



Short communication

Determination of whole genome sequence of human cytomegalovirus circulating in Japan and discovery of geographic genome structure in *UL148* gene

Yuji Wada ^{*}, Ken Ishioka, Tatsuo Suzutani

Department of Microbiology, School of Medicine, Fukushima Medical University, Japan

ARTICLE INFO

Keywords:

Human cytomegalovirus
Phylogenetic analysis

ABSTRACT

Human cytomegalovirus (HCMV), a member of the Herpesviridae family, is prevalent worldwide. HCMV is generally asymptomatic but causes severe disease in immunocompromised patients or infants who are congenitally infected. Recent advances in sequencing technology have led to the rapid expansion of the HCMV genetic database, providing a comprehensive resource for studying viral genetics. Although genetic investigations have been vigorously performed in European countries, information on the whole-genome sequence of HCMV in the East Asian region remains limited. In this study, we determined whole-genome sequences of two clinical isolates of HCMV circulating in Japan. Partial genome sequences of 26 genes in UL/b' region were also identified using additional seven clinical isolates. Phylogenetic analysis of the *UL148* gene revealed a characteristic genetic clade predominantly constructed from HCMV isolates from Japan and China, suggesting a geographic gene structure in the East Asian region. We consider that this research will contribute to expanding the genetic database of HCMV and unveiling novel genetic characteristics of HCMV in Asia.

Human cytomegalovirus (HCMV) is a member of the herpesvirus family and possesses approximately 235 kbp of double-stranded DNA as viral genome [1]. HCMV is commonly transmitted worldwide, and their seroprevalences are estimated to range from 66 to 90 % depending on the region or country [1,2]. Most patients show asymptomatic or self-limiting mild symptoms by HCMV infection; however, the virus causes severe disease in immunocompromised patients, such as transplant recipients [1,3]. In cases HCMV infects to or reactivates in pregnant women, fetuses are transplacentally infected known as congenital cytomegalovirus infection (cCMV), and this disease causes a variety of symptoms represented by sensorineural hearing loss at birth or in childhood [4]. Vaccine development has advanced to prevent injury from CMV infection, but no licensed vaccine is available to date [5].

Recently, long-read next-generation sequencers have been developed and popularized in laboratory and clinical settings, and whole-genome sequences of HCMV have been vigorously analyzed [6–8]. The enrichment of HCMV sequence information unveiled geographic gene structures in its genome, which might be beneficial for designing global vaccines [9]. However, most of the complete genome sequences available in the database have been reported in European countries. To investigate the geographical characteristics of the HCMV genome

further, it is essential to expand the global genetic database.

Here, we analyzed the whole-genome sequences of clinical isolates of HCMV derived from infants with symptomatic cCMV using nanopore sequencing. Briefly, sequencing libraries were prepared using the SQK-LSK114 ligation sequencing kit (Oxford Nanopore Technology (ONT), Oxford, UK) following the manufacturer's protocol. The libraries were loaded onto FLO-MIN114 and sequenced using MinION Mk1C (ONT). The obtained data were base-called using the Guppy software (ONT). The obtained reads were tentatively mapped onto complete HCMV genome sequences available in GenBank using Minimap2 (version 2.17). The mapped reads were output using SAMtools (version 1.9) and visualized using GENETYX-NGS/MAC (version 5.0.5, GENETYX, Tokyo, Japan). In preliminary analysis of nanopore sequencing, it was observed that several genomic regions of the analyzed-HCMV isolates did not align with the complete HCMV genome sequences deposited in GenBank. To understand whole genome structures of the analyzed-HCMV isolates, tentative consensus sequences were determined by manually connecting the unmapped HCMV genomic regions. The samples were then re-analyzed by nanopore sequencing employing an adaptive sampling method using the tentative consensus sequences as a reference. The adaptive sampling method could efficiently analyze mapped reads

^{*} Corresponding author.

E-mail address: y-wada@fmu.ac.jp (Y. Wada).

