1. Supplemental Results and Discussion

Since the submission of the thesis I have succeeded to analyze additional samples which were prepared for the next-generation sequencing but were not sequenced by the thesis submission deadline. The following samples were additionally sequenced: CMV15-2M (cze15), CMV21M (cze21), CMV26M (cze26), CMV33K (cze33k), and CMV 34K (cze34k). Due to low purity of viral DNA in those samples, low coverage of viral DNA sequencing was obtained and sequencing had to be repeated multiple times. Still more missing areas were generated by sequencing these isolates and they had to be compensated for by Sanger sequencing. Therefore, the analysis of these additional isolates was time consuming.

The sequences of the new Czech isolates CMV15-2M (cze15), CMV21M (cze21), CMV26M (cze26), CMV33K (cze33k), CMV34K (cze34k), and the previously sequenced Czech isolate CMV3M (cze3), 24 sequenced Belgian strains, isolates with the full genome sequences available in the public databases, were subjected to phylogenetic analysis. For this purpose, the following genes were selected: RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL73, UL74, UL139, and UL146. The sequence of each of the genes specified above was chosen based on the alignment of HCMV Merlin strain (AY446894.2) with the Czech isolates. Phylogenetic trees were constructed in MEGA 5.05 for each gene separately as specified in detail in the Material and methods section. According to the topology of the trees, genotypes of the isolates (A-H) were determined. As is shown in Table 1 and Figure 1, different strains were clustering together at the gene level but differed in genotypes of individual genes.

Mutations in the UL9 gene were present in 50 % of the Belgian and in 100 % of the Czech isolates. In two Czech isolates (33.3 %), CMV21 and CMV34K, mutations in the RL5A gene were also found. This gene was mutated in 37.5 % of the Belgian isolates. Furthermore, in CMV34K, the RL6 gene was deleted. Such deletion was not observed in other isolates compared. Others studies have shown that mutations frequently occur in genes of the RL11 family (RL5A, RL6, and UL9) (Cunningham *et al.* 2010, Dargan *et al.* 2010), but function of the products of these genes as well as functional changes caused by the mutations are not known yet.

The analysis of the UL73 and UL74 genes showed that 66% and 55% of the Czech isolates belong to genotypes gN4 and gO3, respectively. Comparable frequency of gN4 genotype (41.6 %) was detected in the Belgian isolates, while the frequency of gO3 was much lower (16.7 %). The genotypes gN4 and gO3 were previously linked with a more severe clinical manifestation of HCMV infection (Pignatelli *et al.* 2004, Rossini *et al.* 2005, Roubalová *et al.* 2011). All Czech isolates which belong to these genotypes were obtained from a HSCT recipient with a severe manifestation of HCMV infection.

The analysis of the gene encoding the UL146 chemokine (vCXC-1) revealed its high variability. Only two Czech isolates (CZE34K and cze3) had the same genotype, 14 (N). This genotype was detected only in between Czech but not Belgian isolates. Two other Czech isolates, CMV15-2M and CMV26M, and a Belgian one, 3139, had the same UL146 genotype 7 (H). Protein vCXC-1 participates in the invasion of the host immune system (Penfold *et al.* 1999). The high variability of this gene, shown also by others (Prichard *et al.* 2001, Arav-Boger *et al.* 2006, Aquayo *et al.* 2010), may affect viral escape from the immune surveillance.

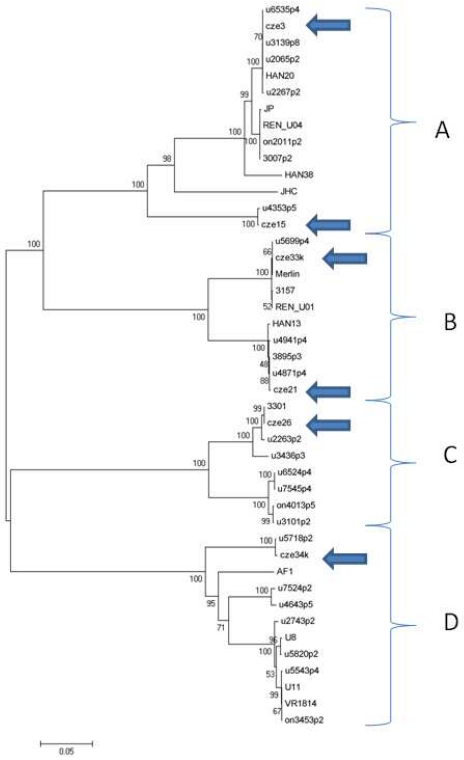
From the preliminary data on the comparison of six new Czech strains with 24 Belgian strains, it is evident that the HCMV genome is highly variable. In the studied set of Czech and Belgian isolates, only two (u2267 and u6535) were identical in all monitored genes (Table 1). This data confirms the results of previous studies which analyzed only a limited number of genes (Manuel *et al.* 2009, Gorzer *et al.* 2010, Pignatelli *et al.* 2010, Ross *et al.* 2011, Renzette *et al.* 2011) and implies an even greater variability of HCMV isolates than expected.

