pathogens

MDPI

Review

Cytomegalovirus Genetic Diversity and Evolution: Insights into Genotypes and Their Role in Viral Pathogenesis

Cristina Venturini * and Judith Breuer

Department of Infection, Immunity and Inflammation, Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK; j.breuer@ucl.ac.uk
* Correspondence: c.venturini@ucl.ac.uk

Abstract: Cytomegalovirus (CMV) is a ubiquitous virus that infects most of the human population and causes significant morbidity and mortality, particularly among immunocompromised individuals. Understanding CMV's genetic diversity and evolutionary dynamics is crucial for elucidating its pathogenesis and developing effective therapeutic interventions. This review provides a comprehensive examination of CMV's genetic diversity and evolution, focussing on the role of different genotypes in viral pathogenesis.

Keywords: human cytomegalovirus; genotype; evolution; viral markers

1. Introduction

Human cytomegalovirus (HCMV; species human betaherpesvirus 5) is a member of the *Betaherpesviriniae* family and is widespread among mammals, establishing a lifelong infection in its specific host. HCMV infection is common worldwide, with the prevalence of specific antibodies ranging from 60% (in developed countries) to 90% in developing countries [1].

HCMV usually causes asymptomatic infections in immunocompetent individuals; however, primary infection can result in a mononucleosis-like syndrome [2]. In addition, several studies have suggested the role of HCMV infection in the development and/or severity of inflammatory cardiovascular diseases [3], certain types of cancers [4,5], and autoimmune diseases [6]. HCMV infection can result in severe disease in immunocompromised individuals. For example, HCMV infection is a significant clinical concern in patients undergoing immunosuppressive therapy, such as solid organ and stem cell transplant recipients, and in patients with acquired immunodeficiency syndrome (AIDS) [2,7]. HCMV is the most common congenital infection in both developed and developing countries, causing sensorineural hearing loss (SNHL) and neurodevelopmental delays [8].

We still do not fully understand the factors that determine the type, duration, and severity of symptoms caused by HCMV infections. The relationship between the infected host and the virus is likely to have significant implications. Current findings indicate the presence of a wide array of HCMV strains, highlighting the complexity of the virus's behaviour [9]. Researchers have been trying to determine how these different strains might affect the course of infection and development of CMV disease. In this review, we explore the genetic variations present in HCMV strains, with a specific focus on their classification and the implications of these differences on the function of the virus and its associated diseases.

check for updates

Academic Editor: Richard D. Dix

Received: 6 September 2024
Revised: 20 December 2024
Accepted: 24 December 2024
Published: 9 January 2025

Citation: Venturini, C.; Breuer, J. Cytomegalovirus Genetic Diversity and Evolution: Insights into Genotypes and Their Role in Viral Pathogenesis. *Pathogens* **2025**, *14*, 50. https://doi.org/10.3390/pathogens14010050

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. Landscape of Genetic Diversity

HCMV is a double-stranded DNA virus (dsDNA) and one of the longest human viruses (235–250 kb), with at least 170 canonical open reading frames [10,11]. It presents a standard herpesvirus class E genome architecture, with two unique regions (unique long UL and unique short US) that are flanked by a pair of inverted repeats (terminal/internal repeat long TRL/IRL and internal/terminal repeat short IRS/TRS), yielding the TRL-UL-IRL-IRS-US-TRS configuration [10,12,13].

2.1. HCMV Variability Is Mostly Found in "Islands" of Diversity

Genetic differences amongst HCMV strains were found over the entire genome, but some regions show a greater variability between strains compared to the rest of the genome, and they are known as "hypervariable". These variable regions did not show nucleotide changes randomly distributed throughout the sequence, but they strongly clustered into well-defined "genotypes", which are stable during the infection [9,14,15]. A genotype is defined as an individual's DNA sequence patterns in a given region or gene, and an allele is one of the possible versions of a DNA sequence [16]. Hypervariable regions with multiple genotypes are also defined as "multiallelic", as more than one allele is present. On the contrary, regions with only one allele are defined as "mono-allelic" rather than conserved, to consider the fact that variants can occur outside the hypervariable regions (Figure 1).

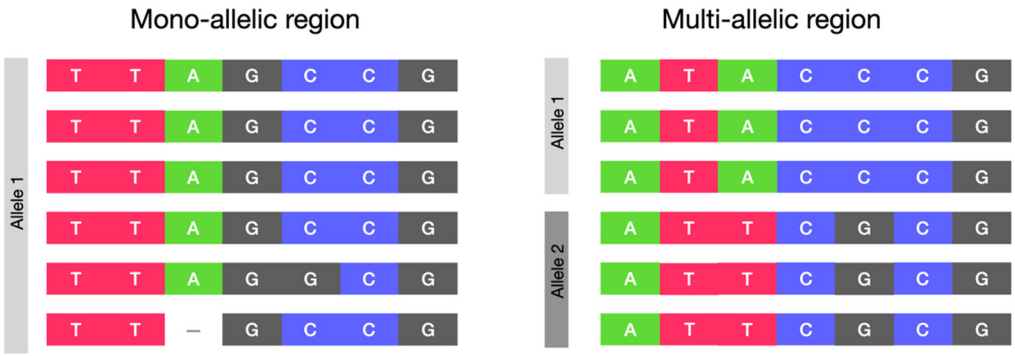


Figure 1. Schematic representation of mono-allelic and multi-allelic regions. Each row represents a different sequence.

The genes used for genotyping are shown in Figure 2. Some of these genes encode glycoproteins essential for the viral life cycle, such as glycoproteins gB (*UL55*), gH (*UL75*), gN (*UL73*), and gO (*UL74*). Other highly variable genes used for genotyping are interesting because they encode human cellular homologues, such as *UL146*, which encodes a viral CXCL chemokine, and the *UL144* gene, which encodes a TNF-receptor homologue. Members of the *RL11* gene family have also been extensively studied because they exhibit considerable variability across different strains. Notably, one member of this family, *RL13*, is highly prone to rapid mutation during in vitro culture [17]. Most genotyping studies have focussed on a few highly variable genes identified through polymerase chain reaction (PCR)-based genotyping [18–20]. Although whole-genome studies are becoming more popular, initial attempts involved isolating the virus in cell culture, making it more susceptible to gene loss and mutations. In addition, researchers have relied on the direct sequencing of PCR amplicons from clinical samples, which could introduce artefacts, due to the high number of PCR cycles [21–25]. Recent studies have overcome these limitations by using target enrichment to facilitate the direct sequencing of strains found in clinical samples [22,26,27]. Another issue in genotyping studies is the identification of genotypes or alleles. Genotyping for candidate genes has often been carried out visually and/or with phylogenetic trees, using a limited number of sequences [18,28–30].

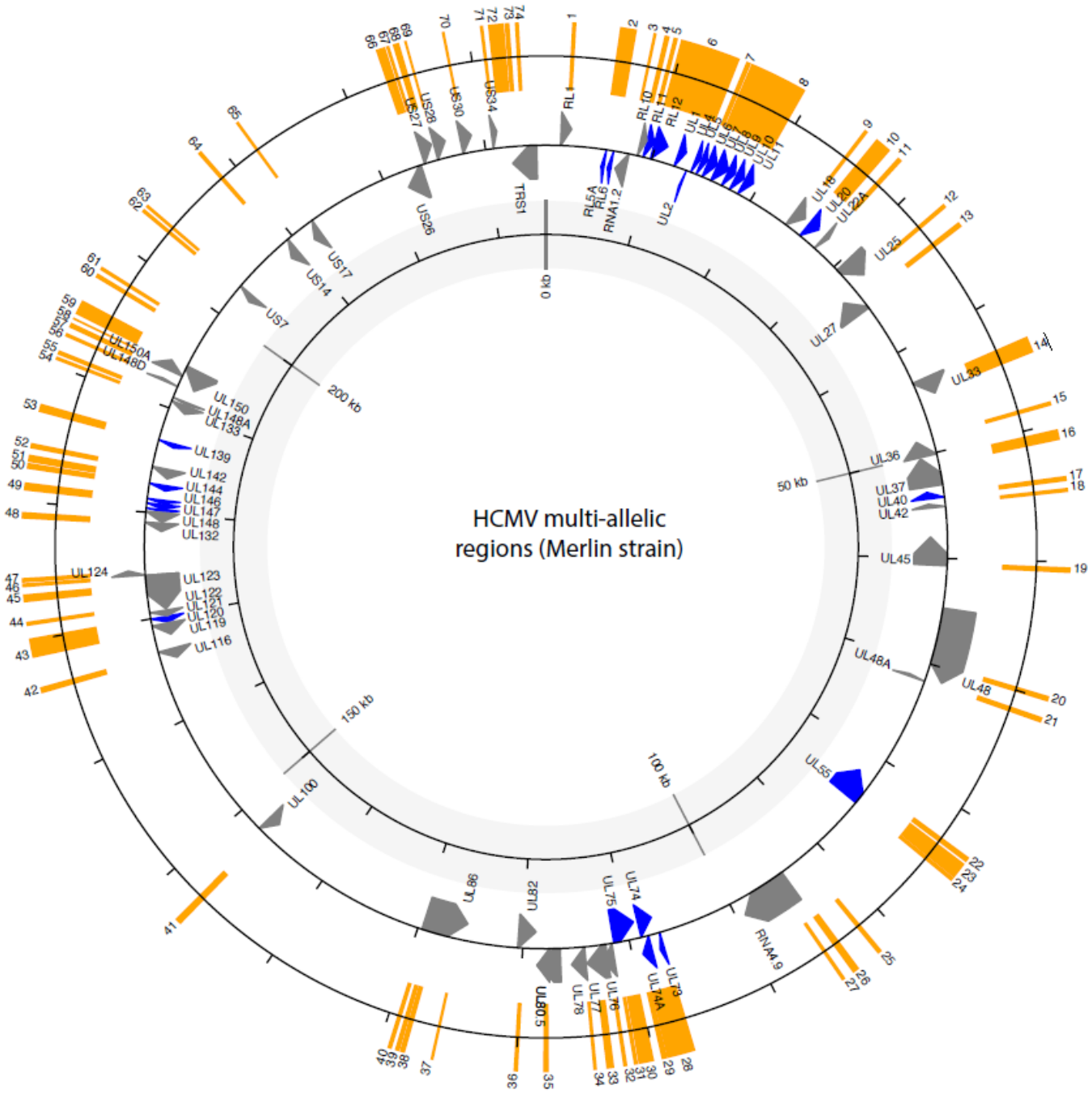


Figure 2. Summary of multi-allelic regions across the HCMV genome. The circle plot shows only the genes overlapping with the multi-allelic regions (in orange). In blue are genes previously used for genotyping.

Suárez et al. recently introduced a new method that addresses some of the previous limitations [31]. They developed genotype-specific motifs from 163 HCMV sequences and confirmed their validity in 243 *UL73* and *UL74* genomes. This approach was later extended to include ten other hypervariable genes (*RL5A*, *RL6*, *RL12*, *RL13*, *UL1*, *UL9*, *UL11*, *UL120*, *UL146*, and *UL139*). Genotyping is accomplished by tallying reads that contain motifs specific to the genotypes of hypervariable genes. Although this method stands out for its use of a large number of genomes and a standardised approach to genotype generation, it has only focussed on a small number of candidate genes identified as hypervariable, and is further limited by its consideration of the whole gene boundaries, rather than the actual

hypervariable region within the gene, as the units of variability and the definition of alleles by eye.

To better identify the boundaries of hypervariable regions and the number of alleles for each region, we utilised Hidden Markov Model (HMM) clustering within a comprehensive dataset of 253 HCMV genomes [32]. HMM can be used to determine the optimal number of sequence clusters or alleles that account for diversity across CMV genomes. Using an unbiased and probabilistic assignment model, our approach accurately identifies the nucleotide positions of multi-allelic and mono-allelic regions. Using this method, we described 74 multi-allelic regions with two to eight alleles each, comprising 14% of the genome, some of which were previously identified as hypervariable, but over 40 regions were novel [32]. These hypervariable regions were more evenly spread across the genome, providing more granular genotyping information.

2.2. CMV Hypervariable Genes

Previous analyses have mostly focussed on variable genes that encode four envelope glycoproteins essential for attachment, cell-to-cell spread, and interaction with the host's immune system: glycoproteins B, H, N, and O [15,33,34]. Owing to their crucial role in initiating signalling transduction cascades in target cells and propagating HCMV infection, glycoproteins have been identified as key HCMV vaccine targets. Interestingly, these glycoproteins often form complexes, facilitating HCMV infection; for example, the glycoprotein M/glycoprotein N (gM/gN) dimer [35], the glycoprotein H/glycoprotein L/glycoprotein O (gH/gL/gO) trimer [36], and the gH/gL/gO/UL128-130 pentameric complex [37].

Three other variable genes were identified in the UL/b region. Two are human homologues; the *UL146* gene encodes a viral CXCL chemokine [38] and the *UL144* gene encodes a TNF-receptor homologue [39]. Less studied, but in the same genomic region, *UL139* has also been reported to be variable and is predicted to encode a membrane protein [15].

Another subset of variable genes includes the *RL11* gene family, whose members encode known or putative glycoproteins [10]. *RL13*, a member of this family of genes, is known to mutate rapidly in vitro [17]. The apparent ease with which *RL13* mutants are selected during cell culture raises the possibility that these mutations pre-existed in the clinical sample, potentially reflecting an expanded cell tropism in vivo [40].

2.2.1. Glycoprotein B

gB is an HCMV envelope glycoprotein encoded by *UL55*, consisting of 907 amino acids (NCBI accession number: YP_081514). Five genotypes have been identified based on variations within the C-terminus, N-terminus, and gp55 cleavage site [28,41]. All five genotypes have been identified in different continents (Asia, Europe, and North America); however, their geographic distributions differ, with gB-1, gB-2, and gB-3 being the most prevalent genotypes in Europe [34,42,43]. In our study of HCMV diversity, we identified three multi-allelic regions (regions 22, 23, and 24, Tables 1 and 2). The first two corresponded to regions identified at the protein level. We also identified a novel region (22) that is conserved at the protein level but shows several changes in the nucleotide alignment.

