Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Biologie
Studijní obor: Genetika, molekulární biologie a virologie

Bc. Jan Dvořák

Studium celogenomové variability lidského cytomegaloviru
The study of the whole genome sequence variability of cytomegalovirus isolates

Diplomová práce

Vedoucí závěrečné práce: RNDr. Ruth Tachezy, Ph.D.
Konzultant: Steven Sijmons, MSc.

Praha, 2013

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 10.8.2013

Podpis

Abstrakt

Tato práce je součástí projektu zabývající se variabilitou HCMV, jejímž dlouhodobým cílem je zmapovat geografické rozložení genotypů, zjistit vliv variability genotypů na průběh HCMV asociovaných onemocnění a vytipovat úseky genomu s potenciálním využitím pro diagnostiku a preventivní terapii. Práce je zaměřena na metodiku přípravy materiálu z klinických izolátů HCMV pro nově vyvinutou metodu celogenomového sekvenování (next-generation sequencing, NGS) a na postupy pro vyhodnocení získaných sekvenačních dat.

V práci byly použity vzorky krve a moči od pacientů po transplantaci hematopoetických kmenových buněk a od kongenitálně infikovaných dětí. Vybrané vzorky vhodné pro NGS byly sekvenovány platformou Illumina a sekvence složeny postupem anglicky označovaným jako *de novo* assembly v kombinaci s mapping assembly.

Vzorky moči ve srovnání s krví vykazovaly vyšší výtěžnost materiálu pro NGS. Ze vzorků pozitivních na HCMV DNA (7 z 50) po amplifikaci na tkáňových kulturách měl pouze jeden vysoký podíl virové buněčné DNA (98 %), u 6 vzorků činil tento podíl méně než 7 %. Ze vzorku obsahujícího 98% virové DNA byla vytvořena sekvence celého genomu a srovnána se sekvencemi dalších klinických izolátů z Belgie v 11 polymorfních oblastech. Analýza těchto oblastí ukázala, že pouze 2 izoláty vykazují shodu ve všech 11 oblastech. Srovnání sekvencí celých genomů klinických izolátů HCMV svědčí o jejich vysoké variabilitě. Analýzou bylo zjištěno, že jednotlivé geny tvoří vzájemně nezávislé shluky. Pro stanovení genetických odlišností mezi izoláty bude třeba sekvenovat obrovské množství izolátů, což umožňuje pouze metodika NGS.

Klíčová slova

Cytomegalovirus, celý genom, variabilita, sekvenování nové generace, kongenitální infekce, imunosuprimovaní pacienti, izolace, klinické izoláty, genotypy

Abstract

This work is part of a project focused on the study of the variability of human cytomegalovirus (HCMV) among clinical isolates with the aim to map the geographical distribution of HCMV genotypes, reveal the relationships between genotypes and the severity of HCMV-associated diseases, and identify regions in the HCMV genome with a potential for use as diagnostic and therapeutic targets. Attention was paid to the development of the methodology for the preparation of the material for next-generation sequencing (NGS) from HCMV clinical isolates and evaluation of the obtained sequencing data.

Blood and urine samples collected from hematopoietic stem cell transplantat recipients and congenitally infected children were analyzed. Samples suitable for NGS were sequenced by the Illumina platform and sequences were created by *de novo* assembly followed by mapping assembly.

Urine samples in comparison to blood samples had higher yield of material for NGS. Of the samples positive for HCMV DNA (7 of 50) after amplification in the cell cultures, only one sample had high purity of the viral DNA (98%) while six samples had purity of less than 7%. The sample containing 98% of the viral DNA was fully sequenced and the sequence was compared to the sequences of other clinical isolates from Belgium in 11 polymorphic regions. Only two isolates were identical in all 11 studied regions. The analysis of the whole genome sequences has shown high variability between HCMV clinical isolates. Futhermore, HCMV genes were found to cluster independently. To find out about the patterns of clinical isolates, a lot of samples have to be sequenced. This is possible thanks to the NGS methodology.

Keywords

Human cytomegalovirus, whole genome, variability, next-generation sequencing, congenital infection, immunosuppresed patients, isolation, clinical isolates, genotypes

Contents

Abbreviations

BAC	Bacterial artificial chromosome
CPE	Cytopathic effect
CLS	Cleavage site
D+, D-	Seropositive donor, Seronegative donor
DE	Delayed – early genes
DNA	Deoxyribonucleic acid
E-MEM	Eagle-Minimum Essential Medium
EBV	Epstein-Barr virus
FBS	Fetal bovine serum
g.e.	Genome equivalents
gB, gO, gN etc.	Glycoprotein B, O, N etc.
GPRC	G protein-coupled receptor
HCMV	Human cytomegalovirus
HHV-5, 6, 7 and 8	Human herpesvirus 5, 6, 7 and 8
HIV	Human immunodeficiency virus
HSCT	Hematopoietic stem cell transplantation
HSV-1, HSV-2	Herpes simplex virus 1 and 2
IE	Immediate – early genes
INF-γ	Interferon gamma
IRL	Long internal repeat
IRS	Short internal repeat
L	Late genes
NGS	Next generation sequencing
MCMV	Mouse cytomegalovirus
MCP	Major capsid protein
MIRA	Mimicking Intelligent Read Assembly
ORF	Open reading frame
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
qPCR	quantitative Real - time PCR
R+, R-	Seropositive recipient, Seronegative recipient
SCP	Smallest capsid protein
rpm	rounds per minute
SOT	Solid organ transplantation
TRL	Long terminal repeat
TRS	Short terminal repeat
TNF-α	Tumor necrosis factor α
UL	Unique long
US	Unique short
VZV	Varicella-Zoster virus

1. Introduction

Human cytomegalovirus (HCMV) is widespread worldwide. The seroprevalence rates in the general population are very high, ranging from 30% to 90%.

The infection of HCMV occurs most often in young age and the major route of transmission is by direct contact with bodily secretions. In an immunocompetent host, the virus usually causes primary asymptomatic infection, after which it establishes lifelong latency and periodically reactivates. Severe complications of HCMV infection can occur in congenitally infected children, immunosuppressed transplant recipients, or persons infected with HIV (human immunodeficiency virus).

These patients are monitored for treatment initiation. Two approaches, i.e. preemptive and prophylactic therapy with antiviral drugs, are applied for the treatment of HCMV infection. But long-term use of these drugs leads to toxicity and resistance of the virus to therapy. There is an intensive research directed to the development of prophylactic vaccines. Several vaccines undergo clinical trials, but none has been available for clinical use.

In clinical isolates of HCMV, genetic polymorphisms in multiple genes were observed. In these genes, diverged clusters of allelic variants (genotypes) exist. The most studied are the genes which are important for virion entry into the host cell, cell tropism, and immune evasion. It is assumed that different genotypes might have different pathogenicity.

Most studies published to date have derived the conclusion from the analysis of a limited number of polymorphic regions. Most studies utilized the restriction fragment length polymorphism (RFLP) method or Sanger sequencing. Only five studies have reported so far using cloning of the HCMV genome into plasmids or bacterial artificial chromosomes (BAC) and/or using overlapping polymerase chain reaction (PCR) products followed by Sanger sequencing. These approaches allowed for the full genome sequencing of the HCMV isolates.

A new chapter was opened by the development of methods for the next-generation sequencing (NGS), which allow for the sequencing of the whole large genomes. These new sequencing methods yield increasing numbers of clinical isolates with fully sequenced genomes. Since 1990 till 2005, the sequences of the whole genome of only 10 HCMV isolates were published. The first reports of data generated with the use of NGS date back to 2005, the first report of full genome sequencing of HCMV

with the use of NGS appeared in 2009. Since 2009, additional seven whole genome sequences generated with NGS methods were reported.

However, the next-generation sequencing has also drawbacks: (a) high acquisition price, (b) high requirements for computer technology, and (c) quality of clinical isolates, particularly in terms of purity and amount of viral DNA. Clinical isolates should contain higher amount of viral DNA than human DNA, ideally more than 60%. Requirements for the amount of input DNA material depend on the type of the kit for preparation of DNA but in general it is at least 50 ng.

Reactivation of the latent infection in immunocompetent people is difficult to detect but in immunocompromised and congenitally infected children, HCMV infection leads to prolonged virus excretion in the urine. In these patients, virus is also present in high amounts in blood and the evaluation of the viral presence and viral load is used for the monitoring of patients and treatment initiation. Therefore, to obtain DNA for sequencing, clinical isolates of HCMV, usually from the urine of immunosuppressed patients or congenitally infected children, are propagated in tissue cultures of human fibroblasts.

Obtaining a high-quality sample for NGS is not an easy task: urine can be toxic for cells due to the therapy, it is not possible to exactly know when uremia occurs, and HCMV DNA isolated from cell cultures is often contaminated by a high proportion of human DNA.

2. Aims of the Thesis

The main objective of this work was the development and improvement of the methodology for the preparation of HCMV DNA for the whole genome sequencing and possibly isolation and characterization of the whole genome of new Czech isolates. These isolates were compared with those obtained in the cooperating Laboratory of Clinical Virology, KU Leuven, Belgium and will be a part of the long-term evaluation and identification of HCMV isolates from different geographical regions. This project is focused on studying geographic distribution of HCMV genotypes, evaluating influence of the genotypes on the type and severity of HCMV disease, and identifying potential regions in the HCMV genome which can be targeted for preventive therapy.

3. Literature review

3.1 Taxonomy

Human cytomegalovirus (HCMV), also known as Human herpesvirus 5 (HHV-5), belongs to the *Herpesviridae* family. This family is divided into three different subfamilies α, β, and γ. The α-subfamily is characterized by a relatively short reproductive cycle, variable host range, destruction of productively infected cells, and establishment of latency in ganglia. The α-herpesviruses have the smallest genome. Two genera from this subfamily infect mammalian hosts: *Simplexvirus* and *Varicellovirus*. Human pathogens - Herpes simplex virus 1 and 2 (HSV-1 and HSV-2) belong to the genus *Simplexvirus* and varicella-zoster virus (VZV) to the genus *Varicellovirus*. Viruses from subfamily γ infect mainly T and B lymphocytes of mammalian hosts. Latent infection is frequently demonstrated in lymphoid tissue. This subfamily includes two human pathogens: Epstein-Barr virus (EBV), belonging to the genus *Lymphocryptovirus*, and human herpesvirus 8 (HHV-8), classified into the genus *Rhadinovirus*. The human cytomegalovirus (HCMV) belongs to the *β-Herpesviridae subfamily*, which is characterized by a long life cycle, narrow host range, enlarged infected cells (cytomegalia), and the establishment of latency in secretory glands, lymphoreticular cells, and other tissues. Subfamily β includes the following genera: *Roseolovirus* with human herpesvirus 6 and 7 (HHV-6, HHV-7), *Cytomegalovirus* with HCMV, and *Muromegalovirus* with mouse cytomegalovirus (MCMV) (Pellett and Roizman 2007).

3.2 Virion structure

The structure of HCMV is typical for herpesviruses (Fig. 1). It consists of a nucleocapsid, tegument, and envelope. The size of the mature virion is 200 -300 nm and depends on the thickness of tegument.

The HCMV capsid is 135 nm in size and has icosahedral symmetry with a triangulation number T = 16. It contains 162 capsomers consisting of 150 hexamers which form the triangular surface and of 12 pentamers which make up the vertices of icosahedron. The capsid is composed of five core proteins: (a) major capsid proteins (MCP, encoded by the UL86 gene), (b) the minor capsid protein (TRI1, UL46 gene), (c)

minor capsid protein-binding protein (TRI2, UL86 gene), (d) the smallest capsid protein (SCP, UL48A gene), and (e) portal protein (PORT, UL104 gene).

Between capsid and envelope, a relatively amorphous substance called the tegument is located. The tegument contains abundant proteins coded by the virus, viral proteins present in a small amount, cellular proteins, and viral RNA. Among the most abundant tegument proteins are pp65 (UL83), pp71 (UL82), pp150 (UL32), pUL47, and pUL48. These proteins are phosphorylated and are important for the initiation and regulation of the virus replication cycle, directing cells for virus replication and virion maturation.

The tegument is surrounded by an envelope which is formed by a lipid bilayer containing virus-encoded envelope glycoproteins gB (UL55), gM (UL100), gN (UL73), gH (UL75), gL (UL115), and gO (UL74). These glycoproteins play an important role in virus attachment and entry into the cell (Mocarski *et al*. 2007)

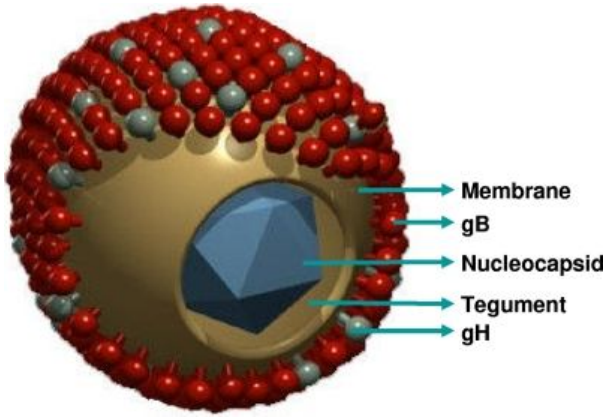


Figure 1. Three-dimensional model of HCMV structure. In this model, various components of the virus are shown. (Adapted from Crough and Khanna 2009)

3.3 Genome

3.3.1 Genome structure

The HCMV genome is formed by linear double stranded DNA of approximately 235 kbp (kilo base pair) in size, which makes it one of the largest viral genomes. The HCMV genome has an E class organization, which is characterized by two unique domains, unique long (UL) and unique short (US), flanked by terminal and internal repeats in the arrangement a_nb-UL-b´a´$_n$c´-US-ca (Fig. 2). Terminal repeat long (TRL) include n copies of the **a** sequence next to the larger **b** sequence. The second terminus is

terminal repeat short (TRS) containing directly repeated **a** sequence next to the **c** sequence. Terminal ab and ca sequences are inserted in an inverted orientation (internal repeat long - IRL and internal repeat short - IRS) separating the genome into two unique domains. The UL and US can invert and create four equimolar isomers of viral DNA with different orientations of the UL and US fragments.

Figure 2. Schematic structure of the HCMV genome. The region $\mathbf{a_n b}$ shows TRL, **b´a´** shows IRL, $\mathbf{a'_n c}$ shows IRS and **ca** is TRS. UL is the unique long domain and US is the unique short domain. (Adapted from Pellet and Roizman 2007)

The HCMV genes encoding three different temporal classes of proteins such as immediate early (IE or α), delayed early (DE or β), and late (L or γ). Products of IE genes regulate transcription of DE and L and gene expression of the host cell and also suppress cell death. Proteins encoded by DE genes regulate replication of viral DNA and products of the L gene are viral structural proteins.

The HCMV genomes contain several cis-elements: replication origin oriLyt is located approximately in the middle of the UL domain and the elements pac-1 and pac-2 in the terminal repeats. These elements are important for encapsidation of the viral DNA (Murphy and Shenk, 2008). Additionally, also the enhancer box and start of transcription belong within the cis elements that are involved in the regulation of transcription of the IE genes (Mocarski *et al.* 2007).

The first conventional ORF (open reading frame) map was constructed for strain AD169 (TRL1-14-UL1-132-IRL14-1-IRS-1-US1-34-TRS1) (Fig. 3) (Chee *et al.* 1990). However, for clinical strains with the known whole-genome sequence, differences in the ORF arrangement were found (RL1-14-UL1-151-IRS-1-US1-34-TRS1) (Fig. 3). Unlike AD169, clinical isolates additionally contain unique DNA segments RL1-14 and UL133–UL151 and lack an IRL repeat (Cha *et al.* 1996, Davison *et al.* 2003, Murphy *et al.* 2003, Dolan *et al.* 2004).

Figure 3. Conventional maps of laboratory strain AD169 and a clinical isolate. (Adapted from Murphy *et al.* 2003)

As a reference strain for clinical isolates, the Merlin strain is usually used. Based on the analyses of the genome of this strain, a consensus genetic map of wild-type HCMV was generated. In this map, genes are divided into gene families named RL11, UL14, UL18, UL82, UL120, UL146, US1, US2, US6, US12, US22, UL25, and GPRC (Fig. 4). Analysis of the genomes of the *Herpesviridae* family has shown that there are genes (core genes) which are conserved among all members of the *Herpesviridae* family (Dolan *et al.* 2004). It is predicted that AD169 has the capacity to contain 208 ORFs. A segment containing 19 additional ORFs was found in the Toledo clinical strain and these additional ORFs have also been detected in other clinical isolates (Cha et al., 1996). A detailed analysis of four clinical strains led to estimates that the maximum number of potentially functional protein-coding ORFs can be 252 (Murphy *et al.* 2003).

3.3.2 Sequenced complete HCMV genome

To date the complete genome sequences of two highly passaged laboratory strains AD169 and Towne have been published. Additionally, sequences of two AD169 variants and one variant of the Towne strain are available: varUK (Chee *et al.* 1990), varATCC, and variant varS (Murphy *et al.* 2003). Complete sequences have also been reported for the low passage clinical strain, Toledo, FIX (VR1814), TR, PH (Murphy *et al.* 2003), Merlin (Dolan *et al.* 2004), TB40E (Sinzger *et al.* 2008), U8, U11, AF1 (Dargan *et al.* 2010), 3157 (Cunningham *et al.* 2010) and strain JP from clinical material (Cunningham *et al.* 2010). Sequences of these strains were determined by random shotgun cloning to standard bacterial plasmid libraries (AD169 varUK) and bacterial artificial chromosomes (AD169 varATCC Toledo, FIX (VR1814), TR, PH). Purified virion DNA was sequenced by PCR sequencing which involves generating numerous overlapping PCR products from the HCMV genome and analyzing them by Sanger sequencing (Merlin, U8, U11, AF1, 3157, and JP).

