. 2016; 8:5–18. [PubMed: 27013807]
- Murdoch C, Finn A. Chemokine Receptors and Their Role in Vascular Biology. *J Vasc Res*. 2000; 37:1–7.
- Noriega V, Gardner T, Redmann V, Bongers G, Lira S, Tortorella D. Human Cytomegalovirus US28 Facilitates Cell-to-Cell Viral Dissemination. *Viruses*. 2014; 6:1202–1218. [PubMed: 24625810]
- O'Connor CM, Shenk T. Human cytomegalovirus pUS27 G protein-coupled receptor homologue is required for efficient spread by the extracellular route but not for direct cell-to-cell spread. *J Virol*. 2011; 85:3700–3707. [PubMed: 21307184]
- O'Connor CM, Skenk T. Human Cytomegalovirus pUS27 G Protein-Coupled Receptor Homologue Is Required for Efficient Spread by the Extracellular Route but Not for Direct Cell-to-Cell Spread. *Journal of Virology*. 2011; 85:3700–3707. [PubMed: 21307184]

- Parczewski M, Urbanska A, Maciejewska K, Clark J, Leszczyszyn-Pynka M. Association of chemokine receptor gene variants with HIV-1 genotype predicted tropism. *HIV Medicine*. 2014; 15:577–586. [PubMed: 24750723]
- Penfold ME, Dairaghi DJ, Duke GM, Saederup N, Mocarski ES, Kemble GW, Schall TJ. Cytomegalovirus encodes a potent alpha chemokine. *Proc Natl Acad Sci U S A*. 1999; 96:9839–9844. [PubMed: 10449781]
- Penfold ME, Schmidt TL, Dairaghi DJ, Barry PA, Schall TJ. Characterization of the rhesus cytomegalovirus US28 locus. *J Virol*. 2003; 77:10404–10413. [PubMed: 12970425]
- Pleskoff O, Treboute C, Brelot A, Heveker N, Seman M, Alizon M. Identification of a chemokine receptor encoded by human cytomegalovirus as a cofactor for HIV-1 entry. *Science*. 1997a; 276:1874–1878. [PubMed: 9188536]
- Pleskoff O, Trébouté C, Brelot A, Heveker N, Seman M, Alizon M. Identification of a Chemokine Receptor Encoded by Human Cytomegalovirus as a Cofactor for HIV-1 Entry. *Science*. 1997b; 276:1874–1879. [PubMed: 9188536]
- Prichard MN, Penfold ME, Duke GM, Spaete RR, Kemble GW. A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. *Rev Med Virol*. 2001; 11:191–200. [PubMed: 11376481]
- Raman D, Sobolik-Delmaire T, Richmond A. Chemokines in health and disease. *Exp Cell Res*. 2011; 317:575–589. [PubMed: 21223965]
- Ramanan P, Razonable RR. Cytomegalovirus infections in solid organ transplantation: a review. *Infect Chemother*. 2013; 45:260–271. [PubMed: 24396627]
- Rawlingson WD, Barrell BG. Spliced Transcripts of Human Cytomegalovirus. *Journal of Virology*. 1993; 67:5502–5513. [PubMed: 7688825]
- Rosa CL, Diamond DJ. The immune response to human CMV. *Future Virology*. 2012; 7:279–293. [PubMed: 23308079]
- Spiess K, Jeppesen MG, Malmgaard-Clausen M, Krzywkowski K, Dulal K, Cheng T, Hjortø GM, Larsen O, Burg JS, Jarvis MA, Garcia KC, Zhu H, Kledal TN, Rosenkilde MM. Rationally designed chemokine-based toxin targeting the viral G protein-coupled receptor US28 potently inhibits cytomegalovirus infection in vivo. *Proceedings of the National Academy of Sciences*. 2015; 112:8427–8432.
- Tu CC, Spencer JV. The DRY box and C-terminal domain of the human cytomegalovirus US27 gene product play a role in promoting cell growth and survival. *PLoS One*. 2014; 9:e113427. [PubMed: 25409008]
- Vieira J, Schall TJ, Corey L, Geballe AP. Functional Analysis of the Human Cytomegalovirus US28 Gene by Insertion Mutagenesis with the Green Fluorescent Protein Gene. *Journal of Virology*. 1998; 72:8158–8165. [PubMed: 9733857]
- Wang N, Baldi PF, Gaut BS. Phylogenetic analysis, genome evolution and the rate of gene gain in the Herpesviridae. *Molecular Phylogenetics and Evolution*. 2007; 43:1066–1075. [PubMed: 17196839]
- Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol Biol Evol*. 2014; 31:2356–2364. [PubMed: 24916030]
- Wilkinson GWG, Davison AJ, Tomasec P, Fielding CA, Aicheler R, Murrell I, Seirafian S, Wang ECY, Weekes M, Lehner PJ, Wilkie GS, Stanton RJ. Human cytomegalovirus: taking the strain. *Med Microbiol Immunol*. 2015; 204:273–284. [PubMed: 25894764]
- Wreghitt TG, Teare EL, Sule O, Devi R, Rice P. Cytomegalovirus Infection in Immunocompetent Patients. *Clinical Infectious Diseases*. 2003; 37:1603–1606. [PubMed: 14689339]
- Zhang J. Evolution by gene duplication: an update. *TRENDS in Ecology and Evolution*. 2003; 18:292–298.

