Published in final edited form as:

Infect Dis Clin North Am. 2015 September ; 29(3): 401–414. doi:10.1016/j.idc.2015.05.009.

Strain Variation and disease severity in congenital CMV infection – in search of a viral marker

Ravit Arav-Boger, MD

Johns Hopkins University School of Medicine, 200 N. Wolfe St, Baltimore, MD 21287

Ravit Arav-Boger: boger@jhmi.edu

Synopsis

The wide spectrum of congenital CMV disease and known differences in the biology and *in vitro* growth of CMV strains continue to drive studies in search for specific viral genetic determinants that may predict severity of congenital CMV disease. Several CMV genes have been studied in detail in congenitally-infected children, but the complexity of the viral genome and differences in the definition of symptomatic disease vs. asymptomatic CMV infection continue to raise questions related to what constitutes a pathogenic CMV strain. In addition, the prevalence and role of multiple CMV strains as opposed to infection with a single strain in disease severity is debated. Although the new era of highly-sensitive next generation sequencing assays may provide detailed information on multiple genetic loci, strict criteria for differentiating between strains will be required. Identifying a viral marker (or combination of markers) for prediction of disease outcome could have a major impact on pre-natal diagnosis and vaccine development. Large, well-controlled studies with long-term follow-up and use of standardized high stringency molecular techniques will be required to move the field forward. Genetic variation of host genes may play an adjunctive role in the outcome of congenital CMV disease.

Keywords

Cytomegalovirus; congenital infection; strains; genotypes; multiple strains; immune evasion genes; population-based sequencing; next generation sequencing

Introduction

Infection with CMV is the leading cause of congenital infection and the most common infectious source of central nervous system (CNS) damage and sensorineural hearing loss in the United states (2,3). The outcome of congenital CMV infection is highly variable. Of infected infants, around 10%-15% exhibit severe symptoms at birth. Of the 85-90% who are asymptomatic at birth, 10-15% will later develop hearing loss and other neurological deficits (4,5). It is not clear why some infants have fatal or multisystem disease while others have no clinical evidence of abnormalities in the neonatal period or later (6). The severity of fetal

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

disease varies widely, ranging from stillbirth due to multi-system disease to no abnormalities. Similarly, the long-term outcome of congenital CMV infection ranges from no apparent impairments to significant CNS damage manifested as global developmental delay, cerebral palsy, hearing loss or impaired vision, appearing alone or in combination.

Interest in the identification of pathogenic CMV strains originated from this wide spectrum of disease manifestations along with laboratory findings of genetic variability and differences in growth characteristics of the very distinctive CMV strains (7,8). Vaccine studies have also contributed to the hypothesis that CMV strains have different pathogenic potential. The laboratory-adapted strains AD169 and Towne, which have been passaged multiple times in human fibroblasts, were found to be attenuated when administered as vaccine candidates. However, the Toledo strain, which had only been passaged several times in tissue culture, caused disease when administered to seropositive individuals (9). Identification of CMV strains, that when acquired, can lead to severe disease outcome would allow for early diagnosis and outcome prediction, and may direct vaccine development targeted to specific viral products. In the era prior to availability of sensitive high-throughput sequencing, only selected genes have been sequenced from the CMV genome, which ranges from 220–240 kB in length. These genes, some of which were discovered to be hypervariable, were selected based on data supporting their potential role in pathogenicity and dissemination of CMV. The more recent introduction of sensitive high-throughput sequencing techniques provides new information on sequence variation among CMV strains, and may lead to better categorization of strains, but requires standardization and stringent criteria for strain definition.

This review will address the following topics amongst others

1. Genetic variability of CMV strains - Genotypic analysis
2. Next generation Sequencing and its implications for identification of a genetic marker
3. The role of multiple CMV strains in outcome of congenital CMV infection

Genetic Definitions (Box 1)

Locus- a specific location DNA sequence

Clade- a grouping that includes a common ancestor and all its descendants. Phylogeny is used to define a clade.

ORF- an open reading frame is a DNA sequence that, when translated into amino acids, contains no stop codons.

Linkage - Linked genes are adjacent to each other on a chromosome, thus they are likely to be transmitted or inherited together.

Molecular mimicry- a strategy used by many viruses to subvert and regulate anti-viral immunity. Identical or similar amino acid sequences shared between pathogens and the host are responsible for molecular mimicry.

NGS- next generation sequencing, non-Sanger-based high-throughput DNA sequencing technologies. Millions or billions of DNA strands are sequenced in parallel, yielding substantially more throughput and minimizing the need for the fragment-cloning methods that are often used in Sanger sequencing of genomes.