<https://doi.org/10.1016/j.virusres.2025.199540>

Received 29 October 2024; Received in revised form 30 January 2025; Accepted 31 January 2025

0168-1702/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

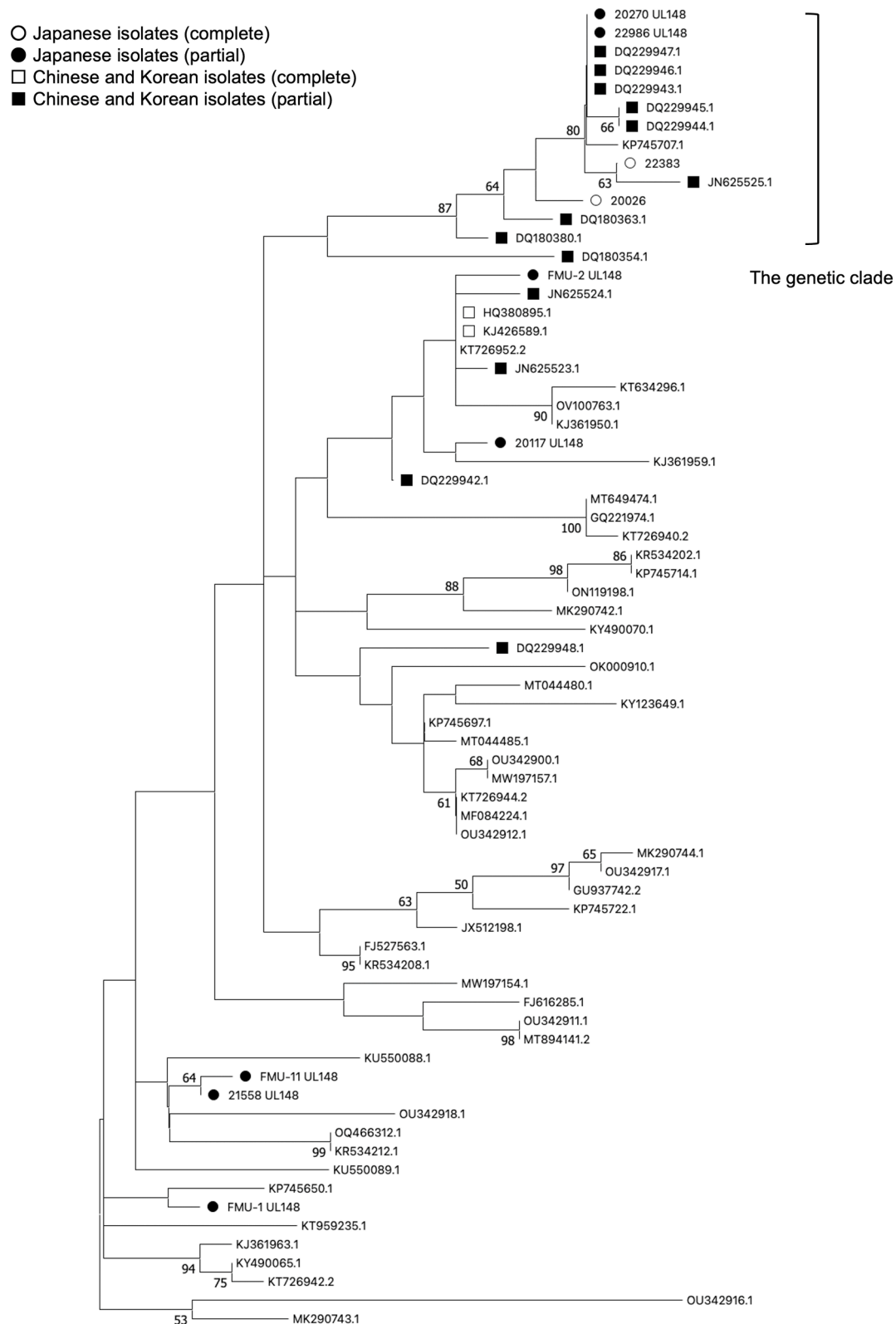


Fig. 1. Phylogenetic analysis of HCMV UL148 gene. A phylogenetic tree was constructed using nucleotide sequences of the HCMV UL148 gene by Maximum Likelihood phylogenetic analysis in the MEGA11 program. A total of 71 HCMV UL148 sequences (951 bp in length) were used for analysis. The percentage of trees in which the associated taxa clustered together was calculated by 1000 replicates, and the frequencies of >50 % were shown at the branches. The two sequences derived from whole genome sequences of HCMV determined in this study are depicted by white circle. The other seven sequences determined in this study are depicted by black circle. The deposited sequences in GenBank, which were isolated in China and Korea, were depicted by white (complete sequences) and black (partial sequences) squares, respectively.