When combined, the multi-allelic regions form 12 haplotypes, some more frequently than others (Supplementary Figure S1A–C). Interestingly, alleles in region 24, which overlap with the antigenic AD2 region, show geographical segregation between European and African sequences [32].

Table 1. Summary of multi-allelic regions identified in our study [32] previously used for genotyping. For each region, the table shows the number of alleles identified, open reading frames overlapping with the regions (genes), coordinates based on Merlin strain (NC_006273.2), geographic allele distribution for European and African sequences (this was calculated with Chi-square for independence and false discovery rate multiple testing correction, as described in [32]. Additional information has been added for linkage disequilibrium (LD), as found in [23]. A summary of previously identified variable regions overlapping with the multi-allelic region is provided: identified genotypes, most variable region of the genes identified, hypervariable regions overlapping with the multi-allelic region, as identified in [27] and geographic genotypes distribution if described). The deletion column is based on [25,27]. A summary of the possible link between HCMV alleles and functions is also provided.

Multi-Allelic Regions							**Genotypes Identified**					**Link Between HCMV Alleles and Function**	
Region	**Genes**	**Start**	**End**	**N of Alleles**	**Geographic Allele Distribution**	**LD**	**Previous Identified Genotypes**	**Most Variable Region**	**Hypervariable Genotypes Regions Suarez et al., 2019**	**Geographic Genotypes Distribution**	**Pseudogenes**	**Phenotypes**	**Disease**
1	*RL1*	1941	2121	2	No						Yes		
2	*RL5A RL6*	5387	6479	5	Yes	Yes			*RL5A*: 6; *RL6*: 7		Yes		
3	*RL9A*	7813	7914	2	No								
4	*RL10*	8620	8868	3	Yes								
5	*RL11*	9286	9479	2	No	Yes							
6	*RL11 RL12 RL13 UL1 UL2 UL4*	9840	14133	5	No	Yes	*UL1*: 3; *UL4*: 4 (Sekulin, 2007)		*RL12*: 10 (+subtypes); *RL13*: 10 (+subtypes); *UL1*: 10	Not described	*UL1, RL13, RL12*	• Putative transmembrane glycoproteins; • Not essential for viral growth in cell culture (Rawlinson, 1996); • *RL13* gene mutates rapidly when HCMV wild-type strains are cultured in different cell culture systems (Stanton, 2010); • Cell tropism (Stanton, 2010; Dolan, 2004).	
7	*UL5 UL6*	14765	14993	2	No	Yes	*UL6*: 4 (Sekulin, 2007)						
8	*UL10 UL11 UL6 UL7 UL8 UL9*	15163	19324	4	Yes	Yes	*UL7*: 3; *UL10*: 3 (Sekulin, 2007)		*UL9*: 9; *UL11*: 7		*UL9, UL11*		
10	*UL20*	25622	26757	3	No	Yes			7				
17	*UL40 UL41A*	53875	54131	2	No			Region encoding the HLA-E-binding peptide (residues 15–23 in AD169) (Heatley, 2013)			Yes	• Viral peptides derived from UL40 and presented on HLA-E are specifically recognised by the activating receptor NKG2C (Hamme, 2018); • UL40 polymorphisms may aid evasion of NK cell immunosurveillance by modulating affinity of the interaction with CD94-NKG2 (Hartley 2013); • NK cell response (Vietzen, 2021).	Not clear (Hartley 2013)

Table 1. *Cont.*

Multi-Allelic Regions						LD	Genotypes Identified				Pseudogenes	Link Between HCMV Alleles and Function	
Region	**Genes**	**Start**	**End**	**N of Alleles**	**Geographic Allele Distribution**		**Previous Identified Genotypes**	**Most Variable Region**	**Hypervariable Genotypes Regions Suarez et al., 2019**	**Geographic Genotypes Distribution**		**Phenotypes**	**Disease**
22	*UL55*	82720	83003	2	No								
23	*UL55*	83278	84403	3	No								
24	*UL55*	84532	84716	3	Yes	Yes	gB-5 genotypes (gB-1 to gB-5) (Wang, 2021)	Codons 26–70, gp55 cleavage site (codon 460)		All five genotypes have been detected in Asia, Europe, and North America; however, their distributions differ (Wang, 2022)		• Essential role in the replication cycle of the virus. Required for virus entry and cell-to-cell spread of HCMV (Isaacson and Compton, 2009); • Women immunised with gB-mF59 had better protection against primary infection with natural strains containing gB-1 compared to viruses with other alleles (Nelson, 2015, see also review Griffiths and Reeves, 2021).	Studies show inconsistent associations between gB genotypes and CMV disease severity or clinical manifestations (Pati, 2013; Yan, 2008; Tarrago, 2003). **cCMV Studies:** Research indicates no consistent link between gB genotypes and symptoms, sensorineural hearing loss, or neurodevelopmental outcomes in congenital CMV cases (Pati, 2013; Arav-Boger, 2002; Bale, 2000). Some studies suggest specific genotypes like gB-3 may be more prevalent, but findings vary (Yan et al., 2008; Dong et al., 2023). **Transmission and Clinical Outcomes:** Studies across HIV and transplant patients show gB genotypes do not correlate with clinical outcomes, though specific genotypes may be associated with complications in transplant recipients (Tarrago, 2003; Torok-Storb, 1997; Dieamant, 2013).
28	*UL73*	107059	109022	7	Yes		gN 4 genotypes gN-3 2 subtypes (Wang, 2021)	N terminal region	4 (+subtypes)	Not described		• Humoral immunity (neutralising response) (Shimamura, 2006); • Anti-gM/gN dimer antibodies possess differential neutralising activities against AD169, Toledo, and TR strains (Burkhardt, 2009; Pati, 2013).	Inconsistent findings (Arav-Boger, 2015) **cCMV Disease Studies**: • No association found between gN genotypes and symptoms or sensorineural hearing loss (SNHL) in cCMV babies (Pati, 2013); • Inconsistent results for gN-4 associated with chorioretinitis in symptomatic cCMV babies (Dong, 2023) and linked to symptoms at birth and sequelae (Pignatelli, 2010). **cCMV transmission:** All genotypes can be transmitted (Pignatelli, 2010). **Transplant Recipients:** No gN genotype was associated with a poorer outcome in solid organ transplant (SOT) recipients with CMV disease (Lisboa, 2012).

Table 1. *Cont.*

Multi-Allelic Regions						LD	Genotypes Identified				Pseudogenes	Link Between HCMV Alleles and Function	
Region	**Genes**	**Start**	**End**	**N of Alleles**	**Geographic Allele Distribution**		**Previous Identified Genotypes**	**Most Variable Region**	**Hypervariable Genotypes Regions Suarez et al., 2019**	**Geographic Genotypes Distribution**		**Phenotypes**	**Disease**
	UL74						5 (gO-1 to gO-5) + subtypes (Wang, 2021)	N-terminal region (codons 1–98), codons 270–313	5 (+ subtypes)	Differences in g) genotypes distribution in Japanese children vs. European samples (Wang, 2021; Yan, 2008)	Deletions at the N-terminus (in the first 90 aa) (Rasmussen, 2002)	• gO-4 genotype showed an increasing tropism for epithelial cells vs. gO-1 genotype (Brait, 2020; Kalser, 2017) • Different gO genotypes have an impact on neutralising antibody response to gH epitopes	
29	*UL75*	109129	109426	2	No	Yes	2	N-terminal region (codons 1–37)		Not described		• gH is the main antigen eliciting a neutralising antibody response (Wang, 2021; Urban, 1996; Freed, 2013:) • Cui, 2017: neutralising ability of certain gH-specific monoclonal antibodies was shown to be strain-specific in fibroblast and epithelial cells	**cCMV Disease Studies**: • No association between gH genotypes and cCMV symptoms or SNHL in cCMV babies (Arav-Boger, 2015; Pati, 2013); • gH-1 genotype was associated with hearing loss in symptomatic cCMV babies in Dong, 2023. **Transplant Recipients**: Renal transplant recipients with mismatched antibodies for gH had a higher incidence of acute transplant rejection and CMV disease (Ishibashi, 2007).
30	*UL75*	110100	111111	2	No								
31	*UL75*	111275	111445	2	No								
43	*UL119* *UL120* *UL121*	168817	170109	4	No	Yes			*UL120*: 4 (+subtypes)				
49	*UL146* *UL147*	180852	181323	8	Yes		14 (G1–G14) Bradley, 2008; Dolan, 2004)		*UL146*: 14	Not described	Deleted in highly passaged lab strains (Cha, 1996)	• Virulence (absent in attenuated strains; only in vivo), receptor binding affinity, signalling efficacy, chemotactic properties (Heo, 2008); • HCMV UL146/UL47 are alpha-chemokine genes and share size and sequence similarity with human alpha-chemokines (Arav-Boger, 2005); • vCXCL-1s differentially activate neutrophils and polymorphisms that affect the binding affinity, receptor usage, and differential peripheral blood neutrophil activation —> HCMV dissemination and pathogenesis (Ho, 2015).	**cCMV Disease Studies**: Inconclusive. • No association between genotypes and disease (Arav-Boger, 2006; Berg, 2021); • G1 more frequent in cCMV cases with CNS damage and hepatomegaly; G7 and G5 were predominant in postnatal CMV (pCMV) (Parawdoska, 2015); • Linked UL146 genotypes G1 and G13 to higher levels of IgG and IgM antibodies, as well as elevated liver enzymes, in babies with cCMV and hepatic involvement (Guo, 2016).

Table 1. *Cont.*

Multi-Allelic Regions							Genotypes Identified					Link Between HCMV Alleles and Function	
Region	**Genes**	**Start**	**End**	**N of Alleles**	**Geographic Allele Distribution**	**LD**	**Previous Identified Genotypes**	**Most Variable Region**	**Hypervariable Genotypes Regions Suarez et al., 2019**	**Geographic Genotypes Distribution**	**Pseudogenes**	**Phenotypes**	**Disease**
50	*UL144*	182416	182725	3	No	Yes	3 (Arav-Boger, 2015)			No	Deleted in highly passaged lab strains (Cha, 1996)	• Tumour necrosis factor alpha-like receptor; • Role in vivo (Cha, 1996).	**cCMV Disease Studies:** Controversial Findings (Arav-Boger, 2015). • Some associations with UL144 type C and symptoms (Pati, 2013) and type A and C with poor outcome in cCMV babies (Arav-Boger, 2002). Another study linked type B to higher enzyme levels in cCMV babies with liver involvement (Guo, 2016); • No association between UL144 types and cCMV disease (Nijman, 2014; Bale, 2001; Picone, 2004). **cCMV Transmission Studies:** All genotypes can be transmitted (Yan, 2008; Bale, 2001; Revello, 2008, Picone, 2004).
53	*UL139*	186573	187057	4	No		3-8 (Qi, 2006; Bradley, 2008)	N-terminal portion (Bradley, 2008)	8 (+subtypes)	Not clear (Bradley 2008)	Deleted in highly passaged lab strains (Cha, 1996)	Shared sequence homology with human CD24 (signal transducer modulating B-cell activation responses). G1c contained a specific attachment site of prokaryotic membrane lipoprotein lipid (Qi, 2006).	
66	*US26* *US27*	223336	223914	2	No	Yes	5				*US27*	Functional beta-chemokine receptor	No association with cCMV disease (Pati, 2013; Arav Boger, 2002)
67	*US27*	224108	224224	3	No								
68	*US27*	224607	224958	2	No								
69	*US27* *US28*	225456	225513	4	No								

Table 2. Multi-allelic regions in UL55.

Multi-Allelic Regions	Previously Identified Variable Regions in gB
reg 24: codons 24–85	codons 26–70
reg 23: codons 128–503	Codons 181–195; 311–317; gp55 cleavage site (codons 460).
reg 22: codons 595–689	Not identified.

Comparisons between previously identified genotypes and our haplotypes identified using HMM are shown in Table 3 and Supplementary Figure S1D. The phylogenetic tree (Supplementary Figure S1D), built with the protein alignment of representative sequences of different haplotypes and genotypes, showed the five clusters used to define the genotypes. However, it also revealed heterogeneity in some clusters, specifically in those identified as gB-2 and gB-4. Haplotype H2 clustered with gB-2, but it was also the closest cluster for H4, even though it had a different allele at region 24 (Supplementary Figure S1A). H1 and H7 clustered closely with gB-4 despite still showing some amino acid differences. The current genotyping system also does not capture H3 and H9, which cluster closer to gB-4 than other genotypes; however, they have a unique combination of alleles in regions 23 and 24. H5 (Merlin strain) and H6 (Towne strain) were classified as gB-1 and differed in region 22 of our analysis. Region 22 harboured many nucleotide changes, but was conserved in the protein alignments used to define the genotypes.

Table 3. Comparison between haplotypes identified in our study (H1–H10) and previously identified genotypes (gB1–gB5). H11 and H12 are not represented here because they were only identified in the reconstructed genomes from mixed infections, and protein sequences were unavailable. This table shows the GenBank accession identifiers for representative sequences for each haplotype H1–H10. Genotypes representative of protein sequences for gB1–gB5 were ACM48044.1, DAA00160.1, ADD39116.1, AAA45925, and AZB53144 [34].

Haplotype	Example Strain	Previously Identified Genotypes
H1	KY490079.1	gB-4
H2	FJ527563.1 AD169	gB-2
H3	KJ361956.1	gB-4
H4	KY490069.1	gB-2
H5	NC_006273.2 Merlin	gB-1
H6	FJ616285.1 Towne	gB-1
H7	KY490067.1	Separate cluster—closer to gB-4
H8	GU179289.1 VR1814	gB-3
H9	KY490088.1	Separate cluster—closer to gB-4
H10	KJ361971.1	gB-5

Many studies have investigated the possible link between gB genotypes and function. gB plays a critical role in CMV infection by mediating the final fusion event between viral and cellular membranes. gB fusogenic function is essential for viral entry and establishing infection [44,45]. Critically, gB is not only essential for viral entry but also for the cell-to-cell spread of the virus [44]. How the mechanisms of virus-to-cell entry and cell-to-cell fusion differ is unclear; however, several publications have reported gB-specific antibodies that are able to block one and not the other [46]. Samples collected from multiple body sites of the same individual showed different gB genotypes. One explanation is the presence of multiple HCMV strains in one host, due to a mixed infection. The partitioning of different

gB genotypes into different body compartments may suggest tropism for different cell types [24,47].