General Information- the CMV genome and CMV strains

The double-stranded DNA (dsDNA) genome of wild-type CMV strains consists of 2 each unique long (UL) and unique short (US) regions, connected at their ends by identical small amino acid sequences (internal repeat sequences) that are arranged in opposite directions (inverted) (Figure 1). The UL and US regions can flip around from right to left during replication, resulting in four possible isomers of the viral genome (10). CMV has evolved with the human host, and the origin of the most recent common root for the α (e.g. Herpes simplex), β (e.g. cytomegalovirus), and γ (e.g. Epstein Barr Virus) herpesviruses is estimated to have occurred about 400 million years ago (11). Under the selection of the host immune response, each herpesvirus developed solutions to allow for its survival. In addition to enzymes and proteins involved in DNA replication and structural proteins, CMV encodes for proteins that can affect virulence through immune evasion and molecular mimicry (12). These functions which can potentially limit host defenses against lytic or latent infection are encoded by CMV genes that play immunomodulatory roles resulting in more efficient virus replication. Genes affecting virulence have been identified in both UL and US regions of the CMV genome (Table 1). The first instance of molecular mimicry in CMV was identified as an MHC-I homolog and is designated as UL18. UL18 encodes a protein with 20% similarity to classical human MHC-I proteins (13). MHC-I molecules present peptides derived from pathogens and cell proteins on the cell surface. Cytotoxic T cells (CTLs) recognize and lyse virus-infected cells through engagement of the T cell receptor with MHC-I molecules presenting viral antigens at the surface of infected cells. The prevention of MHC-I expression allows these cells to escape CTL detection. Viral proteins can interfere with class I antigen presentation. Interference in antigen presentation and consequent MHC-I down-regulation on the cell surface might allow infected cells to evade virus-specific CTLs, but down-regulation of MHC-I molecules makes these cells susceptible to lysis by NK cells. UL18, MHC-I homolog acts as a decoy to block NK cell cytotoxicity.

Sequencing of clinical CMV isolates obtained from different cohorts of CMV-infected patients reveals that several CMV genes have highly different areas of sequences within apparently important genes. These areas of difference within a gene are called variable regions or polymorphisms. When multiple variable sequences are found in a given gene, that has been designated as a hyper variable region.

The first complete sequence of a CMV genome was published in 1990 (21) and differences between the laboratory-adapted strains, known as AD169 and Towne, and the closer-to-wild-type Toledo strain were later localized to open reading frames (ORFs) located in the UL/b' region of the genome (7). These ORFs are thought to play a role in CMV replication and/or disease *in vivo* since they were lost upon extensive passage *in vitro*. While the majority of genetic loci within the UL and US regions are highly conserved, proteins produced by genes in the UL/b' region are among those associated with immune evasion.

These proteins include cytokine and chemokine homologs that can have remarkable sequence variability (hypervariable).

However, these areas did not change much if at all during ongoing replication in any given patient, nor did they seem to change much when spreading from patient to patient. This means that known genotypes are mostly stable (19,22). In addition, linkage between specific genes (linkage: a pair or group of genes being tightly held together during DNA replication so that they are more likely than not to be found together in any CMV in which any one of them is found) was rare, there were exceptions. For example, UL16 and UL147 genes were linked (1,23).

What is the best source for genotypic studies?

CMV replicates in multiple cells and tissues *in vivo*. In organ and bone marrow transplant recipients serial viral load testing from whole blood or plasma guides decisions for initiation of anti-CMV therapy and duration of therapy (24). However, in congenitally-infected children viral loads from blood may rise or fall in the same patient unpredictably and the duration of viremia may be short (25). Generally, saliva and urine samples provide the most reliable and informative sources for CMV genotyping. Saliva was reported to have slightly better yield for CMV detection by PCR in congenitally-infected children compared to urine (26). Multiple passages of virus replication in cell culture can induce mutations in genes or complete loss of some genes that might slow or be unnecessary for growth in human fibroblasts (the cell line in which CMV is usually grown in the lab). So, the question was whether mutations discovered during sequencing were really present in the originally sampled body fluid. Or might they represent an artifact after passing in tissue culture? Sequences of CMV genes have been compared from CMV-infected original samples and tissue-culture isolates from the same original sample to determine whether mutations growth in culture could select for mutations. In general, isolates collected from the same patient showed excellent correlation for several genetic loci between the original and multiple passaged isolates (1). However, in a study of four clinical strains passed in several cell lines in tissue culture, CMV genes RL13 and the UL128 locus (consisting of genes UL128, UL130 and UL131A) mutated quickly during passage in human cells during standard sequencing. Mutations in RL13 occurred in fibroblast, epithelial and endothelial cells, and those in the UL128 locus were limited to fibroblasts and detected later than RL13 (27). Several clinical isolates analyzed directly by next generation sequencing (NGS) did not reveal mutations in these genes (28), but mutations in several gene loci were present in CMV isolated from original body fluids. Mutations were identified in CMV genes RL5A, UL9, UL111A, and UL150 that were culture-independent, i.e. spontaneous mutations *in vivo*. Future analysis of multiple clinical samples by NGS will assist in defining mutations across the CMV genome and their potential role as markers for disease outcome.

The era pre high-throughput /NGS

Two main classes of CMV genes have been studied in detail in samples collected from congenitally-infected children: 1) The surface glycoproteins-B (gB), -H (gH), -N (gN), and -O (gO); and 2) The cytokine/chemokine homologs of human cellular genes that had been hijacked into the CMV genome. The underlying hypothesis was that polymorphisms in

CMV genes that encode for proteins targeted by host immune response, such as envelope glycoproteins, as well as virus-captured human homologs of cytokines and chemokines could be associated with variability in controlling CMV infection.

Genotypes for CMV could be defined by differences in a single gene or some combination of linked genes. There is currently no consensus on what constitutes a clinically relevant genotype. Conflicting data exist for the single gene genotypes investigated to date.

Genetic variability in the tumor necrosis factor receptor like gene, UL144

a truncated tumor necrosis factor- α (TNF- α)-like receptor gene (UL144) was discovered in a region of the CMV genome deleted in highly-passaged laboratory strains, but present in all fresh clinical isolates (7). Sequence variability of UL144 was originally reported from 45 low-passage clinical isolates, revealing 3 major UL144 genotypes which were genetically unlinked with the four previously described major genotypes of the UL55 gene encoding for glycoprotein B (gB) (29). Similar variability was demonstrated in low-passage isolates collected from children attending day care centers and congenitally-infected infants in both Texas and Iowa. No linkage of UL144 genotype with gB subtypes was found (30), meaning that any gB genotype could be present with any UL144 genotype. Children from the two different geographic locations were found to shed the same gB and or UL144.