Table 1. Genotypes of clinical strains and clinical isolates in highly polymorphic genes.

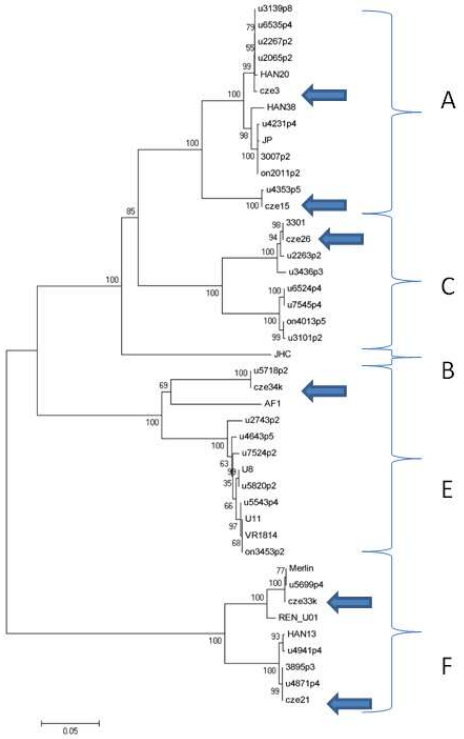
	RL5A	RL6	RL12	RL13	UL1	UL9	UL11	UL73 (gN)	UL74 (gO)	UL146	UL139
U8	B	D	D	E	E	C	A	B	B	A	A
U11	A	B	D	E	E	C	A	A	A	J	A
HAN13	A*	B	B	F	G	A	A	C	C	K	A
HAN20	A	A	A	A	A	H	D	E	A	A	B
HAN38	C	D*	A	A	A	A	C	C	C	I	A
Merlin	A	A	B	F	F	B	A	E	D	J	A
VR1814	A	A	D	E	E	C	A	E	A	B	A
AF1	A	B*	D	E	D	A*	C	B	B	K	A
3301	A	A	C	C	B	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**
cze15	**A**	**B**	**A**	**A**	**A**	**H***	**D**	**E**	**F**	**G**	**A**
cze21	**A ***	**A**	**B**	**F**	**G**	**E ***	**B**	**C**	**C**	**H**	**A**
cze26	**A**	**A**	**C**	**C**	**B**	**E ***	**B**	**E**	**A**	**G**	**A**
cze33k	**A**	**A**	**B**	**F**	**F**	**B***	**A**	**D**	**A**	**A**	**A**
cze34k	**A ***		**D**	**E**	**C**	**B***	**A**	**E**	**F**	**N**	**B**

The shadowed areas indicate missing data for these genes. Stars indicate mutated genes. Individual genotypes are labeled by different colours. Czech isolates are designated cze3, cze15, cze21, cze26, cze33k, and cze34k. k means isolates from blood. The prefix U in Belgian isolates is omitted. Two strains 2267 and 6535 have the same genotypes in all analyzed genes (Adapted from Steven Sijmons, Laboratory of Clinical Virology, Catholic University of Leuven, Belgium).