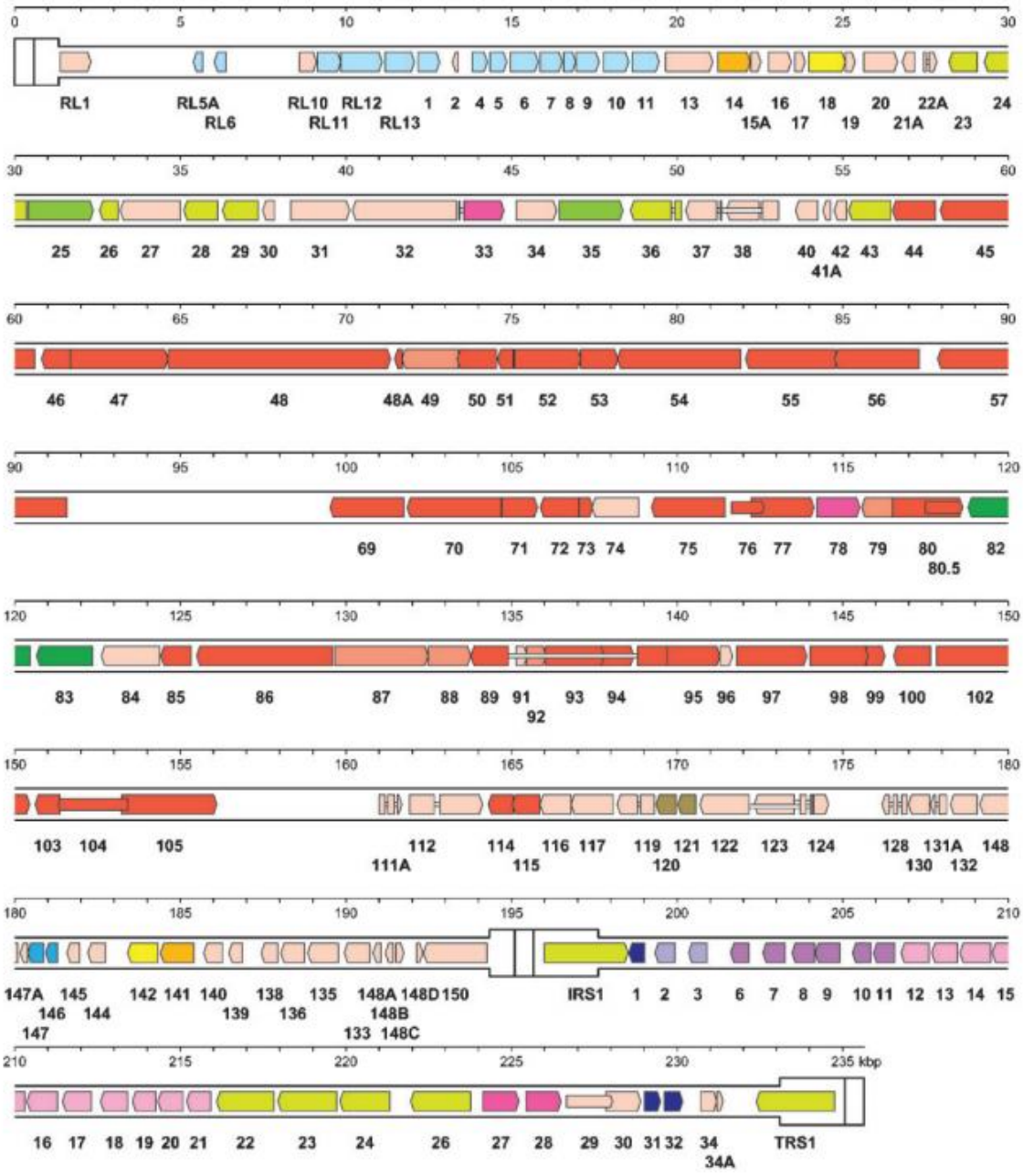


Figure 4. Consensus genetic maps of wild-type HCMV strains based on the Merlin genome. The dark red colour shows genes which are conserved among all members of Herpesviridae (core genes). These genes are located in the central UL domain of protein coding regions. Other color codes indicate the membership in gene families. The US and UL prefixes have been omitted. (Adapted from Dolan *et al.* 2004)

With the development of novel high-throughput sequencing methods, the number of the published complete genomes of both laboratory (AD169 varUC and Towne varL) (Bradley *et al.* 2009) and clinical strains (HAN13, HAN20, HAN38, 3301, JHC) (Cunningham *et al.* 2010, Jung *et al.* 2011) has rapidly increased. All these strains, except for JHC strain which was analyzed using a 454 pyrosequencer, were sequenced by the Illumina genome analyzer.

Most HCMV genome sequences published to date have been derived from strains grown in cell culture. However, growth in cell culture exerts selective pressure and induces genetic adaptation. At least the sequence of the whole genome from two strains derived directly from clinical specimens (JP and 3301) have been published (Dolan *et al.* 2004, Cunningham *et al.* 2010, Dargan *et al.* 2010,).

3.4 Life cycle

3.4.1 Lytic life cycle

Infection in healthy people typically starts with replication in mucosal epithelium at the entry site of infection. In immunocompetent individuals, viral DNA is detected in peripheral blood mononuclear cells (PBMC) but not in the plasma, neutrophils, or endothelial cells. The presence of the virus during the acute phase of the infection has only been studied in immunocompromised people. In these patients, endothelial, epithelial, and hematopoietic cells in the peripheral blood and tissues contained the virus (Mocarski *et al.* 2008).

HCMV replicates in primary fibroblasts, secondary fibroblasts, or myeloid cells. The replication cycle is slow. Immediately after entry of the viral DNA into the nucleus, transcription of IE genes from four different regions occurs: (a) UL36 and 37, (b) IE-1 and IE-2, (c) TRS-1 and IRS-1, (d) and US-3. The main feature of the gene products IE-1 and IE-2 is the activation of transcription of DE and L genes, but these proteins have also other functions such as the inhibition of interferon production or induction of cell cycle arrest. In the later stages of infection, products of DE and L genes regulate the expression of IE genes.

In the next stage of the viral replication cycle, transcription of DE genes occurs. These DE gene products subsequently initiate the replication of viral DNA. Moreover, they prepare the internal environment of the cell for the replication and participate in capsid maturation. The most important are the products of the UL112 and UL113 genes

that encode four proteins which initiate DNA replication by alternative splicing and the UL54 gene which encodes the catalytic subunit of DNA polymerase. HCMV DNA replication starts in oriLyt by transcriptional activation of the participating IE transactivator. At the beginning of the replication, the circularization of viral DNA occurs and directs the transactivator to the point of the origin of replication. At the beginning, the replication is done by the theta mechanism and then by the rolling circle replication mechanism.

The transcription of L genes is divided into transcription γ1 and γ2 on the basis of the time of initiation of transcription. The transcription of the γ1 genes requires the transcription of replication inhibitors and for the transcription of the γ2 genes, inhibition of replication is necessary. Products of L genes, together with the products of DE genes, are involved in the assembly of the capsid, encapsidation of DNA, virion maturation and release from the cells (Mocarski *et al.* 2007).

3.4.2 Latent life cycle

The establishment of latent HCMV infection most likely occurs in everyone who experienced primary infection. Latent infection is usually established in the myeloid stem cells (CD34+) which are located in the bone marrow (Mendelson *et al.* 1996). HCMV causes latent infection also in monocytes CD14 + (Taylor-Wiedeman *et al.* 1991), dendritic cells (Senechal *et al.* 2004), and megakaryocytes (Crapnell *et al.* 2000) located in the peripheral blood of healthy individuals.

In latently infected cells, HCMV may be reactivated and switched to the lytic phase of infection. The triggering stimulus can be immunosuppression, inflammation, or stress. The exact mechanism is not known, but it is assumed that the tumor necrosis factor (TNF-α) plays an important role. TNF-α binds to the TNF receptor of latently infected cells and stimulates the replication of the IE gene through protein kinase C and NF-kB (Stein *et al.* 1993). The TNF-α as well as substances produced during stress (e.g. adrenaline and noradrenaline) (Prosch *et al.* 2000) or during inflammation (Kline *et al.* 1998) also stimulate transcription via the cyclic adenosine monophosphate (cAMP) pathway.

In vitro studies suggest that reactivation of latent HCMV infection is dependent on the activation of specific pathways of differentiation of monocytes which consequently stimulate the replication of HCMV IE genes. The results also indicate that

reactivation of latent HCMV infection in macrophages requires T cell activation and production of INF-γ (Söderberg-Nauclér *et al.* 2001).

3.5 Epidemiology

The HCMV is universally present in the human population. In Europe and North America, the presence of antibodies against HCMV was detected in 30-70 % of the population in reproductive age (20-35 years), but in developing countries, the seroprevalence exceeds 90 % (Rodier *et al.* 1995). Slightly higher prevalence was observed in women than in men. Seroprevalence increases with age and is therefore dependent on the age distribution of the population (Balcarek *et al.* 1990, Roubalová and Seeman 1998, Hecker *et al.* 2004, Staras *et al.* 2006).

Close contact with body fluids (such as saliva, blood, breast milk, etc.) is required for efficient transmission because HCMV is unstable in the environment. Transplacental transmission can result in congenital infection. The primary infection with HCMV occurs most often in early age and most commonly, the virus is transmitted during the childbirth or via breastfeeding. Transmission of the virus from mother to fetus or newborn is common and plays an important role in maintaining infection in the population. Transmission of the virus from child to child and from child to adults is also very effective. Virus can also be transmitted through blood transfusions, solid organ transplantation (SOT), and hematopoietic stem cell transplantation (HSCT). In adolescents, sexual activity is a risk factor for HCMV infection and the high rates of HCMV infection between sex partners suggest sexual transmission of the virus (Mocarski *et al.* 2007).

3.6 Diseases caused by HCMV

In an immunocompetent host, the virus usually causes primary asymptomatic infection, after which it establishes lifelong latency and periodically reactivates. The most common manifestation of symptomatic primary infection in immunologically competent individuals is CMV mononucleosis. Severe complication of CMV infection can occur in pregnant women if they have primary CMV infection in the first trimester. HCMV also causes very severe complications in immunosuppressed transplant recipients or HIV-infected individuals.

3.6.1 Congenital infection

The prevalence of congenital HCMV infection ranges between 0.2 % and 2.2 % (Pignatelli *et al* 2010). The greatest risk of congenital HCMV infection is in women who undergo primary infection during the first trimester of pregnancy. About 45 % of HCMV-seronegative pregnant women are at risk of primary infection and 1-4% of them acquire it. The rate of transplacental (mother-to-fetus) transmission of primary maternal infection during pregnancy is approximately 40%. A congenitally acquired HCMV infection is asymptomatic in 90 % of cases and only 5-15 % of children with asymptomatic prenatally acquired infection develop pathological manifestations. The symptomatic infection occurs in 10-15 % of infected children, with severe consequences for the newborns in up to 90 % of them (Stagno and Whitley 1985). Congenital infection may be a cause of abortion, defects of the fetus, congenital abnormalities, and neonatal death. The most common consequences of congenital HCMV infection are severe neurological defects such as microcephaly, encephalitis, and psychomotor and mental retardation. Congenital HCMV infection is the most common infectious cause of hearing loss. Some other symptoms such as jaundice, hepatosplenomegaly, thrombocytopenia, and hepatitis usually clear spontaneously within a few weeks or months after birth. Symptomatic congenital infection leads to neonatal death in about 15% of cases due to serious damage to vital organs (Boppana *et al* 1992).

3.6.2 Solid organ transplantation

Immunosuppressive therapy, which reduces the risk of transplant rejection, is mostly involved in the reactivation of latent HCMV infection, seroconversion, and development of diseases associated with HCMV. Other important contributors are donor and recipient serostatus and stress during surgery (Cook *et al.* 1998). HCMV infection can be detected in up to 50 % of SOT (solid organ transplantation) recipients, of which 10-50 % will develop symptomatic disease (Rubin 2007). The highest risk for the development of the disease is when the recipient is seronegative (R-) and the donor is seropositive (D+) for HCMV because the seronegative recipient has not developed specific immunity against HCMV. It has been shown that the viral load after SOT is the highest in recipients with the R-/D+ combination (Xia *et al.* 2000). The risk of developing the disease can be reduced by matching seronegative donors (D-) to seronegative recipients (R-), but this approach would greatly restrict the availability of organs.

Primary infection with HCMV in R- patients affects the transplanted organ first as inflammatory reaction, which can cause graft malfunction and rejection. Among the complications caused by systemic spread of the virus are pneumonitis, hepatitis, and enteritis, of which the most severe complication is pneumonitis (Ljungman *et al.* 2002) with a high case-fatality rate (Peterson *et al.* 1980). The incidence of HCMV-associated pneumonitis is dependent on the type of organ transplant (Ho 2008).

3.6.3 Hematopoietic stem cell transplantation

In contrast to SOT recipients, HCMV infection in hematopoietic stem cell transplant (HSCT) recipients is often due to reactivation of latent HMCV (in 80% of subjects) in previously seropositive patients. Immunosuppression and graft deprivation of T-cells in seropositive recipients (R+) prolongs the time required for the revival of HCMV-specific and innate immunity, mainly CD4+ and CD8+ T-lymphocytes. The low number of CD4+ and CD8+ T lymphocytes leads to increased risk of developing the disease due to the inability to prevent reactivation of latent infection (Ozdemir *et al.* 2002, Boeckh *et al.* 2003).

Primary infection with HCMV has been reported in 30% of previously HCMV-seronegative HSCT recipients (R-) (Ljungman 2007). Among seronegative HSCT recipients from HCMV-seropositive donors (D+R-), higher incidence of pneumonitis and opportunistic bacterial and fungal infections has been observed (Ljungman *et al.* 2002). Moreover, the D+R- group has shown higher mortality in comparison to D-R- (Nichols *et al.* 2002). The effect of donor HCMV serostatus on the prognosis of HCMV disease in the D+R+ group is more controversial. Improved survival has been documented in this group (D+R+) compared with D-R+. This observation can be explained by the fact that HCMV-specific immune cells are transferred from the seropositive donor (Ljungman *et al.* 2003). However, other studies have reported no difference (Boeckh and Nichols 2004).

3.7 Polymorphisms in HCMV genome

Even though the genomes of different HCMV strains are 95% similar, they contain polymorphic sequences both in coding and non-coding regions. According to the degree of variation, genes are divided into those with a basal level of polymorphism

and those called hypervariable genes. Individual polymorphic regions can be divided into several highly diverged clusters of alleles called genotypes.

The most extensively studied hypervariable genes are those which are thought to affect the pathogenicity of different genotypes. The first group of hypervariable genes encode envelope glycoproteins gB (UL55), gM (UL100), gN (UL73), gH (UL75), gL (UL115), and gO (UL74). Other highly variable genes are located in the UL/b´ region, e.g. UL144, UL146, and UL139. These genes are missing in the highly passaged laboratory strain AD169. The comparison of AD169 and the Merlin strain, the prototype clinical isolate, has revealed very high variability in genes of the RL11 family (RL5A, RL6, RL12, RL13, UL1, UL9, and UL11), particularly in the RL12 and RL13 genes (Dolan *et al.* 2004, Pignatelli *et al.* 2004).

High rates of divergence in the envelope glycoproteins of the UL/b´ and RL11 gene families can be explained by the exposure to components of the immune system which drive the changes. However, it can be hypothesized that the changes occur slowly because the sequence of a particular genotype is stable over time when sampled from the same patients (Hassan-Walker *et al.* 2004, Bradley *et al.* 2008, Görzer *et al.* 2010) as well as in cell culture (Lurain *et al.* 2006).

Co-infection with multiple genotypes often occurs in transplant recipients, HIV/AIDS patients, and congenitally infected children as reported in many studies (Görzer *et al.* 2010, Manuel *et al.* 2009, Bates *et al.* 2008, Pignatelli *et al.* 2010, Görzer *et al.* 2010, and Ross *et al.* 2011).

Genetic linkage studies have revealed no relationship between genotypes. A comparative analysis of six polymorphic regions (UL55, UL74, UL75, UL115, US9, and US28) has shown that apart from the interaction suggested between the gN and gO genotypes (Mattick *et al.* 2004), genetic linkages are rare and infinite numbers of genetic combinations are theoretically possible (Rasmusen *et al.* 2003). The multiple positivity of patients and absence of genetic linkages between polymorphic regions of different genotypes add to the evidence that genetic diversity among HCMV strains can be explained by recombination of viral variants present in one patient as a result of re-infection or co-infection (Pignatelli *et al.* 2004, Gorzer *et al.* 2010, Renzette *et al.* 2011).

3.7.1 Envelope glycoproteins

Envelope glycoproteins create glycoprotein complexes gCI (gB), gCII (gN + gM), and gCIII (gO + gH + gM) and play an essential role in the viral life cycle.

Glycoprotein B (gCI complex)

This glycoprotein is essential for HCMV attachment and entry, cell fusion, and cell-to-cell spread (Navarro *et al.* 1993, Tugizov *et al.* 1994). It is (906 aa) encoded by ORF UL55 and is transcribed as polyprotein which is cleavage by cellular endoprotease at the cleavage site (CLS) at position 460. This cleavage produces the N-terminal (gb 116) and C-terminal (gb55) products (Spaete *et al.* 1990) which are finally linked by disulfide bonds (Britt *et al.* 1989). The greatest variability has been detected in the CLS area. On the basis of variability in the CLS region, four main genotypes, gB-1, gB-2, gB-3, and gB-4, have been identified (Chou *et al.* 1991, Fries *et al.* 1994, Meyer-König *et al.* 1998) in clinical isolates but other rare genotypes have also been described such as gB-5 (Shepp *et al.* 1998), gB-6, and gB-7 (Trincado *et al.* 2000).

In HIV/AIDS patients, increased occurrence of the gB2 genotype has been observed in the urine and blood (Rasmussen et al., 1997, Drew et al., 2002, Roubalová et al., 2009) and the gB2 genotype has been reportedly connected with homosexual transmission (Zipeto *et al.* 1998, Roubalová *et al.* 2009).

Among SOT recipients, higher prevalence of the gB1 genotype (25%) but also co-infection with multiple genotypes in nearly 50 % of cases have been reported (Aquino et al., 2000, Manuel et al., 2009). In other studies, lower prevalence of multiple infection has been found (Humar *et al.* 2003, Coaquette *et al.* 2004). In patients with multiple infection and D+R+ serostatus, lower incidence of tissue-invasive disease has been reported (Manuel *et al.* 2009). No clear effect of known genotypes on the course and severity of the CMV disease has been established (Aquino *et al.* 2000, Humar *et al.* 2003).

