Spring 5-20-2016

Phylogenetic Analysis of Human Cytomegalovirus pUS27 and pUS28: Ascertaining an Independent or Linked Evolutionary History

Jessica A. Scarborough
University of San Francisco, jess.scarborough12@gmail.com

Phylogenetic Analysis of Human Cytomegalovirus pUS27 and pUS28: Ascertaining an Independent or Linked Evolutionary History

Jessica Scarborough

Biology Department of the University of San Francisco

11 May 2016

Thesis Committee:

Juliet Spencer

John Paul

James Sikes

Abstract

Human cytomegalovirus (HCMV) is a widespread pathogen that is particularly skilled at evading immune detection and defense mechanisms, largely due to extensive co-evolution with its host's immune system. One aspect of this co-evolution involves the acquisition of four virally encoded GPCR chemokine receptor homologs, products of the US27, US28, UL33 and UL78 genes. G protein-coupled receptors (GPCR) are the largest family of cell surface proteins, found in organisms from yeast to humans. In this research, phylogenetic analysis was used to investigate the origins of the US27 and US28 genes, which are adjacent in the viral genome. The results indicate that both US27 and US28 share the same common ancestor, the gene for human chemokine receptor CX3CR1, suggesting that a single human gene was captured and that a viral gene duplication event occurred. It also appears that after the gene duplication event, US27 may have undergone neofunctionalization, while US28 maintained the function of their ancestral gene. While the evolutionary advantage of the gene duplication and neofunctionalization event remains unclear, experimental evidence indicates that each gene has evolved distinct, important functions during virus infection.

Introduction

Herpes Virus Family Evolution

Human cytomegalovirus (HCMV) is a herpes virus with the ability to establish life-long infection. Widespread throughout the population (1), HCMV is adept at evading the host immune system, largely due to having acquired many genes with homology to genes from its human host (2). This propensity for dodging the host immune system is a characteristic trait of many herpes viruses. Several studies have led to the conclusion that the herpes virus family co-evolves with its host, typically through gene capture (3-5). This has led to many independent adaptations including the acquisition of HCMV's US27 and US28 genes, which is only present in primate cytomegaloviruses. Herpes viruses benefit from this lateral transfer in two ways. For one, lateral transfer provides a mechanism for molecular mimicry, which leads to reduced immune detection because the virus has increased host-like proteins that the host immune system is less likely to detect (6). Secondly, it acts as a relatively fast way to introduce new genes and functions into the viral genome, potentially increasing fitness (4, 5).

Human Cytomegalovirus

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 from latency (7). It is believed that HCMV's ability to perform immunomodulation through chemokine and chemokine receptor homologs plays a role in this strong latency protection (7). This research will examine two chemokine receptor homologs, the gene products of US27 (pUS27) and US28 (pUS28), that have various functions

within the virus, one of which is mediation of the host chemoattraction immune defense strategy (7).

Although the virus can reactivate from latency periodically, hosts typically do not experience clinical symptoms due to strong T-cell responses (7, 8). Immunocompromised patients, however, are much more likely to experience observable symptoms upon degradation of the immune system, with flu-like symptoms often being reported (9). HCMV symptoms (malaise, fever, sweats, etc.) may appear minimal, but HCMV infection in an immunocompromised patient is extremely serious and often requires hospitalization for treatment. Additionally, congenital infection can lead to severe birth defects, including hearing loss, mental disability, small head size, seizures, and death (10).

Recently, HCMV has been connected to increased malignancy of cancers through a process known as "oncomodulation" where the virus infects existing tumor cells, leading to increased cellular proliferation, chemotaxis, and other malignancy capabilities (11). This ability to affect the growth and spread of cancer does not rely on an immune compromised host. Oncomodulation properties of HCMV can be seen in glioblastomas, where the virus is capable of promoting the malignant phenotype, making the HCMV a potential target when treating malignant gliomas (12). It also has been recently discovered that inflammatory breast cancer cells have a higher incidence of multiple viral DNAs (including HCMV) than non-inflammatory breast cancer cells (13). Human cytomegalovirus, although originally viewed as primarily benign, undoubtedly has significant impact on all stages of human life, from gestation with congenital birth defects to adulthood with opportunistic attack and oncomodulation.

G-Protein-Coupled Receptors (GPCRs)

This study is focused on US27 and US28, two genes within HCMV that show homology to human chemokine receptors, a class of GPCRs (14). Along with the gene products of UL33 (pUL33) and UL78 (pUL78), these GPCR homologs are key elements in HCMV's ability to establish lifelong latent infection within its host. GPCRs are cell surface receptors, which produce a cellular signal or response upon the binding of a ligand. Humans have over 800 unique GPCRs, with a wide assortment of roles, yet each is similar in physical structure (15). As seen in Figure 1, these receptors are made of a single polypeptide that folds to create seven segments, looping in and out of the cell with the N-terminus and C-terminus being extra- and intracellular, respectively. Once a ligand activates a GPCR, a conformational change occurs causing it to activate a local G-protein, which then leads to the dispatch of signals within the cell.

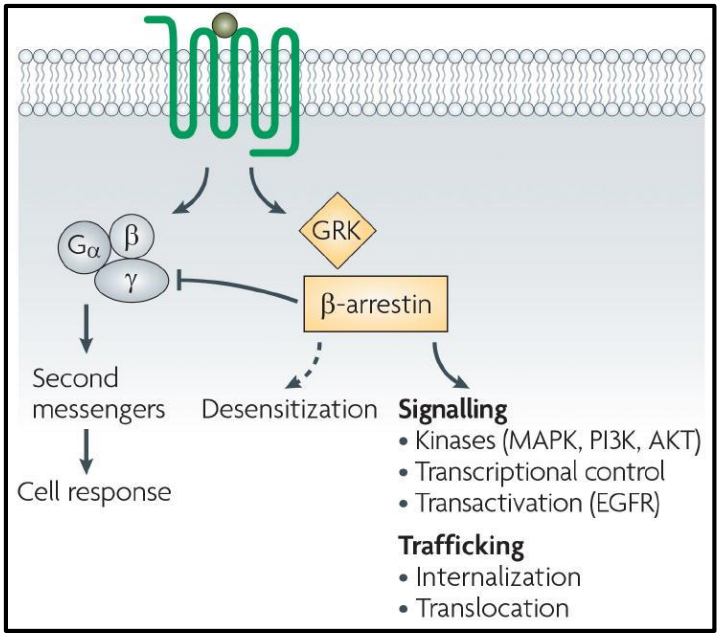

Figure 1. GPCR activation in response to ligand binding. Image taken from Rajagopal S. et al., 2010 (16).

pUS27 and pUS28 have homology to human chemokine receptors, a subset of the GPCR superfamily, which play a crucial role in immune response. Each chemokine receptor binds to specific chemokines, small secreted proteins, and regulates intracellular signaling and immune responses (17). Responses to ligand binding and G-protein activation include pro-inflammation or anti-inflammation, enhanced chemotaxis, and superoxide anion production (17). 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, plays a key role in inflammatory and infectious diseases, including psoriasis and allergies (17).