Highlights

- HCMV encodes two genes that share a common ancestor with host CX3CR1.
- Despite common ancestry, US28 binds CX3CL1 but US27 has no ligand binding activity.
- US27 has diverged significantly and currently has higher similarity to host CCR1.

Figure 1. Schematic diagram of GPCR activation

The GPCR is depicted in brown and has seven transmembrane segments with the amino terminus as the extracellular ligand binding domain. The carboxy-terminal tail serves as an intracellular signaling domain that binds to heterotrimeric G proteins. In response to ligand binding, the receptor undergoes a conformational change that enables the $G\alpha$ protein to exchange GDP for GTP and transmit downstream signals via an intracellular signaling cascade.

Figure 2. pUS27 and pUS28 have high similarity to human chemokine receptors
Reciprocal of E-value from BLASTp alignment results after searching with A) pUS27 and B) pUS28.

Figure 3. Comparison of pUS27 and pUS28 alignment matches with human chemokines
Reciprocal of E-value from BLASTp alignment results for pUS27 and pUS28 with top human chemokine receptor matches and with each other.

Figure 4. Maximum likelihood phylogenetic tree of HCMV pUS27 and pUS28 and human chemokine receptors

Various strains of HCMV are shown in green (pUS27) and purple (pUS28). Human chemokine receptors are in blue. Nodes values display bootstrap values (100 replicates). Inferred using PhyML and Geneious v 8.1.3.

Figure 5. Bayesian phylogenetic tree of HCMV pUS27 and pUS28 human chemokine receptors and human chemokine receptors

Various strains of HCMV are shown in green (pUS27) and purple (pUS28). Human chemokine receptors are in blue. Red indicates likely gene capture event by ancestral virus. Node values display posterior probabilities. Inferred using MrBayes and Geneious v 8.1.3.

Table 1

HCMV pUS28 Alignment Percent Identity Matrix¹

	TB40/E	TR	Toledo	ADI69	Towne	Merlin
TB40/E	100.00	97.46	98.02	98.31	98.31	98.02
TR	97.46	100.00	99.44	98.87	98.87	98.59
Toledo	98.02	99.44	100.00	99.44	99.44	99.15
ADI69	98.31	98.87	99.44	100.00	100.00	99.72
Towne	98.31	98.87	99.44	100.00	100.00	99.72
Merlin	98.02	98.59	99.15	99.72	99.72	100.00

¹ Alignment performed using the MUSCLE multiple sequence alignment tool, version 3.8 (Edgar, 2004). Matrix created using Clustal2.1 formatting.

Table 2

HCMV pUS27 Alignment Percent Identity Matrix¹

	TB40/E	TR	Toledo	ADI69	Towne	Merlin
TB40/E	100.00	98.06	95.29	98.34	95.29	94.74
TR	98.06	100.00	95.30	99.17	95.30	95.30
Toledo	95.29	95.30	100.00	95.58	99.45	98.90
ADI69	98.34	99.17	95.58	100.00	95.58	95.03
Towne	95.29	95.30	99.45	95.58	100.00	98.90
Merlin	94.74	95.30	98.90	95.03	98.90	100.00

¹ Alignment performed using the MUSCLE multiple sequence alignment tool, version 3.8 (Edgar, 2004). Matrix created using Clustal2.1 formatting.