Genetic variations in the gB sequences should be considered when evaluating vaccine efficacy against primary infection, reinfection, or reactivation. The gB-mF59 vaccine is based on the Towne strain, which has gB-1. There is some evidence that women immunised with gB-mF59 have better protection against primary infection with natural strains containing gB-1 compared to with viruses with other alleles [48]. This, along with the finding that the conserved region of Towne gB predominantly carries Africa-segregating SNPs, emphasises the need to test vaccines based on this strain for cross-protective immunity against European strains.

2.2.2. Glycoprotein N

UL73 is a polymorphic locus that encodes viral glycoprotein gN. gN is a type 1 transmembrane protein composed of 135 amino acids (NCBI accession number: YP_081521). Four genotypes (gN 1–4) were identified based on the differences in the N-terminal region (codons 1–87). Two subtypes of gN-3 (gN-3a and gN-3b) and three subtypes of gN-4 (gN-4a, gN-4b, and gN-4c) were identified [49]. These four genotypes are widespread and have similar distributions in Europe, Asia, and North America, except for gN-2, which is primarily found in North America [49]. We identified a multi-allelic region (region 28) that overlapped with *UL73*. Interestingly, this region (including 1990 nucleotides) overlaps with *UL73* and *UL74*, probably because of the high linkage disequilibrium between the two genes. For region 28, we identified seven alleles corresponding to the previously identified genotypes and subtypes (Table 4, Supplementary Figure S2).

Table 4. Comparison between the alleles identified in our study for region 28 (A1–A7) and previously identified genotypes for gN. The GenBank accession identifiers for representative sequences are shown in this table. Genotypes' representative protein sequences were taken from this review [34]. The GenBank accession numbers of published reference genotype sequences are as follows: gN-1 (AD169 strain, P16795.1), gN-2 (Can 2 strain, AAL77763.1), gN-3a (PS strain, AAL77773.1), gN-3b (A8–27F strain, AAO24841.1), gN-4a (ZV strain, AAL77779.1), gN-4b (Towne strain, AGT36491.1), and gN-4c (Toledo strain, AAS48964.1).

Allele	Example Strain	Previously Identified Genotypes	Frequency in Europe	Frequency in America	Frequency in Africa
1	KY490061.1	gN-3a	20.5%	9.1%	23.3%
2	KY490065.1	gN-3b	8.4%	27.3%	3.3%
3	FJ616285.1 Towne	gN-4b	11.6%	18.2%	66.7%
4	NC_006273.2 Merlin	gN-4c	11.6%	0	0
5	KY490062.1	gN-4a	22.8%	0	0
6	FJ527563.1	gN-1	13.5%	9.1%	6.7%
7	KJ361956.1	gN-2	11.6%	36.4%	0

Studies investigating the role of different genotypes on gN function have focussed on humoral immunity to identify strain-specific neutralising activity. Indeed, anti-gM/gN dimer antibodies showed different activities against the AD169, Toledo, and TR strains [50]. This was confirmed in recombinant virus studies, in which four different gN genotypes were reconstructed in the AD169 virus backbone. Viruses with different genotypes were neutralised differently, suggesting that variability in gN could contribute to the evasion of an efficient neutralising antibody response [51,52].

2.2.3. Glycoprotein O

Envelope glycoprotein O (gO) is a soluble protein encoded by *UL74* and is an essential component of the gH/gL/gO trimer. This complex is required for HCMV entry into host cells [53]. gO comprises 457–472 amino acids (NCBI accession number: YP_081522.1) depending on the number of strain-specific deletions. The main variable region is found at codons 1–98 (overlapping with the N-terminal region and where deletions might occur) [54], and minor variation is found at codons 270–313 [34]. This variation has led to the identification of five gO genotypes (gO-1 to gO-5) in HIV-positive and HIV-negative immunocompromised patients [54] and renal transplant recipients [55]. Further analysis identified subtypes of gO-1 (gO-1a, gO-1b, and gO-1c) and gO-2 (gO-2a and gO-2b) genotypes [20,56]. All five genotypes have been identified in Europe, North America, and East Asia. Recently, we identified seven alleles in European sequences, four of which were also found in African genomes, with a significant difference in distribution [32]. Our allele classification reflected the different genotypes and subtypes identified in previous studies (Table 5, Supplementary Figure S3). The only difference was that the subtypes gO-1a and gO-1c were both classified as A1 using our method. This could be due to the selection of sequences used for our HMM methods, where all wild-type strains available in GenBank were included, but not all lab-passaged strains were included.

Table 5. Comparison between the alleles identified in our study for region 28 and previously identified genotypes of gO. The GenBank accession identifiers for representative sequences are shown in this table. Genotypes representative of protein sequences were taken from this review [34]. GenBank accession numbers of published reference genotype sequences: gO-1a (AD169 strain, ACL51143.1), gO-1b (Cincy 2strain, ACS93309.1), gO-1c (Toledo strain, AAS48965.1), gO-2a (FUK19U strain, ABY48952.1), gO-2b (SW1102 strain, AAN40063.1), gO-3 (SW5 strain, AAN40074.1), gO-4 (Towne strain, AGT36493.1), and gO-5 (Merlin strain, YP_081522.1).

Allele	Example Strain	Previously Identified gO Genotypes	Frequency in Europe	Frequency in America	Frequency in Africa
1	KY490061.1	gO-1b	20.5%	9.1%	23.3%
2	KY490065.1	gO-2a	8.4%	27.3%	3.3%
3	FJ616285.1 Towne	gO-4	11.6%	18.2%	66.7%
4	NC_006273.2 Merlin	gO-5	11.6%	0	0
5	KY490062.1	gO-3	22.8%	0	0
6	FJ527563.1 AD169	gO-1a	13.5%	9.1%	6.7%
7	KJ361956.1	gO-2b	11.6%	36.4%	0

Our study observed that the gO and gN genes are located within the same multi-allelic region, designated as region 28. This co-location can be attributed to their close spatial proximity and the strong degree of linkage disequilibrium between the two genes [23,57].

Several studies have investigated the association between different gO genotypes and function. HCMV recombinant studies, in which different gO genotypes were reconstructed into the TB40E HCMV strain backbone, have shown that some genotypes elicit increased tropism for epithelial cells [58,59]. They also suggested that variability in gO can have a dramatic impact on cell-free and cell-to-cell spread, as well as antibody neutralisation [60].

2.2.4. Glycoprotein H

Another key HCMV glycoprotein with substantial genetic diversity is glycoprotein H, which is encoded by *UL75*. gH is an essential component of the gH/gL/gO trimer and

is part of the pentameric complex. In addition to gB, the pentameric complex is a vaccine target because of its role in entering epithelial, endothelial, and monocytic cells [34,61].

gH is 742 amino acids long (NCBI accession number: YP_081523.1) and is less variable than the other glycoproteins described. Two genotypes have been identified based on the N-terminal region (codon 1–37), both of which are found in Asia, Europe, and North America [34]. Although *UL75* is typically regarded as less variable than other glycoproteins, our research revealed the presence of three multi-allelic regions within the gene (regions 29, 30, and 31) (Tables 1, 6 and 7).

Table 6. Multi-allelic regions in *UL75*.

Multi-Allelic Regions	Previously Identified Most Variable Regions in gH
reg 31: codons 3–60	codons 1–37
reg 30: codons 114–451	Not identified.
reg 29: codons 676–742	Not identified.

Table 7. Comparison between haplotypes (H1–H6) identified in our study for regions 29, 30, and 31, and previously identified genotypes for gH (gH1 and gH2). The GenBank accession identifier for the representative sequences of each haplotype is shown in this table. Genotype representative protein sequences were taken from this review [34]. The GenBank accession numbers of published reference genotype sequences are gH-1 (AD169 strain, ACL51144.1) and gH-2 (Towne strain, AGT36494.1).

Allele	Strain	Genotype	Frequency in Europe	Frequency in America	Frequency in Africa
H1	KY490061.1	gH-1	26.05%	0	30%
H2	NC_006273.2 Merlin/FJ616285.1 Towne	gH-2	27.3%	40%	40%
H3	FJ527563.1 AD169	gH-1	20%	9.1%	13%
H4	JX512206.1	hybrid	1.86%	0	13%
H5	KJ361946.1	gH-2	10.70%	36.4%	0
H6	KP745640.1	hybrid	1.40%	27.3%	0

Two alleles were identified in similar proportions in the European and African sequences of each region (Supplementary Figure S4B) [32]. The two genotypes, gH-1 and gH-2, partially overlapped with our two alleles in each region (Supplementary Figure S4C). However, after combining these three regions, we obtained six haplotypes (H1–H6) (Supplementary Figure S4A), probably because of recombination. Because two of these regions (regions 30 and 31) produce changes in the protein, this data analysis method allows us to assess the functionality of the interactions between different alleles.

gH is considered to be one of the main antigens for eliciting neutralising antibody responses [62]. This response appears to be strain-specific in fibroblasts and epithelial cells [63]. Cui et al. have shown that sequence SNPs within residues 27–48 (region 31) govern both the binding and neutralisation of virus entry into epithelial cells and fibroblasts. In addition, it has also been reported that the T-helper cell response to gH is strain-specific (aa 284–302, overlapping with our region 30) [64].

2.3. Viral Cytokine/Chemokine Proteins (Human Cellular Homologues)

2.3.1. *UL144*

UL144 is a tumour necrosis factor-α (TNF-α)-like receptor gene [10,65].

The sequence variability of *UL144* was initially reported in 45 low-passage HCMV clinical samples from congenitally infected infants, where three major genotypes were

identified [65]. These three genotypes were confirmed by further analyses [14,66,67]. Our study identified a multi-allelic region overlapping *UL144* (4–107 codons) with three alleles corresponding to the genotypic groups identified previously (Table 8, Supplementary Figure S5B) [32]. The three alleles are present in Europe, America, and Africa, although they have slightly different prevalence rates (Table 8, Supplementary Figure S5A).

Table 8. Comparison between alleles identified in our study for region 50 and previously identified genotypes for *UL144*. The GenBank accession identifiers for representative sequences are shown in this table. Genotypes' representative protein sequences were taken from Lurain et al. [65]. The GenBank accession numbers of the published reference genotype sequences are as follows: group 1, AAF13363.1; group 2, AAF09111.1; and group 3, AAF09096.1.

Allele	Strain	Genotype	Frequency in Europe	Frequency in America	Frequency in Africa
A1	NC_006273.2 Merlin	Group 1	39.5%	9%	60%
A2	FJ616285.1 Towne/MF084224.1	Group 3	45.6%	55%	23.3%
A3	KY490064.1	Group 2	14.9%	36%	16.7%

UL144 likely plays multiple roles in regulating immunity to HCMV infection.

Molecular mimicry of cytokines and cytokine receptors is a strategy HCMV uses to modulate host immunity. The *UL144* gene, found in the UL/b' region of the HCMV genome, has amino acid sequence similarity with members of the tumour necrosis factor receptor superfamily [68], and helps HCMV to evade the host immune system by inhibiting T cell activation, by binding to the B and T lymphocytes attenuator (BTLA) [69]. UL144 is a potent activator of NF-κB via a TRAF6-dependent mechanism. This activation enhances the expression of the chemokine CL22 through NFκB-responsive elements found in its promoter [70]. In addition, UL144 can also be anti-inflammatory by evading the CD160-mediated activation of NK cells [71].

Extensively passaged laboratory strains lack the UL/b' region and do not encode UL144. The genes in this specific region (designated ULb') are not deemed necessary for growth within fibroblast cell cultures [72].

2.3.2. *UL146* and *UL147*

HCMV encodes two genes, *UL146* and *UL147*, whose protein products (vCXCL-1 and vCXCL-2) exhibit limited identity with CXC-chemokines [38]. Both genes show a consistently extreme variability of over 60% [29,66] at the amino acid level. Fourteen distinct *UL146* and *UL147* genotypes were identified (G1–G14) based on the variation throughout *UL146* and within a small region of *UL147* corresponding to a possible signal peptide [10,29]. Minor differences in genotypic frequencies have been identified among continents (Africa, Australia, Asia, Europe, and North America), but there has been no clear geographical separation [73]. Our study confirmed the high variability between strains, identifying eight alleles in a region overlapping *UL146* and *UL147*, with different frequencies between European and African sequences (Table 9, Supplementary Figure S6A). Our eight alleles corresponded to the eight most common genotypes (2, 5, 7, 8, 9, 11, 12, and 13) (Table 9, Supplementary Figure S6B). The absence of genotypes 1, 3, 4, 6, and 10 in our analysis is due to the unavailability of complete genome sequences for these genotypes in GenBank. Most studies describing these genotypes focus on sequencing specific genes of interest rather than the entire genomes. Since our method relies on whole-genome sequences, these genotypes could not be included.

Table 9. Comparison between alleles identified in our study for region 49 and previously identified genotypes for *UL144*. The GenBank accession identifiers for representative sequences are shown in this table. Genotypes' representative protein sequences were taken from Lurain et al. [29]. The GenBank accession numbers of published reference genotype sequences: 1 AAZ91734.1; 2 AAZ91735.1; 3 AAZ91718.1; 5 ABA02110.1; 7 AAZ91719.1; 8 AAZ91724.1; 9 AAZ91727.1; 10 ABA02161.1; 11 AAZ91729.1; 12 ABA02131.1; 13 ABA02122.1; and 14 ABA02092.1. Genotypes 4 and 6 have not been identified by Lurain et al. [29].