As summarized by Beisser et al. (2008), the key hallmark features of a chemokine receptor homologue 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 (18). In addition to putative chemokine receptors pUS27, pUS28, pUL33 and pUL78, HCMV contains several genes that could code for a seven transmembrane region. However, only pUS27, pUS28, and pUL33 meet these criteria, with pUS27 and pUS28 showing the greatest similarity to chemokine receptors.

US27 and US28

This study aims to examine the evolutionary history of HCMV's US27 and US28, two genes that code for human chemokine receptor homologs, currently hypothesized to have arisen via gene duplication. In an analysis by Beisser et al. (2002), US27 and US28 showed the closest similarity to the human chemokine receptor gene CX3CR1, but the study did not evaluate the

evolutionary history of these two genes (19). The goal of this research is to provide a rooted, phylogenetic analysis to determine the ancestral origins of US27 and US28. Additionally, although these viral proteins have been determined to be homologs of human GPCRs, it is unclear what parent gene US27 and US28 came from (18). HCMV, like other herpes viruses, is known to use gene duplication and co-evolution with its host as a means of diversifying its genome (4). Phylogenetic analysis may indicate whether US27 and US28 evolved independently (showing two separate parent genes) or via gene duplication (with one parent gene).

pUS27 is a putative chemokine receptor, and it has no known ligands to date. Furthermore, this protein can be found intracellularly and within the transmembrane region of an infected cell (7). It increases the extracellular spread of the HCMV virus by 10-fold in fibroblasts and endothelial cells in comparison to a US27 knockout strain (20). pUS27 may be responsible for enhancing chemotaxis of virus-infected cells by increasing signaling through CXCR4, a human chemokine receptor (21). Additionally, pUS27 has shown to increase cellular proliferation in human embryonic kidney cells (HEK293) by suppressing host negative growth regulators, although the complete mechanism is unknown (21).

In contrast to pUS27, pUS28 is a functional chemokine receptor that is known to bind to an extensive variety of ligands (22). pUS28 appears to play a key role in HCMV infection efficacy and potency along with increased cell-to-cell spread of the HCMV virus (23, 24). pUS28 has immunoevasion properties, due to its ability to internalize host chemokines, decreasing the host immune system's capabilities (7, 19). Some chemokines that pUS28 plays a role in binding include, but are not limited to: CCL2, CCL4, CCL5, CCL7, and CX3CL1 (7, 25). By binding to these chemokines, pUS28 is responsible for limiting their dissemination,

interfering with host immune activity (7). Despite the differences between pUS27 and pUS28, both proteins play an important role in the virus' ability to infect its host.

Methods

HCMV Strains

The US27 and US28 genes come from the AD169, Towne, Toledo, Merlin, TR, and TB40/E strains of HCMV, and all of these strains were used for the phylogenetic analysis portion of this study. The GenBank accession numbers of these viral genes can be found in Table 1. These strains were chosen because they are prominent laboratory strains (AD169, TB40/E, Towne) and clinical isolates (Toledo, TR, Merlin) (26, 27). Gross gene alignments were done using HCMV's strain AD169, the first HCMV strain to be sequenced and annotated, which is extensively used during laboratory research (26).

Gross Gene Comparison

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 gene to the entire humane genome, displaying the top 100 matches by E-value (28). 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. The outgroup used in phylogenetic analysis is a human G-protein coupled receptor that is not part of the chemokine receptor family. Accession numbers for the human receptors can be found in Table 2.

Table 1. HCMV protein accession numbers for HCMV strains[1]

Name	Strain	Accession Number
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
US28	TB40/E	ABV71518.1

[1]obtained from GenBank.

Table 2. Human GPCR accession numbers[1]

Receptor Name	Accession Number
CCR1	NP_001286.1
CCR4	NP_005499.1
CCR9	AAH95516.1
CX3CR1	NP_001328.1
CCR5	NP_000570.1
CCR3	NP_001828.1
CXCR3	NP_001495.1
CCR6	NP_004358.2
CCR8	NP_005192.1
CXCR4	CAA12166.1
CCR2	AGC02843.1
Oxoeicosanoid Receptor 1[2]	NP_062545.1

[1]Obtained from GenBank.

[2]Oxoeicosanoid Receptor 1 is the outgroup used in later phylogenetic analysis.

Alignment

Protein alignments between HCMV strains were performed using the MUSCLE multiple sequence alignment tool, version 3.8 (29), and alignment matrices were created by Clustal2.1 (30, 31) within Geneious v. 9.1, a bioinformatics phylogenetics software, with the default settings used (32). The alignments were then further edited by hand to replace blanks at the beginning and end of the alignment with missing data question marks.

Phylogenetic Analysis

Using Geneious v. 9.1, this research inferred phylogenetic trees using Bayesian and maximum likelihood method to analyze the evolutionary history of the pUS27 and pUS28 proteins in relation to select human chemokine receptors. The maximum likelihood analysis was inferred using PhyML v. 3.0 plugin (33, 34) 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 (35).

Maximum likelihood estimations determine which tree most likely depicts an accurate evolutionary history. Bayesian analysis also examines the likelihood of a given tree based on given models of evolution. Parsimony analysis was not used due to its higher sensitivity to long-branch attraction than either maximum likelihood or Bayesian models (36).

Protein Analysis

Using GenomeNet's MOTIF search, this study searched pUS27 and pUS28 in the National Center for Biotechnology Information-Conserved Domain Database (NCBI-CDD). The ExPASy Bioinformatics Resource Portal was also utilized to search the Prosite database using ScanProsite. Both HCMV genes analyzed came from the AD169 strain.

Results

Protein Alignments

In order to determine which HCMV strain sequence to use for subsequent analyses of pUS27 and pUS28, variability among different strains was examined. A Clustal comparison matrix is shown in Tables 3 and 4. Here, it is observable that the 6 strains examined in this study show minimal variation within the pUS27 and pUS28 primary protein structure. The pUS28 comparisons are displayed in Table 3, where it is shown that the lowest percent identity is between the strains TB40 and TR (both clinical strains) with 97.46% identity match within the pUS28 primary protein structure. The closest identity is seen between AD169 and Towne, both laboratory strains that share 100% identity match within the pUS28 protein. Table 4 shows pUS27 percent identity comparisons, where there is slightly more variation between strains. The highest variation is between TB40/E and Merlin, a laboratory and clinical strain respectively, with 94.74% identity match within the pUS27 protein. The closest similarity is again seen between Towne and AD169 where they show 95.58% identity match within the pUS27 protein. Thus, both genes are highly conserved among distinct virus strains, indicating their function is likely beneficial for the virus.

Table 3. HCMV pUS28 alignment percent identity matrix.[1]

	TB40/E	**TR**	**Toledo**	**AD169**	**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
AD169	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 (29). Matrix created using Clustal2.1 formatting.

Table 4. HCMV pUS27 alignment percent identity matrix. [1]

	TB40/E	**TR**	**Toledo**	**AD169**	**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
AD169	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 (29). Matrix created using Clustal2.1 formatting.