Whether more severe outcomes of congenital CMV are related to specific UL144 genotypes remains controversial. For example, we showed an association with UL144 genotypes A and C, and recombinants (A/B and A/C) ($p=0.04$) (31) using isolates retrospectively collected from congenitally infected children who were school-aged and whose outcome was definitively known. This is important since some children who appear asymptomatic early in life may develop progressive hearing loss (symptomatic congenital CMV) that is diagnosed later, e.g. near school entry. We also found that UL144 genotype C was associated with other severe outcomes, for example pregnancy termination was associated with genotype C from amniotic fluids collected in Italy ($p=.03$) (18) and genotype C was marginally associated symptomatic disease in newborns (odds ratio, 8.81 [95% confidence interval, 0.48-164.02]; $p=.05$) – note the wide confidence interval encompassing 1.0.

Despite this, we found that overall UL144 genotype C was not necessarily associated with higher viral loads in amniotic fluids. This was unexpected, given that higher viral loads are thought to predict worse outcomes in congenital CMV infection. In fact, genotype B (the most common UL144 genotype found in congenitally-infected children) seemed more predictive of higher viral loads in symptomatic newborns than asymptomatic infants ($p=.003$).

Another study of UL144 genotypes in congenitally-infected children had clinical follow-up at ages 12 - 18 months. It revealed that UL144 genotypes A and C were associated with a high plasma viral load ($p < 0.04$) and both genotypes A and C were significantly associated with worse developmental outcome (32). Further follow-up of this cohort for outcome of congenital CMV infection should assist in confirming these genotype data.

Genotyping during the CMV and Hearing Multicenter Screening (CHIMES) study also revealed that UL144 genotype C was more common in symptomatic children but interestingly not specifically associated with sensorineural hearing loss (16).

Despite these studies implicating UL144 genotype C with more severe outcomes, several other reports did not support an association between specific UL144 genotypes and outcome of congenital CMV infection (16,33-36).

A large sample collection from multicenter trials of congenitally-infected infants that has several years of follow-up, strict definition of symptomatic vs asymptomatic CMV disease and outcome measures, in conjunction with whole genome sequencing might clarify the role of UL144 genotypes or CMV proteins that may interact with UL144 genotypes as potential diagnostic markers.

Genetic variability in UL146 and UL147

Two α -chemokine genes, UL146 and UL147, are physically adjacent in the CMV genome. They have consistently shown extreme hypervariability (polymorphism) in all tested CMV isolates regardless of the patient population or geographical location from which the CMV isolates were obtained (19,20,36-43). The 12-14 distinct UL146 and UL147 genotypes differ by over 60% at the amino acid level (19,40). Despite there being many described genotypes of UL146 and UL147, new genotypes did not arise over timeframe analyzed to date (19,36). This indicates that rapid sequence drift in UL146 and 147 is not the underlying mechanism of this gene variability.

Polymorphisms have been noted throughout UL146, and also within a small region of UL147 corresponding to a possible signal peptide (19,37). CMV isolates obtained from multiple members of families with a congenitally-infected infant were sequenced. The UL146 genotype from different isolates cultured from any given family was the same within that family (41). In each family, an older sibling of the congenitally infected infant rather than the father was the virus source for most pregnant women.

We reported extreme hypervariability of UL146 and UL147 in 23 congenitally-infected children (20). Because of the small sample size and large number of genotypes for these genes, no specific UL146 or UL147 genotype was associated with disease outcome. In a study of 121 infants from Poland (32 of whom were congenitally-infected newborns), genotype G1 was more prevalent in symptomatic infants; and an association between UL146 genotype and viruria was identified (38). Nevertheless, much larger studies with hundreds of isolates will be required to identify a correlation between UL146/UL147 genotypes and outcome of congenital infection.

US28

CMV encodes for a functional β -chemokine receptor (US28) that binds and sequesters extracellular chemokines and may alter protective host immune responses (44). Although there are much fewer (only 5) US28 genotypes compared to UL144, UL146/UL147, no

association between a specific US28 genotype and outcome of congenital CMV infection is apparent (16,18).

Glycoprotein N (UL73)

In Italian cohorts four genotypes gN (UL73) have been identified and all were transmitted from mother to child. Glycoprotein N-1 (gN-1) was associated with improved long-term outcome (45). In a later study from the same group gN-1 and gN-3a genotypes were associated with no symptoms at birth and a favorable long-term outcome, while gN-4 genotype was associated with symptoms at birth, abnormal imaging results, and sequelae ($p < .05$) (46). However, in the CHIMES study, the 7 gN genotypes did not correlate with symptomatic CMV disease (16).

Glycoprotein O (UL74)

Four genotypes of UL74 have been identified, with high level of sequence variability among the genotypes. Sequence variability among clinical strains was up to 46% in amino acids, including deletions and amino acid changes (47). Similar genotypes were found in HIV and non-HIV-infected individuals. Genetic linkage between gN and gO has been reported (48).

Glycoprotein H (UL75)

The gH gene was sequenced in the CHIMES study along with gN and gB. No particular genotype (5 for UL55, 2 for UL75 and 7 for UL73) was associated with symptomatic infection or sensorineural hearing loss (SNHL) (16).