Allele	Strain	Genotype	Frequency in Europe	Frequency in America	Frequency in Africa
A1	KY490068.1	13	13.49%	27.3%	3.3%
A2	KY490084.1	7	2.33%	18.2%	3.3%
A3	MK290742.1	12	17.67%	9.1%	20%
A4	KY490088.1	9	12.09%	0	50%
A5	MK290743.1	5	13.95%	9.1%	13.3%
A6	MF084224.1	8	16.28%	9.1%	0
A7	NC_006273.2 Merlin	2	13.02%	18.2%	0
A8	KY490067.1	11	11.16%	9.1%	10%

Although both proteins are potential homologues for CXC chemokines, only the functional effects of pUL146 (vCXCL-1) have been reported in detail. Even though there is limited homology with the host, the UL146 acts as a functional CXC chemokine that binds to CXCR1 and CXCR2, and induces neutrophil chemotaxis and calcium mobilisation [38,74,75]. Disruption/deletion of the *UL146* gene from high-passage lab strains limited the ability of HCMV-infected fibroblast to promote neutrophil chemotaxis [74]. Early vaccine trials suggested the role of xCXCL-1 in infections in vivo, where the Toledo strain was found to be more virulent than the Towne strain [76].

2.4. Other Hypervariable Regions

Several other genes also have multiple alleles. Table 1 presents a complete list of all the multi-allelic regions and genes. *UL139*, for example, has been described in several studies and three to eight alleles have been identified [31,32,73,77]. The protein encoded by each HCMV *UL139* genotype contained a putative signal peptide sequence and a transmembrane region. A limited region of sequence identity (15 amino acids) has been identified between HCMV UL139 and human CD24, a highly glycosylated protein involved in B cell activation that is overexpressed in cancer [78,79]. While this raises the possibility that *UL139* may encode a CD24 homologue, the evidence remains preliminary, and further research is required to substantiate this claim.

Another subset of hypervariable genes is included in the *RL11* domain. The *RL11* gene family includes *RL11-RL13*, *UL1*, *UL4–11*, *RL6*, and *RL5A*, located near the genome's N-terminus. Previous studies have identified two to five genotypes for some of these genes, with *UL1* exhibiting the highest degree of variation [32,80]. These genes are believed to encode putative transmembrane glycoproteins that are not essential for viral growth in cell culture. They are also absent in murine CMV [81] and *UL1* is absent in chimpanzee CMV [82]. A recent study looking at *RL11* evolutionary history confirmed that these genes are unique to Old World monkeys and Great Apes CMVs, and suggested that some human CMV-specific *RL11* genes emerged before the divergence of humans and chimpanzees, but were subsequently lost in the latter [83].

Some of the multi-allelic regions identified in our study were novel (n = 49) (Supplementary Table S1). Of these, 26 showed alleles with a different geographical preva-

lence, such as tegument proteins (*UL48* and *UL82*), capsid proteins (*UL86, UL48A*, and *UL80*), and envelope glycoproteins (*UL37* and *UL100*). The rest showed the same proportion of alleles in European and African sequences and included several membrane and envelope glycoproteins (*UL18*, *UL33*, *UL132*, *UL142*, and *US7*), tegument proteins (that is, *UL25*, *UL36*, *IRS1*, and *TRS1*), and membrane proteins (that is, *US14* and *US17*). Interestingly, five multi-allelic regions were identified in genes with no well-defined functions (*UL27*, *UL41/UL42*, *UL116*, *UL133*, and *UL148A/UL150A*), which contain potential transmembrane domains or signal peptides, and five non-coding regions overlapping with repeats (TATA box) and regulatory RNAs (Supplementary Table S1).

2.5. Outside the Hypervariable Genes

Although SNPs are typically located in multi-allelic regions, some can be found elsewhere in the genome. We identified several SNPs (n = 440) in the mono-allelic portion of HCMV, which showed geographical segregation between European and African sequences. Only 15% of these SNPs were non-synonymous. Of these, 16% overlapped with known B and T cell epitopes, providing little evidence that the geographic population structure in CMV, unlike EBV [84], is driven by a unique host immune pressure [32].

Other types of SNPs outside multi-allelic regions can appear during infection in a host. For example, administering antiviral drugs to treat HCMV viraemia disrupts the viral population, leading to the selection of drug-resistant variants [85]. Recent studies by our group and others [86,87] have shown that antiviral resistance variants can be found in transplant recipients with CMV viremia. These variants are especially prevalent in DNA polymerase *UL54* and protein kinase *UL97*, which are the primary targets for common antiviral drugs such as ganciclovir, cidofovir, and foscarnet [88,89]. Additional drug targets have been identified in *UL51*, *UL56*, and *UL89* for letermovir and *UL27* for maribavir [88,90,91] (our summary of genes with antiviral resistance variants can be found here http://cmv-resistance.ucl.ac.uk/herpesdrg/, accessed on 26 July 2024). All but one gene was found in the mono-allelic regions. *UL27*, an HCMV gene of unknown function that confers low-level resistance to maribavir [92], is the only gene with drug-resistance variants overlapping with a small multi-allelic region of 275 nt (nucleotides) (*UL27* is 1827 nt in Merlin strain NC_006273.2) (Table 1) [32].

2.6. Clinical Mutants with Nonfunctional Genes (Pseudogenes)

The genomes of some HCMV strains exhibit disruptions in their open reading frames (ORFs), which can lead to "pseudogenes". This results from mutations that cause premature translational termination, such as SNPs introducing in-frame stop codons, splice sites, or structural variations (insertions and deletions), leading to frameshifting or a loss of protein-coding regions. In contrast to significant gene loss in highly passaged lab strains, more subtle mutations leading to pseudogenes have also been found in strains isolated from clinical samples. Sijmons et al. [25] showed that 75% of clinical strains are not genetically intact, but contain disruptive mutations in a diverse set of 26 genes. Only one out of four clinical isolates has the complete set of intact genes, with the other isolates having one (33%), two (27%), three (13%), or four (3%) mutated genes. None of these 26 genes are essential for the growth of fibroblast cells. Interestingly, most overlapped with multi-allelic regions (Table 1), except for *US9*, *UL111A*, *UL128*, *US13*, *UL136*, *UL30*, *UL145*, *US6*, and *US12*. However, all but one of the strains used in this study were passaged in cell culture. Therefore, some mutations might be the artefact of culture adaptation. More recently, Suárez et al. demonstrated that the distribution of pseudogenes in 91 strains sequenced directly from clinical materials was similar to the previous study. The most frequently mutated genes were *UL9*, *RL5A*, *UL1*, and *RL6* (members of the *RL11* family); *US7* and *US9*

(*US6* gene family); and *UL111A* (encoding viral interleukin 10) [27]. It is not clear what the impact of pseudogenes is on the phenotype. Some pseudogenes originate from genes involved in immune modulation, such as *UL111A*, *UL40*, and *UL9*. *UL111A* encodes a viral interleukin 10 homologue, cmvIL-10. cmvIL-10 can bind to the human IL-10 receptor and compete with the human IL-10 for binding sites, despite the two proteins being only 27% identical [75,93,94]. The UL40 protein in human cytomegalovirus (HCMV) plays a crucial role in modulating the immune response, particularly in evading natural killer (NK) cell-mediated cytotoxicity. It achieves this by interacting with the HLA-E molecule, which is a ligand for NK cell receptors [95,96]. Interestingly, certain *UL40* variants are associated with higher levels of viremia and can affect the proliferation and activation of NK cells differently, impacting the clinical outcomes in transplant recipients [97–99]. *UL9* was predicted to be an immunoglobulin-binding domain [33]. However, the sample size was limited, and its involvement in pathogenesis is still speculative. Further studies are required to investigate the presence of pseudogenes in HCMV samples from different types of individuals (immunocompetent and immunocompromised) and tissues. The timing of deletions and their evolution over time also need to be investigated.

2.7. Repeats

Another source of sequence variation is the heterogeneity in the copy number of adjacently repeated elements or tandem repeats (TRs). Short tandem repeats, also known as "microsatellites", are patterns of short motifs consisting of one–six bases.

Several studies have found that TR variations may affect the functionality and pathogenicity of viruses [100–102]. TRs in HCMV have been previously described, where insertion and deletion polymorphisms can differentiate between different viral strains and can be used as epidemiological markers [102,103]. Many of these microsatellites are found in non-coding regions. However, some are also found in coding areas, promoters, and other functional DNA, such as oriLyt (origin of DNA replication) [104].

Studying these microsatellites can provide valuable insights into the genetic variability, viral evolution, and gene regulation of viruses. Further research is needed, and long-read sequencing will help to reconstruct repeat regions [105].

3. Within Host Diversity

Multiple infections with diverse HCMV genotypes (also referred to as "mixed infections") are common and have been documented in various patient cohorts, including those with intact immune systems [9].

In immunocompetent individuals, the presence of mixed infections with multiple HCMV strains was first identified in women attending a sexually transmitted disease clinic, as reported by Chandler et al. [106]. Subsequent studies have confirmed the existence of multiple gB and gN genotypes in samples from other immunocompetent populations, such as adults analysed post mortem [107], healthy children [108,109], and seropositive healthy women [110,111]. These studies suggest that reinfection with different HCMV genotypes is common throughout a person's lifetime. This phenomenon has significant implications for pregnant women, the risk of congenital infections (see "Clinical Significance of Multi-Allelic Regions" section), and vaccine development, highlighting the importance of understanding the frequency and impact of reinfection [112].

Mixed infections are extremely common in transplant recipients [56,86,87,113–117]. Several studies have revealed a wide frequency of mixed infections in these patients, ranging from 15% to 90% [9]. However, this difference is likely due to variations in the patient populations and the range of methods used for genotype analysis. Mixed infections have been associated with poor outcomes [118,119]. However, more recent studies using

whole-genome sequencing have found no association between multiple-strain infection and particular virological or clinical features, including mortality [87,115].

Multiple strains in clinical samples significantly overestimate the HCMV genome variability within an individual host [24,47]. The detected variants primarily represent genetic differences between strains, rather than the evolution of a single strain within the host [86]. When considering mixed infections, several studies [86,87,114,116] have shown that in infections with a single strain, the HCMV genome is highly stable in patients over time and during different reactivation episodes.

Infections involving multiple HCMV strains may disrupt genome stability by providing opportunities for homologous recombination [86,114,120], which plays a significant role in generating CMV diversity. Recombination has been demonstrated in a laboratory setting when two HCMV strains infect the same cell and interact during replication. This interaction generates progeny, whose genomes consist of new haplotypes formed by a mix of genotypes/alleles obtained from both parental strains. This process leads to the creation of a greater variety of haplotypes of different genotypes [121,122]. The occurrence of recombination in vivo is also supported by CMV sequencing in immunocompromised patients [114,123,124].

Many genotyping methods have been used to detect multiple strains in a sample. Most of these methods are PCR-based, and assume that sequence diversity occurs in a limited number of alleles/genotypes [9]. However, these studies only focussed on specific genes and did not allow for the detection of low-abundance alleles. As previously discussed, Suarez et al. [31] developed a method that uses whole-genome sequencing to detect genotype-specific motifs in 12 hypervariable genes. These motifs can be used to detect infections with multiple strains. However, this method uses only 12 genes and does not allow for the reconstruction of the entire genome. Reconstructing the whole genome is helpful, as it identifies which genotyped regions are part of the same genome, thus providing the potential to study epistatic interactions. Therefore, we developed a HaROLD in our laboratory. This programme uses a probabilistic framework to reconstruct the genomes of mixed infections and performs validation with high accuracy, using simulated and real mixtures of HCMV genomes [125]. Using HaROLD to reconstruct the whole genome allows for the accurate identification of the sequences of different CMV viruses present in one sample. This can then be used to explore viral dynamics over time, and identify within-host recombination [87,114].

4. Clinical Significance of Multi-Allelic Regions

Finding evidence supporting the connections between different HCMV alleles and pathology has been challenging. Multiple studies have investigated the link between HCMV genotypes of different genes, the presence or absence of disease, the severity of clinical symptoms, and transmissibility [9,14,15,66]. This type of study is useful because it could help identify prognostic factors for the likelihood and severity of disease in clinical settings.

HCMV mainly affects people with weakened immune systems and newborns, and most studies have focussed on these two groups of patients. In healthy individuals, primary infections typically do not exhibit symptoms, or only result in mild infections. Consequently, there has been limited focus on understanding the relationship between different HCMV genotypes and the likelihood or seriousness of primary infections in healthy individuals [9].

Infections during pregnancy are more concerning. HCMV is the most common infectious cause of congenital, acquired disability, ranging from sensorineural hearing loss to severe neurocognitive impairment [126]. Primary maternal infection during pregnancy confers a 30–40% risk of transmission to the foetus [126]. However, maternal immunity to HCMV before gestation does not prevent transmission to the foetus, and even women

with long-standing immunity to HCMV can shed and transmit the virus [112,127]. In contrast to other congenital infections, such as rubella, parvovirus, and toxoplasmosis, the highest rate of congenital HCMV infection is found in populations in which women of childbearing age have the highest prevalence of serological immunity to HCMV, such as in Africa, Asia, and South America [112]. The reactivation of HCMV, but also reinfection with antigenically distinct strains, is also possible in immunocompetent women [128] and has also been observed in rhesus macaque models (RhCMV), where macaques with robust pre-existing adaptive immunity could be readily reinfected with another wild type or lab strain [129].

Researchers have studied how certain HCMV genotypes are transmitted in the womb, with a focus on gB and *UL144*. In short, viruses transmitted from mother to child share the same genetic sequences, and placental transmission appears to be independent of a specific viral strain [66]. Several studies have shown that all gB and *UL144* genotypes can be transmitted from mothers to foetuses and that the distribution of genotypes in infants infected before and after birth is similar [67,130–133]. Similar results were found for other genes investigated, such as gN and *UL149* [133,134].