In HSCT recipients, the most often detected genotype is gB1, followed by gB3. The gB1 genotype seems to be associated with lower viral load, milder infection, and higher risk of graft versus host disease (Roubalová *et al.* 2010). A study of Woo *et al.* has shown that the gB2 genotype is more prevalent in patients with worse prognosis or even fatal CMV disease (Woo *et al.* 1997).

In congenitally infected children, gB1 and gB3 have been found by several authors as the most common genotypes (Barbi *et al.* 2001, Jin *et al.* 2007, Roubalová *et al.* 2009). In European countries, gB4 has also been reported (Arista *et al.* 2003, Roubalová *et al.* 2009) but has not been detected in other geographical areas (Yamamoto *et al.* 2007). All four gB genotypes can cause congenital infection; however, association with the outcome of intrauterine HCMV infection for each of them is not clear (Barbi *et al.* 2001, Arista *et al.* 2003). In the study of Yan et al., no relationship between genotypes and viral load in the urine (Yan *et al.* 2008) and the outcome of the congenital infection has been found. Co-infection with multiple gB genotypes is very common in congenitally infected children. The presence of different genotypes in specimens collected from different anatomic sites in newborns has been documented (Ross et al., 2011).

gCII complex (gM and gN)

Products of the gCII gene complex are involved in the virus attachment to the host cell.

Glycoprotein M is encoded by ORF UL100 and is highly conserved because most nucleotide changes are silent mutations. In this glycoprotein, only a small variable region was detected (about 25 nucleotides).

Glycoprotein N is a transmembrane protein encoded by ORF UL73. This glycoprotein is expressed in the late stage of the viral replication cycle. It has been shown that it is able to induce neutralizing antibody (Mach *et al.* 2000). Detailed sequence analysis of ORF UL73 and its gene products from clinical isolates and laboratory-adapted strains have shown that gN is highly polymorphic especially in the N-terminal region. UL73 hypervariability is clearly clustered into seven distinct genotypes: gN-1, gN-2, gN-3a, gN-3b, gN-4a, gN-4b, and gN-4c (Pignatelli *et al.* 2001 and 2003).

Studies into the relation between gN genotypes and clinical outcome have suggested that gN-1 can represent a less virulent phenotype, while the gN-4 genotype is predominantly associated with severe manifestation (Pignatelli *et al.* 2003 and 2004, Rossini *et al.* 2005). The gN-1 has been shown to be less virulent also in congenitally infected children. Children with gN-1 infection have been reported to be at lower risk of sequelae in comparison to gN-4 (Pignatelli *et al.* 2010). Co-infection with multiple gN genotypes has been detected in both congenitally infected children (Ross *et al.* 2011) and SOT recipients (Görzer *et al.* 2010).

gCIII complex (gH, gL and gO)

Glycoprotein H (encoded by ORF UL75) is an immunologically dominant HCMV envelope glycoprotein which induces virus neutralizing antibodies. gH is important for mediating cell-host membrane fusion and is also essential for replication in cell culture. Variability of this glycoprotein is mainly restricted to the first 37 N-terminal amino acids, but overall it is very conserved, the overall amino acid variability is only 4.6%. Two genotypes have been identified so far, gH-1 and gH-2 (Chou *et al.* 1992).

Glycoprotein L (encoded by ORF UL115) is necessary for transport of gH from the endoplasmic reticulum to the cell surface. Major variability was detected in the first 45 amino acids but the overall amino acid difference among clinical isolates is only 6.5%. According to sequence similarity, four genotypes, gL-1, gL-2, gL-3, and gL-4, were characterized.

Glycoprotein O (encoded by ORF UL74) is important for virus-mediated cell fusion (Paterson *et al.* 2002). This glycoprotein is not required for propagation of the virus in fibroblasts as other envelope glycoproteins but enhances viral replication (Hobom *et al.* 2000). Comparison of the sequence of gO from one clinical isolate with that of the HCMV strain AD169 showed high degree of nucleotide variability, which is particularly pronounced in the N-terminal domain (40%) (Fig. 5). The observed amino acid sequence similarity between AD169 and the Toledo strain was 66% (Paterson et al., 2002), while comparison of clinical isolates with AD169 revealed 46% similarity. On the basis of the phylogenetic analysis of 90 clinical isolates, eight distinct genotypes, i.e. gO-1a, gO-1b, gO-1c, gO-2a, gO-2b, gO-3, gO-4, and gO-5, were identified (Rasmussen *et al.* 2002, Mattick *et al.* 2004). The highest variability was detected in the gO3 and gO2b genotypes (Roubalová *et al.* 2011).

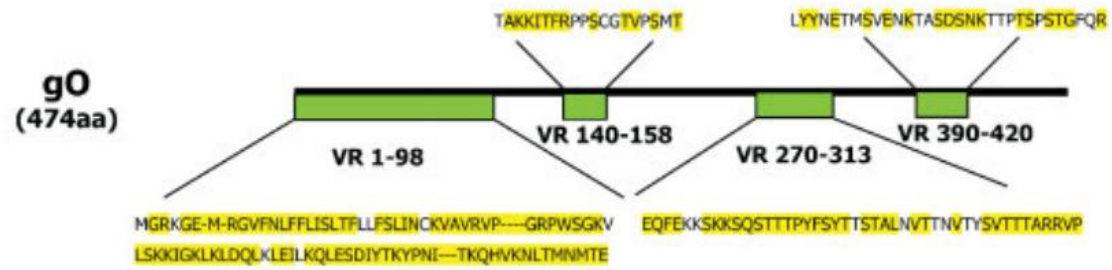


Figure 5. Amino acid variability in membrane glycoprotein gO from complex gCIII. The green boxes show the areas of major variability. Within each region, highlighted areas show amino acids that differ between clinical HCMV isolates. (Adaptated from Pignatelli *et al.* 2004)

In HSCT recipients, gO1 and gO2 are more prevalent than other genotypes. The gO3 genotype has been identified as an independent risk factor for the development of symptomatic CMV infection. Furthermore, in the same study, patients with gO4 had less frequent recurrences and a better survival rate (Roubalová *et al.* 2011).

Analysis of gO genotypes in SOT recipients has confirmed frequent multiple infection as in case of other genotypes (gB, gN, UL139, etc.) (Görzer *et al.* 2010). There is a good evidence that multiple infection in SOT recipients is associated with a worse course (Puchhammer-Stöckl *et al.* 2006, Manuel *et al.* 2009). The gO-2b genotype seems to be most frequent in SOT recipients.

The 454 pyrosequencing detected one or two high abundance genotypes, with a frequency above 10% and several low abundance genotypes. The most diverse genotype population was seen in patients with the D+R+ serostatus (Görzer *et al.* 2010).

3.7.2 UL-b´ region

UL144

This ORF encodes a homolog of the herpes simplex virus entry mediator, a member of the tumor necrosis factor (TNF)-α-like receptor superfamily (Lurain *et al.* 1999). The UL144 region is deleted in highly passaged laboratory strains, but is present in the Toledo strain and in low-passage clinical isolates (Cha *et al.* 1996). It was revealed that the product of UL144 causes upregulation of the Th2 atractant and regulatory T cells and can also decrease lymphocyte immune response to HCMV, which may help HCMV evade immune surveillance (Poole *et al.* 2006). The majority of nucleotide changes were detected in the 5´ half of the gene. Amino acid sequence

variability in protein pUL144 is particularly concentrated in the N-terminus and the overall amino acid variability among clinical strains was 33%. The phylogenetic analysis indicated that based on the differences in the nucleotide and amino acid sequences, it is possible to characterize three major genotypes (A, B, C) and two minor variants (A/B and A/C) (Lurain *et al.* 1999, Arav-Boger *et al.* 2002)

Several studies have shown a lack of correlation between symptoms and UL144 clades (Heo *et al.* 2008, Mao *et al.* 2007, Picona *et al.* 2005), but Arav-Boger *et al.* have suggested that UL144 is predictive of congenital infection outcomes (Arav- Boger *et al.* 2002 and 2006).

UL139

UL139 is predicted to encode type 1 membrane glycoprotein but its function is still not clear. However, this putative protein shares sequence homology with human CD24, a signal transducer modulating B-cell activation response, and may be likely to affect the immune response to HCMV infection. A highly polymorphic region was found particularly at the 5′ half of the ORF. The phylogenetic analysis identified variants which can be sorted into three major groups (G1, G2, and G3), with G1a, G1b, G1c, G2a, and G2b subgroups (Qi *et al.* 2006). Bradley *et al.* proposed a new nomenclature and sorted identified variants on the basis of the phylogenetic analysis into eight genotypes (G1-G8) (Bradley *et al.* 2008).

UL146

ORF UL 146 encodes a chemokine (vCXC-1) protein which functions as a neutrophil attractant. It has been shown that this protein is able to induce calcium mobilization, chemotaxis and degranulation of neutrophils (Penfold *et al.* 1999). It contains a motif which has been reported to be essential for receptor binding and IL-8 activity (Clarck-Lewis *et al.* 1991, Hassan-Walker *et al.* 2004). This gene is highly hypervariable throughout its whole length as evidenced in many studies (Prichard *et al.* 2001, Arav-Boger *et al.* 2006, Aquayo *et al.* 2010). On the basis of the observed variability, 14 genotypes have been defined (Dolan *et al.* 2004).

3.7.3 RL11 family

Genes of the RL11 family (RL5A, RL6, RL11, RL12, RL13, UL1, UL4, UL5, UL6, UL7, UL8, UL9, UL10, and UL11) are hypervariable and rapidly mutated in viruses adapted to fibroblast culture. In RL5A, RL6, UL9, and RL13 genes, termination, frameshift, or deletion mutations have been identified (Dolan *et al.* 2004, Cunningham *et al.* 2010, Dargan *et al.* 2010). The majority of these genes are predicted to encode class1 membrane glycoproteins (Chee *et al.* 1990; Davison et al., 2003). Proteins encoded by RL12 and RL13 are able to bind the Fc end of human IgG antibodies (Cortes *et al.* 2012). The UL11 ORF encodes a putative transmembrane protein, which has been detected on the surface of infected cells (Hitomi *et al.* 1997).

3.8 Detection of HCMV

3.8.1 Serology

Serological tests can be used to identify seronegative blood donors and blood products. Measurement of CMV IgG antibody is used to determine the serological status of the individual or as an indicator for acute or previous infection, but is not suitable for monitoring reactivation of latent infection in transplant patients. The first step in the diagnosis of acute infection with HCMV is usually conducted via detection of anti-HCMV-specific IgG and IgM antibodies with recombinant protein pp150 as an antigen. Samples reactive to IgM antibodies indicate acute, recent, or reactivated infection. Seroconversion to CMV IgM and IgG also confirms a recent diagnosis of HCMV infection. In infants, the presence of IgM antibodies indicates prenatal infection. The titer of IgM antibodies increases 2-6 weeks after infection and persists for about 2 years. IgG persists throughout life, but may decline with age (Mocarski *et al.* 2007).

For the confirmation of primary CMV infection, measurement of CMV IgG avidity is used. A positive IgM result in combination with low avidity of IgG is a strong indicator of primary CMV infection in the last 4 months and therefore, this method is suitable for the identification of pregnant women whose children are at risk of mother-to-child transmission of CMV infection in the first trimester (Bodeus *et al.* 2002).

3.8.2 Tissue Culture

Detection of the presence of the virus is carried out on human fibroblast cell cultures by monitoring the cytopathic effect (CPE). The monitoring time must be at least 21 days, which is a considerable limitation of this method. For routine diagnosis, the method called "Shell vials" is more widely used. This method is based on the increase of probability of infection of the cells by centrifugation of the clinical material to a monolayer of fibroblasts on the slide. The detection of the virus is done by immunofluorescence or immunoperoxidase methods using monoclonal antibodies against IE-1 protein after 24-36 hours (Gleaves *et al.* 1984).

3.8.3 Detection of viral antigens

The most commonly used method for the detection of viral antigens (antigenemia) and their quantification is based on the detection of tegument protein pp65 (UL83) in the peripheral blood leukocytes (PBL). The leukocytes are isolated from the blood by cytocentrifugation and indirect immunofluorescence detecting pp65 is performed (The *et al.* 1990). The method is mainly helpful in monitoring the effectiveness of antiviral therapy (Hebart and Einsele, 2004).

3.8.4 Detection of viral DNA

For the detection of viral DNA, qualitative and quantitative amplification methods are used. Qualitative PCR is able to detect the presence of viral DNA in the blood, plasma, and leukocytes. For the diagnosis of active HCMV infection and CMV disease, quantitative PCR is used. In transplanted patients, the detection by PCR can be done in the plasma, blood, peripheral blood leukocytes (PBL), and peripheral blood mononuclear cells (PBMC) (Razonable *et al.* 2002). In neonates and infants, PCR detection of HCMV DNA can also be performed in the urine (Schalasta *et al.* 2000).

Primers for the detection of HCMV are mainly targeted to the conserved regions of the genome, such as UL55, MIE, UL54, UL83, etc. By comparing the primers for the detection of HCMV in clinical samples, the highest sensitivity has been exhibited by the primers targeted to the UL55 and UL54 genes and therefore, these primers are mostly utilized in the routine testing (Habbal *et al.* 2009).

For research, nested PCR for the detection of HCMV DNA can be used. This method also allows for a partial quantification of HCMV DNA (Schafer et al., 1993).

Real-time quantitative PCR (qPCR) analysis is used for research and clinical monitoring of HCMV infection in transplant recipients as well. The commercially available system LightCycler® (Roche) utilizes FRET (fluorescence resonance energy transfer) hybridization probes. Other systems utilize TaqMan® assays (Razonable *et al.* 2002, Pang *et al.* 2003, Pumannová *et al.* 2006)

3.9 Next generation sequencing (NGS)

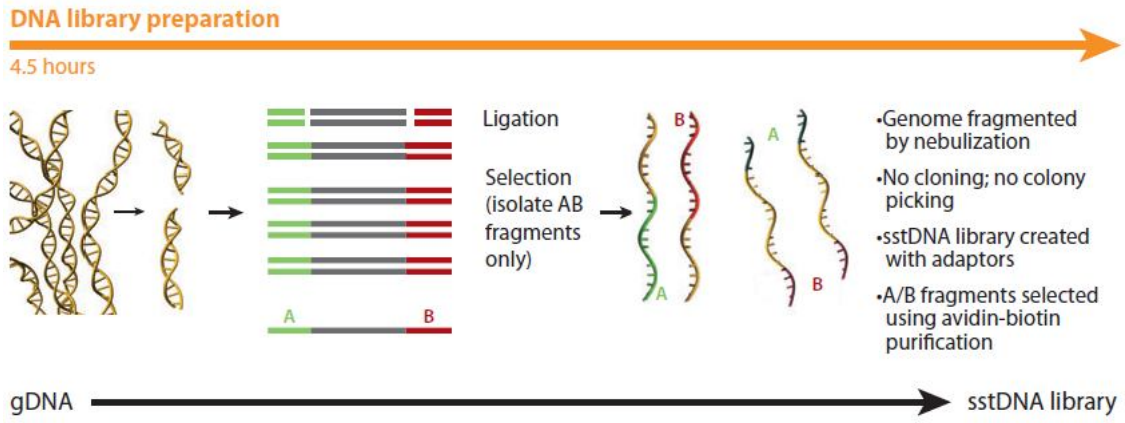
Two platforms are most widely used for the whole genome sequencing: (a) 454 pyrosequencing (Roche) represented by the 454 GS FLX Titanium and (b) Illumina (Illumina) represented by the Illumina Genome Analyzer and HiSeq systems, which are now dominating the NGS market. These two methods differ in template preparation, sequencing chemistry, and imaging. The preparation of the library for sequencing is based on the same principle for both systems. It starts with fragmentation of DNA and ligation of specific adapters (Mardis 2008, Metzeker 2010, Beerenwinkel *et al.* 2012).

3.9.1 Template preparation

454 pyrosequencing

The 454 sequencer utilizes emulsion PCR. Prior to the PCR amplification, library fragments are mixed with beads containing oligonucleotides complementary to the specific adapter sequences present on the library fragments. . Subsequently, the beads with fragments are mixed with a water-in-oil emulsion in ratio allowing the hybridization of only one DNA molecules to one bead and aqueous microreactors containing single and PCR reactants are formed (Fig. 6). But the majority of beads do not contain any DNA molecule. After emulsion PCR, every bead contains approximately one million copies of the fragments (Mardis 2008, Metzeker 2010).

A/



B/

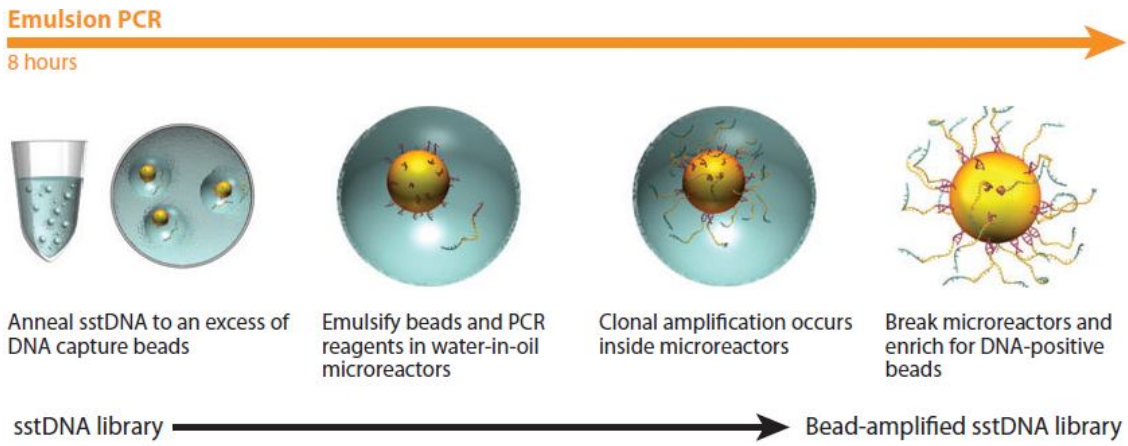


Figure 6. Library and template preparation for 454 pyrosequencing. Picture A/ shows DNA library preparation and B/ shows emulsion PCR. gDNA - genomic DNA, sstDNA – single-strand template DNA. (Adapted from Mardis 2008)

Illumina Sequencing

Two types of DNA libraries are possible to use in the Illumina sequencer: single-read library and paired-end library. The paired-end library allows sequencing both the forward and reverse template strands of each cluster during one paired-end read resulting in an increase in coverage against single-read libraries. Illumina utilizes bridge amplification PCR. Library DNA fragments with adapters are attached to the solid surface of the flow cells containing oligonucleotides complementary to adapters. Consequently, these fragments are extended, the original template is washed away, and the bridge amplification PCR is performed with newly synthesized strands covalently

attached to the flow cells. This bridge PCR creates clusters of the original DNA fragments with about a million of copies (Fig. 7) (Mardis 2008, Metzeker 2010).