Glycoprotein B (UL55)

Glycoprotein B (gB) mediates virus attachment and entry into infected cells, cell-to-cell transmission and syncytium formation (49). A significant proportion of neutralizing antibody to CMV in human serum is specific for epitopes on gB, and essentially all CMV-infected individuals have antibody to gB(50). Thus, it is not surprising that the gB gene has been commonly sequenced from both adults and children. Despite some suggested correlation, it appears that none of the five major gB genotypes are predictors of congenital CMV outcome (16,18,51), although genotype 3 was more common in congenitally-infected neonates with sensorineural hearing loss ($p = 0.009$) (35).

Mixed infection with multiple CMV strains

Mixed infections have traditionally been reported from immunocompromised hosts. Multiple strains were detected in AIDS patients (52), and transplant recipients (53,54). Previously CMV-seropositive transplant recipients became infected with new CMV strains from their donors (54). Multiple CMV strains were detected in blood from 5 of 11 CMV-infected renal transplant patients (22), all of whom were seronegative prior to transplantation. Thus, infection with multiple strains in these patients could have occurred through acquisition of strains from normal seropositive donors who harbored and transmitted multiple CMV strains or the transplant recipients were susceptible to multiple exposures at different time points. In addition to acquisition of new strains over time,

replacement of strains can occur (23). Immunocompromised hosts can become a reservoir for generation of new CMV strains *in-vivo* through recombination (55).

Multiple CMV strains have been reported far less commonly in normal hosts and usually when repeated exposures to new sources of CMV would be expected. For example, four of eight women attending a sexually transmitted disease (STD) clinic excreted distinct CMV strains simultaneously at different sites or shed new strains when studied serially (56). Infection with new CMV strains has also been reported in children attending day care centers (57,58). Previously immune mothers were reported to acquire a new CMV strain during pregnancy and transmit it to their offspring (59). Multiple CMV strains were identified in autopsy tissues after in-utero death as a result of disseminated CMV infection (18). One recent report on strain diversity in healthy CMV-seropositive women suggested that 15/16 women who tested PCR positive (total 2.9% PCR positivity in that cohort) were infected with more than one strain based on genotyping gN and gB from blood or urine samples (60). However, the time of primary infection was unknown in these women, so it is not possible to draw conclusions as to diversity of CMV strains acquired around seroconversion. In addition, the detection of multiple genotypes from blood samples in that cohort may suggest that this specific population was at risk for infection with multiple strains. However, information on potential risks or immunodeficiency was not available. The detection of a single or multiple CMV strains was investigated in a cohort of healthy seronegative women who participated in a CMV glycoprotein B (gB) vaccine clinical trial, followed until they acquired CMV, and then for up to 3 years afterwards (1). CMV genotyping was based on direct PCR sequencing at several distinct gene loci including UL55, UL144, UL146 and UL9 from different body fluids collected at intervals after infection. CMV genotypes were also compared between original body fluids and after cell culture isolation, to address whether growth in culture might select for strains that grow more efficiently and therefore mask the detection of multiple strains. In 53/55 women one CMV strain was detected, based on sequencing of UL55 and UL144. There was no evidence for acquisition of new strains or of sequence changes based on testing of multiple genetic loci. All known genotypes of gB, UL144 and UL146 were identified similar to those reported from other studies (19,61). DNA sequence data obtained from cultured isolates matched the sequences obtained from original fluids, suggesting that original urine or saliva samples in which high virus load is achieved may provide the most useful or efficacious source for genotyping without the need for culturing. Although most women had their own unique strain, 25% of women enrolled in the study at different times and with no known epidemiologic links shed the same CMV strain based on sequencing of four variable genetic loci. Since this CMV strain was found to be shared based on comparison of hypervariable genotypes (UL55, UL144, UL146, UL9) it is unlikely that sequencing of additional loci would have further distinguished the strains. These data suggest that a common CMV strain can be shed and circulate in a specific geographical region during a short time.

Despite these data in young women with recent primary CMV infection, results from the CHIMES study have shown that congenitally-infected children may be infected with more than one CMV strain. In that study, mixed infections were detected in 45% of infants, both symptomatic and asymptomatic. These data need additional confirmation and inter-

laboratory validation (16). It is also possible that the women from the gB vaccine trial could acquire new CMV strains in longer follow-up studies. Altogether, the identification of multiple CMV strains from seemingly normal asymptomatic hosts raises questions about the clinical significance of infection with multiple strains.

The era of Next Generation Sequencing (NGS)

This era is marked by a massive increase in the throughput and resolution of the genomics field. The NGS approach analyzes the total DNA present in an isolate, thus, it is not only targeting viral DNA. DNA extracted from cells grown in culture contains mainly cellular DNA and low DNA copies of viral DNA. In one study only 3% of sequence reads collected directly from the sample were of viral origin, but it was possible to reconstruct the complete genome using these 3% of reads (62). This approach may be relevant to situations in which CMV DNA is abundant in original samples. DNA amplification steps may assist in sequencing the CMV genome from samples low in CMV DNA.

A study using 454 GS FLX (Roche®) technology analyzed CMV populations by sequencing PCR amplicons of the hypervariable genes UL73, UL74, and UL139 from lung transplant recipients (63). Viral populations were identified with mixtures of up to six genotypes. Usually, one or two genotypes represented the majority of the population while the other genotypes were present at frequencies below 10%. Testing of serial samples collected from the patients revealed fluctuation in the genotype abundance although the sequence itself did not change, again confirming the stability of CMV genes. The authors appropriately noted that PCR amplification may result in artificial recombinants and cautioned about problems in interpretation of deep sequencing experiments (64,65).