22383	1	MLRLFLTLVLLALHGPSVNA	SRDYVHVRLLSYRGDPLVFK	HTFSGVRRPFTELGWAACRD	WDSMHCTPFWSTDLEQITDS	80
BE/13/2012	1	80
20026	1	80
JHC	1M.....	80
HAN	1M.....	80
HANChild4	1M.....	80
SYD-SCT2	1P..M.....	80
UK/Lon3/Plasma/2012	1P..M.....	80
BE/29/2010	1M.....	80
JER2282	1M.....	80
AD169	1M.....	80
DUR10	1G.....M.....	80
LUS283	1Q..G.....M.....	80
NAN2LA	1Y...D.....M.....	80
PRA3	1Y...D.....M.....	80
Towne	1Y.....M.....	80
NL/Rot1/Urine/2012	1VF.....	80
22383	121	ENVGLYVAYVVDGERPQQF	FTPQVDVVRFAlyLETLSRI	VEPLESGRLAVEFDTPDLTL	APDLVSSLYVAGHGETDFYM	200
BE/13/2012	121	200
20026	121A.....	200
JHC	121A.....F.....	200
HAN	121A.....F.....	200
HANChild4	121A.....F.....	200
SYD-SCT2	121A.....F.....	200
UK/Lon3/Plasma/2012	121A.....F.....	200
BE/29/2010	121V.....A.....F.....	200
JER2282	121V.....A.....F.....	200
AD169	121A.....F.....	200
DUR10	121A.....F.....	200
LUS283	121K.....A.....F.....	200
NAN2LA	121A.....F.....	200
PRA3	121A.....F.....	200
Towne	121A.....F.....	200
NL/Rot1/Urine/2012	121A.....V.....F.....	200

Fig. 2. Amino acid sequence alignment of HCMV UL148. Multiple alignment analysis of HCMV UL148 are shown. The amino acid sequences used in this analysis were randomly corrected by regions or countries. As exceptions, Towne and NL/Rot1/Urine/2012 strains were manually employed in this analysis, because amino acid BLAST analysis showed these strains possessed M77I and F189Y mutations. Dots indicate identical amino acid. Non-synonymous mutations specific in the genetic clade of the *UL148* gene are highlighted in gray.

with no unfavorable procedures or effects, and enough amounts of data have been obtained to determine whole genome sequences of HCMV [10].

Finally, the complete genome sequences of two HCMV isolates, 22,383 and 20,026 (accession numbers: LC846338 and LC846339), were determined. The obtained data comprised 29,462 mapped reads ranging up to 22,893 nt with an average coverage depth of 70.85 reads/nt in isolate 22,383, and 27,299 mapped reads ranging up to 37,232 nt with an average coverage depth of 44.15 reads/nt in isolate 20,026. Sanger sequencing was performed to confirm the genomic regions where frameshift mutations were suggested by nanopore sequencing and to correct them as needed.

The sequences of the *UL148* gene in 22,383 and 20,026 were similar to HCMV isolates from China. To investigate the genetic character of the *UL148* gene further, we additionally exploited seven clinical isolates of HCMV, which had also been isolated from infants with symptomatic cCMV in Japan, and the nucleotide sequences of the *UL148* gene were determined (accession numbers are listed in the Supplemental Table). Phylogenetic analysis was performed using MEGA11 software (version 11.0.13) using Maximum Likelihood method and Tamura-Nei model [11]. The phylogenetic tree of the *UL148* gene showed that four of nine analyzed-HCMV isolates (44.4 %) formed a characteristic genetic clade predominantly constructed from clinical isolates in Japan and China (Fig. 1). As an exception, the BE/13/2012 strain (KP745707) isolated in Belgium also belongs to this clade. Charles *et al.* reported that some HCMV strains isolated in Europe were derived from an Asian population; therefore, the BE/13/2012 strain might originate from the Asian region [9]. Alternatively, it would be also possible that Asian type of the *UL148* gene was introduced into origin of the BE/13/2012 strain in Europe

through superinfection and recombination event [7,12]. Considering the genetic features of HCMV, it would be understandable some European HCMV strains possessed geographically characteristic gene structures in Asian region. Nucleotide BLAST analysis confirmed that no other exceptional strains belonged to this clade. Based on these observations, the characteristic gene structures in the East Asian region are considered to be conserved in the *UL148* gene.