Models with RhCMV looking at gB and gL have been used to investigate CMV viral populations in intrauterine transmission [135]. Rhesus macaques were inoculated with a mixture of the three strains. One of the three strains in the inoculum was dominant in all maternal and foetal CMV samples. However, the viral populations were still diverse, with minor haplotypes related to the dominant strain. These were consistently detected within the samples' maternal tissues at multiple time points, indicating persistence over time and transmission between different maternal compartments. Some maternal haplotypes were also present in foetal and maternal–foetal interface tissues, supporting the idea of a mother-to-foetus transmission bottleneck [135]. Multiple closely related haplotypes were also identified in a small study from our group in five HIV-infected mothers, with compartmentalisation of CMV populations between the cervix and breast milk [136]. Babies were initially infected with one strain, but they commonly acquired a different strain from breast milk. In congenitally infected infants, the viruses that passed from mother to baby were similar to the dominant strain in the cervix, but had specific genetic differences compared to the strains in the breast milk and cervix of mothers whose babies were infected after birth. These genetic differences were found in 19 genes, notably in members of the *RL11* family, *UL40*, *UL74*, and *US27/US26* [136].

The second question is whether HCMV genotypes are associated with symptomatology at birth and neurological sequelae in congenital HCMV infections. Studies aiming to identify a viral marker of CMV disease outcomes have caused more debate than reached a consensus [66]. Most studies have focussed on gB and *UL144*; however, the results are inconsistent. Several studies have analysed the connection between gB and symptoms at birth and neurological consequences, such as hearing loss. However, none of the gB genotypes reliably predicted congenital CMV outcomes. The few studies that have found such a correlation are contradictory. For example, the gB-3 genotype was more prevalent in congenitally infected Japanese babies, especially those with SNHL [133]. Recently, Dong et al. [137] found that gB-3 was associated with a higher risk of skin petechiae in 42 cCMV-symptomatic infants than in 140 babies with postnatal infection, but they did not find a correlation with hearing loss. Furthermore, Bale et al. found that gB-3 was the most common genotype in babies from Iowa and common in asymptomatic infections, and they found no correlation with neurodevelopmental outcomes [67]. The results for *UL144* are also controversial. Some studies have found associations between types A and C of *UL144* and symptoms and unfavourable outcomes of CMV infection [52,138]. However, in cCMV infants with hepatic involvement, type B is associated with higher levels of hepatic

enzymes [139]. In contrast, other studies did not find any link between *UL144* types and disease [67,131,132].

The results of investigations of other glycoproteins indicate no correlation between gN and gH genotypes and disease [52,66]. In contrast, Pignatelli et al. [134] monitored 74 congenitally infected newborns for symptoms of CMV disease at birth and during long-term follow-up, and revealed that newborns with symptoms at birth, abnormal imaging results, and sequelae were associated with the gN-4 genotype. Additionally, the genotype gN-4 has been linked to chorioretinitis in 42 cCMV Chinese babies [137]. The same study also found a relationship between gH-1 and hearing loss, although this finding lacks the support of other studies.

Some studies have shown a possible connection between the *UL146*/*UL147* types (G5 and G7) and symptomatic congenital CMV infection. G1, on the other hand, is more common in children with CNS damage and hepatomegaly [140]. G1 and G13 have also been linked to higher levels of IgG and IgM, and increased levels of hepatic enzymes in babies with cCMV and hepatic involvement [139]. However, other studies have shown no association between *UL146*/*UL147* genotypes and clinical manifestations of congenital infections [12,141].

Various factors have hindered the establishment of a clear relationship between the HCMV alleles and disease development. First, most studies were based on limited descriptions of alleles and genotypes. For example, most studies have focussed on only one gene at a time without considering that, due to the high degree of recombination, there are many possible allelic combinations (haplotypes) [23]. However, the link between haplotypes and clinical outcomes has not been adequately investigated. More comprehensive descriptions of multi-allelic regions [31,32] will likely help in designing better studies to investigate the link between alleles and HCMV clinical manifestations. Investigating the entire genome has proven to be challenging, owing to the limited sample sizes of these studies [66], and the problem of multiple testing and collective effort will be needed to overcome this. Another factor is the relationship between HCMV strains and the genetic background of the host. For instance, the host's HLA status could impact the recognition of specific peptides for HCMV alleles, potentially influencing an individual's immunological response [63,109].

Previous studies have explored the potential correlation between specific HCMV genotypes, primarily gB genotypes, and the clinical manifestation or severity of disease in immunosuppressed patients. However, only a limited number of these investigations have suggested a possible link between certain gB genotypes and disease severity in transplant recipients or individuals with AIDS [136–138]. A more recent study of 59 patients treated with allogeneic hematopoietic stem cell transplant has found that specific gB genotypes can have beneficial (i.e., earlier engraftment) and adverse effects (i.e., shorter overall survival) on the transplant host. Further studies are needed to validate these findings.

5. Evolution of Diversity

Despite the significant variability among HCMV strains, it is remarkable that a limited number of alleles can be defined for each multi-allelic region. Although the alleles are highly diverse, there is a much lower level of genetic variation within individual alleles. In addition, HCMV sequences are stable over short timescales in patients with single infections, indicating that alleles do not change over time, even in immunocompromised patients with chronic infections [55,86,87,130]. This is also true for cell culture [29]. Taken together, these findings suggest that alleles have a long history, perhaps emerging during the evolution of populations of early humans or their predecessors [10,73].

Geographical segregation has been described for other human herpesviruses, such as herpes simplex virus (HSV-1) [126], varicella-zoster virus (VZV) [103], and Epstein–Barr

virus (EBV) [31]. However, most HCMV studies have indicated genetic similarity within viral sequences, regardless of geographic location [9,23,25,117]. In contrast, our recent findings suggest a distinct geographical separation of CMV genomes between Europe and Africa [32]. Our data indicated that most geographically informative SNPs were in mono-allelic genomic portions, which were under purifying selection [25]. Unlike EBV, in which host immune selection appears to drive local viral adaptation to different human host populations [84], our study showed no evidence that selection within known immunogenic regions of the HCMV genome is the dominant driver of the observed genetic variability. Instead, we postulate that bottleneck events, such as founder viruses, are plausible explanations. Interestingly, 32 of the identified multi-allelic regions followed the same pattern as the mono-allelic regions.

The majority of multi-allelic regions (n = 42) showed no geographical segregation, but maintained the full allele palette across Europe and Africa. According to our analyses, multi-allelic regions that were not geographically segregated were significantly enriched for genes encoding immunomodulatory functions. Many of these variable regions correspond to those previously identified to be in strong linkage disequilibrium (LD) by Lassalle et al. [23] (Table 1). For example, region 6 shows no geographical segregation and contains the nonrecombinant haplotype *RL11D* block, whose members have been proven or predicted to be virion membrane glycoproteins [129]. Variability in this region might be crucial for CMV adaptation to different primate species [23].

Our data strongly support the idea that the CMV genome is the result of two distinct evolutionary forces. The first type, supported by geographical segments of the genome, is a founder event. These events are characterised by genetic drift occurring in separate viral populations as a result of human movement and migration. Second, multi-allelic regions with similar allele frequencies worldwide reflect negative frequency-dependent balancing selection, a form of adaptation that maintains pre-existing diversity in the face of genetic drift [127]. This is similar to what happens with the major histocompatibility complex in humans, where the maintenance of multiple alleles at certain frequencies contributes to the ability of the immune system to respond to pathogens [22].

However, the study of the evolutionary history and phylogeographic origins of HCMV is complicated by the pervasive genome-wide recombination that occurs in HCMV. This leads to a large variety of possible haplotypes and hybrid strains [9]. An intriguing example of a hybrid strain is represented by the lab strain Towne, with 79% of SNPs in the mono-allelic portions segregating with African sequences and the rest segregating with European [32]. Apart from specific multi-allelic regions in high local linkage disequilibrium, the rest of the genome recombines freely [23]. Recombination is mostly observed in the mono-allelic part of the genome, where a high degree of conservation is likely an effect of strong purifying selection. Purifying selection is common in viral pathogens, indicating long-standing environmental constraints [137]. Such constraints in HCMV may be represented by a long co-evolutionary history with its host [77]. Indeed, recombination is known to increase the effect of negative selection by unlinking selected sites from the genomic background, providing a means to achieve the levels of purifying selection required to maintain HCMV genome functionality [23].

6. Conclusions

This review summarises the knowledge regarding genetic diversity in HCMV. Our recent study identified 74 regions, providing a more comprehensive understanding of HCMV genetic variability compared to previous research. Notably, our study provided more granularity for certain regions (i.e., *UL55*) and uncovered additional regions of interest. These findings provide valuable insights into the evolution and geographical

population structure of HCMV. However, to gain a more complete picture, future studies should incorporate sequences from diverse populations worldwide. The investigation of viral markers associated with CMV-related diseases has generated more debate than consensus in the field. To address this, future research should focus on well-controlled populations and standardise genotype/allele classification. These approaches will help to clarify the relationship between viral genetic diversity and disease outcomes, ultimately advancing our understanding of HCMV pathogenesis and potentially informing more effective diagnostic and therapeutic strategies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pathogens14010050/s1, Figure S1. (A) All possible alleles' combinations (haplotypes) of regions 22, 23 and 24 in *UL55*. Three alleles were identified in each region. Twelve haplotypes were identified. (B) Frequency of haplotypes in *UL55* in 253 GenBank sequences. Haplotypes 11 and 12 were only found in three African sequences (each) reconstructed from mixed infection with HaROLD [32,125] [24,131]. (C) Frequency of haplotypes by continent in *UL55* in 253 GenBank sequences. (D) Phylogenetic tree (Neighbour joining) of representative sequences for the haplotypes (H1-H10) and genotypes (gB-1- gB-5); Figure S2. (A) Frequency of alleles in regions 28 comprising *UL73/UL74* in 253 GenBank sequences. (B) Frequency of alleles in regions 28 comprising *UL73/UL74* in 253 GenBank sequences. (C) Phylogenetic tree (Neighbour joining) of representative sequences for the alleles in region 28 (A1–A7) and gN genotypes (gN-1- gN-4 + subtypes); Figure S3. Phylogenetic tree (Neighbour joining) of representative sequences for the alleles in region 28 (A1–A7) and gO genotypes; Figure S4. (A) All possible alleles' combinations (haplotypes) of regions 29, 30 and 31 in *UL75*. Three alleles were identified in each region. Seven haplotypes were identified. (B) Frequency of haplotypes of regions 29, 30 and 31 in *UL75* by continent in 253 GenBank sequences. (C) Phylogenetic tree (Neighbour joining) of representative sequences for the haplotypes in regions 29,30 and 31 and gH genotypes; Figure S5. (A) Frequency of alleles in region 50 comprising *UL144* in 253 GenBank sequences. (B) Phylogenetic tree (Neighbour joining) of representative sequences for *UL144* alleles and genotypes; Figure S6. (A) Frequency of alleles in *UL146* by continent in 253 GenBank sequences. (B) Phylogenetic tree (Neighbour joining) of representative sequences for *UL146-UL147* alleles and genotypes; Table S1. Novel multi-allelic regions and genes identified by Charles, Venturini et al. [32]. References [24,131] are cited in the supplementary materials.

Author Contributions: C.V. researched data for the article and wrote the first draft, which was revised by J.B. All authors have read and agreed to the published version of the manuscript.

Funding: C.V. is funded by the Wellcome Trust (224530/Z/21/Z). J.B. receives funding from the NIHR UCL/UCLH Biomedical Research Centre.

Conflicts of Interest: The authors declared no conflicts of interest.

References

1. Altevogt, P.; Sammar, M.; Hüser, L.; Kristiansen, G. Novel insights into the function of CD24: A driving force in cancer. *Int. J. Cancer* **2021**, *148*, 546–559. [CrossRef] [PubMed]
2. Aquino, V.H.; Figueiredo, L.T.M. High prevalence of renal transplant recipients infected with more than one cytomegalovirus glycoprotein B genotype. *J. Med. Virol.* **2000**, *61*, 138–142. [CrossRef]
3. Arav-Boger, R. Strain Variation and Disease Severity in Congenital Cytomegalovirus Infection: In Search of a Viral Marker. *Infect. Dis. Clin.* **2015**, *29*, 401–414. [CrossRef]
4. Arav-Boger, R.; Foster, C.B.; Zong, J.-C.; Pass, R.F. Human Cytomegalovirus–Encoded α-Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns. *J. Infect. Dis.* **2006**, *193*, 788–791. [CrossRef]
5. Arav-Boger, R.; Willoughby, R.E.; Pass, R.F.; Zong, J.-C.; Jang, W.-J.; Alcendor, D.; Hayward, G.S. Polymorphisms of the Cytomegalovirus (CMV)–Encoded Tumor Necrosis Factor–α and β-Chemokine Receptors in Congenital CMV Disease. *J. Infect. Dis.* **2002**, *186*, 1057–1064. [CrossRef]
6. Bale, J.F.; Petheram, S.J.; Robertson, M.; Murph, J.R.; Demmler, G. Human cytomegalovirus *a* sequence and *UL144* variability in strains from infected children. *J. Med. Virol.* **2001**, *65*, 90–96. [CrossRef]