The complete CMV genome was assembled using overlapping PCR amplicons (66,67). BAC clones of AD169 and Toledo were re-sequenced to establish a quality filtering threshold for distinguishing real intra-host variants from PCR errors. In a first study, the viral populations of three congenitally-infected infants were characterized. The genetic diversity of viral populations was comparable to quasi-species RNA viruses such as HIV. Variants present at high frequencies (> 90%) accounted for 20% of reads, while low-frequency variants (< 10%) represented 73% of reads, which corroborated with the study of UL73, UL74 and UL139 amplicons in lung transplant recipients (63). Compared to the diversity found in the congenitally-infected infants, intra-host diversity was lower in malignant gliomas (68). One might suggest that during early life there might be higher replication capacity of CMV, which might lead *de novo* mutations. However, this is not supported by other studies showing stability of hypervariable CMV genes in immunocompromised hosts. It is also important to distinguish between the finding of random variants and subtype variants, the latter being the subject of investigation for a genetic marker for disease severity. The concept of intra-host diversity requires testing of additional samples from other patient populations, as well as development of criteria for sequence standardization of viral genomes. In another study, serial urine and plasma isolates were collected from five congenitally-infected infants (59). In the same compartment the consensus sequences differed only by 0.2% at the nucleotide level, indicating stability of the viral populations. However, in one child comparison of isolates collected concurrently from

urine and plasma revealed a 1% variability. The observed dynamics were modeled using demographic variables (population size and structure) and selective pressures. While the effects of positive selection in a specific compartment were small, evidence for positive selection was found when comparing different compartments. It is possible that this child was infected with two different strains, but additional studies may provide data related to the importance of sequencing samples from different compartments using NGS.

The ability to analyze outcomes from a large data set of genetic polymorphisms was demonstrated by use of artificial neural network. Using sequence data from the hypervariable genes UL144, UL146, UL147 and US28 the network was trained and correctly predicted congenitally-infected infants to be symptomatic or asymptomatic at birth in 90% of cases (69). This technique can be applied to analyze outcomes from whole genome sequences.

Summary

Studies for the identification of a viral marker of CMV disease outcome has generated more debates than consensus, but several themes have been reproducible (Box 2). There is no doubt that investigations should continue involving well-controlled populations, inter-laboratory validation and standardization of both genotypes and strains. NGS can be useful for subtype categorization and may assist in clear definitions of strains.

Acknowledgments

I thank my mentor Dr. Gary Hayward, for ongoing teaching and wonderful advice and support over the years.

References

1. Murthy S, Hayward GS, Wheelan S, Forman MS, Ahn JH, Pass RF, Arav-Boger R. Detection of a single identical cytomegalovirus (CMV) strain in recently seroconverted young women. *PLoS One*. 2011; 6:e15949. [PubMed: 21264339]
2. Yow MD, Demmler GJ. Congenital cytomegalovirus disease--20 years is long enough. *N Engl J Med*. 1992; 326:702–703. [PubMed: 1310526]
3. Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The “silent” global burden of congenital cytomegalovirus. *Clin Microbiol Rev*. 2013; 26:86–102. [PubMed: 23297260]
4. Demmler GJ. Infectious Diseases Society of America and Centers for Disease Control. Summary of a workshop on surveillance for congenital cytomegalovirus disease. *Rev Infect Dis*. 1991; 13:315–329. [PubMed: 1645882]
5. Stagno, S. Cytomegalovirus. In: K, J.; Remington, JS., editors. *Infectious disease of the fetus and newborn infant*. 4. Saunders WB; Philadelphia: 1995. p. 312–353.
6. Stagno S, Pass RF, Dworsky ME, Henderson RE, Moore EG, Walton PD, Alford CA. Congenital cytomegalovirus infection: The relative importance of primary and recurrent maternal infection. *N Engl J Med*. 1982; 306:945–949. [PubMed: 6278309]
7. Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol*. 1996; 70:78–83. [PubMed: 8523595]
8. Brown JM, Kaneshima H, Mocarski ES. Dramatic interstrain differences in the replication of human cytomegalovirus in SCID-hu mice. *J Infect Dis*. 1995; 171:1599–1603. [PubMed: 7769298]
9. Quinnan GV Jr, Delery M, Rook AH, Frederick WR, Epstein JS, Manischewitz JF, Jackson L, Ramsey KM, Mittal K, Plotkin SA. Comparative virulence and immunogenicity of the Towne strain

and a nonattenuated strain of cytomegalovirus. *Ann Intern Med.* 1984; 101:478–483. [PubMed: 6089634]