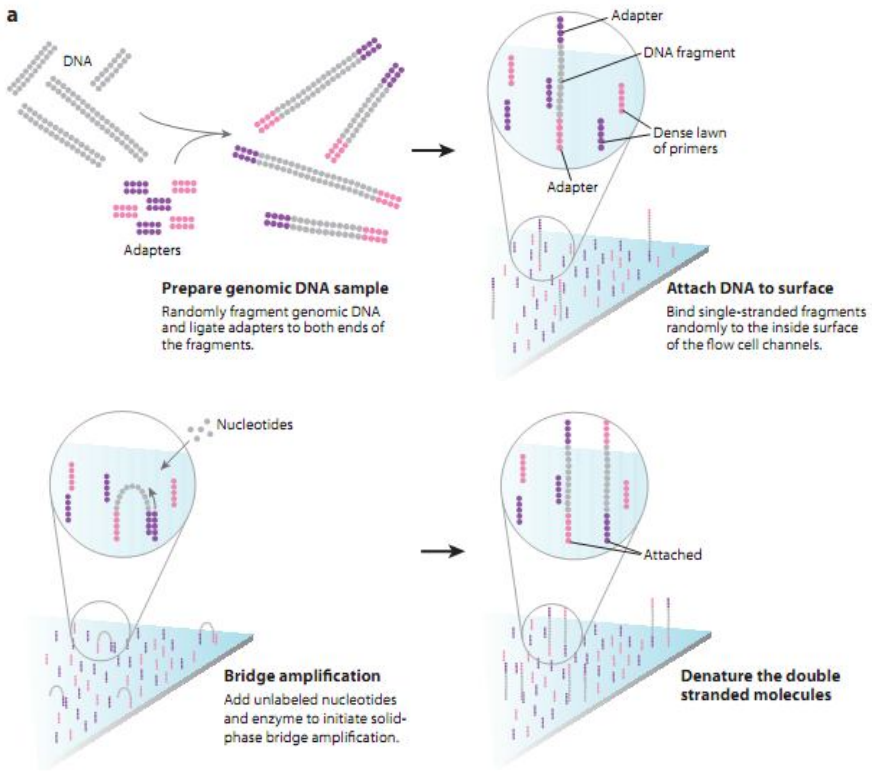


Figure 7. Library and template preparation for Illumina sequencing. (Adapted from Mardis 2008)

3.9.2 Sequencing and Imaging

454 Pyrosequencing

The 454 sequencer utilizes the method called Pyrosequencing. Before sequencing, beads with amplified DNA fragments are placed into a PicoTiter plate where each well contains one bead. Subsequently, beads coupled with important enzymes (sulphurylase and firefly luciferase) for pyrosequencing are added. This method is based on the release of pyrophosphate after the incorporation of each nucleotide. The release of the pyrophosphate initiates a series of downstream reactions that ultimately produce light by the firefly enzyme luciferase. The nucleotides are added in cycling order. The light signal is detected by CCD (charge-coupled device) camera and is recorded as a series of peaks (flowgram) where the size of the peak (intensity of the light signal) is approximately proportional to the number of nucleotides that have been incorporated (Fig. 8) (Mardis 2008, Metzeker . 2010).

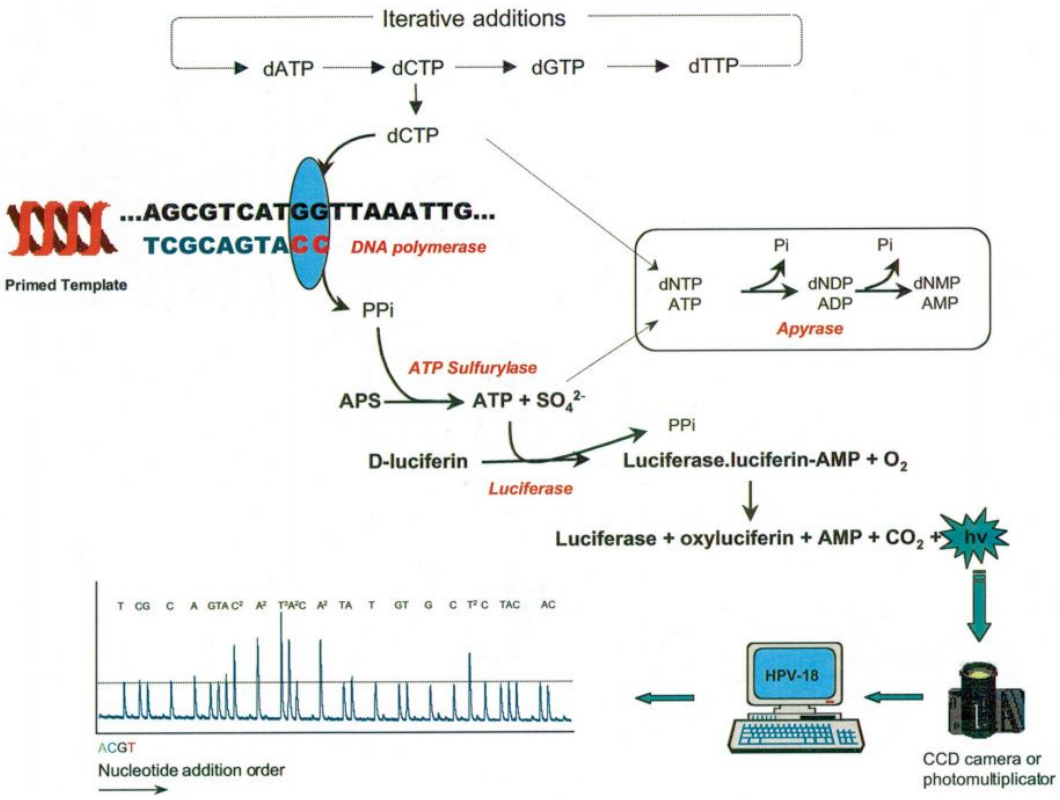
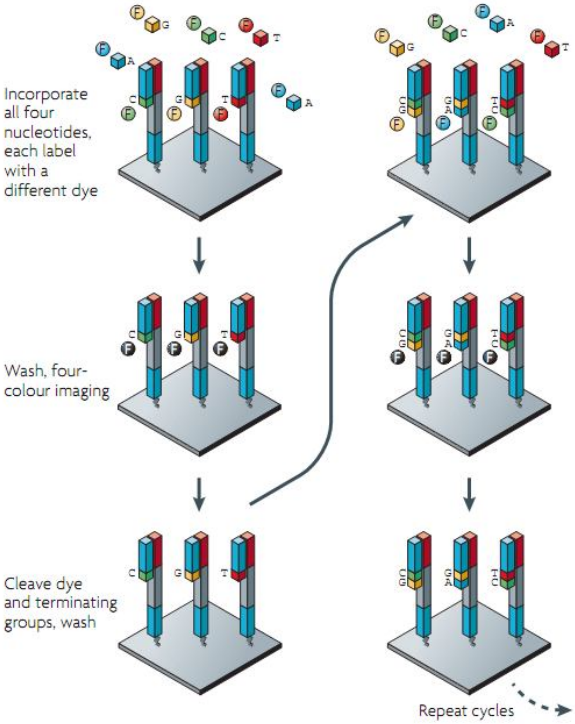


Figure 8. Schematic representation of the 454 pyrosequencing chemistry and imaging. (Adapted from Gharizadeh et al., 2001)

Illumina Sequencing

Illumina chemistry relies on the principle of cyclic reverse termination uses modified nucleotides with a reversibly chemically blocked 3´-OH group. The incorporation of this modified nucleotide results in termination of DNA synthesis, thus ensuring the incorporation of only one modified nucleotide. Because these nucleotides are added simultaneously, each nucleotide is labeled by different fluorescence dye. After incorporation of each base, the image of emitted fluorescence from each cluster on the flow cells is captured by laser excitation. Consequently, the fluorescence dye is removed, the 3´-OH group is regenerated and the cycle is repeated (Fig. 9) (Mardis 2008, Metzeker 2010).

A/



B/

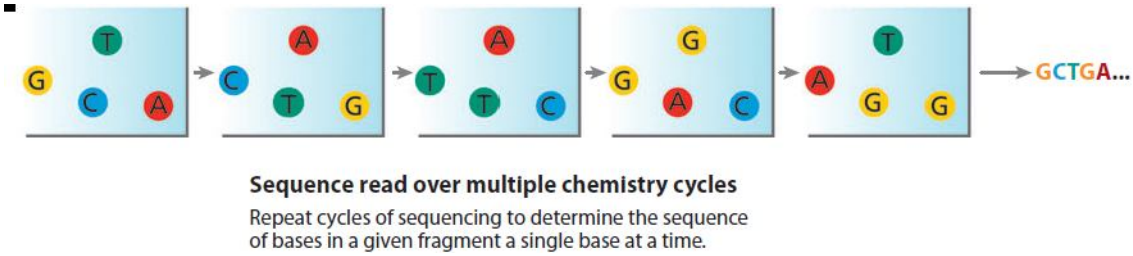


Figure 9. Schematic visualization of chemistry and imaging of the Illumina platform. Picture A/ shows sequencing steps. Picture B/ shows sequence reads. (Adapted from Mardis 2008 and Metzeker 2010).

3.9.3 Other differences and Error rates

These two platforms are also different in other aspects such as throughput, run time, costs, read lengths and error pattern (Table 1).

454 sequencing has a high error rate in readings of homopolymer repeats. In these regions, the linear relationship between intensity and number of incorporated nucleotides fails. Insertions and deletions are common type of errors. Insertions errors can be caused by nucleotides which are not completely removed by washing while deletions can be caused by incomplete extension. In Illumina sequencing, substitutions

have similar frequencies as in the 454 system. The mismatch rates depend on the length of the reads, sequence context, and type of substitutions (Beerenwinkel *et al.* 2012).

Table 1. Differences between the most frequently used next-generation sequencing platforms.

Platform	Read Length (bases)	Run time (days)	Gb per Run	Reads per Run
454 GS FLX Titanium XLR70 (Roche)	450	0.35	0.45	1 million
Hiseq 2000 (Illumina)	2x 100	11	600	6 billion

3.9.4 Other platforms

Besides the two main NGS platforms mentioned above, a lot of other platforms, which have different sequencing chemistry, exist. Because they have higher error rates, higher cost and/or longer run time compared with the Illumina and 454 systems or have been just recently put on the market, they are much less frequently utilized. To name To name some other platforms, we have SOLiD (Applied Biosystems), HeliScope (Helicos BioSciences), Ion Torrent (Applied Biosystems) and Real-time sequencing (Pacific BioSciences). The SOLiD platform does not use DNA polymerase like other platforms but instead uses DNA ligase. The Ion Torrent uses the principle of hydrogen ion release after incorporation of every base. Sequencing is conducted on a special chip, which detects very small changes of pH that occur after release of the hydrogen ion. The HeliScope uses reversible terminators like Illumina with the difference that reversible terminators are labeled by only one dye and dispensed individually. Real time sequencing uses four different fluorescence labeled hexaphosphate nucleotides and a polymerase which is attached to the surface. After the incorporation of the nucleotide, a fluorescence pulse is detected in real time (Metzeker 2010, http://www.invitrogen.com.).

3.9.5 Assembly methods

For the creation of a long sequence (contig) from individual short reads, different assembly methods can be used: (a) mapping assembly, (b) *de novo* assembly, or (c) a combination of both assembly strategies (hybrid assembly). In the mapping assemblies, reads are assembled against an existing backbone sequence - reference genome. *De novo* assembly does not use a backbone sequence. It creates contigs by

searching for overlaps between short reads. *De novo* assembly does not necessarily reconstruct the complete genome at once. The hybrid assembly proceeds as follows: *de novo* assembly is performed first to create a hybrid sequence, followed by the mapping assembly.

3.9.6 HCMV whole genome variability

NGS technology for sequencing the whole HCMV genome of clinical strains has been so far used in only three studies (Cunningham *et al.* 2010, Jung *et al.* 2011, Renzette *et al.* 2011).

Cunningham *et al.* used Illumina NGS sequencing for analysis of clinical strains after short cell culture passaging and the one strain derived from clinical specimen. The study confirmed data from several previous studies, which, however, analyzed only part of the HCMV genome of particular strains. The process of adaptation of HCMV to fibroblasts is strongly associated with the generation of mutations in a region encompassing three small genes termed the UL128 locus (UL128, UL130, and UL131A) and in RL11genes, especially in RL13. Other mutations were observed in RL5A, RL6, and UL9 genes (Dolan *et al.* 2004, Cunningham *et al.* 2010, Dargan *et al.* 2010).

In this study, it has also been shown for the first time that Illumina NGS can be successfully performed in clinical specimens containing only 3% of viral DNA. However, the challenge remains before sequencing of complete HCMV genomes from clinical specimens can become part of the routine diagnosis.. Many samples contain such low levels of HCMV DNA that without prior enrichment and amplification steps they are not be suitable for NGS. Possible approaches include the isolation of virion DNA or the purification of viral DNA by buoyant density centrifugation, followed by the whole-genome HCMV amplification (Cunningham *et al.* 2010).

In studies dealing with characterization of polymorphic areas from clinical samples, so far less than 5% of the HCMV genome has been sequenced. Renzette et al. were the first to which utilized the NGS technology to study the intra-host HCMV variability. In this study, clinical specimens from three congenitally infected children were used. Purified HCMV DNA was amplified before Illumina sequencing by primers spanning the entire genome. It was shown that genotyping on the basis of just several polymorphic regions may not be appropriate for measuring HCMV diversity in clinical samples. It was determined that in congenitally infected children one major genotype is

present together with many minor variants. The observed intra-host variability of HCMV was comparable with that of RNA viruses but the source of this diversity for HCMV is currently unknown. It is hypothesized that even though the mutation rates are low, the replication level is high and this could lead to accumulation of mutations. Reinfection and co-infection events as a source of the variability cannot be excluded either (Renzette *et al.* 2011).

4. Material and Method

4.1 Patients

The samples were obtained from 32 patients who underwent allogeneic hematopoietic stem cell transplantation (HSCT) and from two congenitally infected children. Selected were patients with viremia of 100 copies/10000 g.e. (genome equivalents) as assessed by qPCR in their peripheral blood. From the two congenitally infected children, urine samples were collected in the Motol University Hospital. From 15 HSCT recipients patients, urine samples were obtained and from additional 16 patients, urine and blood samples were taken. From one patient, only a blood sample was available. One patient was treated in the General University Hospital, Prague, Czech Republic and the rest were patients treated at the Institute of Hematology and Blood Transfusion, Prague, Czech Republic. The urine, if not processed immediately, was stored at -70°C. Blood was processed immediately.

4.2 Preparation of clinical samples

For the multiplication of the HCMV, human embryonal lung fibroblasts (LEP) were used. The cells were grown in 1x E-MEM (Eagle-Minimum Essential medium) consisting of 10 ml of 10x E-MEM per 100 ml (Sigma), 1ml of 100x Penstrep-glutamin per 100 ml (Gibco), 5 ml of FBS (fetal bovine serum) per 100 ml (Gibco), 1.5 ml of 7.5% $NaHCO_3$ per 100 ml (Gibco), and 1 ml of 100x Vitamins (PAA) per 100 ml.

4.2.1 Urine

Centrifugation

For infection of the cells, a mixture of 1 ml of urine, 100 µl of 10x PBS, 100 µl of amphotericin, and 5.7 µl of gentamicin was used. The mixture was vortexed and applied to the monolayer of human embryonal fibroblasts in a 25 cm^3 flask without 1x

E-MEM. Subsequently, the cells were incubated at 37 °C for 2 hours. After incubation, the flask with the sample was centrifuged for 30 min at 1000 rpm (rounds per minute) (Heraeus Megafuge 16R with rotor TS 75003629, Thermo Scientific) and 10 ml of the fresh medium was added and cells were incubated at 37 °C and 5% CO_2. The next day the medium was replaced with the fresh medium.

Filtration

Urine was filtered through a 220μm filter. Filtered urine was added directly to the medium in the 25 cm^3 flask with monolayer human embryonal fibroblasts and incubated at 37 °C and 5% CO_2.

4.2.2 Blood

Venous anticoagulated blood from patients was diluted in 1x E-MEM without FBS (Gibco) in a ratio of 1:1. Diluted blood was carefully (to avoid mixing of these two components) pipetted onto Histopaque 1077 (Sigma) in a ratio of 1:1. Before adding blood, Histopaque was incubated at room temperature for 30 minutes. The sample was centrifuged for 30 min at 400 g at room temperature with slow braking. After centrifugation, the plasma layer was carefully aspirated with a Pasteur pipet. About 5 mm of the plasma layer was kept above the opaque interface containing mononuclear cells. The layer with mononuclear cells was transferred with a Pasteur pipet to a new 50 ml Falcon tube, added with 30 ml 1x PBS, and centrifuged. The supernatant was removed, the pellet was resuspended in 10 ml 1x PBS, and centrifuged at 1000 rpm for 10 min at 4°C (Hettich® UNIVERSAL 320R, Hettich Zentrifugen). Ten millilitres of 1x PBS were added, followed by centrifugation at 800 rpm for 10 min at 4°C. The pellet was dissolved in 2 ml of 1x E-MEM without FBS. Cells were counted in a hemocytometer. The mononuclear cells were then directly introduced on the monolayer of human embryonic fibroblasts into a 25 cm^3 flask with medium. The whole volume of isolated mononuclear cells was used if their number was less than 10 million.