Figures 2 and 3 display the closest chemokine receptors to pUS27 and pUS28, respectively. Alignments between complete primary protein structures were evaluated using e-value. Figures 2-4 use the reciprocal e-value in order to show a clear visualization of a larger bar equating to a closer alignment match. Figure 2 shows that 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 extraordinarily different. pUS28 on the other hand, shows a very strong correlation to CX3CR1 in particular. The difference in alignments between pUS28 and CX3CR1 and CCR4 (the second most closely matched to US27) is very significant. As seen in Figure 4, the alignment of pUS27 and pUS28 results in an e-value of 1E-49. This shows that pUS27 is closer to pUS28 than any human chemokine receptor; however, pUS28 is closer to CX3CR1 than pUS27.

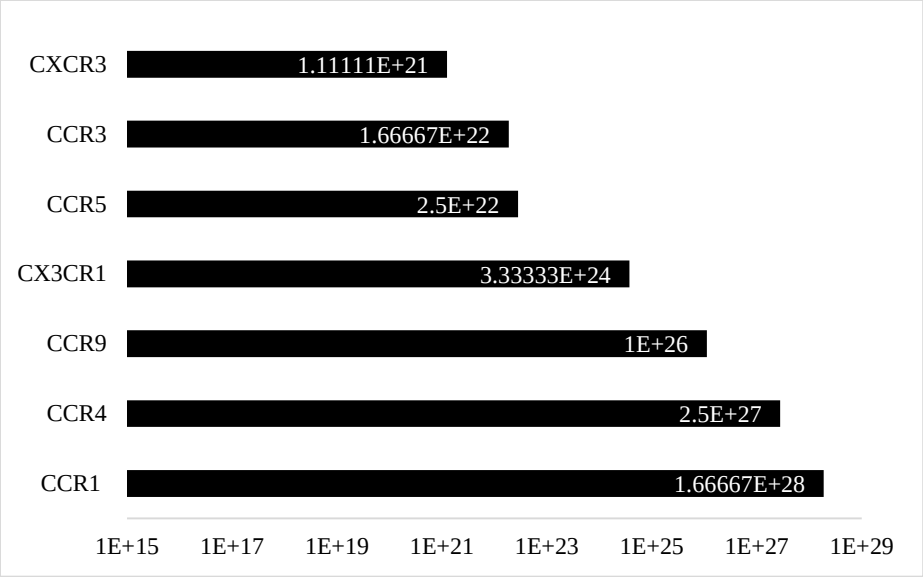

Figure 2. US27's closest chemokine receptor matches, displaying reciprocal e-value for alignment results from BLASTp results after searching pUS27 and accumulating the closest chemokine receptor matches (28).

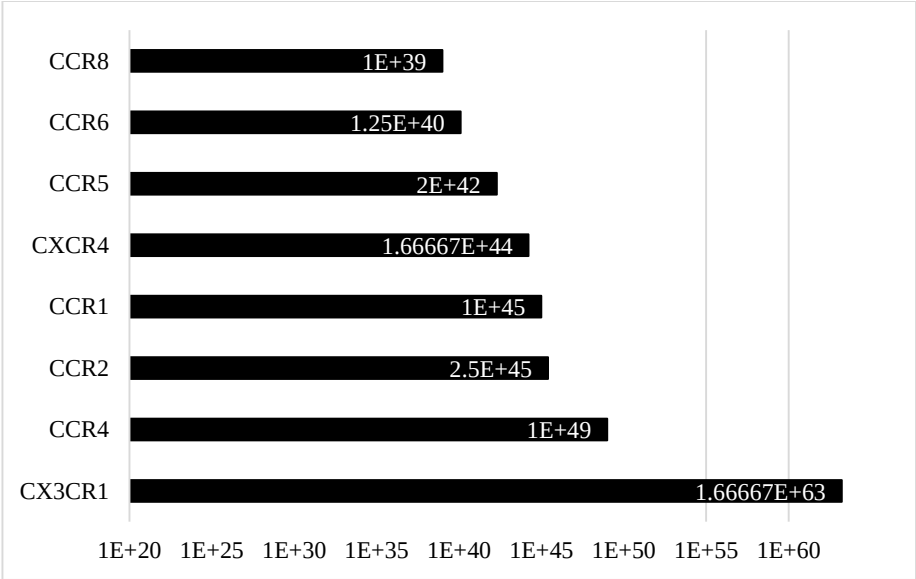

Figure 3. US28's closest chemokine receptor matches, displaying reciprocal e-value for alignment results from BLASTp results after searching pUS27 and accumulating the closest chemokine receptor matches (28).

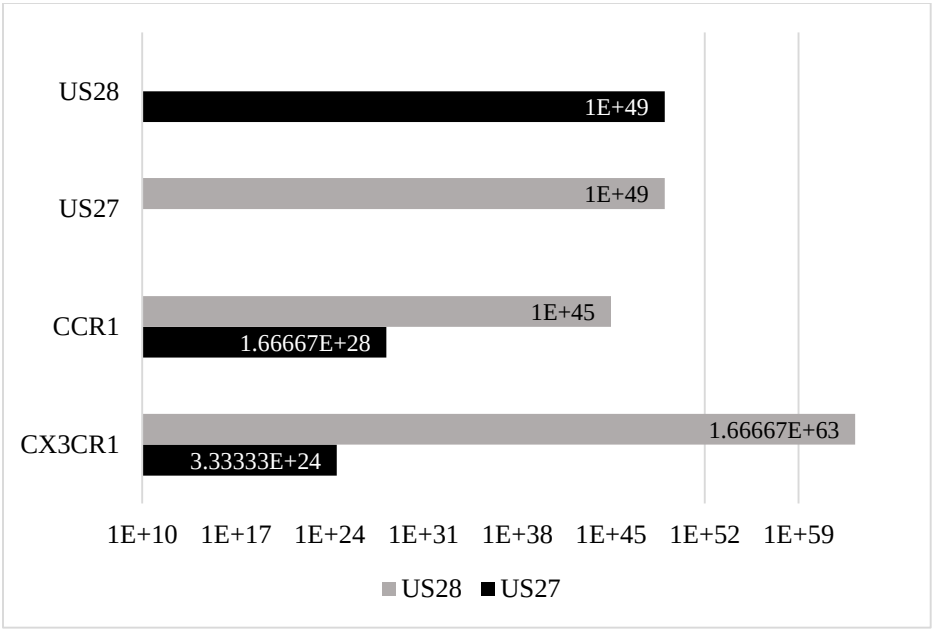

Figure 4. Comparison of pUS27 and pUS28 closest alignment matches in addition to added alignment of pUS27 and pUS28.

Phylogenetic Analysis

Phylogenetic analysis was used to determine if US27 and US28 do or do not share a common ancestor. 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

Figure 5 displays a maximum likelihood model phylogeny, which supports the hypothesis that US28 and US27 share a common ancestor. Additionally, the common ancestor of these two viral genes shares a common ancestor with human CX3CR1. Here, the bootstrap support values show strong support for a common ancestor between pUS27 and pUS28. There is very minimal support, however, towards pCX3CR1 being the closest human protein to pUS27 and pUS28's common ancestor. The Bayesian phylogenetic analysis in Figure 6 concurs 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. The posterior probabilities are extremely 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 HCMV virus. Then, within the HCMV virus, there was another gene duplication event that led to US28 and US27 evolving from one common ancestor. Using Bayesian and maximum likelihood phylogeny, it was determined that US27 and US28 most likely arose from a gene duplication event, and their closest human ancestor gene is CX3CR1.