10. Mocarski, ES. Cytomegaloviruses and their replication. In: Fields, BN.; Knipe, DM.; Howley, PM., editors. *Fields Virology*. third. Lippincott Williams & Wilkins; 1996. p. 2447-2492.
11. McGeoch DJ, Gatherer D, Dolan A. On phylogenetic relationships among major lineages of the Gammaherpesvirinae. *J Gen Virol.* 2005; 86:307–316. [PubMed: 15659749]
12. Prichard MN, Penfold ME, Duke GM, Spaete RR, Kemble GW. A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. *Rev Med Virol.* 2001; 11:191–200. [PubMed: 11376481]
13. Beck S, Barrell BG. Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens. *Nature.* 1988; 331:269–272. [PubMed: 2827039]
14. Garrigue I, Faure-Della CM, Magnin N, Recordon-Pinson P, Couzi L, Lebrette ME, Schrive MH, Roncin L, Taupin JL, Dechanet-Merville J, Fleury H, Lafon ME. UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. *Transplantation.* 2008; 86:826–835. [PubMed: 18813108]
15. Heatley SL, Pietra G, Lin J, Widjaja JM, Harpur CM, Lester S, Rossjohn J, Szer J, Schwarer A, Bradstock K, Bardy PG, Mingari MC, Moretta L, Sullivan LC, Brooks AG. Polymorphism in human cytomegalovirus UL40 impacts on recognition of human leukocyte antigen-E (HLA-E) by natural killer cells. *J Biol Chem.* 2013; 288:8679–8690. [PubMed: 23335510]
16. Pati SK, Pinninti S, Novak Z, Chowdhury N, Patro RK, 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. [PubMed: 23694837]
17. Deckers M, Hofmann J, Kreuzer KA, Reinhard H, Edubio A, Hengel H, Voigt S, Ehlers B. High genotypic diversity and a novel variant of human cytomegalovirus revealed by combined UL33/UL55 genotyping with broad-range PCR. *Virol J.* 2009; 6:210. [PubMed: 19941648]
18. Arav-Boger R, Willoughby RE, Pass RF, Zong JC, Jang WJ, Alcendor D, Hayward GS. Polymorphisms of the cytomegalovirus (CMV)-encoded tumor necrosis factor- α and beta-chemokine receptors in congenital CMV disease. *J Infect Dis.* 2002; 186:1057–1064. [PubMed: 12355354]
19. Lurain NS, Fox AM, Lichy HM, Bhorade SM, Ware CF, Huang DD, Kwan SP, Garrity ER, 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. [PubMed: 16409621]
20. Arav-Boger R, Foster CB, Zong JC, Pass RF. Human Cytomegalovirus-Encoded alpha - Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns. *J Infect Dis.* 2006; 193:788–791. [PubMed: 16479512]
21. Chee MS, Bankier AT, Beck S, Bohni R, Brown CM, Cerny R, Horsnell T, Hutchison CA III, Kouzarides T, Martignetti JA. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol.* 1990; 154:125–169. [PubMed: 2161319]
22. Stanton R, Westmoreland D, Fox JD, Davison AJ, Wilkinson GW. Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J Med Virol.* 2005; 75:42–46. [PubMed: 15543586]
23. Rasmussen L, Geissler A, Winters M. Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis.* 2003; 187:809–819. [PubMed: 12599055]
24. Kotton CN, Kumar D, Caliendo AM, Asberg A, Chou S, Danziger-Isakov L, Humar A. Updated international consensus guidelines on the management of cytomegalovirus in solid-organ transplantation. *Transplantation.* 2013; 96:333–360. [PubMed: 23896556]
25. Arav-Boger R, Pass RF. Diagnosis and management of cytomegalovirus infection in the newborn. *Pediatr Ann.* 2002; 31:719–725. [PubMed: 12455480]
26. Ross SA, Ahmed A, Palmer AL, Michaels MG, Sanchez PJ, Bernstein DI, Tolan RW Jr, Novak Z, Chowdhury N, Fowler KB, Boppana SB. Detection of congenital cytomegalovirus infection by real-time polymerase chain reaction analysis of saliva or urine specimens. *J Infect Dis.* 2014; 210:1415–1418. [PubMed: 24799600]

27. Dargan DJ, Douglas E, Cunningham C, Jamieson F, Stanton RJ, Baluchova K, McSharry BP, Tomasec P, Emery VC, Percivalle E, Sarasini A, Gerna G, Wilkinson GW, Davison AJ. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J Gen Virol.* 2010; 91:1535–1546. [PubMed: 20479471]
28. Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, Thomson M, Griffiths PD, Wilkinson GW, Schulz TF, Davison AJ. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol.* 2010; 91:605–615. [PubMed: 19906940]
29. Lurain NS, Kapell KS, Huang DD, Short JA, Paintsil J, Winkfield E, Benedict CA, Ware CF, Bremer JW. Human cytomegalovirus UL144 open reading frame: sequence hypervariability in low-passage clinical isolates. *J Virol.* 1999; 73:10040–10050. [PubMed: 10559318]
30. Bale JF Jr, Petheram SJ, Robertson M, Murph JR, Demmler G. Human cytomegalovirus a sequence and UL144 variability in strains from infected children. *J Med Virol.* 2001; 65:90–96. [PubMed: 11505449]
31. Arav-Boger R, Battaglia CA, Lazzarotto T, Gabrielli L, Zong JC, Hayward GS, Diener-West M, Landini MP. Cytomegalovirus (CMV)-encoded UL144 (truncated tumor necrosis factor receptor) and outcome of congenital CMV infection. *J Infect Dis.* 2006; 194:464–473. [PubMed: 16845629]
32. Waters A, Hassan J, De GC, Kissoon G, Knowles S, Molloy E, Connell J, Hall WW. Human cytomegalovirus UL144 is associated with viremia and infant development sequelae in congenital infection. *J Clin Microbiol.* 2010; 48:3956–3962. [PubMed: 20810771]
33. Picone O, Costa JM, Chaix ML, Ville Y, Rouzioux C, Leruez-Ville M. Human cytomegalovirus UL144 gene polymorphisms in congenital infections. *J Clin Microbiol.* 2005; 43:25–29. [PubMed: 15634946]
34. Nijman J, Mandemaker FS, Verboon-Macielek MA, Aitken SC, van Loon AM, de Vries LS, Schuurman R. Genotype distribution, viral load and clinical characteristics of infants with postnatal or congenital cytomegalovirus infection. *PLoS One.* 2014; 9:e108018. [PubMed: 25268349]
35. 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. [PubMed: 18273679]
36. Heo J, Petheram S, Demmler G, Murph JR, Adler SP, Bale J, Sparer TE. Polymorphisms within human cytomegalovirus chemokine (UL146/UL147) and cytokine receptor genes (UL144) are not predictive of sequelae in congenitally infected children. *Virology.* 2008; 378:86–96. [PubMed: 18556037]
37. Arav-Boger R, Zong JC, Foster CB. Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes.* 2005; 31:65–72. [PubMed: 15965610]
38. Paradowska E, Jablonska A, Plociennikowska A, Studzinska M, Suski P, Wisniewska-Ligier M, Dzierzanowska-Fangrat K, Kasztelewicz B, Wozniakowska-Gesicka T, Lesnikowski ZJ. Cytomegalovirus alpha-chemokine genotypes are associated with clinical manifestations in children with congenital or postnatal infections. *Virology.* 2014; 462-463:207–217. [PubMed: 24999045]
39. Aguayo F, Murayama T, Eizuru Y. UL146 variability among clinical isolates of human cytomegalovirus from Japan. *Biol Res.* 2010; 43:475–480. [PubMed: 21526275]
40. Bradley AJ, Kovacs IJ, Gatherer D, Dargan DJ, Alkharsah KR, Chan PK, Carman WF, Dedicoat M, Emery VC, Geddes CC, Gerna G, Ben Ismaeil B, Kaye S, McGregor A, Moss PA, Pusztai R, Rawlinson WD, Scott GM, Wilkinson GW, Schulz TF, Davison AJ. Genotypic analysis of two hypervariable human cytomegalovirus genes. *J Med Virol.* 2008; 80:1615–1623. [PubMed: 18649324]
41. Revello MG, Campanini G, Piralla A, Furione M, Percivalle E, Zavattoni M, Gerna G. Molecular epidemiology of primary human cytomegalovirus infection in pregnant women and their families. *J Med Virol.* 2008; 80:1415–1425. [PubMed: 18551604]