4.3 Cell passaging

Cryovials with cells ($5x10^6$) stored in liquid nitrogen were thawed rapidly (60 seconds at 37°C) and transferred into a 25 cm^3 flask with 9 ml of E-MEM. Two weeks after infection, the cells were passaged into two 75 cm^3 flasks. Medium was removed from the 25 cm^3 flask and the cells were washed with 0.5 ml Try-EDTA (Trypsin-EDTA) (0.05 % trypsin (Gibco), 0.53 mM EDTA (Sigma)). Then, 1 ml of 0.05% Try-

EDTA was added to the flask and incubated at 37 °C. Immediately after detachment, 5 ml of 1x E-MEM were added and a homogeneous suspension of cells was obtained. Finally, the cells were divided into prearranged 75 cm^3 flasks with 17 ml of 1x E-MEM. The medium was changed once a week. Two weeks after passage into two 75 cm^3 flasks cells were transferred to four 150 cm^3 flasks. Passage to 150 cm^3 flasks was performed in the same way as that to 75 cm^3 flasks. The medium was changed once a week. After 14 days, regardless of the presence of the CPE, the flasks were transferred to -70 °C.

4.4 Purification of virions

The cells were scraped and resuspended in the medium by pipetting up and down. Cell suspensions were transferred to two 50 ml falcons and centrifuged at 4 °C for 5 min at 250 g (Hettich® UNIVERSAL 320R, Hettich Zentrifugen). The supernatant was carefully removed and the pellets from two 50 ml Falcon flasks were combined. The pellet was dissolved in 10 ml of ice-cold 1xPBS and centrifuged at 4 °C and 250 g for 5 min. The supernatant was carefully removed, 5 ml of cold permeabilization buffer (1.28 M sucrose, 20 mM $MgCl_2$, 140 mM Tris-HCl pH 7.5, 4 % Triton X-100) were added, and the pellet was resuspended and incubated 10 min on ice. After incubation, the sample was centrifuged for 15 min at 4 °C and 1300g. The supernatant was carefully removed, resuspended, 5 ml of permeabilization buffer were added, and the sample was centrifuged again for 15 min at 4°C and 1300g. Fifty microlitres of a mixture of 40 mM PIPES pH 7.0, 7% sucrose, 20 mM NaCl, 2 mM $CaCl_2$, 10 mM 2-mercaptoethanol, 200 µM PMSF, 150 U of micrococcal nuclease (300 U/µL, Fermentas), 1 µL of RNase A (100 mg/mL, Qiagen), and protease inhibitor cocktail (Sigma) were added. After centrifugation, the supernatant was removed and the pellet was resuspended in 50 µl of nuclei buffer (10 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, and 10 % sucrose). The resuspended pellet was transferred to a 1.5 ml tube with 50 µl of a mixture containing 2x nuclease, RNase A, micrococcal nuclease, and protease inhibitor and incubated at 37 °C for 30 min. The reaction was stopped by adding 2.4 µl of 0,5M EDTA. The sample was then stored at -20 °C.

4.5 DNA isolation

4.5.1 Isolation from the urine

One and half ml of urine was centrifuged at 14000 rpm for 20 min at room temperature (Hettich® UNIVERSAL 320R, Hettich Zentrifugen). The supernatant was removed except for 200 µl and the pellet was resuspended. DNA was isolated by a QIAamp® DNA Mini kit (Qiagen) according to the manufacturer's instructions and eluted with 50 µl of the elution buffer.

4.5.2 DNA isolation from the purified virion

DNA from the virions was isolated by means of QIAamp® DNA Mini kit (Qiagen) according to the manufacturer's instructions.

4.6 Nested PCR

For the detection of HCMV, nested PCR with primers amplifying the UL55 region encoding envelope glycoprotein gB was used (for primers sequences, see Table 2). Forty-five microlitres of the reaction mixture for the first step contained 1x PCR Buffer, 1.5 mM $MgCl_2$, 250 µM dNTPs, 0.5 pmol/µl of each primer, 1.25 U Taq DNA polymerase (Fermentas), and 5 µl of DNA. For the second reaction, 1 µl of the product generated in the first reaction and 49 µl of reaction mixture of the same composition as for the first reaction were used. Thermal profiles for the amplification are specified in Table 3.

Table 2. The sequences of primers used for nested PCR

Outer primers	**First reaction**
Forward CMV OF	5´-TCCAACACCCACAGTACCCGT-3´
Reverse CMV OR	5´-CGGAAACGATGGTGTAGTTCG-3´
Inner primers	**Second reaction**
Forward CMV IF	5´-GTCAAGGATCAGTGGCACAGC-3´
Reverse CMV IR	5´-GTAGCTGGCATTGCGATTGGT-3´

Table 3. Thermal profiles for the nested PCR

First reaction	
	95°C for 3 min
35x	94°C for 30 s, 45°C for 30 s, 72°C for 45 s
	72°C for 5 min
Second reaction	
	95°C for 3 min
20x	94°C for 30 s, 45°C for 30 s, 72°C for 45 s
	72°C for 5 min

4.7 Whole genome amplification

A REPLI-g® Mini Kit (Qiagen) was used to increase the amount of the purified viral DNA for Illumina® sequencing. This kit works on the principle of the Multiple Displacement Amplification technology (Spits *et al.* 2006). Amplification was performed according to the manufacturer's instructions.

4.8 Quantitative real-time PCR

The TaqMan probe was used to determine the ratio of viral and cellular DNA in the clinical isolates. This ratio is crucial for the suitability of the sample for the whole genome sequencing. CMV-specific primers and a probe (for primers and probe sequences, see Table 4) were localized in the UL86 region (major capsid protein). The CMV- specific TaqMan probe was labeled on the 5´ end with FAM™ (6-carboxy-fluorescein) and on the 3´end with TAMRA™ (6-carboxy-tetramethyl-rhodamine). The amplification was carried out in a Applied Biosystems 7500 Fast Real-Time PCR System (Applied Biosystems) in a final volume of 25 µl. The reaction mixture contained TaqMan® universal Master Mix (Applied Biosystems), forward and reverse primers, and the TaqMan probe. An aliquot of 5 µl of the 1000x diluted REPLI-g product was used for each reaction and all samples were run in duplicates. The PCR was performed in 0.2 ml tubes under the following conditions: after 2 min at 50 °C and 10 min at 95 °C, the samples were submitted to 45 cycles of 15 s at 95 °C and 1 min at 60 °C for the amplification of both HCMV and the β-globin gene. The fluorescence was detected on different channels. The human genomic DNA was quantified with

PC03/KM38 primers specific for the β-globin gene and the TaqMan probe with covalently attached reporter dye JOE™ (6-Carboxy-4',5'-Dichloro-2',7'-Dimethoxyfluorescein, succinimidyl ester) on the 5´ end and a quencher dye TAMRA™ on the 3´end (for primer sequences, see Table 4). The PCR specific for the β-globin gene was performed under the same PCR conditions as described above for HCMV.

Table 4. The sequences of primers and probes used for the qPCR

CMV	
CMV_TM_FW	5´-CACGGTCCCGGTTTAGCA-3´
CMV_TM_RV	5´-CGTAACGTGGACCTGACGTTT-3´
CMV_TM_Probe	5´-TGTAACCGCGATCCTCGGGCAGATA-3´
β-globin	
FW_PC03	5´-ACACAACTGTGTTCACTAGC-3´
RV_KM38	5´-TGGTCTCCTTAAACCTGTCTTG-3´
b_GLO_TP	5´-CCACCAACTTCATCCACGTTCACCTT-3´

4.9 Next-generation sequencing

Based on the results of qPCR samples with a ratio of viral to cellular DNA higher than 60 % were preferably selected for NGS. However, the samples with very low purity of viral/cellular DNA (lower than 7 %) were also taken. But the samples with very low purity of viral/cellular DNA provide low coverage and the assembly is very difficult or even impossible.

The whole HCMV genome was sequenced by Illumina® paired-end sequencing. Illumina® sequencing was performed by HiSeq2000 at the KU Leuven Genomics Core, Belgium. For generating new HCMV sequence, the short reads were assembled by *de novo* assembly approach, followed by mapping assembly. *De novo* assembly was performed by using the MIRA3 (Mimicking Intelligent Read Assembly) software (http://sourceforge.net/projects/mira-assembler/files). The results of this assembly were contiguous sequences called contigs. The longest contigs were analyzed by NCBI BLAST optimized for megablast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for selection of the reference genome. As a reference genome, the HCMV JP Strain (GQ221975.1), which had the highest similarity (higher than 97%) with each contig, was chosen. This strain was used for construction of a hybrid backbone sequence by mapping *de novo* contigs on the reference strain. This sequence contains the sequences of *de novo* contigs

supplemented with parts of the reference genome where there is no coverage of contig sequences. The program NUCmer (NUCleotide MUMmer) included with MUMmer 3.0 (http://sourceforge.net/projects/mummer) was used to create an alignment of the reference strain and analyzed contigs. MEGA5.05 software (http://www.megasoftware.net/mega.php) was used for making the backbone. Contigs were placed in the appropriate areas on the basis of alignment with the JP strain. Missing areas, located between contigs, were completed with the corresponding areas of the JP strain.

Consequently, mapping assembly of the original sequence reads on the created hybrid backbone was performed with MIRA3. The assembly was visualized in Tablet (Next Generation Sequence Assembly Graphical Viewer) (http://bioinf.scri.ac.uk/tablet/download.shtml) and the consensus sequence was simultaneously viewed in MEGA5.05. In Tablet, the assembly was reviewed to find assembly errors. These errors were corrected in the consensus sequence in MEGA5.05 or gaps were introduced if an erroneous assembly was not immediately correctable. In cases where the ratio of two different nucleotides in a single column for one base in the consensus sequence was approximately 50:50, this base was substituted by the relevant IUPAC (International Union of Pure and Applied Chemistry) nucleotide ambiguity code. The repaired consensus sequence was divided into several parts bounded by gaps and sequence reads were mapped on these new contigs by MIRA3. After the MIRA3 analysis, newly created contigs were compared with original contigs from the repaired consensus sequence in Tablet. When the parts were extended by several nucleotides, extensions were added to the parts of the original sequence in MEGA5.05. In MEGA5.05, it was investigated whether the extensions helped to create an overlap. When the overlap was created, the gap was closed. This procedure was iterated several times. If the gaps were not closed after these steps, Sanger sequencing was applied.

4.10 Sanger Sequencing

4.10.1 PCR amplification

The PCR assay PCR was performed in a total volume of 25 µl (20 µl master mix and 5 µl DNA). Master mix contained 5 µl of QIAGEN OneStep RT-PCR buffer, 1 µl of QIAGEN OneStep RT-PCR Enzyme mix, 1.5 µl of a 10µM solution of appropriate forward and reverse primers (for primer sequences, see Table 5) and 1 µl of dNTPs.

Amplification started by denaturation for 15 min at 95 °C, followed by 35 cycles of 10 s at 94 °C, 30 s at 56 °C and 2 min at 72 °C and a final elongation of 10 min at 72 °C. The products were detected by polyacrylamide gel electrophoresis after EtBr (Ethidium bromide) staining.

Table 5. Sequences and melting temperatures of primers designed for gaps.

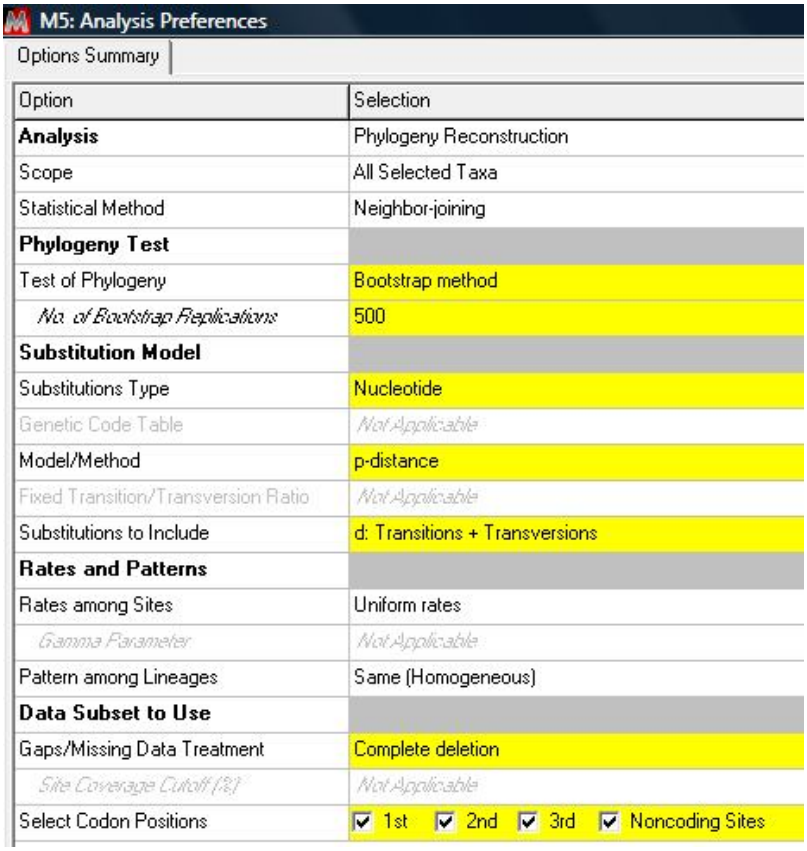
gap1		T_m
CZE3_1FW	5´-TATACCGGATGCTAGGCGAC-3´	60,5
CZE3_1RV	5´-GAAAATTTGCCCCGACTGCG-3´	60,5
gap2		
CZE3_2FW	5´-TTTACCTCCGCAGCCGTACG-3´	60,5
CZE3_2RV	5´-TGTGAAGAGATAGAGTGTGAGC-3´	62,5
gap3		
CZE3_3FW	5´-TTTACCTCCGCAGCCGTACG-3´	60,1
CZE3_3RV	5´-ATCACTTATGGGGTCACCGC-3´	61,2

4.10.2 Sequencing

Purification of PCR products was performed by ExoSAP-IT ™ (USB). Two microlitrers of ExoSAP were added to 5 µl of PCR products, incubated for 15 minutes at 37°C, and inactivated for 15 minutes at 80°C. After this step, 4 µl of mixture containing 2 µl of Sequencing buffer and 2 µl of BigDye® Terminator v3.1 (Applied Biosystems) were combined with 1 µl of forward or reverse primer specific for the particular gap (for primer sequences, see Table 5). Sequencing products were precipitated with 90 µl of mixture containing 62.5 µl of 96% ethanol (EtOH), 24.5 µl of distilled water, and 3 µl of 3M NaOAc. Subsequently, mixture with products was centrifuged at 13000 rpm for 30 minutes at room temperature (Hettich® MICRO 120, Hettich Zentrifugen). After centrifugation, the supernatant was removed, 150 µl of 70% EtOH was added and centrifuged at 13000 rpm for 5 minutes at room temperature. After that, the supernatant was carefully removed, the sample was dried at 55 °C, and 16 µl of formamide was added and incubated for 2 min at 95 °C. Sequencing products were analyzed with by an ABI PRISM® 3120 sequencer (Applied Biosystems).

4.11 Phylogenetic analysis of highly polymorphic regions

On the basis of the HCMV Merlin strain (NC_006273.2), the individual genes of our clinical isolate were annotated. Phylogenetic analysis was performed on highly polymorphic regions such as RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL73, UL74, UL139, and UL146. The database of genes of HCMV for both laboratory strains and clinical isolates was constructed from available sequences. The gene sequence from the new isolate was than compared with all other available gene sequences. In MEGA 5.05, sequences were translated to protein sequences, a protein alignment was made and back-translated to a nucleotide sequence. Phylogenetic trees were constructed by the neighbor - joining method in MEGA 5.05. Phylogenetic analysis was set up as shown in Figure 10.



M5: Analysis Preferences

Options Summary

Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	500
Substitution Model	
Substitutions Type	Nucleotide
Genetic Code Table	*Not Applicable*
Model/Method	p-distance
Fixed Transition/Transversion Ratio	*Not Applicable*
Substitutions to Include	d: Transitions + Transversions
Rates and Patterns	
Rates among Sites	Uniform rates
Gamma Parameter	*Not Applicable*
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	*Not Applicable*
Select Codon Positions	☑ 1st ☑ 2nd ☑ 3rd ☑ Noncoding Sites

Figure 10. Settings of HCMV phylogenetic analysis

5. Results

5.1 Cell culture multiplication of the virus

Of 33 urine samples subjected to nested PCR, 14 were positive (42 %). Although the selected patients had high viral load in the blood, the virus was not detectable in urine at the time of sampling. Of these 14 positive samples, the cytopathic effect (CPE) (Fig. 11) in the tissue culture developed only in five samples. The presence of HCMV DNA was confirmed in all of these samples as reveled by nested PCR of the DNA isolated from infected cells. The remaining eight samples negative for CPE were also screened for HCMV DNA but all were negative.

The CPE was observed in two of the samples infected by filtrated urine (CMV1M - CMV8M) (25 %). To test for the possibility that filtration causes decrease of the viral load for cell infection, the method of centrifugation of the urine to the cell monolayer was also used. Using this method, CPE was observed in three of 25 samples (12 %). The method of urine filtration and inoculation to the tissue culture was more efficient in our hands. Moreover, while using the second method we noted an increased number of contaminated samples. The contamination was detected in five of 25 samples and toxicity effects were observed in three of 25 samples. Samples with contamination in the cell culture were inoculated again using the filtration method. All but one samples became contaminated again.

Since the yield of CPE-positive urine samples was low, we also used blood samples where available. Altogether 17 blood samples were tested. CPE was detected in 2 (12 %) of them. The viral load of these samples at the time of sampling was higher than 230 copies/10000 g.e. (genome equivalents).Both samples were HCMV DNA-positive. Even though the yield of HCMV DNA positive samples was low we confirmed that both types of clinical specimens – urine and blood – can be used for HCMV multiplication prior to DNA extraction for NGS (Table 6).

Table 6. Methods of inoculation. Table shows number of samples inoculated by various methods, number of samples positive by PCR, CPE (Cytopathic effect) positive samples and PCR positive samples after isolation from cell culture. An asterisk indicates samples, which were positive by qPCR in the routine virological laboratory.