Among the HCMVs constructing the genetic clade of the *UL148* gene, 13–15 nucleotide substitutions, including three non-synonymous mutations (M77I, A179T, and F189Y), were observed compared to the HAN (KJ426589) and JHC (HQ380895) strains, whose complete genome sequences were available in GenBank as East Asian HCMV strains (Fig. 2). The M77I and F189Y mutations were common in all strains constructing the genetic clade of the *UL148* gene, but they were also observed in several clade-unrelated strains. Although several HCMV strains possessed the M77I and F189 mutations in *UL148*, amino acid BLAST analysis confirmed that these mutations were quite uncommon comparing to available HCMV sequences in GenBank. In contrast, the A179T mutation was exclusively observed in the genetic clade of the *UL148* gene, even though it was not observed in 20,026. The *UL148* gene encodes an endoplasmic reticulum-resident glycoprotein, named as UL148, which is involved in the evasion of host immunity and alteration of viral cell tropism through the regulation of envelope glycoprotein expression [13,14]. Based on the role of the UL148 protein, M77I, A179T, and F189Y mutations may have emerged as a result of ethnicity-dependent immunological pressure. To assess whether these amino acid substitutions were positively selected, the dN/dS at the codon level was estimated using the FEL method in the HyPhy package [15,16]. No significant differences were observed in these three codons,

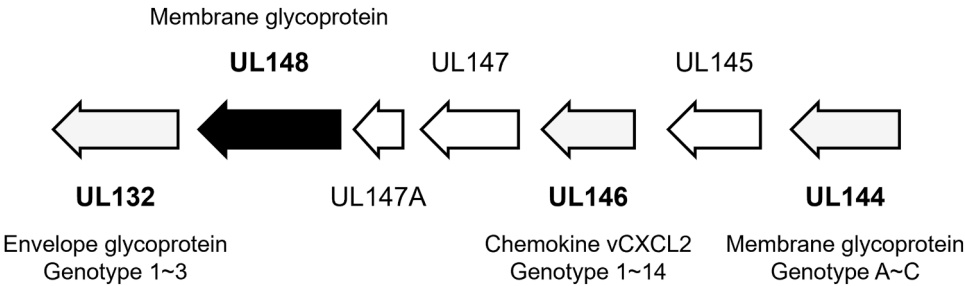


Fig. 3. A diagram of protein-coding ORF in UL/b' region of HCMV. The genetic organization between the UL132 gene to the UL144 gene of HCMV are shown. Protein coding region and direction are indicated by arrows. The *UL148* gene is highlighted in black. The genes analyzed their genotype (the *UL132*, *UL146*, and *UL144* genes) are highlighted in grey.

Table 1
Genotyping analysis of the *UL132*, *UL146*, and *UL144* genes.

Sample ID	Sequence availability	AA mutations in <i>UL148</i>	Genotype (group ID)		
			<i>UL132</i>	<i>UL146</i>	<i>UL144</i>
22,383	complete	M77I	2	6	B
22,986	partial	A179T	3	12	B
20,270	partial	F189Y	2	12	C
20,026	complete	M77I	1	11	B
		F189Y			
FMU-1	partial	Not	1	7	A
FMU-2	partial	Detected	3	1	C
FMU-11	partial		1	8	C
20,117	partial		3	9	A
21,558	partial		1	1	C

suggesting no evidence of positive selection, at least in the statistical analysis (M77I: $p = 0.846$, A179T: $p = 0.667$, and F189Y: $p = 0.379$). Further investigations are needed to understand the role of these mutations, including the biological activity of *UL148*.