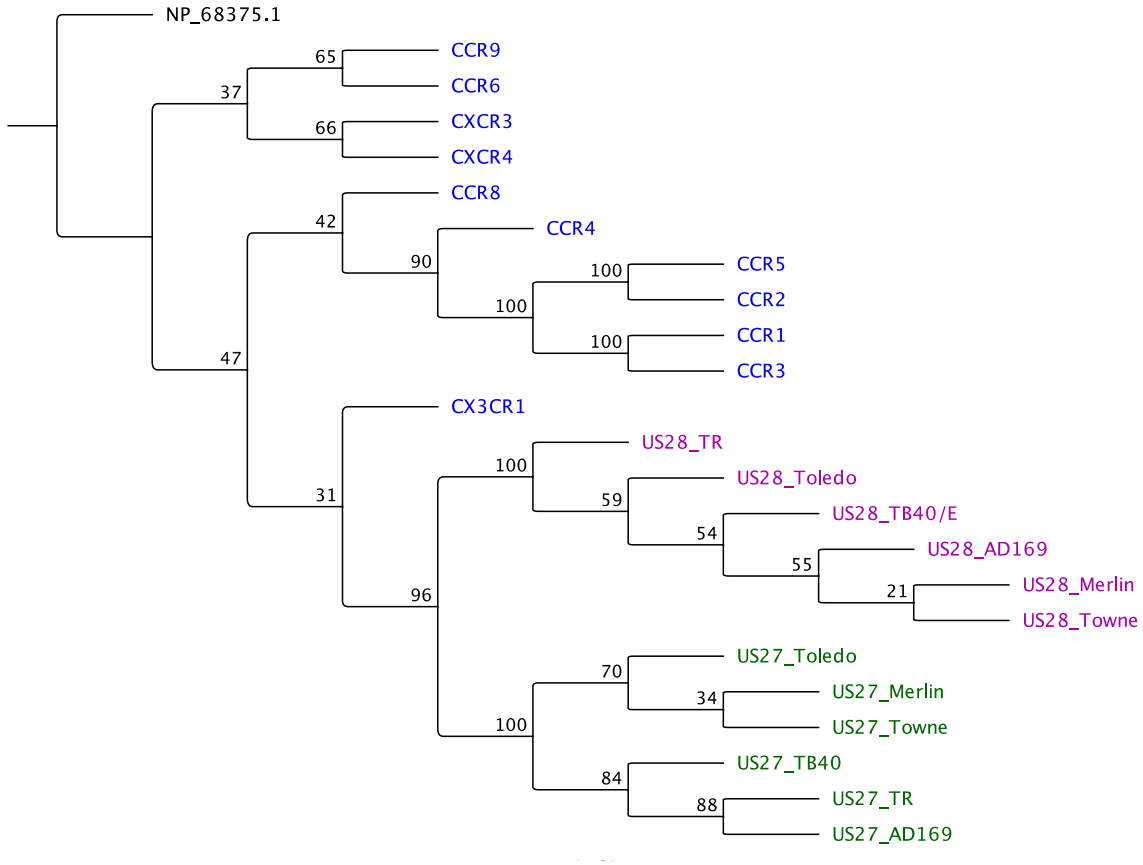

Figure 5. Maximum likelihood phylogenetic tree of human chemokine receptors and pUS27 and pUS28. Nodes values display bootstrap values (100 replicates). Inferred using PhyML and Geneious v 8.1.3 (33).

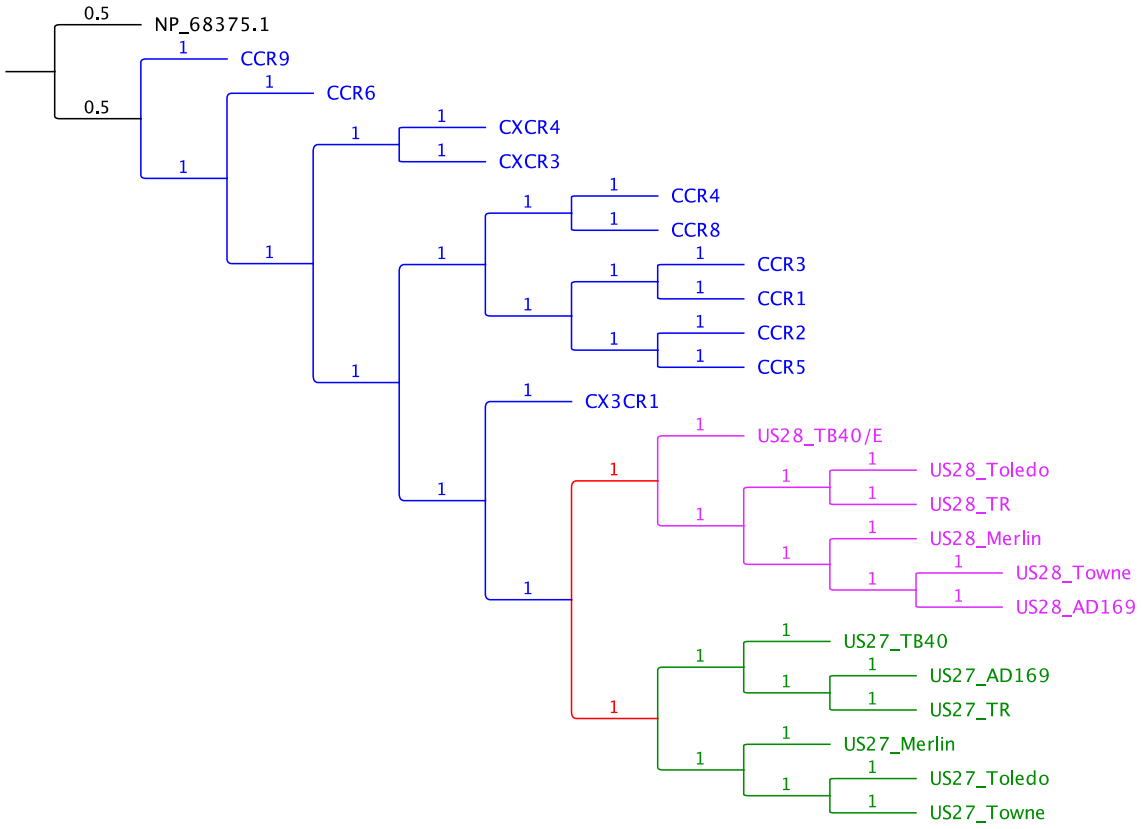

Figure 6. Bayesian phylogenetic tree of human chemokine receptors and pUS27 and pUS28. Node values display posterior probabilities. Inferred using MrBayes and Genious v 8.1.3 (35).