	Urine		Blood
Method of inoculation	**Filter**	**Centrifugation**	
# samples	8	25	17
# PCR positive	5	9	17*
# CPE positive	2	3	2
# PCR positive after isolation	2	3	2

A/ B/

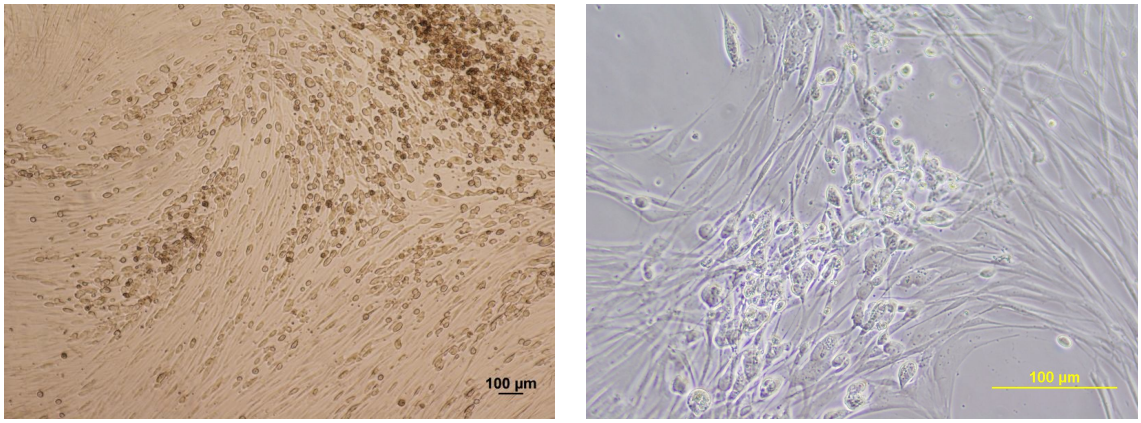


Figure 11. Cytopathic effect in two selected samples, CMV3M and CMV15M. Panel A illustrates the cytopathic effect in CMV3M on post-infection day 16, panel B in CMV15M on post-infection day 35.

5.2 Real-time qPCR

The whole genome amplification by REPLI-g was done on seven DNAs extracts from tissue cultures which were confirmed by the nested PCR as being HCMV DNA positive. One sample was surprisingly negative (CMV1M) after qPCR. Of the remaining six samples, only one (CMV3M) had purity (HCMV/cellular) of viral DNA of 98% viral DNA. The other samples had purity of less than 7%. Because the sample CMV15M had low purities of HCMV DNA, it was amplified again on the cell culture and isolated (CMV15-2M); however, the purity did not improve (Tab. 7).

Table 7. The purity of HCMV DNA (HCMV/cellular). This table shows the total amount and purity of HCMV DNA in samples after qPCR.

	# CMV DNA	purity CMV/cellular
	µg	%
CMV1M	0,00	0
CMV3M	0,49	93
CMV15M	1,28	4
CMV15-2M	0,15	4
CMV21M	0,20	7
CMV26M	0,06	4
CMV33K	0,05	4
CMV34K	0,38	6

5.3 Assembly of sequence

After de novo assembly of short Illumina reads of sample CMV3M, longer contigs were created by MIRA 3. The longest contigs were consequently analyzed by NCBI Blast. On the basis of Blast analysis, strain JP (GQ221975.1) was chosen as a reference strain for construction of the backbone (consensus) sequence. The construction of the backbone was performed by an aligment of JP strain with analyzed contigs using NUCmer. Subsequently, based on this alignment, the contigs were placed in the appropriate areas using the MEGA5.05 software. The missing areas, located between contigs, were filled in based on the JP strain sequence (Fig. 12).

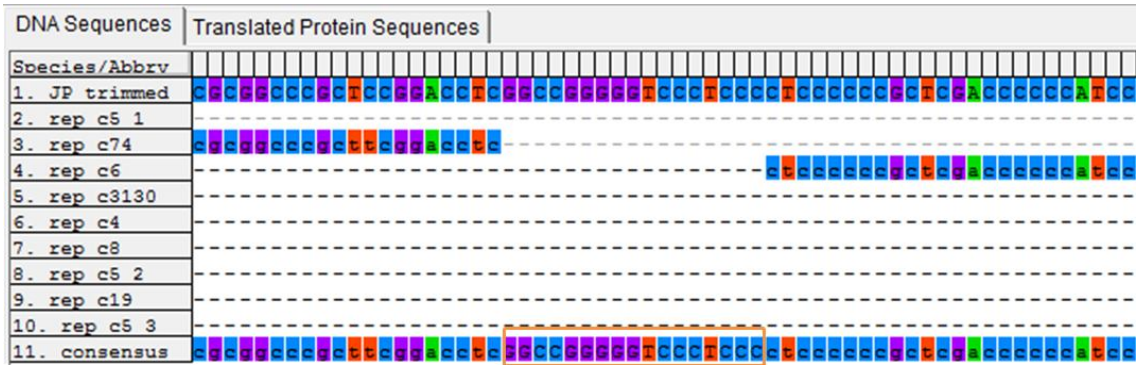


Figure 12. Creation of backbone sequence. The orange lines bounding the area show an complemented area between two contigs from the reference genome JP (JP trimmed) in the backbone (consensus) sequence.

Subsequently, mapping assembly of the sequence reads on the created backbone was performed by MIRA 3. The assembly was visualized in Tablet. After reviewing the sequence, assembly errors were corrected and problematic areas were replaced by gaps and repaired in MEGA5.05 (Table X.). The repaired consensus sequence was divided into seven contigs and these contigs were analyzed by MIRA 3.

Table 8. Problematic areas in newly created consensus sequence

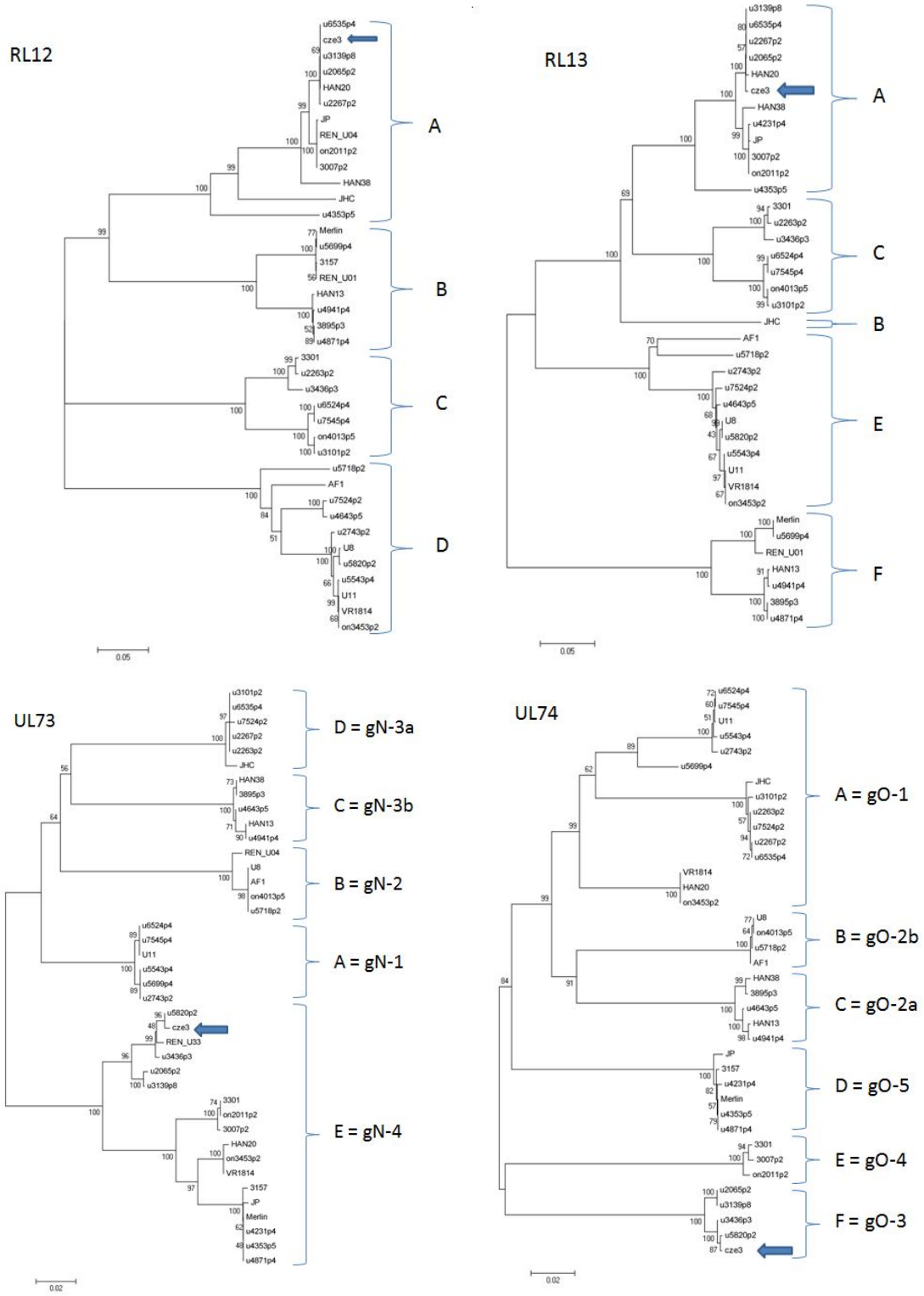
Position of the genome	Type of problem
92577-92581	gap
93186	in 1 column 3A:3C substituted by M
93314	1 base gap
area about 117984	problematic area substituted by gap
193216-193491	gap
196379 -196494	gap
230886-231635	gap

After the next round of MIRA 3 mapping, overlap between contigs 3 and 4 was found. Between other contigs, gaps were still present. After 3 rounds of MIRA 3 mapping, the gaps (gap1, gap2, and gap3) were not closed and therefore, Sanger sequencing was applied to close these gaps. These parts of the genome were sequenced from both sites. This sequence and sequences of contigs containig gaps were assembled and analyzed in the SeqMan Pro™ program (LASERGENE®) and finally, the whole sequence of the Czech HCMV clinical isolate was created.

5.4 Phylogenetic analysis

The sequences of the Czech isolate CMV3M (cze3), all Belgian newly sequenced strains, and other isolates with the full genome sequences available in the public databases were subjected to phylogenetic analysis. The following genes were analyzed: RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL73, UL74, UL139, and UL146. Phylogenetic trees were constructed in MEGA 5.05 for each gene separately. According to the topology of the trees, isolates were classified in different genotypes designated A-H. As is shown in Table 9 and Fig. 13, the different strains are clustering together in genotypes in different genes. That means that there is no clear correspondence between genotypes in one gene and another.. The UL9 gene which is

frequently mutated in Belgian, was also mutated in CMV3M. CMV3M from the Czech patient belongs to gN4 and gO3 genotypes, as revealed by analyzing the UL73 and UL74 genes. Comparison of UL146 has shown that this strain belongs to genotype 14 (Fig. 13). In RL5A, the Czech strain showed a similarity level of 83% with Belgian isolates, in other genes, lower similarity level was found: 13% in RL6, 17% in UL74, and 25% in UL9. From the preliminary data on the comparison of XX Belgian strains and one Czech strain, it is evident that complete homogeneity between isolates is very rare. Here only two isolates were identical in all monitored genes (u2267 and u6535) (Table 9).


RL12
u6535p4
cze3
u3139p8
u2065p2
HAN20
u2267p2
JP
REN_U04
on2011p2
3007p2
HAN38
JHC
u4353p5
Merlin
u5699p4
3157
REN_U01
HAN13
u4941p4
3895p3
u4871p4
3301
u2263p2
u3436p3
u6524p4
u7545p4
on4013p5
u3101p2
u5718p2
AF1
u7524p2
u4643p5
u2743p2
U8
u5820p2
u5543p4
U11
VR1814
on3453p2
A
B
C
D
0.05
RL13
u3139p8
u6535p4
u2267p2
u2065p2
HAN20
cze3
HAN38
u4231p4
JP
3007p2
on2011p2
u4353p5
3301
u2263p2
u3436p3
u6524p4
u7545p4
on4013p5
u3101p2
JHC
AF1
u5718p2
u2743p2
u7524p2
u4643p5
U8
u5820p2
u5543p4
U11
VR1814
on3453p2
Merlin
u5699p4
REN_U01
HAN13
u4941p4
3895p3
u4871p4
A
C
B
E
F
0.05
UL73
u3101p2
u6535p4
u7524p2
u2267p2
u2263p2
JHC
HAN38
3895p3
u4643p5
HAN13
u4941p4
REN_U04
U8
AF1
on4013p5
u5718p2
u6524p4
u7545p4
U11
u5543p4
u5699p4
u2743p2
u5820p2
cze3
REN_U33
u3436p3
u2065p2
u3139p8
3301
on2011p2
3007p2
HAN20
on3453p2
VR1814
3157
JP
Merlin
u4231p4
u4353p5
u4871p4
D = gN-3a
C = gN-3b
B = gN-2
A = gN-1
E = gN-4
0.02
UL74
u6524p4
u7545p4
U11
u5543p4
u2743p2
u5699p4
JHC
u3101p2
u2263p2
u7524p2
u2267p2
u6535p4
VR1814
HAN20
on3453p2
U8
on4013p5
u5718p2
AF1
HAN38
3895p3
u4643p5
HAN13
u4941p4
JP
3157
u4231p4
Merlin
u4353p5
u4871p4
3301
3007p2
on2011p2
u2065p2
u3139p8
u3436p3
u5820p2
cze3
A = gO-1
B = gO-2b
C = gO-2a
D = gO-5
E = gO-4
F = gO-3
0.02

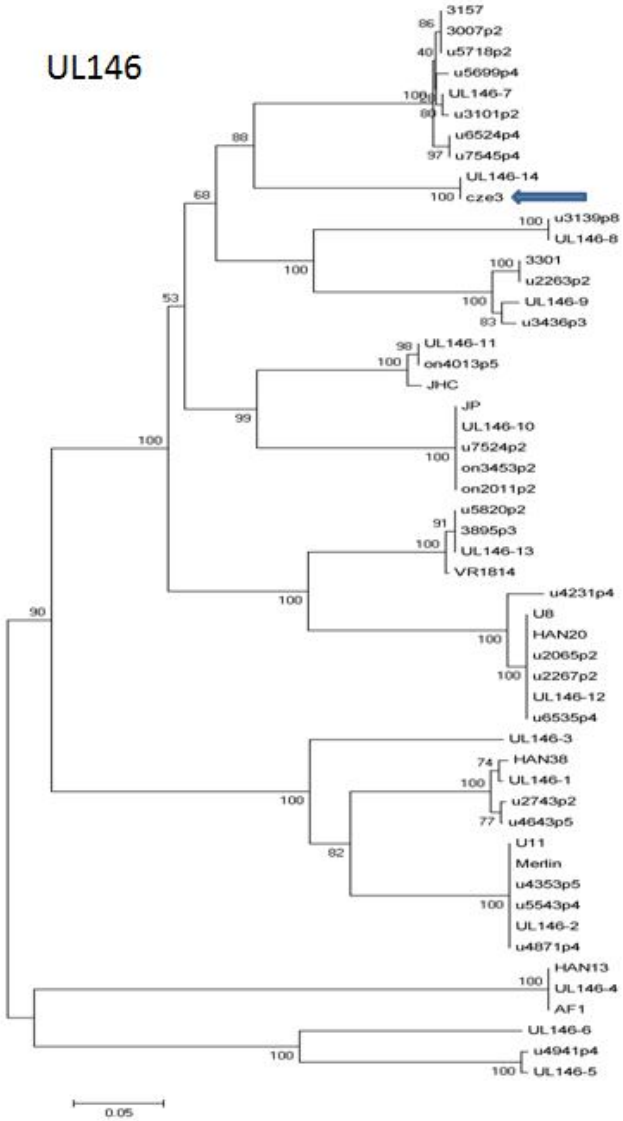


Figure 13. Phylogenetic trees of HCMV RL12, RL13, UL73, UL74 and UL146 genes. In these trees, isolates are clustered into subgroups (genotypes). The blue arrow shows the Czech isolate (cze3).

Table 9. Genotypes of clinical strains and clinical isolates in highly polymorphic genes. The shadowed areas indicate missing data for these genes. Stars indicate mutated genes. Individual genotypes are labeled by different colors. The prefix U in Belgium isolates is omitted. Czech isolate is marked as cze3. Two strains 2267 and 6535 have the same genotypes at all analyzed genes (Adapted from Steven Sijmons, Laboratory of Clinical Virology, Catholic University of Leuven, Belgium).

	RL5A	RL6	RL12	RL13	UL1	UL9	UL11	UL73 (gN)	UL74 (gO)	UL146	UL139
U8	B	D	D	E	E	C	A	B	B	A	A
U11	A	B	D	E	E	C	A	A	A	J	A
HAN13	A*	B	B	F	G	A	A	C	C	K	A
HAN20	A	A	A	A	A	H	D	E	A	A	B
HAN38	C	D*	A	A	A	A	C	C	C	I	A
Merlin	A	A	B	F	F	B	A	E	D	J	A
VR1814	A	A	D	E	E	C	A	E	A	B	A
AF1	A	B*	D	E	D	A*	C	B	B	K	A
3301	A	A	C	C	B	H	B	E	E	E	A
3157	A	C	B	F*	F	B	A	E	D	H	A
JHC	A	C	A	B		H	D	D	A	C	A
REN_U01	A	A	B	F	F						
REN_U04		A	A	A*	A			B			
REN_U33	A	A		A*	A	D	A	E			
JP	A*	A	A	A	A	E	D	E	D	D	A
2011	A	A	A	A	A	E*	D	E	E	D	B
2065	A*	C	A	A	A	H*	D	E	F	A	A
2263	C	C*	C	C	B	F	C	D	A	E	A
2267	A*	C	A	A	A	H*	D	D	A	A	B
2743	A	C	D	E	E	A	A	A	A	I	A
3007	A	A	A	A	A	E*	D	E	E	H	A
3101	A*	C	C	C		G	D	D	A	H	B
3139	A*	C	A	A	A	H	D	E	F	G	A
3436	A	B	C	C	B	H	C	E	F	F	A
3453	C	C*	D	E	E	C	A	E	A	D	B
3895	A	A	B	F	G	A*	A	C	C	B	A
4013	A	B	C	C		A	C	B	B	C	A
4231	A*	C	A*	A	A	E	D	E	D	A	A
4353	A	A	A	A	A	H	D	E	D	J	A
4643	A	A	D	E	E	C*	A	C	C	I	A
4871	A*	C*	B	F	G	A*	A	E	D	J	A
4941	A*	B	B	F	G	A*	A	C	C	L	A
5543	A	C	D	E	E	C	A	A	A	J	A
5699	A	A	B	F	F*	B*	A	A	A	H	A
5718	A	A	D	E	C	B	A	B	B	H	A
5820	B	D	D	E	E	C	A	E	F	B	B
6524	A*	C	C	C			D	A	A	H	B
6535	A*	C	A	A	A	H*	D	D	A	A	B
7524	B	D	D	E	E	C*	A	D	A	D	A
cze3	**A**	**B**	**A**	**A**	**A**	**H***	**D**	**E**	**F**	**N**	**B**

6. Discussion

My project was aimed at HCMV isolation and preparation of clinical isolates for the whole genome sequencing. Until recently, whole genome sequences of only a few HCMV isolates were reported (Murphy *et al.* 2003, Dolan *et al.* 2004, Sinzger *et al.* 2008, Cunningham *et al.* 2010, Dargan *et al.* 2010, Jung *et al.* 2011). NGS is an approach which allows to study the variability of viral isolates of large size. Due to high variability of HCMV strains, a number of isolates need to be sequenced to eventually predict targets for the diagnosis and therapy of HCMV-associated diseases and to link a particular sequence of the viral genome to a specific pathology.