7. Bale, J.F.; Petheram, S.J.; Souza, I.E.; Murph, J.R. Cytomegalovirus reinfection in young children. *J. Pediatr.* **1996**, *128*, 347–352. [CrossRef]
8. Barbi, M.; Binda, S.; Caroppo, S.; Primache, V.; Didò, P.; Guidotti, P.; Corbetta, C.; Melotti, D. CMV gB genotypes and outcome of vertical transmission: Study on dried blood spots of congenitally infected babies. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **2001**, *21*, 75–79. [CrossRef]
9. Bates, M.; Monze, M.; Bima, H.; Kapambwe, M.; Kasolo, F.C.; Gompels, U.A. High human cytomegalovirus loads and diverse linked variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. *Virology* **2008**, *382*, 28–36. [CrossRef]
10. Benedict, C.A.; Butrovich, K.D.; Lurain, N.S.; Corbeil, J.; Rooney, I.; Schneider, P.; Tschopp, J.; Ware, C.F. Cutting edge: A novel viral TNF receptor superfamily member in virulent strains of human cytomegalovirus. *J. Immunol.* **1999**, *162*, 6967–6970. [CrossRef]
11. Beninga, J.; Kalbacher, H.; Mach, M. Analysis of T Helper Cell Response to Glycoprotein H (gpUL75) of Human Cytomegalovirus: Evidence for Strain-Specific T Cell Determinants. *J. Infect. Dis.* **1996**, *173*, 1051–1061. [CrossRef] [PubMed]
12. Berg, C.; Rosenkilde, M.M.; Benfield, T.; Nielsen, L.; Sundelin, T.; Lüttichau, H.R. The frequency of cytomegalovirus non-ELR *UL146* genotypes in neonates with congenital CMV disease is comparable to strains in the background population. *BMC Infect. Dis.* **2021**, *21*, 386. [CrossRef] [PubMed]
13. Boeckh, M.; Geballe, A.P. Cytomegalovirus: Pathogen, paradigm, and puzzle. *J. Clin. Investig.* **2011**, *121*, 1673–1680. [CrossRef] [PubMed]
14. Bonavita, C.M.; Cardin, R.D. Don't Go Breaking My Heart: MCMV as a Model for HCMV-Associated Cardiovascular Diseases. *Pathogens* **2021**, *10*, 619. [CrossRef]
15. Boppana, S.B.; Ross, S.A.; Fowler, K.B. Congenital Cytomegalovirus Infection: Clinical Outcome. *Clin. Infect. Dis.* **2013**, *57* (Suppl. S4), S178–S181. [CrossRef]
16. Bradley, A.J.; Kovács, I.J.; Gatherer, D.; Dargan, D.J.; Alkharsah, K.R.; Chan, P.K.S.; Carman, W.F.; Dedicoat, M.; Emery, V.C.; Geddes, C.C.; et al. Genotypic analysis of two hypervariable human cytomegalovirus genes. *J. Med. Virol.* **2008**, *80*, 1615–1623. [CrossRef]
17. Brait, N.; Stögerer, T.; Kalser, J.; Adler, B.; Kunz, I.; Benesch, M.; Kropff, B.; Mach, M.; Puchhammer-Stöckl, E.; Görzer, I. Influence of Human Cytomegalovirus Glycoprotein O Polymorphism on the Inhibitory Effect of Soluble Forms of Trimer- and Pentamer-Specific Entry Receptors. *J. Virol.* **2020**, *94*, e00107-20. [CrossRef]
18. Britt, W.J. Congenital Human Cytomegalovirus Infection and the Enigma of Maternal Immunity. *J. Virol.* **2017**, *91*, e02392-16. [CrossRef]
19. Burkhardt, C.; Himmelein, S.; Britt, W.; Winkler, T.; Mach, M. Glycoprotein N subtypes of human cytomegalovirus induce a strain-specific antibody response during natural infection. *J. Gen. Virol.* **2009**, *90*, 1951–1961. [CrossRef]
20. Cha, T.A.; Tom, E.; Kemble, G.W.; Duke, G.M.; Mocarski, E.S.; Spaete, R.R. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **1996**, *70*, 78–83. [CrossRef]
21. Chandler, S.H.; Handsfield, H.H.; McDougall, J.K. Isolation of Multiple Strains of Cytomegalovirus from Women Attending a Clinic for Sexually Transmitted Diseases. *J. Infect. Dis.* **1987**, *155*, 655–660. [CrossRef] [PubMed]
22. Janeway, C.A., Jr.; Travers, P.; Walport, M.; Shlomchik, M.J. The major histocompatibility complex and its functions. In *Immunobiology: The Immune System in Health and Disease*, 5th ed.; Garland Science: New York City, NY, USA, 2001. Available online: https://www.ncbi.nlm.nih.gov/books/NBK27156/ (accessed on 26 July 2024).
23. Charles, O.J.; Venturini, C.; Breuer, J. cmvdrg—An R package for Human Cytomegalovirus Antiviral Drug Resistance Genotyping. *bioRxiv* **2020**. [CrossRef]
24. Charles, O.J.; Venturini, C.; Gantt, S.; Atkinson, C.; Griffiths, P.; Goldstein, R.A.; Breuer, J. Genomic and geographical structure of human cytomegalovirus. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2221797120. [CrossRef] [PubMed]
25. Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; Hutchison, C.A.; Kouzarides, T.; Martignetti, J.A.; et al. Analysis of the Protein-Coding Content of the Sequence of Human Cytomegalovirus Strain AD169. In *Cytomegaloviruses*; McDougall, J.K., Ed.; Springer: Berlin/Heidelberg, Germany, 1990; pp. 125–169. [CrossRef]
26. Chou, S. Reactivation and Recombination of Multiple Cytomegalovirus Strains from Individual Organ Donors. *J. Infect. Dis.* **1989**, *160*, 11–15. [CrossRef] [PubMed]
27. Chou, S. Comparative analysis of sequence variation in gp116 and gp55 components of glycoprotein B of human cytomegalovirus. *Virology* **1992**, *188*, 388–390. [CrossRef]
28. Chou, S. Diverse Cytomegalovirus *UL27* Mutations Adapt to Loss of Viral *UL97* Kinase Activity under Maribavir. *Antimicrob. Agents Chemother.* **2009**, *53*, 81–85. [CrossRef]
29. Chou, S.; Dennison, K.M. Analysis of Interstrain Variation in Cytomegalovirus Glycoprotein B Sequences Encoding Neutralization-Related Epitopes. *J. Infect. Dis.* **1991**, *163*, 1229–1234. [CrossRef]
30. Coaquette, A.; Bourgeois, A.; Dirand, C.; Varin, A.; Chen, W.; Herbein, G. Mixed Cytomegalovirus Glycoprotein B Genotypes in Immunocompromised Patients. *Clin. Infect. Dis.* **2004**, *39*, 155–161. [CrossRef]

31. Correia, S.; Bridges, R.; Wegner, F.; Venturini, C.; Palser, A.; Middeldorp, J.M.; Cohen, J.I.; Lorenzetti, M.A.; Bassano, I.; White, R.E.; et al. Sequence Variation of Epstein-Barr Virus: Viral Types, Geography, Codon Usage, and Diseases. *J. Virol.* **2018**, *92*, e01132-18. [CrossRef]
32. Cruz, D.V.; Nelson, C.S.; Tran, D.; Barry, P.A.; Kaur, A.; Koelle, K.; Permar, S.R. Intrahost cytomegalovirus population genetics following antibody pretreatment in a monkey model of congenital transmission. *PLoS Pathog.* **2020**, *16*, e1007968. [CrossRef]
33. Cudini, J.; Roy, S.; Houldcroft, C.J.; Bryant, J.M.; Depledge, D.P.; Tutill, H.; Veys, P.; Williams, R.; Worth, A.J.J.; Tamuri, A.U.; et al. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5693–5698. [CrossRef] [PubMed]
34. Cui, X.; Freed, D.C.; Wang, D.; Qiu, P.; Li, F.; Fu, T.-M.; Kauvar, L.M.; McVoy, M.A. Impact of Antibodies and Strain Polymorphisms on Cytomegalovirus Entry and Spread in Fibroblasts and Epithelial Cells. *J. Virol.* **2017**, *91*, e01650-16. [CrossRef] [PubMed]
35. Cunningham, C.; Gatherer, D.; Hilfrich, B.; Baluchova, K.; Dargan, D.J.; Thomson, M.; Griffiths, P.D.; Wilkinson, G.W.G.; Schulz, T.F.; Davison, A.J. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J. Gen. Virol.* **2010**, *91*, 605–615. [CrossRef]
36. Dargan, D.J.; Douglas, E.; Cunningham, C.; Jamieson, F.; Stanton, R.J.; Baluchova, K.; McSharry, B.P.; Tomasec, P.; Emery, V.C.; Percivalle, E.; et al. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J. Gen. Virol.* **2010**, *91*, 1535–1546. [CrossRef]
37. Davis, C.L.; Field, D.; Metzgar, D.; Saiz, R.; Morin, P.A.; Smith, I.L.; Spector, S.A.; Wills, C. Numerous Length Polymorphisms at Short Tandem Repeats in Human Cytomegalovirus. *J. Virol.* **1999**, *73*, 6265–6270. [CrossRef]
38. Davison, A.J.; Dolan, A.; Akter, P.; Addison, C.; Dargan, D.J.; Alcendor, D.J.; McGeoch, D.J.; Hayward, G.S. The human cytomegalovirus genome revisited: Comparison with the chimpanzee cytomegalovirus genome. *J. Gen. Virol.* **2003**, *84 Pt 1*, 17–28. [CrossRef]
39. Day, L.Z.; Stegmann, C.; Schultz, E.P.; Lanchy, J.-M.; Yu, Q.; Ryckman, B.J. Polymorphisms in Human Cytomegalovirus Glycoprotein O (gO) Exert Epistatic Influences on Cell-Free and Cell-to-Cell Spread and Antibody Neutralization on gH Epitopes. *J. Virol.* **2020**, *94*, e02051-19. [CrossRef]
40. Depledge, D.P.; Palser, A.L.; Watson, S.J.; Lai, I.Y.-C.; Gray, E.R.; Grant, P.; Kanda, R.K.; Leproust, E.; Kellam, P.; Breuer, J. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS ONE* **2011**, *6*, e27805. [CrossRef]
41. Dhingra, A.; Götting, J.; Varanasi, P.R.; Steinbrueck, L.; Camiolo, S.; Zischke, J.; Heim, A.; Schulz, T.F.; Weissinger, E.M.; Kay-Fedorov, P.C.; et al. Human cytomegalovirus multiple-strain infections and viral population diversity in haematopoietic stem cell transplant recipients analysed by high-throughput sequencing. *Med. Microbiol. Immunol.* **2021**, *210*, 291–304. [CrossRef]
42. Dobbins, G.C.; Patki, A.; Chen, D.; Tiwari, H.K.; Hendrickson, C.; Britt, W.J.; Fowler, K.; Chen, J.Y.; Boppana, S.B.; Ross, S.A. Association of CMV genomic mutations with symptomatic infection and hearing loss in congenital CMV infection. *BMC Infect. Dis.* **2019**, *19*, 1046. [CrossRef]
43. Dolan, A.; Cunningham, C.; Hector, R.D.; Hassan-Walker, A.F.; Lee, L.; Addison, C.; Dargan, D.J.; McGeoch, D.J.; Gatherer, D.; Emery, V.C.; et al. Genetic content of wild-type human cytomegalovirus. *J. Gen. Virol.* **2004**, *85*, 1301–1312. [CrossRef] [PubMed]
44. Dong, N.; Cao, L.; Zheng, D.; Su, L.; Lu, L.; Dong, Z.; Xu, M.; Xu, J. Distribution of CMV envelope glycoprotein B, H and N genotypes in infants with congenital cytomegalovirus symptomatic infection. *Front. Pediatr.* **2023**, *11*, 1112645. [CrossRef] [PubMed]
45. Fang, X.; Zheng, P.; Tang, J.; Liu, Y. CD24: From A to Z. *Cell. Mol. Immunol.* **2010**, *7*, 100–103. [CrossRef] [PubMed]
46. Faure-Della Corte, M.; Samot, J.; Garrigue, I.; Magnin, N.; Reigadas, S.; Couzi, L.; Dromer, C.; Velly, J.-F.; Déchanet-Merville, J.; Fleury, H.J.A.; et al. Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. *J. Clin. Virol.* **2010**, *47*, 161–169. [CrossRef]
47. Freed, D.C.; Tang, Q.; Tang, A.; Li, F.; He, X.; Huang, Z.; Meng, W.; Xia, L.; Finnefrock, A.C.; Durr, E.; et al. Pentameric complex of viral glycoprotein H is the primary target for potent neutralization by a human cytomegalovirus vaccine. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4997–E5005. [CrossRef]
48. Fulkerson, H.L.; Nogalski, M.T.; Collins-McMillen, D.; Yurochko, A.D. Overview of Human Cytomegalovirus Pathogenesis. In *Human Cytomegaloviruses: Methods and Protocols*; Yurochko, A.D., Ed.; Springer: New York, NY, USA, 2021; pp. 1–18. [CrossRef]
49. Garrigue, I.; Corte, M.F.-D.; Magnin, N.; Recordon-Pinson, P.; Couzi, L.; Lebrette, M.-E.; Schrive, M.-H.; Roncin, L.; Taupin, J.-L.; Déchanet-Merville, J.; et al. *UL40* Human Cytomegalovirus Variability Evolution Patterns Over Time in Renal Transplant Recipients. *Transplantation* **2008**, *86*, 826. [CrossRef]
50. Görzer, I.; Guelly, C.; Trajanoski, S.; Puchhammer-Stöckl, E. Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time. *J. Virol.* **2010**, *84*, 7195–7203. [CrossRef]
51. Görzer, I.; Kerschner, H.; Jaksch, P.; Bauer, C.; Seebacher, G.; Klepetko, W.; Puchhammer-Stöckl, E. Virus load dynamics of individual CMV-genotypes in lung transplant recipients with mixed-genotype infections. *J. Med. Virol.* **2008**, *80*, 1405–1414. [CrossRef]
52. Gretch, D.R.; Kari, B.; Rasmussen, L.; Gehrz, R.C.; Stinski, M.F. Identification and characterization of three distinct families of glycoprotein complexes in the envelopes of human cytomegalovirus. *J. Virol.* **1988**, *62*, 875–881. [CrossRef]