Protein Analysis

When analyzing pUS27 and pUS28 using the Prosite protein database (37), both pUS27 and pUS28 showed GPCR similarities (due to their seven transmembrane hydrophobic regions), and their N-terminal and C-terminal domains showed no statistically significant motifs. Although both pUS27 and pUS28 contain sites for cAMP- and cGMP-dependent protein kinase phosphorylation, N-myristoylation, N-glycosylation, and Casein kinase II phosphorylation, these motifs are not considered significant due to their "high probability of occurrence" (37). They are motifs that are found in many proteins, frequently enabling post-translational modification.

Hydrophobicity plots of CX3CR1, pUS27, and pUS28 can be found in Figures 7, 8, and 9 respectively. Here, the 7 regions of hydrophobicity correlating to the 7 transmembrane regions of each of these proteins can be observed. These transmembrane regions are sites of amino acid similarity between all GPCRs due to the limited number of hydrophobic amino acids and conserved nature of the GPCR structure. Even so, US27 and US28 displayed a closer evolutionary history to CX3CR1, as seen in Figures 5 and 6. These hydrophobicity plots are important for determining the intracellular, extracellular, and transmembrane regions of the protein. A future extension of this project should examine which segments are constrained within the transmembrane regions, compared to the intracellular and extracellular regions. It is expected that the transmembrane regions will have a lower amino acid substitution rate, while the intracellular and extracellular regions will have a higher amino acid substitution rate. Extracellularly, an increased amino acid substitution rate will allow the receptors to bind to different chemokines. Intracellular changes in the primary protein structure could allow for variations in cell signaling pathways. In short, examining the hydrophobic regions of the

CX3CR1, pUS27, and pUS28 will provide insight into the significant differences between these receptors.

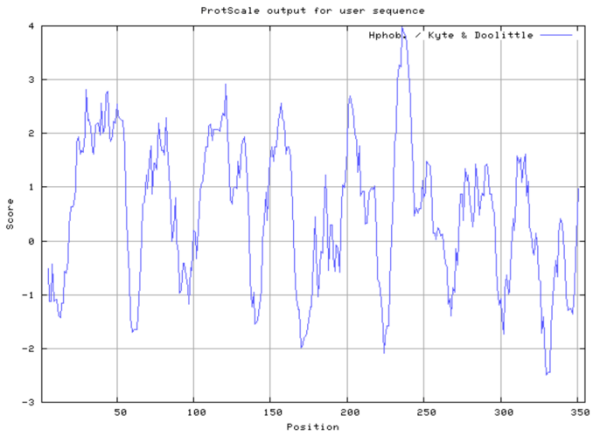

Figure 7. CX3CR1 hydrophobicity plot built utilizing the Kyte and Doolittle amino acid scale (38) and the ProtScale software (39).

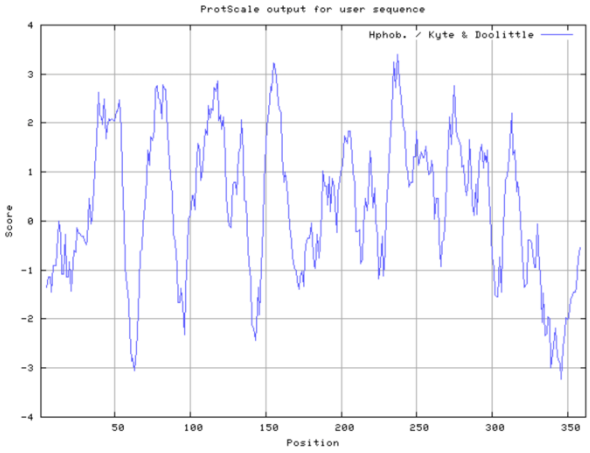

Figure 8. pUS27 hydrophobicity plot built utilizing the Kyte and Doolittle amino acid scale (38) and the ProtScale software (39).

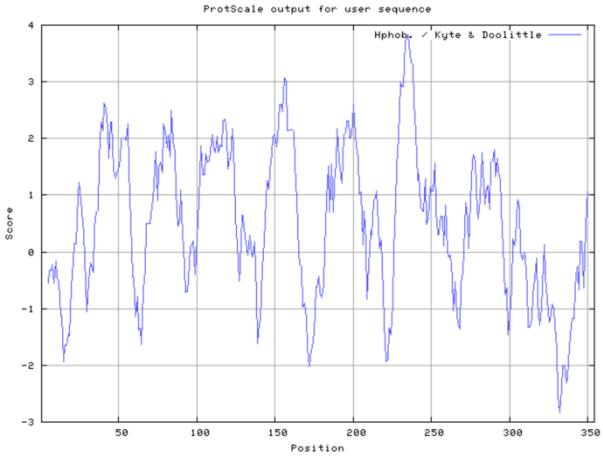

Figure 9. pUS28 hydrophobicity plot built utilizing the Kyte and Doolittle amino acid scale (38) and the ProtScale software (39).

Discussion

Although the evolutionary history of viruses is difficult to determine due to significant variation between strains and high rates of mutation, this evidence points towards the conclusion of a gene duplication leading to the presence of the US27 and US28 genes in HCMV. It is likely that an ancestral HCMV captured CX3CR1 from its human host, then integrated this gene into its own genome. This newly integrated gene would have then undergone a gene duplication event which led to the creation US27 and US28. Some of the numerous possible outcomes of a gene duplication are displayed in Figure 10.

Labeled "b," neofunctionalization occurs when one gene from a gene duplication lacks selective pressure, while the other gene is constrained by it (40). In the neofunctionalization scenario, one protein has a low amino acid substitution rate due to purifying selection pressure, causing it to stay much more similar to its parent gene. On the other hand, the second gene may take on a new function because of the lack of selective pressure (41). Subfunctionalization, labeled "c" in Figure 10, occurs when there is no evolutionary advantage to the extra gene copy (41). 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 shows a much higher similarity to pCX3CR1 than pUS27 (as seen in Figure 4), it is likely that US27 underwent neofunctionalization after divergence, although subfunctionalization is also a possibility. With this model, it is proposed that after gene duplication occurred, the virus was able to keep the functionality of pUS28, while pUS27 could be less constrained and eventually gain a new function. Other outcomes for duplicated genes

include gene loss (Fig. 10a) or regulatory differentiation (Fig. 10d) (42). Although regulatory differentiation could have occurred, it is not the focus of this research. Since regulatory sequences are not transcribed into mRNA, the protein amino acid sequences examined in this research is not sufficient for examining the differences between US27 and US28's regulatory patterns (41). What can be concluded from this research is that pUS27 likely underwent neofunctionalization after a gene duplication event while US28 had a lower amino acid substitution rate, keeping its original function. Further research should examine the intracellular and extracellular regions of pUS27, pUS28, and CX3CR1. If the neofunctionalization model is correct, then pUS28 should show closer similarity to CX3CR1 in these regions, because it was more conserved after the gene duplication event than pUS27. Further analysis could confirm whether neofunctionalization is the most likely outcome of the US27/US28 gene duplication.