NGS methods need quite a lot of starting viral DNA. In case of HCMV, the preparation of this material is not an easy task. First of all, about 70% of the healthy population are infected with HCMV (Roubalová and Seeman 1998, Hecker *et al.* 2004, Staras *et al.* 2006), but in these individuals, HCMV DNA cannot be found. It can be detected in the peripheral blood and urine upon reactivation of the virus. This is obviously a rare event in healthy individuals. Futhermore, primary infection of healthy adults when HCMV DNA can be temporally detected is also not common. Our data show that seroconversion occurred in only 2/500 healthy blood donors who were followed up for 2 years (personal observation, not published). The reactivation of the virus, viremia, and prolonged urinary shedding of the virus can be observed in immunosuppressed patients and congenitally infected children (Lazzarroto *et al.* 2008, de Vries *et al.* 2012).

My work was done at the Institute of Hematology and Blood Transfusion. The samples were collected from HSCT recipients. These patients are routinely monitored for the presence and viral load of HCMV by qPCR in the virological laboratory. Additionally, we were able to obtain two samples from congenitally infected children.

Patients with a HCMV viral load in the blood higher than 100 copies/10,000 g.e. were asked to provide a urine sample. It did not add to the burden of severely ill patients. The urine was used for the infection of a human fibroblast cell line and subsequently was tested for the presence of HCMV by nested PCR. Not surprisingly, only those samples positive on PCR revealed also CPE upon the infection of the human fibroblast cell line. Others have observed that a HCMV viral load of at least 1000 – 10,000 copies per milliliter of urine is needed for efficient infection of the cell culture (personal communications). For some HCMV strains, a prolonged time is needed for

the CPE to appear. We passaged the cells for a defined time not to introduce mutations caused by keeping the virus in the cell culture for a prolonged period. Viral DNA was extracted from all samples grown in the cell culture but a sufficient amount of it was obtained only from those samples where the CPE had appeared. Prolonged incubation of the cell culture might improve the DNA viral yield but at the expense of an increased mutation rate. Therefore, this issue needs a future investigation.

I have additionally attempted the cultivation of the virus from peripheral blood leukocytes. It is not used in the routine diagnosis. The yield of the successfully cultivated samples was not superior to that obtained from urine samples. A reason for this can be that in some of these patients, the treatment was already started and consequently, their viral load decreased considerably.

Two protocols were used to infect the human fibroblast cells with HCMV. One procedure was spinning the virus onto the cells ("shell vial"), as used for the detection of HCMV antigens in routine settings (Gleaves *et al.* 1984). This approach has been shown before to increase the infectivity of the virus by one magnitude. However, I failed to observe a higher CPE rate in the cell culture. For samples positive for HCMV DNA in the urine, the CPE rate was 40% when inoculated via filtration and 33% when inoculated by centrifugation.

Furthermore, this protocol cannot be used for samples with massive bacterial contamination which is regularly present in the urine of immunosuppressed patients (Donnelly 1995, Ninin *et al.* 2001). For some of these samples, the second protocol with the urine filtered before inoculation onto the cell culture was successfully used. Other samples had to be excluded due to toxicity for cells.

Even though the yield of HCMV DNA-positive samples was low regardless of the clinical specimens used and the method of inoculation applied, we confirmed that both types of clinical specimens - urine and blood – can be used for HCMV multiplication and DNA extraction for NGS. Inoculation of the filtered urine seems, in our hands, to be a superior approach because it provides a higher yield of viral DNA and is less demanding. In comparison to blood, urine collection places less burden on the patient.

As was mentioned before, NGS sequencing needs quite a lot of viral DNA. So far, only Cunningham *et al.* have shown on one HCMV strain (3301) that it is possible to obtain the whole genome sequence from a sample which contains as little as 3% of the viral DNA. However, in such case, a low coverage is obtained by NGS and the

assembly of the sequence is very difficult (Cunningham *et al.* 2010). Therefore, this approach is not feasible when sequencing many isolates. To overcome this problem, amplification of the viral DNA is performed prior to NGS sequencing.

Another problem with the preparation of HCMV DNA for NGS is the purity. Of seven samples positive for viral HCMV DNA after multiplication in tissue culture, only one had high purity (98%) allowing for relatively straightforward NGS and easy creation of the whole genome sequence. This isolate was sequenced on the Illumina platform using a combination of two approaches for composing short readings produced by the Illumina sequencer and de novo assembly followed by mapping assembly as is commonly used for NGS of clinical isolates(Cunningham *et al.* 2010, Beerenwinkel *et al*. 2011, Renzette *et al.* 2011). The smaller purity (less than 7%) of the remaining 6 samples, can be explained by the fact that micrococcal nuclease used for removal of the excess of cell DNA did not work properly. Higher purity of these samples could possibly also be achieved by increasing the number of passages but we have been currently working with the Belgian group under the same protocol and at this point, it was not desirable to make changes to it. The samples with low purity were also subjected to NGS and are currently under analyses. It can be expected that in this case, NGS will have a low coverage due to the presence of contaminating cell DNA and therefore, the assembly will be much more difficult.

The phylogeny analyzes of the Czech isolate showed that based on the sequence of each of the analyzed genes, the isolate is assigned to a different group. Comparing 11 polymorphic regions between 14 Czech clinical strains and 24 Belgian clinical isolates, only two of these samples were identical in all 11 polymorphic regions. This data confirms the results of previous studies analyzing only a limited number of genes (Bates *et al.* 2008, Manuel *et al.* 2009, Gorzer *et al.* 2010, Pignatelli *et al.* 2010, Ross *et al.* 2011, Renzette *et al.* 2011) and imply an even bigger variability of HCMV isolates than expected.

In the Czech isolate, mutations in the UL9 gene were detected. Others have shown that mutations frequently occur in genes of the RL11 family such as RL5A, RL6, and UL9, and in the UL128 locus (Dolan *et al.* 2004, Cunningham *et al.* 2010, Dargan *et al.* 2010). Analysis of the selected polymorphic regions revealed that the Czech isolate is the most dissimilar in the gene encoding the UL146 chemokine (vCXC-1). This protein functions as a neutrophil attractant and thus participates in the invasion of the host immune system (Penfold *et al.* 1999). The high variability of this gene,

confirmed in many other studies (Prichard *et al.* 2001, Arav-Boger *et al.* 2006, Aquayo *et al.* 2010), may affect viral escape from the immune surveillance.

As for UL73 and 74, this study has shown that most isolates from Belgium and also the Czech isolate belong to the gN-4 and gO-3 genotypes, respectively. gN-4 and gO-3 were previously linked with a more severe clinical manifestation of HCMV infection (Pignatelli *et al.* 2003 and 2004, Rossini *et al.* 2005, Roubalová *et al.* 2011). The Czech sample was obtained from a HSCT recipient with a severe manifestation of HCMV infection.

7. Conclusions

HCMV usually causes asymptomatic infections in the immunocompetent individuals but infection in immunosuppresed patients has severe consequences. Additionally, about 45% of women in child bearing age are at risk of primary HCMV infection which can lead to congenital infection with a severe pathological manifestation. Therefore, new preventive, diagnostic, and therapeutic methods are needed. To meet these needs, it is important to study in detail the sequences of many HCMV isolates. So far only some genes of multiple clinical isolates were studied. With the advent of a new technology – NGS, the possibilities opened up for sequencing the complete genomes of multiple isolates of HCMV. In this project, I was participating in a study conducted by the Laboratory of Clinical Virology, KU Leuven, Belgium. The preliminary data have shown that the preparation of the material from clinical samples for NGS still needs improvement and standardization. However, more than 30 new isolates from Belgium have already been analyzed. One of the Czech isolates was also fully sequenced and other six are being sequenced now. The comparison of numerous polymorphic regions of the sequenced isolates suggests even bigger variability in HCMV viruses than expected. Therefore, it is obvious that many samples must be sequenced before it is possible to specify the relationship between genotypes and disease severity, to map the geographical variation in the distribution of HCMV genotypes, or possibly to identify specific genes as diagnostic and therapeutic targets.

8. Literature

Aquayo F, Murayama T, Eizuru Y. (2010). UL146 variability among clinical isolates of human cytomegalovirus from Japan. Biol. Res. 43(4), 475-480.

Aquino V and Figueiredo LT. (2000). High prevalence of renal transplant recipients infected with more than one cytomegalovirus glycoprotein B genotype. J. Med. Virol. 61, 138-142.

Arav-Boger R, Foster CB, Zong JC, Pass RF. (2006). Human cytomegalovirus- encoded a-chemokines exhibit high sequence variability in congenitally infected newborns. J. Infect Dis. 193, 788–791.

Arav-Boger R, Willoughby RE, Pass RF, Zong JC, Jang WJ, Alcendor D, Hayward GS. (2002). Polymorphisms of the cytomegalovirus (CMV)-encoded tumor necrosis factor-alpha and beta-chemokine receptors in congenital CMV disease. J. Infect. Dis. 186(8), 1057-1064

Arista S, De Grazia S, Giammanco G, Di Carlo P, Iannitto E. (2003). Human cytomegalovirus glycoprotein B genotypes in immunocompetent, immunocompromised, and congenitally infected Italian populations. Arch Virol ,148, 547-554.

Balcarek KB, Bagley R, Cloud GA, Pass RF. (1990). Cytomegalovirus infection among employees of a children's hospital. No evidence for increased risk associated with patient care. JAMA. 263(6), 840-844.

Barbi, M., Corbetta, C., Primache, V., Binda, S., Caroppo, Primache V, Didò P, Guidotti P, Corbetta C, Melotti D. (2001). CMV gB genotypes and outcome of vertical transmission: study on dried blood spots of congenitally infected babies. J. Clin. Virol. 21, 75-79.

Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front. Microbiol. 3, 329.

Bradley AJ, Kovács 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. (2008). Genotypic analysis of two hypervariable human cytomegalovirus genes. J. Med. Virol. 80(9), 1615-1623

Bradley AJ, Lurain NS, Ghazal P, Trivedi U, Cunningham C, Baluchova K, Gatherer D, Wilkinson GWG, Dargan D, Davison AJ. (2009). High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. J. Gen. Virol. 90, 2375–2380.

Bodeus M, Van Ranst M, Bernard P, Hubinont C, Goubau P. (2002). Anticytomegalovirus IgG avidity in pregnancy: a 2-year prospective study. Fetal. Diagn. Ther. 17, 362–366

Boppana SB, Pass RF, Britt WJ, Stagno S, Alford CA. (1992). Symptomatic congenital cytomegalovirus infection:neonatal morbidity and mortality. Pediatr.Infect. Dis. J. 11, 93–99.

Boeckh M, Leisenring W, Riddell SR, Bowden RA, Huang ML, Myerson D, Stevens-Ayers T, Flowers ME, Cunningham T, Corey L. (2003). Late cytomegalovirus disease and mortality in recipients of allogeneic hematopoietic stem cell transplants: importance of viral load and T-cell immunity. Blood. 101(2), 407-414

Boeckh M and Nichols WG. (2004). The impact of cytomegalovirus serostatus of donor and recipient before hematopoietic stem cell transplantation in the era of antiviral prophylaxis and preemptive therapy. Blood. 103(6), 2003-2008.

Britt WJ, Vugler L. (1989). Processing of the gp55-116 envelope glycoprotein complex (gB) of human cytomegalovirus. J. Virol. 63, 403–410.

Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J. Virol. 70, 78–83.

Chee MS, Bankier AT, Beck S, Bohni R, Brown CM, Cerny R, Horsnell T, Hutchison CA., III, Kouzarides T and other authors (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. 154, 125–169.

Chou S. (1992). Effect of interstrain variation on diagnostic DNA amplification of the cytomegalovirus major mmediate-early gene region. J. Clin. Microbiol. 30, 2307–2310.

Clark-Lewis I, Schumacher C, Baggiolini M, Moser B. (1991). Structure activity relationships of interleukin-8 determined using chemically synthesized analogs. Critical role of NH2-terminal residues and evidence for uncoupling of neutrophil chemotaxis, exocytosis, and receptor binding activities. J. Biol. Chem. 266, 23128–23134.

Coaquette A, Bourgeois A, Dirand C, Varin A, Chen W, Herbein, G. (2004). Mixed cytomegalovirus glycoprotein B genotypes in immunocompromised patients. Clin. Infect. Dis. 39, 155-161.

Cook CH, Yenchar JK, Kraner TO, Davies EA, Ferguson RM. (1998). Occult herpes family viruses may increase mortality in critically ill surgical patients. Am. J. Surg. 176, 357–360.

Crapnell K, Zanjani ED, Chaudhuri A, Ascensao JL, St Jeor S, Maciejewski JP. (2000). In vitro infection of megakaryocytes and their precursors by human cytomegalovirus. Blood 95, 487–493.

Crough T and Khanna R. (2009). Immunobiology of human cytomegalovirus: from bench to bedside. Clin. Microbiol. Rev. 22(1), 76-98.

Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, Thomson M, Griffiths PD, Wilkinson GWG, Schulz TF, Davison AJ. (2010). Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. J. Gen. Virol. 91, 605–615.

Dargan DJ, Hassan-Walker AF, Emery VC, Griffiths PD, Wilkinson GWG. (2003). Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. J. Gen. Virol. 84, 657–663.

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. (2004). Mach M, Kropff B, Dal Monte P, Britt WJ. (2000). Complex formation of human cytomegalovirus glycoprotein M (gpUL100) and glycoprotein N (gpUL73). J Virol. 74, 11881–11892.

Donnelly JP. (1995). Bacterial complications of transplantation: diagnosis and treatment. J. Antimicrob. Chemother. 36 Suppl B, 59-72.

Drew WL, Chou S, Miner RC, Mohr BA, Busch, M.P, van der Horst CM, Asmuth DM, Kalish LA. (2002). Cytomegalovirus glycoprotein B groups in human immunodeficiency virus-infected patients with incident retinitis. J. Infect Dis. 186, 114-117.

Fries BC, Chou S, Boeckh M, Torok-Storb B. (1994). Frequency distribution of cytomegalovirus envelope glycoprotein genotypes in bone marrow transplant recipients. J. Infect. Dis. 169, 769–774.

Gambarotto K, Ranger-Rogez S, Aubard Y, Piver P, Duffetelle B, Delpeyroux C, Roussanne MC, Nicot T, Denis F. (1997). Primary cytomegalovirus infection and pregnant women: epidemiological study on 1100 women at Limoges. Pathol. Biol. (Paris). 45(6), 453-461.

Gharizadeh B, Kalantari M, Garcia CA, Johansson B, Nyrén P. (2001). Typing of human papillomavirus by pyrosequencing. Lab. Invest. 81(5),673-679.

Gleaves CA, Smith TF, Shuster EA, Pearson GR. (1984). Rapid detection of cytomegalovirus in MRC-5 cells inoculated with urine specimens by using low-speed centrifugation and monoclonal antibody to an early antigen. J. Clin. Microbiol. 19(6), 917-919.

Görzer I, Guelly C, Trajanoski S, Puchhammer-Stöckl E. (2010). Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. J. Virol. 84(14), 7195-7203

Habbal W, Monem F, Gärtner BC. (2009). Comparative evaluation of published cytomegalovirus primers for rapid real-time PCR: which are the most sensitive? J. Med. Microbiol. 58(7), 878-883.

Hassan-Walker AF, Okwuadi S, Lee L, Griffiths PD, Emery VC. (2004). Sequence variability of the a-chemokine UL146 from clinical strains of human cytomegalovirus. J. Med. Virol. 74, 573–579.

Hecker M, Qiu D, Marquardt K, Bein G, Hackstein H. (2004). Continuous cytomegalovirus seroconversion in a large group of healthy blood donors. Vox. Sang. 86(1), 41-44.

Hebart H and Einsele H. (2004). Clinical aspects of CMV infection after stem cell transplantation. Hum. Immunol. 65(5), 558-564

Heo J, Petheram S, Demmler G, Murph JR, Adler SP, Bale J, Sparer TE. (2008). Polymorphisms within human cytomegalovirus chemokine (UL146/UL147) and cytokine receptor genes (UL144) are not predictive of sequelae in congenitally infected children. Virology. 378(1), 86-96.

Hitomi S, Kozuka-Hata H, Chen Z, Sugano S, Yamaguchi N, Watanabe S. (1997). Human cytomegalovirus open reading frame UL11 encodes a highly polymorphic protein expressed on the infected cell surface. Arch. Virol. 142(7), 1407-1427.

Ho M. (2008). The history of cytomegalovirus and its diseases. Med. Microbiol. Immunol. 197(2), 65-73.

Hobom U, Brune W, Messerle M, Hahn G, Koszinowski U. (2000). Fast screening procedures for randomtransposon libraries of cloned herpesvirus genomes: mutational analysis of human cytomegalovirus envelope glycoprotein genes. J. Virol. 74, 7720–7729.

Humar A, Kumar D, Gilbert C, Boivin G. (2003). Cytomegalovirus (CMV) glycoprotein B genotypes and response to antiviral therapy, in solid- organ-transplant recipients with CMV disease. J. Infect. Dis. 188, 581–584.