53. Grosjean, J.; Hantz, S.; Cotin, S.; Baclet, M.C.; Mengelle, C.; Trapes, L.; Virey, B.; Undreiner, F.; Brosset, P.; Pasquier, C.; et al. Direct genotyping of cytomegalovirus envelope glycoproteins from toddler's saliva samples. *J. Clin. Virol.* **2009**, *46*, S43–S48. [CrossRef]
54. Gugliesi, F.; Pasquero, S.; Griffante, G.; Scutera, S.; Albano, C.; Pacheco, S.F.C.; Riva, G.; Dell'Oste, V.; Biolatti, M. Human Cytomegalovirus and Autoimmune Diseases: Where Are We? *Viruses* **2021**, *13*, 260. [CrossRef] [PubMed]
55. Guo, G.; Zhang, L.; Ye, S.; Hu, Y.; Li, B.; Sun, X.; Mao, C.; Xu, J.; Chen, Y.; Zhang, L.; et al. Polymorphisms and features of cytomegalovirus *UL144* and *UL146* in congenitally infected neonates with hepatic involvement. *PLoS ONE* **2017**, *12*, e0171959. [CrossRef] [PubMed]
56. Haberland, M.; Meyer-König, U.; Hufert, F.T. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *J. Gen. Virol.* **1999**, *80*, 1495–1500. [CrossRef] [PubMed]
57. Hage, E.; Wilkie, G.S.; Linnenweber-Held, S.; Dhingra, A.; Suárez, N.M.; Schmidt, J.J.; Kay-Fedorov, P.C.; Mischak-Weissinger, E.; Heim, A.; Schwarz, A.; et al. Characterization of Human Cytomegalovirus Genome Diversity in Immunocompromised Hosts by Whole-Genome Sequencing Directly From Clinical Specimens. *J. Infect. Dis.* **2017**, *215*, 1673–1683. [CrossRef]
58. Hansen, S.G.; Powers, C.J.; Richards, R.; Ventura, A.B.; Ford, J.C.; Siess, D.; Axthelm, M.K.; Nelson, J.A.; Jarvis, M.A.; Picker, L.J.; et al. Evasion of CD8+ T Cells Is Critical for Superinfection by Cytomegalovirus. *Science* **2010**, *328*, 102–106. [CrossRef]
59. Hassan-Walker, A.F.; Okwuadi, S.; Lee, L.; Griffiths, P.D.; Emery, V.C. Sequence variability of the α-chemokine *UL146* from clinical strains of human cytomegalovirus. *J. Med. Virol.* **2004**, *74*, 573–579. [CrossRef]
60. Heo, J.; Petheram, S.; Demmler, G.; Murph, J.R.; Adler, S.P.; Bale, J.; Sparer, T.E. 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. [CrossRef]
61. Houldcroft, C.J.; Beale, M.A.; Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **2017**, *15*, 183–192. [CrossRef]
62. Isaacson, M.K.; Compton, T. Human Cytomegalovirus Glycoprotein B Is Required for Virus Entry and Cell-to-Cell Spread but Not for Virion Attachment, Assembly, or Egress. *J. Virol.* **2009**, *83*, 3891–3903. [CrossRef]
63. Ishibashi, K.; Tokumoto, T.; Shirakawa, H.; Hashimoto, K.; Kushida, N.; Yanagida, T.; Shishido, K.; Aikawa, K.; Yamaguchi, O.; Toma, H.; et al. Association between antibody response against cytomegalovirus strain-specific glycoprotein H epitopes and HLA-DR. *Microbiol. Immunol.* **2009**, *53*, 412–416. [CrossRef]
64. Kalser, J.; Adler, B.; Mach, M.; Kropff, B.; Puchhammer-Stöckl, E.; Görzer, I. Differences in Growth Properties among Two Human Cytomegalovirus Glycoprotein O Genotypes. *Front. Microbiol.* **2017**, *8*, 1609. [CrossRef] [PubMed]
65. Kaufer, B.B.; Jarosinski, K.W.; Osterrieder, N. Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. *J. Exp. Med.* **2011**, *208*, 605–615. [CrossRef] [PubMed]
66. Lassalle, F.; Depledge, D.P.; Reeves, M.B.; Brown, A.C.; Christiansen, M.T.; Tutill, H.J.; Williams, R.J.; Einer-Jensen, K.; Holdstock, J.; Atkinson, C.; et al. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. *Virus Evol.* **2016**, *2*, vew017. [CrossRef]
67. Lurain, N.S.; Chou, S. Antiviral Drug Resistance of Human Cytomegalovirus. *Clin. Microbiol. Rev.* **2010**, *23*, 689–712. [CrossRef]
68. Lurain, N.S.; Fox, A.M.; Lichy, H.M.; Bhorade, S.M.; Ware, C.F.; Huang, D.D.; Kwan, S.-P.; Garrity, E.R.; Chou, S. Analysis of the human cytomegalovirus genomic region from *UL146* through *UL147A* reveals sequence hypervariability, genotypic stability, and overlapping transcripts. *Virol. J.* **2006**, *3*, 4. [CrossRef]
69. Lurain, N.S.; Kapell, K.S.; Huang, D.D.; Short, J.A.; Paintsil, J.; Winkfield, E.; Benedict, C.A.; Ware, C.F.; Bremer, J.W. Human cytomegalovirus *UL144* open reading frame: Sequence hypervariability in low-passage clinical isolates. *J. Virol.* **1999**, *73*, 10040–10050. [CrossRef]
70. Mach, M.; Kropff, B.; Dal Monte, P.; Britt, W. Complex Formation by Human Cytomegalovirus Glycoproteins M (gpUL100) and N (gpUL73). *J. Virol.* **2000**, *74*, 11881–11892. [CrossRef]
71. Manicklal, S.; Emery, V.C.; Lazzarotto, T.; Boppana, S.B.; Gupta, R.K. The 'silent' global burden of congenital cytomegalovirus. *Clin. Microbiol. Rev.* **2013**, *26*, 86–102. [CrossRef]
72. Manuel, O.; Åsberg, A.; Pang, X.; Rollag, H.; Emery, V.C.; Preiksaitis, J.K.; Kumar, D.; Pescovitz, M.D.; Bignamini, A.A.; Hartmann, A.; et al. Impact of Genetic Polymorphisms in Cytomegalovirus Glycoprotein B on Outcomes in Solid-Organ Transplant Recipients with Cytomegalovirus Disease. *Clin. Infect. Dis.* **2009**, *49*, 1160–1166. [CrossRef]
73. Manuel, O.; Pang, X.L.; Humar, A.; Kumar, D.; Doucette, K.; Preiksaitis, J.K. An Assessment of Donor-to-Recipient Transmission Patterns of Human Cytomegalovirus by Analysis of Viral Genomic Variants. *J. Infect. Dis.* **2009**, *199*, 1621–1628. [CrossRef]
74. Masse, M.J.; Karlin, S.; Schachtel, G.A.; Mocarski, E.S. Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5246–5250. [CrossRef] [PubMed]
75. Mattick, C.; Dewin, D.; Polley, S.; Sevilla-Reyes, E.; Pignatelli, S.; Rawlinson, W.; Wilkinson, G.; Dal Monte, P.; Gompels, U.A. Linkage of human cytomegalovirus glycoprotein gO variant groups identified from worldwide clinical isolates with gN genotypes, implications for disease associations and evidence for N-terminal sites of positive selection. *Virology* **2004**, *318*, 582–597. [CrossRef] [PubMed]

76. McFaline-Figueroa, J.R.; Wen, P.Y. The Viral Connection to Glioblastoma. *Curr. Infect. Dis. Rep.* **2017**, *19*, 5. [CrossRef] [PubMed]
77. McGeoch, D.J.; Rixon, F.J.; Davison, A.J. Topics in herpesvirus genomics and evolution. *Virus Res.* **2006**, *117*, 90–104. [CrossRef] [PubMed]
78. McSharry, B.P.; Avdic, S.; Slobedman, B. Human cytomegalovirus encoded homologs of cytokines, chemokines and their receptors: Roles in immunomodulation. *Viruses* **2012**, *4*, 2448–2470. [CrossRef]
79. Meyer-König, U.; Ebert, K.; Schrage, B.; Pollak, S.; Hufert, F.T. Simultaneous infection of healthy people with multiple human cytomegalovirus strains. *Lancet* **1998**, *352*, 1280–1281. [CrossRef]
80. Meyer-König, U.; Vogelberg, C.; Bongarts, A.; Kampa, D.; Delbrück, R.; Wolff-Vorbeck, G.; Kirste, G.; Haberland, M.; Hufert, F.T.; von Laer, D. Glycoprotein B genotype correlates with cell tropism in vivo of human cytomegalovirus infection. *J. Med. Virol.* **1998**, *55*, 75–81. [CrossRef]
81. Michaelis, M.; Doerr, H.W.; Cinatl, J. The Story of Human Cytomegalovirus and Cancer: Increasing Evidence and Open Questions. *Neoplasia* **2009**, *11*, 1–9. [CrossRef]
82. Murphy, E.; Shenk, T.E. Human Cytomegalovirus Genome. In *Human Cytomegalovirus*; Shenk, T.E., Stinski, M.F., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19. [CrossRef]
83. Nanamiya, H.; Tanaka, D.; Hiyama, G.; Isogai, T.; Watanabe, S. Detection of four isomers of the human cytomegalovirus genome using nanopore long-read sequencing. *Virus Genes* **2024**, *60*, 377–384. [CrossRef]
84. Nelson, C.S.; Huffman, T.; Jenks, J.A.; Cisneros de la Rosa, E.; Xie, G.; Vandergrift, N.; Pass, R.F.; Pollara, J.; Permar, S.R. HCMV glycoprotein B subunit vaccine efficacy mediated by nonneutralizing antibody effector functions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6267–6272. [CrossRef]
85. Nijman, J.; Mandemaker, F.S.; Verboon-Maciolek, M.A.; Aitken, S.C.; van Loon, A.M.; de Vries, L.S.; Schuurman, R. Genotype Distribution, Viral Load and Clinical Characteristics of Infants with Postnatal or Congenital Cytomegalovirus Infection. *PLoS ONE* **2014**, *9*, e108018. [CrossRef] [PubMed]
86. Novak, Z.; Ross, S.A.; Patro, R.K.; Pati, S.K.; Kumbla, R.A.; Brice, S.; Boppana, S.B. Cytomegalovirus Strain Diversity in Seropositive Women. *J. Clin. Microbiol.* **2008**, *46*, 882–886. [CrossRef] [PubMed]
87. Pang, J.; Slyker, J.A.; Roy, S.; Bryant, J.; Atkinson, C.; Cudini, J.; Farquhar, C.; Griffiths, P.; Kiarie, J.; Morfopoulou, S.; et al. Mixed cytomegalovirus genotypes in HIV-positive mothers show compartmentalization and distinct patterns of transmission to infants. *eLife* **2020**, *9*, e63199. [CrossRef] [PubMed]
88. Paradowska, E.; Jabłońska, A.; Płóciennikowska, A.; Studzińska, M.; Suski, P.; Wiśniewska-Ligier, M.; Dzierżanowska-Fangrat, K.; Kasztelewicz, B.; Woźniakowska-Gęsicka, T.; Leśnikowski, Z.J. Cytomegalovirus alpha-chemokine genotypes are associated with clinical manifestations in children with congenital or postnatal infections. *Virology* **2014**, *462–463*, 207–217. [CrossRef]
89. Pati, S.; Pinninti, S.; Novak, Z.; Chowdhury, N.; Patro, R.; Fowler, K.; Ross, S.; Boppana, S. Genotypic Diversity and Mixed Infection in Newborn Disease and Hearing Loss in Congenital Cytomegalovirus Infection. *Pediatr. Infect. Dis. J.* **2013**, *32*, 1050–1054. [CrossRef]
90. Penfold, M.E.T.; Dairaghi, D.J.; Duke, G.M.; Saederup, N.; Mocarski, E.S.; Kemble, G.W.; Schall, T.J. Cytomegalovirus encodes a potent α chemokine. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9839–9844. [CrossRef]
91. Perdue, M.L.; García, M.; Senne, D.; Fraire, M. Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res.* **1997**, *49*, 173–186. [CrossRef]
92. Picone, O.; Costa, J.-M.; Ville, Y.; Chaix, M.-L.; Rouzioux, C.; Leruez-Ville, M. Genetic polymorphism of cytomegalovirus strains responsible of congenital infections. *Pathol.-Biol.* **2004**, *52*, 534–539. [CrossRef]
93. Pignatelli, S.; Dal Monte, P.; Landini, M.P. gpUL73 (gN) genomic variants of human cytomegalovirus isolates are clustered into four distinct genotypes. *J. Gen. Virol.* **2001**, *82*, 2777–2784. [CrossRef]
94. Pignatelli, S.; Dal Monte, P.; Rossini, G.; Chou, S.; Gojobori, T.; Hanada, K.; Guo, J.J.; Rawlinson, W.; Britt, W.; Mach, M.; et al. Human cytomegalovirus glycoprotein N (gpUL73-gN) genomic variants: Identification of a novel subgroup, geographical distribution and evidence of positive selective pressure. *J. Gen. Virol.* **2003**, *84*, 647–655. [CrossRef]
95. Pignatelli, S.; Lazzarotto, T.; Gatto, M.R.; Dal Monte, P.; Landini, M.P.; Faldella, G.; Lanari, M. Cytomegalovirus gN Genotypes Distribution among Congenitally Infected Newborns and Their Relationship with Symptoms at Birth and Sequelae. *Clin. Infect. Dis.* **2010**, *51*, 33–41. [CrossRef] [PubMed]
96. Pignatelli, S.; Monte, P.D.; Rossini, G.; Landini, M.P. Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev. Med. Virol.* **2004**, *14*, 383–410. [CrossRef] [PubMed]
97. Plotkin, S.A.; Starr, S.E.; Friedman, H.M.; Gönczöl, E.; Weibel, R.E. Protective effects of Towne cytomegalovirus vaccine against low-passage cytomegalovirus administered as a challenge. *J. Infect. Dis.* **1989**, *159*, 860–865. [CrossRef] [PubMed]
98. Prichard, M.N.; Penfold, M.E.T.; Duke, G.M.; Spaete, R.R.; Kemble, G.W. A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. *Rev. Med. Virol.* **2001**, *11*, 191–200. [CrossRef]