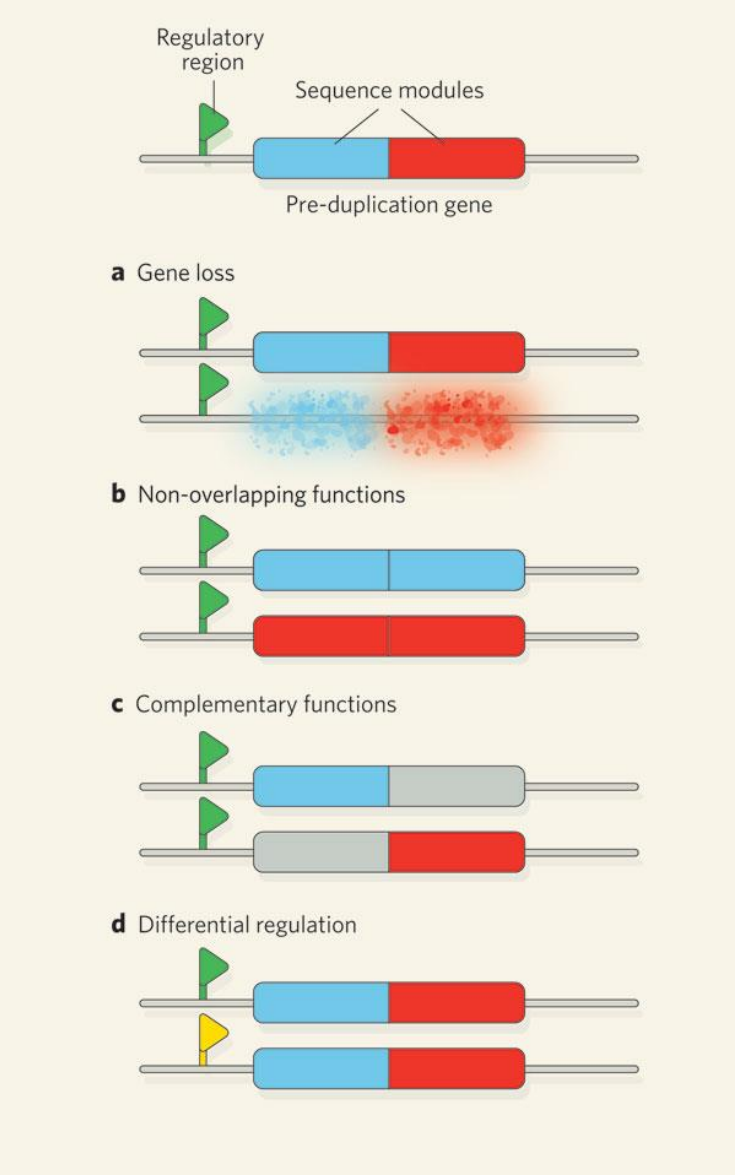

Figure 10. Representation of the possible fates of a duplicated gene. Image taken from Louis, 2007 (42).

Both maximum likelihood and Bayesian phylogenetic hypotheses a common ancestor between US27 and US28, suggesting a gene duplication event led to their existence. Their closest human ancestral gene appears to be Chemokine (C-X3-C Motif) Receptor 1 (CX3CR1). The effectiveness of these methods can be observed when examining the clades containing CCR1, CCR2, CCR3, CCR4, CCR5, and CCR8 in Figures 5 and 6. Five out of six of these proteins are coded for on Chromosome 3, likely suggesting that they share an evolutionary history. The fact that they were grouped together 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, located on Chromosome 3, codes for a chemokine receptor that binds fractalkine, also known as CX3C Ligand 1. Activation of the CX3CR1 protein leads to mediation ofthe cell's adhesion and movement functions (43). This receptor is largely associated with expression on lymphocytes, but CX3CR1 has also been observed in leukocytes and migroglial cells (44). In the central nervous system, CX3CR1 and CX3CL1 play an important role, and the binding of fractalkine to CX3CR1 leads to the attraction of intraepithelial lymphocytes (45). Additionally, 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 (46). When it is in a soluble state, it 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, fractalkine and CX3CR1 form strong binding interactions (46).

The human gene CX3CR1 contains 4 exons and 3 introns, displayed in Figure 11 (47). On the other hand, US27 and US28 do not contain any introns (48). In this figure, the DNA of CX3CR1 is shown with 3 introns, surrounded by 4 exons, while the processed mRNA only contains the 4 exons. This is important to note, because the exact mechanism of gene capture has not been determined; however, since US27 and US28 do not contain introns, it is possible that HCMV gained the common ancestor to US27 and US28 from a retrovirus genome during a mixed infection (49). Further research could experimentally determine if HCMV is capable of gaining genetic information from a retrovirus during the co-infection of a cell.

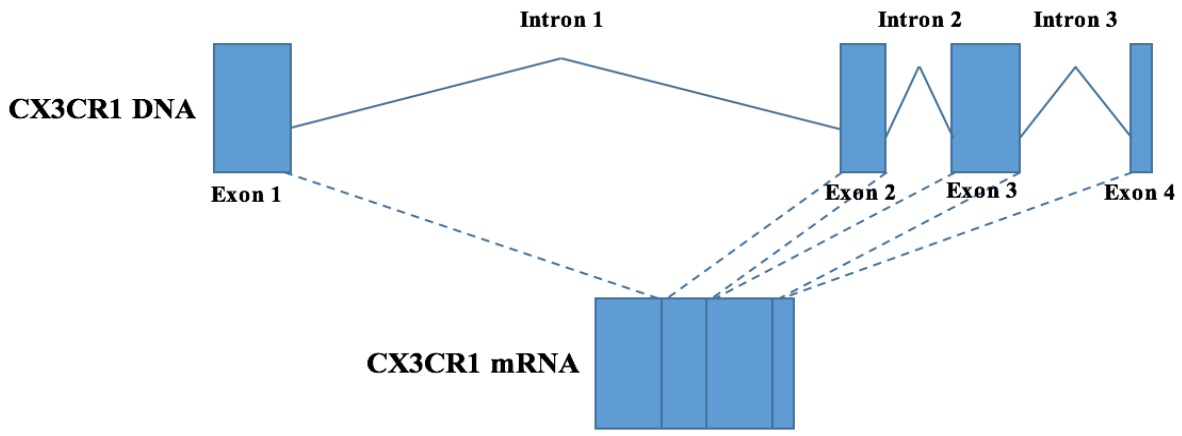

Figure 11. Representation of CX3CR1 gene displaying exons and introns compared to mature CX3CR1 mRNA

One disease commonly associated with CX3CR1 is human immunodeficiency virus infection and its progression to acquired immune deficiency syndrome (HIV/AIDS). CX3CR1 is a co-receptor for HIV-1, and some variations of the gene have been linked to an increase in susceptibility of infection of HIV and rapid progression of HIV infection to AIDS (50). There are also increased levels of expression of the CX3CR1 within HIV-positive monocytes (51). Moreover, active HCMV infection in HIV-infected patients has been connected with perpetually low CD4+ T cell counts (52). It's possible that HCMV co-infection is an increased risk factor for the progression of HIV to AIDS, because pUS28 has similar binding properties to CX3CR1. In fact, US28 has been identified as a co-receptor for HIV, and HCMV has been connected to HIV's disease progression to AIDS (52, 53). There is certainly a noteworthy connection between pUS27 and pUS28, whose ancestral protein is CX3CR1, and the role of HCMV in the progression of HIV-infection to AIDS.