42. Hassan-Walker AF, Okwuadi S, Lee L, Griffiths PD, Emery VC. Sequence variability of the alpha-chemokine UL146 from clinical strains of human cytomegalovirus. *J Med Virol.* 2004; 74:573–579. [PubMed: 15484281]
43. He R, Ruan Q, Qi Y, Ma YP, Huang YJ, Sun ZR, Ji YH. Sequence variability of human cytomegalovirus UL146 and UL147 genes in low-passage clinical isolates. *Intervirology.* 2006; 49:215–223. [PubMed: 16491016]
44. Pleskoff O, Casarosa P, Verneuil L, Ainoun F, Beisser P, Smit M, Leurs R, Schneider P, Michelson S, Ameisen JC. The human cytomegalovirus-encoded chemokine receptor US28 induces caspase-dependent apoptosis. *FEBS J.* 2005; 272:4163–4177. [PubMed: 16098198]
45. Pignatelli S, Dal Monte P, Rossini G, Lazzarotto T, Gatto MR, Landini MP. Intrauterine cytomegalovirus infection and glycoprotein N (gN) genotypes. *J Clin Virol.* 2003; 28:38–43. [PubMed: 12927749]
46. Pignatelli S, Lazzarotto T, Gatto MR, Dal MP, Landini MP, 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. [PubMed: 20504230]
47. 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. [PubMed: 12368327]
48. 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. [PubMed: 18753237]
49. Navarro D, Paz P, Tugizov S, Topp K, La Vail J, Pereira L. Glycoprotein B of human cytomegalovirus promotes virion penetration into cells, transmission of infection from cell to cell, and fusion of infected cells. *Virology.* 1993; 197:143–158. [PubMed: 7692667]
50. Pass RF, Zhang C, Evans A, Simpson T, Andrews W, Huang ML, Corey L, Hill J, Davis E, Flanigan C, Cloud G. Vaccine prevention of maternal cytomegalovirus infection. *N Engl J Med.* 2009; 360:1191–1199. [PubMed: 19297572]
51. Yamamoto AY, Mussi-Pinhata MM, de Deus Wagatsuma VM, Marin LJ, Duarte G, Figueiredo LT. Human cytomegalovirus glycoprotein B genotypes in Brazilian mothers and their congenitally infected infants. *J Med Virol.* 2007; 79:1164–1168. [PubMed: 17597488]
52. Baldanti F, Sarasini A, Furione M, Gatti M, Comolli G, Revello MG, Gerna G. Coinfection of the immunocompromised but not the immunocompetent host by multiple human cytomegalovirus strains. *Arch Virol.* 1998; 143:1701–1709. [PubMed: 9787655]
53. Puchhammer-Stockl E, Gorzer I, Zoufaly A, Jaksch P, Bauer CC, Klepetko W, Popow-Kraupp T. Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation.* 2006; 81:187–194. [PubMed: 16436961]
54. Chou SW. Acquisition of donor strains of cytomegalovirus by renal-transplant recipients. *N Engl J Med.* 1986; 314:1418–1423. [PubMed: 3010114]
55. Chou SW. Reactivation and recombination of multiple cytomegalovirus strains from individual organ donors. *J Infect Dis.* 1989; 160:11–15. [PubMed: 2543706]
56. Chandler SH, Handsfield HH, McDougall JK. Isolation of multiple strains of cytomegalovirus from women attending a clinic for sexually transmitted disease. *J Infect Dis.* 1987; 155:655–660. [PubMed: 3029241]
57. Bale JF Jr, Petheram SJ, Souza IE, Murph JR. Cytomegalovirus reinfection in young children. *J Pediatr.* 1996; 128:347–352. [PubMed: 8774502]
58. Bale JF Jr, O'Neil ME, Fowler SS, Murph JR. Analysis of acquired human cytomegalovirus infections by polymerase chain reaction. *J Clin Microbiol.* 1993; 31:2433–2438. [PubMed: 8408568]
59. Boppana SB, Rivera LB, Fowler KB, Mach M, Britt WJ. Intrauterine transmission of cytomegalovirus to infants of women with preconceptual immunity. *N Engl J Med.* 2001; 344:1366–1371. [PubMed: 11333993]

