Strain Variation and Disease Severity in Congenital Cytomegalovirus Infection: In Search of a Viral Marker

Ravit Arav-Boger, MD

KEYWORDS

• Cytomegalovirus • Congenital infection • Strains • Genotypes • Multiple strains
• Immune evasion genes • Population-based sequencing
• Next-generation sequencing

KEY POINTS

- Genetic diversity of cytomegalovirus (CMV) 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 has 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.

INTRODUCTION

Infection with cytomegalovirus (CMV) is the leading cause of congenital infection and the most common infectious source of central nervous system (CNS) damage and sensorineural hearing loss (SNHL) in the United States.[1,2] The outcome of

Division of Infectious Diseases, Department of Pediatrics, Johns Hopkins University School of Medicine, 200 North Wolfe Street, Baltimore, MD 21287, USA
E-mail address: boger@jhmi.edu

Infect Dis Clin N Am 29 (2015) 401–414
http://dx.doi.org/10.1016/j.idc.2015.05.009 **id.theclinics.com**
0891-5520/15/$ – see front matter

congenital CMV infection is highly variable. Of infected infants, around 10% to 15% exhibit severe symptoms at birth. Of the 85% to 90% who are asymptomatic at birth, 10% to 15% will later develop hearing loss and other neurologic deficits.[3,4] 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.[5] The severity of fetal disease varies widely, ranging from stillbirth due to multisystem 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.[6,7] 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.[8] 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 before availability of sensitive high-throughput sequencing, only selected genes have been sequenced from the CMV genome, which ranges from 220 to 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 addresses the following topics, among others:

1. Genetic variability of CMV strains—genotypic analysis
2. Next-generation sequencing (NGS) and its implications for identification of a genetic marker
3. The role of multiple CMV strains in outcome of congenital CMV infection.

GENETIC DEFINITIONS

Locus: a specific location of DNA sequence (**Box 1**)

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

Open reading frame (ORF): 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 antiviral immunity. Identical or similar amino acid sequences shared between pathogens and the host are responsible for molecular mimicry.

NGS: 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.