Recent research has also shown a functional role of CX3CR1 in the metastasis of breast cancer, while inhibition of the receptor impairs breast cancer cells' ability to seed in their host (54). Furthermore, primary breast cancer tumors have shown to overexpress CX3CR1 (54). Our research has shown that US27 and US28 likely evolved from a CX3CR1 homolog. Again, if pUS27 and pUS28 express similar binding properties to CX3CR1, they may be creating the same cellular signaling pathways that cause CX3CR1 to play a role in the metastasis of breast cancer. Although this has not been experimentally proven, it is a hypothesis that should be explored further, because it provides an explanation for HCMV's link to the metastasis of patients with primary breast cancers (55, 56).

The US27 gene is only found in primate CMV strains. Additionally, a gene duplication event with US28 homologs has been documented within RhCMV, where there are five copies of

the US28 homolog, Rh214, Rh215, Rh216, Rh218, and Rh220 (57). Further research should examine the evolutionary advantage of these 5 copies, potentially shedding light on US27's function within HCMV. Another strategy to examine pUS27 should include comparative analyses between US27/pUS27 and other mammal CMV genomes and proteomes. Identifying common motifs, especially within the N- or C-terminal domains, could indicate pUS27's role within HCMV. Understanding the role pUS27 and pUS28 have in ensuring the success of HCMV infection and proliferation can shed light on their function within the progression of diseases such as HIV/AIDS and cancer. It also furthers research regarding potential antiviral targets in treating such disease.

Although the binding and signaling mechanisms of pUS27 remain unclear, there is no doubt that both pUS27 and pUS28 have a fundamental role in the functioning, infection, and spread of HCMV. BLAST searching and alignment features were able to elucidate the closest human chemokine receptor proteins to pUS27 and pUS28, while phylogenetic analyses examined the evolutionary history. Here, this research found that a gene capture event likely occurred between CMV and its host, along with an intermediate retrovirus. The gene captured in this event became the common ancestor of US27 and US28. From there, neofunctionalization occurred, leaving pUS28 with a similar function to its common ancestor and pUS27 with a new function that research has yet to fully elucidate. Further research should focus on elucidating the differences between pUS27 and pUS28, particularly within intracellular and extracellular regions of the amino acid code, where the most meaningful variation is likely to occur. The impact of this research includes identifying key components to HCMV function that could be targeted in antiviral therapies. These therapies could treat a myriad of diseases that HCMV plays a role in aggravating, including breast cancer and HIV/AIDS.

References:

1. **Cannon MJ, Schmid DS, Hyde TB.** 2010. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. Rev Med Virol **20:**202-213.
2. **McSharry B, Avdic S, Slobedman B.** 2012. Human Cytomegalovirus Encoded Homologs of Cytokines, Chemokines and their Receptors: Roles in Immunomodulation. Viruses **4:**2448-2470.
3. **Alcendor DJ, Zong J, Dolan A, Gatherer D, Davison AJ, Hayward GS.** 2009. Patterns of divergence in the vCXCL and vGPCR gene clusters in primate cytomegalovirus genomes. Virology **395:**21-32.
4. **Wang N, Baldi PF, Gaut BS.** 2007. Phylogenetic analysis, genome evolution and the rate of gene gain in the Herpesviridae. Molecular Phylogenetics and Evolution **43:**1066-1075.
5. **McGeoch DJ, Rixon FJ, Davison AJ.** 2006. Topics in herpesvirus genomics and evolution. Virus Research **117:**90-104.
6. **Hughes AL, Friedman R.** 2005. Poxvirus genome evolution by gene gain and loss. Molecular Phylogenetics and Evolution **35:**186-195.
7. **McSharry BP, Avdic S, Slobedman B.** 2012. Human Cytomegalovirus Encoded Homologs of Cytokines, Chemokines and their Receptors: Roles in Immunomodulation. Viruses **4:**2448-2470.
8. **Rosa CL, Diamond DJ.** 2012. The immune response to human CMV. Future Virology **7:**279-293.
9. **Wreghitt TG, Teare EL, Sule O, Devi R, Rice P.** 2003. Cytomegalovirus Infection in Immunocompetent Patients. Clinical Infectious Diseases **37:**1603-1606.
10. **Tabata T, Petitt M, Zydek M, Fang-Hoover J, Larocque N, Tsuge M, Gormley M, Kauvar LM, Pereira L.** 2015. Human Cytomegalovirus Infection Interferes with the Maintenance and Differentiation of Trophoblast Progenitor Cells of the Human Placenta. J Virol **89:**5134-5147.
11. **Michaelis M, Doerr HW, Cinatl J.** 2009. The Story of Human Cytomegalovirus and Cancer: Increasing Evidence and Open Questions. Neoplasia **11:**1-9.
12. **Dziurzynski K, Chang SM, Heimberger AB, Kalejta RF, McGregor Dallas SR, Smit M, Soroceanu L, Cobbs CS.** 2012. Consensus on the role of human cytomegalovirus in glioblastoma. Neuro-Oncology **14:**246-255.
13. **El-Shinawi M, Mohamed HT, Abdel-Fattah HH, Ibrahim SAA, El-Halawany MS, Nouh MA, Schneider RJ, Mohamed MM.** 2015. Inflammatory and Non-inflammatory Breast Cancer: A Potential Role for Detection of Multiple Viral DNAs in Disease Progression. Annals of Surgical Oncology **23:**494-502.
14. **Vieira J, Schall TJ, Corey L, Geballe AP.** 1998. Functional Analysis of the Human Cytomegalovirus US28 Gene by Insertion Mutagenesis with the Green Fluorescent Protein Gene. Journal of Virology **72:**8158-8165.
15. **Chattopadhyay A.** 2014. GPCRs: Lipid-Dependent Membrane Receptors That Act as Drug Targets. Advances in Biology **2014:**1-12.
16. **Rajagopal S, Rajagopal K, Lefkowitz RJ.** 2010. Teaching old receptors new tricks: biasing seven-transmembrane receptors. Nature Reviews Drug Discovery **9:**373-386.
17. **Murdoch C, Finn A.** 2000. Chemokine Receptors and Their Role in Vascular Biology. J Vasc Res **37:**1-7.