60. Novak Z, Ross SA, Patro RK, Pati SK, Kumbla RA, Brice S, Boppana SB. Cytomegalovirus Strain Diversity in Seropositive Women. *J Clin Microbiol.* 2008; 46:882–886. [PubMed: 18216215]
61. Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, Lee L, Addison C, Dargan DJ, McGeoch DJ, Gatherer D, Emery VC, Griffiths PD, Sinzger C, McSharry BP, Wilkinson GW, Davison AJ. Genetic content of wild-type human cytomegalovirus. *J Gen Virol.* 2004; 85:1301–1312. [PubMed: 15105547]
62. Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, Thomson M, Griffiths PD, Wilkinson GW, Schulz TF, Davison AJ. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol.* 2010; 91:605–615. [PubMed: 19906940]
63. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *J Virol.* 2010; 84:7195–7203. [PubMed: 20463084]
64. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. *J Virol Methods.* 2010; 169:248–252. [PubMed: 20691210]
65. Zong JC, Arav-Boger R, Alcendor DJ, Hayward GS. Reflections on the interpretation of heterogeneity and strain differences based on very limited PCR sequence data from Kaposi's sarcoma-associated herpesvirus genomes. *J Clin Virol.* 2007; 40:1–8. [PubMed: 17698410]
66. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog.* 2011; 7:e1001344. [PubMed: 21625576]
67. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, Jensen JD, Kowalik TF. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLoS Genet.* 2013; 9:e1003735. [PubMed: 24086142]
68. Bhattacharjee B, Renzette N, Kowalik TF. Genetic analysis of cytomegalovirus in malignant gliomas. *J Virol.* 2012; 86:6815–6824. [PubMed: 22496213]
69. Arav-Boger R, Boger YS, Foster CB, Boger Z. The use of artificial neural networks in prediction of congenital CMV outcome from sequence data. *Bioinform Biol Insights.* 2008; 2:281–289. [PubMed: 19812782]

Box 1**Definitions**

We have used the following definitions to describe the genetic content of CMV (1):

Strain: Overall description of viral genomic structure based on the sum of all loci tested. Strains can be distinguished from one another by differences in one locus or multiple loci.

Genotype: A combination of subtype designations based on sequence data obtained from two or more gene loci

Subtype: Cluster pattern based on sequence data at any one gene locus

Variant: Minor nucleotide changes observed within a subtype

Isolate: CMV recovered from a human specimen and passaged in culture a limited number of times

Clustering: represents a clade of closely-related genomes

Box 2**The genetics of Congenital CMV**

- The strain detected from mother is the same strain transmitted to her child
- Transmission has not been reported to depend on a specific strain
- The same strains are detected in the same human family
- In most congenital infections one strain is detected
- Mixed infection have been reported with unclear relevance to disease progression
- Urine or saliva isolates that have been minimally passaged in cell culture can provide could source for CMV genotyping
- NGS can be used in the future for genotype classification in large cohorts

Key Points

- Genetic diversity of Cytomegalovirus strains has been reported in children and adults.
- Sequence variability exists in CMV genes that may affect immune responses and viral dissemination.
- Viruses transmitted from mother to child share the same genetic content. Placental transmission appears to be independent of a specific virus strain.
- Sequencing of several hypervariable genes from original samples or low-passage virus isolates have generally yielded the same genetic information.
- Overall, there is no linkage between the different variable CMV genes, resulting in a very large number of CMV strains circulating in the population.
- Next generation sequencing may allow better classification of strains once standardization of sequencing between laboratories is accomplished.

Figure 1.
The double-stranded DNA (dsDNA) genome of wild-type CMV strains.

Table 1
Potential Virulence Factors in the CMV Genome

Gene/Homology	Function	Number of Genotypes (Reference)
US3/ none	Interferes with MHC-I/II processing so CMV antigen not presented efficiently to host immune system	11 (14)
UL40/ HLA-E homolog	Protects infected cells against NK cytotoxicity	12 (14), (15)
US28/ G-protein coupled receptor	Promotes chemotaxis	5 (16)
US27/ G- protein coupled receptor	Extracellular virus spread	N/A
UL33, UL78/chemokine receptors	Regulation of cellular chemokine receptors (CXCR4 and CCL5)	4 (17)
UL144/cytokine receptor	Inhibition of T-cell proliferation, induction of chemokines via NF-KB	3 (18)
α -chemokine	neutrophil chemotaxis	14 (19) (20)
UL147/ α -chemokine	unknown	11 (19,20)
UL36/ *	Anti-apoptotic	N/A
UL37/ *	Viral mitochondria inhibitor of apoptosis (vMIA)	N/A
UL18/MHC-I homolog	NK cell decoy	10 (14)
UL111A/ IL10 homolog	Inhibits MHC class II expression	12 (14)

N/A- not reported from clinical isolates.

* Functional (not sequence) homology to host anti-apoptotic genes