99. Prod'homme, V.; Tomasec, P.; Cunningham, C.; Lemberg, M.K.; Stanton, R.J.; McSharry, B.P.; Wang, E.C.Y.; Cuff, S.; Martoglio, B.; Davison, A.J.; et al. Human cytomegalovirus *UL40* signal peptide regulates cell surface expression of the NK cell ligands HLA-E and gpUL18. *J. Immunol.* **2012**, *188*, 2794–2804. [CrossRef]
100. Puchhammer-Stöckl, E.; Görzer, I. Cytomegalovirus and Epstein-Barr virus subtypes—The search for clinical significance. *J. Clin. Virol.* **2006**, *36*, 239–248. [CrossRef]
101. Puchhammer-Stöckl, E.; Görzer, I. Human cytomegalovirus: An enormous variety of strains and their possible clinical significance in the human host. *Future Virol.* **2011**, *6*, 259–271. [CrossRef]
102. Qi, Y.; Mao, Z.-Q.; Ruan, Q.; He, R.; Ma, Y.-P.; Sun, Z.-R.; Ji, Y.-H.; Huang, Y. Human cytomegalovirus (HCMV) *UL139* open reading frame: Sequence variants are clustered into three major genotypes. *J. Med. Virol.* **2006**, *78*, 517–522. [CrossRef]
103. Quinlivan, M.; Hawrami, K.; Barrett-Muir, W.; Aaby, P.; Arvin, A.; Chow, V.T.; John, T.J.; Matondo, P.; Peiris, M.; Poulsen, A.; et al. The molecular epidemiology of varicella-zoster virus: Evidence for geographic segregation. *J. Infect. Dis.* **2002**, *186*, 888–894. [CrossRef]
104. Rasmussen, L.; Geissler, A.; Cowan, C.; Chase, A.; Winters, M. The Genes Encoding the gCIII Complex of Human Cytomegalovirus Exist in Highly Diverse Combinations in Clinical Isolates. *J. Virol.* **2002**, *76*, 10841–10848. [CrossRef]
105. Rasmussen, L.; Geissler, A.; Winters, M. Inter- and Intragenic Variations Complicate the Molecular Epidemiology of Human Cytomegalovirus. *J. Infect. Dis.* **2003**, *187*, 809–819. [CrossRef] [PubMed]
106. Rawlinson, W.D.; Farrell, H.E.; Barrell, B.G. Analysis of the complete DNA sequence of murine cytomegalovirus. *J. Virol.* **1996**, *70*, 8833–8849. [CrossRef] [PubMed]
107. Renzette, N.; Gibson, L.; Bhattacharjee, B.; Fisher, D.; Schleiss, M.R.; Jensen, J.D.; Kowalik, T.F. Rapid Intrahost Evolution of Human Cytomegalovirus Is Shaped by Demography and Positive Selection. *PLoS Genet.* **2013**, *9*, e1003735. [CrossRef]
108. Renzette, N.; Pokalyuk, C.; Gibson, L.; Bhattacharjee, B.; Schleiss, M.R.; Hamprecht, K.; Yamamoto, A.Y.; Mussi-Pinhata, M.M.; Britt, W.J.; Jensen, J.D. Limits and patterns of cytomegalovirus genomic diversity in humans. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4120–E4128. [CrossRef]
109. Retière, C.; Lesimple, B.; Lepelletier, D.; Bignon, J.-D.; Hallet, M.-M.; Imbert-Marcille, B.-M. Association of glycoprotein B and immediate early-1 genotypes with human leukocyte antigen alleles in renal transplant recipients with cytomegalovirus infection. *Transplantation* **2003**, *75*, 161. [CrossRef]
110. Revello, M.G.; Gerna, G. Human cytomegalovirus tropism for endothelial/epithelial cells: Scientific background and clinical implications. *Rev. Med. Virol.* **2010**, *20*, 136–155. [CrossRef]
111. Ross, S.A.; Arora, N.; Novak, Z.; Fowler, K.B.; Britt, W.J.; Boppana, S.B. Cytomegalovirus Reinfections in Healthy Seroimmune Women. *J. Infect. Dis.* **2010**, *201*, 386–389. [CrossRef]
112. Ross, S.A.; Novak, Z.; Pati, S.; Patro, R.K.; Blumenthal, J.; Danthuluri, V.R.; Ahmed, A.; Michaels, M.G.; Sánchez, P.J.; Bernstein, D.I.; et al. Mixed Infection and Strain Diversity in Congenital Cytomegalovirus Infection. *J. Infect. Dis.* **2011**, *204*, 1003–1007. [CrossRef]
113. Sakaue, S.; Gurajala, S.; Curtis, M.; Luo, Y.; Choi, W.; Ishigaki, K.; Kang, J.B.; Rumker, L.; Deutsch, A.J.; Schönherr, S.; et al. Tutorial: A statistical genetics guide to identifying HLA alleles driving complex disease. *Nat. Protoc.* **2023**, *18*, 2625–2641. [CrossRef]
114. Sapuan, S.; Theodosiou, A.A.; Strang, B.L.; Heath, P.T.; Jones, C.E. A systematic review and meta-analysis of the prevalence of human cytomegalovirus shedding in seropositive pregnant women. *Rev. Med. Virol.* **2022**, *32*, e2399. [CrossRef]
115. Sekulin, K.; Görzer, I.; Heiss-Czedik, D.; Puchhammer-Stöckl, E. Analysis of the variability of CMV strains in the *RL11D* domain of the *RL11* multigene family. *Virus Genes* **2007**, *35*, 577–583. [CrossRef] [PubMed]
116. Shimamura, M.; Mach, M.; Britt, W.J. Human Cytomegalovirus Infection Elicits a Glycoprotein M (gM)/gN-Specific Virus-Neutralizing Antibody Response. *J. Virol.* **2006**, *80*, 4591–4600. [CrossRef] [PubMed]
117. Sijmons, S.; Thys, K.; Mbong Ngwese, M.; Van Damme, E.; Dvorak, J.; Van Loock, M.; Li, G.; Tachezy, R.; Busson, L.; Aerssens, J.; et al. High-Throughput Analysis of Human Cytomegalovirus Genome Diversity Highlights the Widespread Occurrence of Gene-Disrupting Mutations and Pervasive Recombination. *J. Virol.* **2015**, *89*, 7673–7695. [CrossRef]
118. Sijmons, S.; Van Ranst, M.; Maes, P. Genomic and Functional Characteristics of Human Cytomegalovirus Revealed by Next-Generation Sequencing. *Viruses* **2014**, *6*, 1049–1072. [CrossRef]
119. Sinzger, C.; Hahn, G.; Digel, M.; Katona, R.; Sampaio, K.L.; Messerle, M.; Hengel, H.; Koszinowski, U.; Brune, W.; Adler, B. Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J. Gen. Virol.* **2008**, *89*, 359–368. [CrossRef]
120. Stanton, R.J.; Baluchova, K.; Dargan, D.J.; Cunningham, C.; Sheehy, O.; Seirafian, S.; McSharry, B.P.; Neale, M.L.; Davies, J.A.; Tomasec, P.; et al. Reconstruction of the complete human cytomegalovirus genome in a BAC reveals *RL13* to be a potent inhibitor of replication. *J. Clin. Investig.* **2010**, *120*, 3191–3208. [CrossRef]
121. Stanton, R.; Westmoreland, D.; Fox, J.D.; Davison, A.J.; Wilkinson, G.W.G. Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J. Med. Virol.* **2005**, *75*, 42–46. [CrossRef]
122. Stern-Ginossar, N.; Weisburd, B.; Michalski, A.; Le, V.T.K.; Hein, M.Y.; Huang, S.-X.; Ma, M.; Shen, B.; Qian, S.-B.; Hengel, H.; et al. Decoding Human Cytomegalovirus. *Science* **2012**, *338*, 1088–1093. [CrossRef]

123. Suárez, N.M.; Blyth, E.; Li, K.; Ganzenmueller, T.; Camiolo, S.; Avdic, S.; Withers, B.; Linnenweber-Held, S.; Gwinner, W.; Dhingra, A.; et al. Whole-Genome Approach to Assessing Human Cytomegalovirus Dynamics in Transplant Patients Undergoing Antiviral Therapy. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 267. [CrossRef]
124. Suárez, N.M.; Musonda, K.G.; Escriva, E.; Njenga, M.; Agbueze, A.; Camiolo, S.; Davison, A.J.; Gompels, U.A. Multiple-Strain Infections of Human Cytomegalovirus With High Genomic Diversity Are Common in Breast Milk From Human Immunodeficiency Virus–Infected Women in Zambia. *J. Infect. Dis.* **2019**, *220*, 792–801. [CrossRef]
125. Suárez, N.M.; Wilkie, G.S.; Hage, E.; Camiolo, S.; Holton, M.; Hughes, J.; Maabar, M.; Vattipally, S.B.; Dhingra, A.; Gompels, U.A.; et al. Human Cytomegalovirus Genomes Sequenced Directly From Clinical Material: Variation, Multiple-Strain Infection, Recombination, and Gene Loss. *J. Infect. Dis.* **2019**, *220*, 781–791. [CrossRef] [PubMed]
126. Szpara, M.L.; Gatherer, D.; Ochoa, A.; Greenbaum, B.; Dolan, A.; Bowden, R.J.; Enquist, L.W.; Legendre, M.; Davison, A.J. Evolution and diversity in human herpes simplex virus genomes. *J. Virol.* **2014**, *88*, 1209–1227. [CrossRef] [PubMed]
127. Takahashi, Y.; Kawata, M. A comprehensive test for negative frequency-dependent selection. *Popul. Ecol.* **2013**, *55*, 499–509. [CrossRef]
128. Urban, M.; Klein, M.; Britt, W.J.; Haßfurther, E.; Mach, M. Glycoprotein H of human cytomegalovirus is a major antigen for the neutralizing humoral immune response. *J. Gen. Virol.* **1996**, *77*, 1537–1547. [CrossRef]
129. Van Damme, E.; Van Loock, M. Functional annotation of human cytomegalovirus gene products: An update. *Front. Microbiol.* **2014**, *5*, 218. [CrossRef]
130. Venturini, C.; Colston, J.M.; Charles, O.; Best, T.; Atkinson, C.; Forrest, C.; Williams, C.; Rao, K.; Worth, A.; Thorburn, D.; et al. Persistent low-level variants in a subset of HCMV genes are highly predictive of poor outcome in immunocompromised patients with cytomegalovirus infection. *medRxiv* **2022**. [CrossRef]
131. Venturini, C.; Pang, J.; Tamuri, A.U.; Roy, S.; Atkinson, C.; Griffiths, P.; Breuer, J.; Goldstein, R.A. Haplotype assignment of longitudinal viral deep-sequencing data using co-variation of variant frequencies. *Virus Evol.* **2022**, *8*, veac093. [CrossRef]
132. Wada, K.; Mizuno, S.; Kato, K.; Kamiya, T.; Ozawa, K. Cytomegalovirus Glycoprotein B Sequence Variation among Japanese Bone Marrow Transplant Recipients. *Intervirology* **2008**, *40*, 215–219. [CrossRef]
133. Walker, A.; Petheram, S.J.; Ballard, L.; Murph, J.R.; Demmler, G.J.; Bale, J.F. Characterization of Human Cytomegalovirus Strains by Analysis of Short Tandem Repeat Polymorphisms. *J. Clin. Microbiol.* **2001**, *39*, 2219–2226. [CrossRef]
134. Wang, D.; Shenk, T. Human Cytomegalovirus *UL131* Open Reading Frame Is Required for Epithelial Cell Tropism. *J. Virol.* **2005**, *79*, 10330–10338. [CrossRef]
135. Wang, H.-Y.; Valencia, S.M.; Pfeifer, S.P.; Jensen, J.D.; Kowalik, T.F.; Permar, S.R. Common Polymorphisms in the Glycoproteins of Human Cytomegalovirus and Associated Strain-Specific Immunity. *Viruses* **2021**, *13*, 1106. [CrossRef] [PubMed]
136. Wegner, F.; Lassalle, F.; Depledge, D.P.; Balloux, F.; Breuer, J. Coevolution of Sites under Immune Selection Shapes Epstein–Barr Virus Population Structure. *Mol. Biol. Evol.* **2019**, *36*, 2512–2521. [CrossRef] [PubMed]
137. Wertheim, J.O.; Kosakovsky Pond, S.L. Purifying Selection Can Obscure the Ancient Age of Viral Lineages. *Mol. Biol. Evol.* **2011**, *28*, 3355–3365. [CrossRef] [PubMed]
138. Wu, Y.; Prager, A.; Boos, S.; Resch, M.; Brizic, I.; Mach, M.; Wildner, S.; Scrivano, L.; Adler, B. Human cytomegalovirus glycoprotein complex gH/gL/gO uses PDGFR-α as a key for entry. *PLoS Pathog.* **2017**, *13*, e1006281. [CrossRef] [PubMed]
139. Yan, H.; Koyano, S.; Inami, Y.; Yamamoto, Y.; Suzutani, T.; Mizuguchi, M.; Ushijima, H.; Kurane, I.; Inoue, N. Genetic linkage among human cytomegalovirus glycoprotein N (gN) and gO genes, with evidence for recombination from congenitally and post-natally infected Japanese infants. *J. Gen. Virol.* **2008**, *89*, 2275–2279. [CrossRef]
140. Yan, H.; Koyano, S.; Inami, Y.; Yamamoto, Y.; Suzutani, T.; Mizuguchi, M.; Ushijima, H.; Kurane, I.; Inoue, N. Genetic variations in the gB, *UL144* and *UL149* genes of human cytomegalovirus strains collected from congenitally and postnatally infected Japanese children. *Arch. Virol.* **2008**, *153*, 667–674. [CrossRef]
141. Zuhair, M.; Smit, G.S.A.; Wallis, G.; Jabbar, F.; Smith, C.; Devleesschauwer, B.; Griffiths, P. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev. Med. Virol.* **2019**, *29*, e2034. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.