Hutto C, Little EA, Ricks R, Lee JD, Pass RF. (1986). Isolation of cytomegalovirus from toys and hands in a day care center. J Infect. Dis. 154(3), 527-530.

Jin H, Wang X, Li S. (2007). Human cytomegalovirus glycoprotein B genotype correlates with different symptoms of infected infants. Intervirology. 50, 219-223.

Jung GS, Kim YY, Kim JI, Ji GY, Jeon JS, Yoon HW, Lee GC., Ahn JH., Lee KM, Lee CH. (2011). Full genome sequencing and analysis of human cytomegalovirus strain JHC isolated from a Korean patient. Virus Res. 156 (1-2), 113-120

Kim WR, Badley AD, Weisner RH, Porayko MK, Seaberg EC, Keating MR, Evans RW, Dickson ER, Krom RA and Paya CV. (2000). The economic impact of cytomegalovirus infection after liver transplantation. Transplantation. 69(3), 357-361

Kline JN, Hunninghake GM, He B, Monick MM, Hunninghake GW. (1998). Synergistic activation of the human cytomegalovirus major immediate early promoter by prostaglandin E2 and cytokines. Exp. Lung Res. 24, 3–14.

Lazzarotto T, Guerra B, Lanari M, Gabrielli L, Landini MP. (2008). New advances in the diagnosis of congenital cytomegalovirus infection. J. Clin .Virol. 41, 192–197.

Ljungman P. (2007). Risk assessment in haematopoietic stem cell transplantation: viral status. Best Pract. Res. Clin. Haematol. 20(2), 209-217.

Ljungman P, Brand R, Einsele H, Frassoni F, Niederwieser D, Cordonnier C. (2003). Donor CMV serologic status and outcome of CMV-seropositive recipients after unrelated donor stem cell transplantation: an EBMT megafile analysis. Blood. 102(13), 4255-4260.

Ljungman P, Griffiths P, Paya C (2002). Definitons of cytomegalovirus infection and disease in transplant recipient. Clin. Infect. Dis. 34, 1094-1097.

Lurain NS, Fox AM, Lichy HM., Bhorade SM, Ware CF, Huang DD, Kwan SP, Garrity ER and Chou S. (2006). Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. Virol. J. 3, 4.

Lurain NS, Kapell KS, Huang DD, Short JA, Paintsil J, Winkfield E, Benedict CA, Ware CF and Bremer JW. (1999). Human cytomegalovirus UL144 open reading frame: sequence hypervariability in low-passage clinical isolates. J. Virol. 73, 10040-10050.

Manuel O, Asberg A, Pang X, Rollag H, Emery VC, Preiksaitis JK, Kumar D, Pescovitz MD, Bignamini AA, Hartmann A, Jardine AG and Humar A. (2009). Impact of genetic polymorphisms in cytomegalovirus glycoprotein B on outcomes in solid-organ transplant recipients with cytomegalovirus disease. Clin. Infect. Dis. 49, 1160–1166.

Mao ZQ, He R, Sun M, Qi Y, Huang YJ, Ruan Q. (2007). The relationship between polymorphisms of HCMV UL144 ORF and clinical manifestations in 73 strains with congenital and/or perinatal HCMV infection. Arch. Virol. 152(1), 115-124.

Mardis ER. (2008). Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9, 387-402.

Mattick C, Dewin D, Polley S, Sevilla-Reyes E, Pignatelli S, Rawlinson W, Wilkinson G, Dal Monte P, Gompels UA. (2004). 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 318, 582-597.

Mendelson M, Monard S, Sissons P, Sinclair J. (1996). Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. J. Gen. Virol. 77, 3099–3102.

Meyer-Konig U, Haberland M, von Laer D, Haller O, Hufert FT.(1998). Intragenic variability of human cytomegalovirus glycoprotein B in clinical strains. J. Infect. Dis. 177, 1162–1169.

Metzker ML. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46.

Mocarski ES Jr., Shenk T, Pass RF. (2007). Cytomegaloviruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE. (eds.). Fields Virology, 5th edition, Volume 2. Lippincott Williams and Wilkins, a Wolters Kluwer Business, Philadelphia, USA, 2702-2772.

Murphy E, Yu D, Grimwood J, Schmutz J, Dickson M, Jarvis MA, Hahn G, Nelson JA, Myers RM, Shenk TE. (2003). Coding potential of laboratory and clinical strains of human cytomegalovirus. Proc. Natl. Acad. Sci. USA. 100, 14976-14981.

Murphy E and Shenk T. (2008). Human cytomegalovirus genome. In: Shenk T and Stinski MS. (eds.). Human Cytomegalovirus. Current Topics in Microbiology and Imunology 325. Springer-Verlag Berlin Heidelberg, Germany, 1-20.

Navarro D, Paz P, Tugizov Z, Toppk K, La Vail J, Pereira L. (1993). Glycoprotein B of human cytomegalovirus promotes virion penetration into cells, transmission of infection from cell to cell and fusion of infected cells. Virology. 197, 143–158.

Ninin E, Milpied N, Moreau P, André-Richet B, Morineau N, Mahé B, Vigier M, Imbert BM, Morin O, Harousseau JL, Richet H. (2001). Longitudinal study of bacterial, viral, and fungal infections in adult recipients of bone marrow transplants. Clin. Infect. Dis. 33, 41-47.

Nichols WG, Corey L, Gooley T, Davis C, Boeckh M. (2002). High risk of death due to bacterial and fungal infection among cytomegalovirus (CMV)-seronegative recipients of stem cell transplants from seropositive donors: evidence for indirect effects of primary CMV infection. J.Infect. Dis. 185(3), 273-282.

Pang XL, Chui L, Fenton J, LeBlanc B, Preiksaitis JK. (2003). Comparison of LightCycler-based PCR, COBAS amplicor CMV monitor, and pp65 antigenemia assays for quantitative measurement of cytomegalovirus viral load in peripheral blood specimens from patients after solid organ transplantation. J. Clin. Microbiol. 41(7), 3167-3174.

Pass RF (1985). Epidemiology and transmission of cytomegalovirus. J. Infect. Dis. 152, 243–248.

Paterson DA, Dyer AP, Milne RS, Sevilla-Reyes E, Gompels UA. (2002). A role for human cytomegalovirus glycoprotein O (gO) in cell fusion and a new hypervariable locus. Virology. 239, 281–294.

Pellet PE, Roizman B. (2007). The Family Herpesviridae: A Brief Introduction. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE. (eds.). Fields Virology, 5th edition, Volume 2. Lippincott Williams and Wilkins, a Wolters Kluwer Business, Philadelphia, USA, 2702-2772.

Penfold ME, Dairaghi DJ, Duke GM, Saederup N, Mocarski ES, Kemble GW, Schall TJ. (1999). Cytomegalovirus encodes a potent alpha chemokine. Proc. Natl. Acad. Sci. USA. 96, 9839–9844.

Peterson PK, Balfour HH, Marker SC, Fryd DS, Howard RJ, Simmons RL. (1980). Cytomegalovirus disease in renal allograft recipients: a prospective study of the clinical features, risk factors and impact on renal transplantation. Medicine. 59, 283-300.

Picone O, Costa JM, Chaix ML, Ville Y, Rouzioux C, Leruez-Ville M. (2005). Human cytomegalovirus UL144 gene polymorphisms in congenital infections. J. Clin. Microbiol. 43(1), 25-29.

Pignatelli S, Dal Monte P, Rossini G and Landini MP (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. Rev. Med. Virol. 14, 383–410.

Pignatelli S, Lazzarotto T, Gatto MR, Dal Monte P, Landini MP, Faldella G, Lanari M. (2010). Cytomegalovirus gN genotypes distribution among congenitally infected newborns and their relationship with symptoms at birth and sequelae. Clin. Infect. Dis.51(1), 33-41

Poole E, King CA, Sinclair JH, Alcami A. (2006). The UL144 gene product of human cytomegalovirus activates NFkappaB via a TRAF6-dependent mechanism. EMBO. J. 25(18), 4390-4399.

Prichard MN, Penfold MET, Duke GM, Spaete RR, Kemble GW. (2001). A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. Rev. Med. Virol. 11(3),191-200.

Prosch S, Wendt CE, Reinke P, Priemer C, Oppert M, Kruger DH, Volk HD, Docke WD. (2000). A novel link between stress and human cytomegalovirus (HCMV) infection: sympathetic hyperactivity stimulates HCMV activation. Virology. 272, 357–365.

Puchhammer-Stöckl E, Görzer I, Zoufaly A, Jaksch P,. Bauer CC, Klepetko W, Popow-Kraupp T. (2006). Emergence of multiple cytomegalovirus

strains in blood and lung of lung transplant recipients. Transplantation. 81, 187–194

Pumannova M, Roubalova K, Vitek A, Sajdova J. (2006). Comparison of quantitative competitive polymerase chain reaction–enzyme-linked immunosorbent assay with LightCycler-based polymerase chain reaction for measuring cytomegalovirus DNA in patients after hematopoietic stem cell transplantation. Diagn. Microbiol. Infect. Dis. 54(2), 115-120

Ozdemir E, St John LS, Gillespie G, Rowland-Jones S, Champlin RE, Molldrem JJ, Komanduri KV. (2002). Cytomegalovirus reactivation following allogeneic stem cell transplantation is associated with the presence of dysfunctional antigen-specific CD8+ T cells. Blood. 100(10), 3690-3697.

Qi Y, Mao ZQ, Ruan Q, He R, Ma YP, Sun ZR, Ji YH, Huang Y. (2006). Human cytomegalovirus (HCMV) UL139 open reading frame: Sequence variants are clustered into three major genotypes. J. Med. Virol. 78, 517–522.

Rasmussen L, Geissler A, Cowan C, Chase A, Winters M. (2002). The genes encoding the gCIII komplex of human cytomegalovirus exist in highly diverse combinations in clinical isolates. J. Virol. 76, 10841–10848.

Rasmussen L, Hong C, Zipeto D, Morris S, Sherman D, Chou S, Miner R, Drew WL, Wolitz R, Dowling A, Warford A, Merigan TC. (1997). Cytomegalovirus gB genotype distribution differs in human immunodeficiencyvirus- infected patients and immunocompromised allograft recipients. J. Infect. Dis. 175, 179-184.

Razonable RR, Brown RA, Wilson J, Groettum C, Kremers W, Espy M, Smith TF, Paya CV. (2002). The clinical use of various blood compartments for cytomegalovirus (CMV) DNA quantitation in transplant recipients with CMV disease. Transplantation. 73(6), 968-973.

Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. (2011). Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog. 7(5)

Rodier MH, Berthonneau J, Bourgoin A, Giraudeau G, Agius G, Burucoa C, Hekpazo A, Jacquemin JL. (1995). Seroprevalences of Toxoplasma, malaria, rubella, cytomegalovirus, HIV and treponemal infections among pregnant women in Cotonou, Republic of Benin. Acta. Trop. 59(4), 271-277.

Ross SA, Novak Z, Pati S, Patro RK, Blumenthal J, Danthuluri VR, Ahmed A, Michaels MG, Sánchez PJ, Bernstein DI, Tolan RW, Palmer AL, Britt WJ, Fowler KB, Boppana SB. (2011). Mixed infection and strain diversity in congenital cytomegalovirus infection. J Infect Dis. 204(7), 1003-1007.

Rossini G, Pignatelli S, Dal Monte P, Camozzi D, Lazzarotto T, Gabrielli L, Gatto MR, Landini MP. (2005). Monitoring for human cytomegalovirus infection in solid organ transplant recipients through antigenemia and glycoprotein N (gN) variants: evidence of correlation and potential prognostic value of gN genotypes. Microbes. Infect. 7(5-6), 890-896.

Roubalová K and Seeman J. (1998). Serologický přehled protilátek proti herpetickým virům CMV, EBV a VZV. Zprávy CEM. příloha 1: 29.

Roubalová K, Strunecký O, Vitek A, Žufanová S, Procházka B. (2011). Genetic variability of cytomegalovirus glycoprotein O in hematopoietic stem cell transplant recipients. Transpl. Infect. Dis. 13(3), 237-243.

Roubalová K, Žufanová S, Vítek A, Staňková M. (2009). Zastoupení různých genotypů glykoproteinu gB u pacientů s vysokým rizikem symptomatické infekce lidským cytomegalovirem (CMV). Epidemiol. Mikrobiol. Imunol. 58, 148-153.

Rubin RH. (2007). The pathogenesis and clinical management of cytomegalovirus infection in the organ transplant recipient: the end of the 'silo hypothesis'. Curr. Opin. Infect. Dis. 20(4), 399-407.

Schafer P, Braun RW, Mohring K, Henco K, Kang J, Wendland T, Kuhn JE. (1993). Quantitative determination of human cytomegalovirus target sequences in peripheral blood leukocytes by nested polymerase chain reaction and temperature gradient gel electrophoresis. J Gen Virol 74, 2699– 2707.

Schalasta G, Eggers M, Schmid M, Enders G. (2000). Analysis of human cytomegalovirus DNA in urines of newborns and infants by means of a new ultrarapid real-time PCR-system. J. Clin. Virol. 19(3),175-185.

Senechal B, Boruchov AM, Reagan JL, Hart DN, Young JW. (2004). Infection of mature monocyte-derived dendritic cells with human cytomegalovirus inhibits stimulation of T-cell proliferation via the release of soluble CD83. Blood. 103(11), 4207-4215.

Shepp D, Match M, Lipson S, Pergolizzi, RG. (1998). A fifth human cytomegalovirus glycoprotein B genotype. Res. Virol. 149, 109-114

Sinzger C, Hahn G, Digel M, Katona R, Sampaio KL, Messerle M, Hengel H, Koszinowski U, Brune W and Adler B. (2008). Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. J. Gen. Virol. 89, 359–368.

Söderberg-Nauclér C, Streblow DN, Fish KN, Allan-Yorke J, Smith PP, Nelson JA. (2001). Reactivation of latent human cytomegalovirus in CD14(+) monocytes is differentiation dependent. J. Virol. 75(16), 7543-7554.

Sohn YM, Oh MK, Balcarek KB, Cloud GA, Pass RF. (1991). Cytomegalovirus infection in sexually active adolescents. J. Infect. Dis. 163, 460–463.

Spaete RR, Saxena A, Scott PI, Song GJ, Probert WS, Britt WJ, Gibson W, Rasmussen L, Pachl C. (1990). Sequence requirements for proteolytic processing of glycoprotein B of human cytomegalovirus strain Towne. J Virol. 64, 2922–2931.

Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. (2006). Whole-genome multiple displacement amplification from single cells. Nat. Protoc. 1(4),1965-1970.

Stagno S and Whitley RJ. (1985). Herpesvirus infections of pregnancy. Part I: Cytomegalovirus and Epstein-Barr virus infections. N. Engl. J. Med. 313(20), 1270-1274.

Stanberry LR, Rosenthal SL, Mills L, Succop PA, Biro FM, Morrow RA, Bernstein DI. (2004). Longitudinal risk of herpes simplex virus (HSV) type 1, HSV type 2, and cytomegalovirus infections among young adolescent girls. Clin. Infect. Dis. 39(10), 1433-1438.

Staras SA, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. (2006). Seroprevalence of cytomegalovirus infection in the United States, 1988-1994. Clin. Infect. Dis.43(9), 1143-1151.

Stein J, Volk HD, Liebenthal C, Kruger DH, Prosch S. (1993). Tumour necrosis factor alpha stimulates the activity of the human cytomegalovirus major immediate early enhancer/promoter in immature monocytic cells. J. Gen. Virol. 74, 2333–2338.

Taylor-Wiedeman J, Sissons JG, Borysiewicz LK, Sinclair JH. (1991). Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. J. Gen. Virol. 72, 2059–2064.

The TH, van der Bij W, van den Berg AP, van der Giessen M, Weits J, Sprenger HG, van Son WJ. (1990). Cytomegalovirus antigenemia. Rev. Infect. Dis. 12 Suppl. 7, S734-744.

Trincado D, Scott G, White, P, Hunt C, Rasmussen L, Rawlinson WD. (2000). Human cytomegalovirus strains associated with congenital and perinatal infections. J. Med. Virol. 61, 481-487.

Tugizov S, Navarro D, Paz P, Wang Y, Quadri I, Pereira L. (1994). Function of human cytomegalovirus glycoprotein B: syncytium formation in cells constitutively expressing gB is blocked by virus neutralizing antibodies. Virology.; 201: 263–276.

de Vries JJ, van der Eijk AA, Wolthers KC, Rusman LG, Pas SD, Molenkamp R, Claas EC, Kroes AC, Vossen AC. (2012). Real-time PCR versus viral culture on urine as a gold standard in the diagnosis of congenital cytomegalovirus infection. J. Clin. Virol. 53(2), 167-170.

Woo PC, Lo CY, Lo SK, Siau H, Peiris JS, Wong SS, Luk WK, Chan TM, Lim WW and Yuen KY. (1997). Distinct genotypic distributions of cytomegalovirus (CMV) envelope glycoprotein in bone marrow and renal transplant recipients with CMV disease. Clin. Diagn. Lab. Immunol. 4, 515-518.

Yamamoto AY, Mussi-Pinhata MM, De Deus Wagatsuma VM, Marin LJ, Duarte G and Figuereido LT. (2007). Human cytomegalovirus glycoprotein B genotypes in Brazilian mothers and their congenitally infected infants. J. Med. Virol. 79, 1164-1168.

Yan H, Koyano S, Inami Y, Yamamoto Y, Suzutani T, Mizuguchi M, Ushijima H, Kurane I, Inoue N. (2008). Genetic variations in the gB, UL144 and UL149 genes of human cytomegalovirus strains collected from congenitally and postnatally infected Japanese children. Arch. Virol. 153, 667-674.

Zipeto D, Hong C, Gerna G, Zavattoni M, Katzenstein D, Merigan TC, Rasmussen L. (1998). Geographic and demographic differences in the frequency of human cytomegalovirus gB genotypes 1-4 in immunocompromised patients. AIDS. Res. Hum. Retroviruses. 14, 533-536.

http://bioinf.scri.ac.uk/tablet/download.shtml

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://sourceforge.net/projects/mira-assembler/files

http://sourceforge.net/projects/mummer

http://www.invitrogen.com

http://www.megasoftware.net/mega.php