18. **Beisser PS, Lavreysen H, Bruggeman CA, Vink C.** 2008. Chemokines and Chemokine Receptors Encoded by Cytomegaloviruses doi:10.1007/978-3-540-77349-8_13, p 221-242. Springer Science + Business Media.
19. **Beisser PS, Goh C-S, Cohen FE, Michelson S.** 2002. Viral Chemokine Receptors and Chemokines in Human Cytomegalovirus Trafficking and Interaction with the Immune System, p 203-234, Viral Proteins Counteracting Host Defenses, vol 269. Springer Berlin Heidelberg.
20. **O'Connor CM, Skenk T.** 2011. 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 **85:**3700-3707.
21. **Arnolds KL, Lares AP, Spencer JV.** 2013. The US27 gene product of human cytomegalovirus enhances signaling of host chemokine receptor CXCR4. Virology **439:**122-131.
22. **Kuhn DE, Beall CJ, Kolattukudy PE.** 1995. The Cytomegalovirus US28 Protein Binds Multiple CC Chemokines with High Affinity. Biochemical and Biophysical Research Communications **211:**325-330.
23. **Noriega V, Gardner T, Redmann V, Bongers G, Lira S, Tortorella D.** 2014. Human Cytomegalovirus US28 Facilitates Cell-to-Cell Viral Dissemination. Viruses **6:**1202-1218.
24. **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.** 2015. 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 **112:**8427-8432.
25. **Kledal TN, Rosenkilde MM, Schwartz TW.** 1998. Selective recognition of the membrane-bound CX 3 C chemokine, fractalkine, by the human cytomegalovirus-encoded broad-spectrum receptor US28. FEBS Letters **441:**209-214.
26. **Wilkinson GWG, Davison AJ, Tomasec P, Fielding CA, Aicheler R, Murrell I, Seirafian S, Wang ECY, Weekes M, Lehner PJ, Wilkie GS, Stanton RJ.** 2015. Human cytomegalovirus: taking the strain. Med Microbiol Immunol **204:**273-284.
27. **Hosogai M, Shima N, Nakatani Y, Inoue T, Iso T, Yokoo H, Yorifuji H, Akiyama H, Kishi S, Isomura H.** 2015. Analysis of human cytomegalovirus replication in primary cultured human corneal endothelial cells. British Journal of Ophthalmology **99:**1583-1590.
28. **Madden T.** 2002. The BLAST Sequence Analysis Tool. *In* McEntyre J, Ostell J (ed), The NCBI Handbook. National Center for Biotechnology information, Bethesda, MD.
29. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:**1792-1797.
30. **Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R.** 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Research **38:**W695-W699.
31. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. ClustalW and ClustalX version 2.0. Bioinformatics **23:**2947-2948.
32. **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.**

2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647-1649.

33. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696-704.
34. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology **59:**307-321.
35. **Huelsenbeck J, Ronquist F.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-755.
36. **Bergsten J.** 2005. A review of long-branch attraction. Cladistics **21:**163-193.
37. **Sigrist CJA, Castro Ed, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I.** 2012. New and continuing developments at PROSITE. Nucleic Acids Research doi:10.1093/nar/gks1067
38. **Kyte J, Doolittle RF.** 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology **157:**105-132.
39. **Gastieger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A.** 2005. Protein Identification and Analysis Tools on the ExPASy Server, p 571-607. *In* Walker JM (ed), The Proteomics Protocols Handbook doi:10.1385/1592598900. Humana Press.
40. **Gibson TA, Goldberg DS.** 2009. Questioning the Ubiquity of Neofunctionalization. PLoS Computational Biology **5**.
41. **Zhang J.** 2003. Evolution by gene duplication: an update. TRENDS in Ecology and Evolution **18:**292-298.
42. **Louis EJ.** 2007. Evolutionary genetics: Making the most of redundancy. Nature **449:**673-674.
43. **Imai T, Hieshima K, Haskell C, Baba M, Magira M, Nishimura M, Kakizaki M, Takagi S, Nomiyama H, Schall TJ, Yoshie O.** 1997. Identification and molecular characterization of fractalkine receptor CX3CR1, which mediates both leukocyte migration and adhesion. Achivum Immunologiae et Therapia Experimentalis **91:**521-530.
44. **Limotala C, Ransohoff RM.** 2014. Modulating neurotoxicity through CX3CL1/CX3CR1 signaling. Frontiers in Cellular Neuroscience **8**.
45. **Isse K, Harada K, Zen Y, Kamihira T, Shimoda S, Harada M, Nakanuma Y.** 2005. Fractalkine and CX3CR1 Are Involved in theRecruitment of Intraepithelial Lymphocytes ofIntrahepatic Bile Ducts. Hepatology **41:**506-516.
46. **Haskell CA, Cleary MD, Charo IF.** 1999. 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 **274:**10053-10058.
47. **Garin A, Pellet P, Deterre P, Debré P, Combadiére C.** 2002. Cloning and functional characterization of the human fractalkine receptor promoter regions. Biochemical Journal **368:**753-760.
48. **Rawlingson WD, Barrell BG.** 1993. Spliced Transcripts of Human Cytomegalovirus. Journal of Virology **67:**5502-5513.
49. **McGeoch DJ, Davison AJ, Dolan A, Gatherer D, Sevilla-Reyes EE.** 2008. Molecular Evolution of the Herpesvirales, p 447-475. *In* Domingo E, Parrish CR, Holland JJ (ed), Origin and Evolution of Viruses. Elsevier.

50. **Parczewski M, Urbańska A, Maciejewska K, Clark J, Leszczyszyn-Pynka M.** 2014. Association of chemokine receptor gene variants with HIV-1 genotype predicted tropism. HIV Medicine **15:**577-586.
51. **Krishnan S, Wilson EMP, Sheikh V, Rupert A, Mendoza D, Yang J, Lempicki R, Migueles SA, Sereti I.** 2014. Evidence for Innate Immune System Activation in HIV Type 1–Infected Elite Controllers. Journal of Infection Diseases **209:**931-939.
52. **Munawwar A, Singh S.** 2016. Human Herpesviruses as Copathogens of HIV Infection, Their Role in HIV Transmission, and Disease Progression. Journal of Laboratory Physicians **8:**5-18.
53. **Pleskoff O, Tréboute C, Brelot A, Heveker N, Seman M, Alizon M.** 1997. Identification of a Chemokine Receptor Encoded by Human Cytomegalovirus as a Cofactor for HIV-1 Entry. Science **276:**1874-1879.
54. **Shen F, Zhang Y, Jernigan DL, Feng X, Yan J, Garcia FU, Meucci O, Salvino JM, Fatatis A.** 2016. Molecular Cancer Research doi:10.1158/1541-7786.
55. **Taher C, Frisk G, Fuentes S, Religa P, Costa H, Assinger A, Vetvik KK, Bukholm IRK, Yaiw K-C, Smedby KE, Bäcklund M, Söderberg-Naucler C, Rahbar A.** 2014. High Prevalence of Human Cytomegalovirus in Brain Metastases of Patients with Primary Breast and Colorectal Cancers. Translational Oncology **7:**732-740.
56. **Mohamed HT, El-Shinawi M, Nouh MA, Bashtar A-R, Elsayed ET, Schneider RJ, Mohamed MM.** 2014. Inflammatory breast cancer: high incidence of detection of mixed human cytomegalovirus genotypes associated with disease pathogenesis. Frontiers in Oncology **4**.
57. **Hansen SG, Strelow LI, Franchi DC, Anders DG, Wong SW.** 2003. Complete Sequence and Genomic Analysis of Rhesus Cytomegalovirus. Journal of Virology **77:**6620-6636.