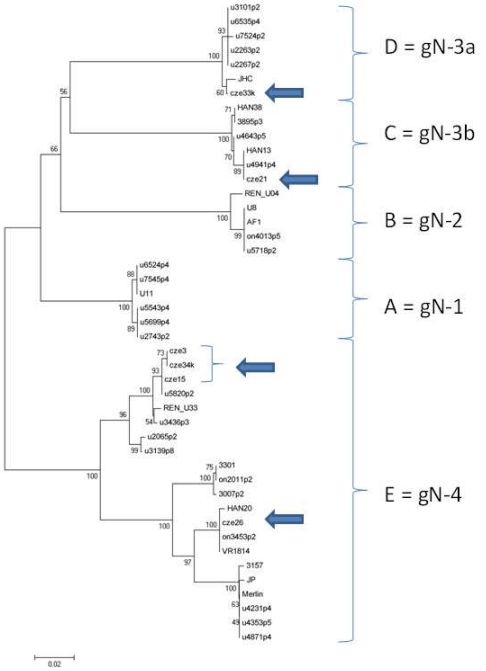
RL12



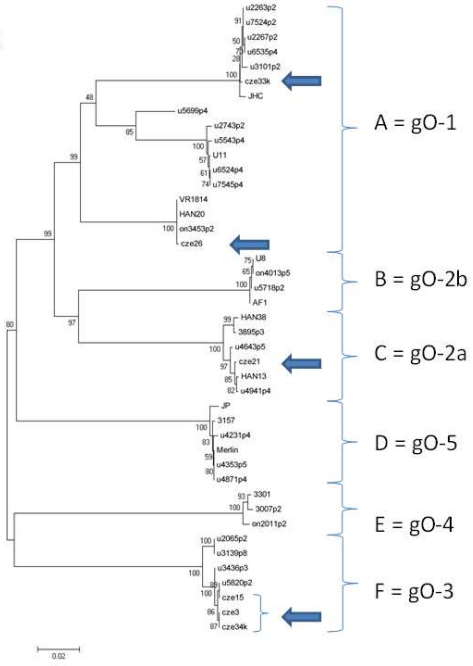
RL13



UL73



UL74



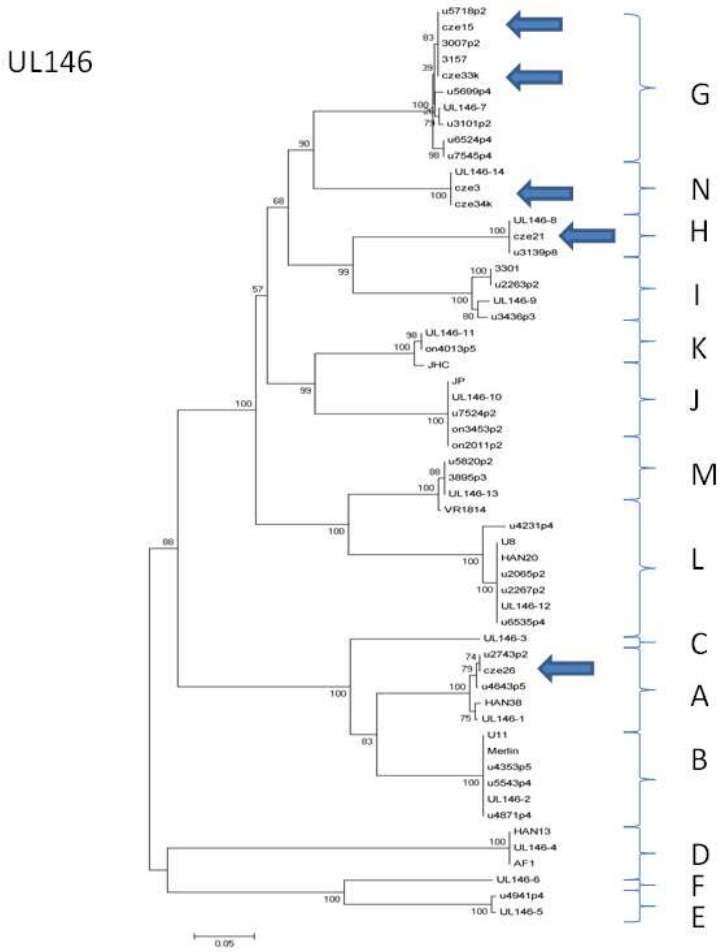


Figure 1. Phylogenetic trees of HCMV isolates for the RL12, RL13, UL73, UL74, and UL146 genes. In these trees, isolates are clustered into subgroups (genotypes) labelled by letters. The blue arrows show the positions of the Czech isolates.

2. Conclusions

HCMV usually causes asymptomatic infections in immunocompetent individuals but infection in immunosuppressed patients can have 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 and 6 new Czech isolates have already been analyzed. The comparison of numerous polymorphic regions of the sequenced isolates suggests even greater variability in HCMV viruses than expected. Therefore, it is obvious that many samples must be sequenced before it might be possible to reveal 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.

3. References

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

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.

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, Douglas E, Cunningham C, Jamieson F, Stanton RJ, Baluchova K, McSharry BP, Tomasec P, Emery VC, Percivalle E, Sarasini A, Gerna G, Wilkinson GW, Davison AJ. (2010). Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. J. Gen. Virol. 91, 1535-1546.

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

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.

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.

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

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.

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)

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, 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.