Finally, the analysis target was expanded to the UL/b' region of HCMV, where the *UL148* gene was located. Besides the *UL148* gene, nucleotide sequences of 25 genes in the UL/b' region were determined using the 7 clinical isolates of HCMV in Japan, and phylogenetic analysis was performed for all of the analyzed-genes (Supplemental Table and Figure). Genetic clades predominantly constructed from HCMV isolates in Japan and China were also observed in the *UL132* and *UL147* genes, which were located surrounding the *UL148* gene. However, these genetic clades were constructed from fewer HCMV isolates in Japan than the *UL148* gene (two of nine analyzed-HCMV isolates) and additional non-Asian HCMV strains. Genetic variabilities were also analyzed by targeting the *UL132*, *UL146* and *UL144* genes, which have been previously investigated for the genotyping of HCMV (Fig. 3) [17–19]. Various genotype sets were observed in the analyzed-isolates; however, no consistent genotypes were observed among respective pairs of the *UL132*, *UL146*, and *UL144* genes in 22,383, 22,986, 20,270, and 20,026 isolates (Table 1). Therefore, the genotypes of the *UL132*, *UL146* and *UL144* genes would not be associated with the M77I, A179T, and F189Y mutations in the *UL148* gene. These observations suggest that the characteristic gene structures in the East Asian region were partially common in *UL148* and the surrounding genes but were mainly conserved in the *UL148* gene. There was low nucleotide diversity and no evidence of recombination in the *UL148* gene [7]. Considering its independence and stability, it is possible that the synonymous and non-synonymous mutations accumulated in the *UL148* gene through local virus circulation and reflected the geographic viral evolution. On the other hand, the phylogenetic analysis of the *UL148* gene showed that five of nine analyzed-HCMV isolates were not classified into the genetic clade and dispersed through the phylogenetic tree with non-Asian HCMVs (Fig. 1). This phenomenon suggested that two types of HCMVs having distinct origins have been circulating in Japan, which were

locally evolved one and globally spread one.
In this study, we determined the whole and partial genome sequences of clinical isolates of HCMV circulating in Japan. We also identified a novel form of the *UL148* gene in clinical isolates from Japan and China. This report will contribute to overcoming the obstacles to the genetic investigation of HCMV, which is limited to whole-genome sequence information of HCMV in the East Asian region.

CRediT authorship contribution statement

Yuji Wada: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ken Ishioka:** Writing – review & editing, Resources. **Tatsuo Suzutani:** Writing – review & editing, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Prof. Ikuo Wada (Department of Cell Science, School of Medicine, Fukushima Medical University) for supporting the nanopore sequencing analyses. We thank Prof. Kazufumi Ikuta (Department of Clinical Laboratory Science, Medical Science and Technology, School of Health Science, College of Medical, Pharmaceutical and Health Sciences, Kanazawa University) for carefully reading and giving advices on this manuscript.

This work was supported by JSPS KAKENHI Grant Number JP24K18454.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2025.199540.

Data availability

Data will be made available on request.

References

[1] Grgic, I., Gorenec, L., 2024. Human Cytomegalovirus (HCMV) genetic diversity, drug resistance testing and prevalence of the resistance mutations: a literature review. *Trop. Med. Infect. Dis.* 9 (2).
[2] Zuhair, M., Smit, G.S.A., Wallis, G., Jabbar, F., Smith, C., et al., 2019. Estimation of the worldwide seroprevalence of cytomegalovirus: a systematic review and meta-analysis. *Rev. Med. Virol.* 29 (3), e2034.
[3] Meesing, A., Razonable, R.R., 2018. New developments in the management of Cytomegalovirus infection after transplantation. *Drugs* 78 (11), 1085–1103.

- [4] Chiopris, G., Veronese, P., Cusenza, F., Procaccianti, M., Perrone, S., et al., 2020. Congenital Cytomegalovirus infection: update on diagnosis and treatment. *Microorganisms*. 8 (10).
- [5] Boppana, S.B., van Boven, M., Britt, W.J., Gantt, S., Griffiths, P.D., et al., 2023. Vaccine value profile for cytomegalovirus. *Vaccine* 41 (Suppl 2), S53–S75.
- [6] Li, K.K., Lau, B., Suarez, N.M., Camiolo, S., Gunson, R., et al., 2023. Direct nanopore sequencing of Human Cytomegalovirus genomes from high-viral-load clinical samples. *Viruses*. 15 (6).
- [7] Sijmons, S., Thys, K., Mbong Ngwese, M., Van Damme, E., Dvorak, J., et al., 2015. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J. Virol.* 89 (15), 7673–7695.
- [8] Karamitos, T., van Wilgenburg, B., Wills, M., Klennerman, P., Magiorkinis, G., 2018. Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. *BMC Genomics*. 19 (1), 577.
- [9] Charles, O.J., Venturini, C., Gantt, S., Atkinson, C., Griffiths, P., et al., 2023. Genomic and geographical structure of human cytomegalovirus. *Proc. Natl. Acad. Sci. U.S.A.* 120 (30), e2221797120.
- [10] Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M.D., et al., 2022. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* 23 (1), 11.
- [11] Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027.
- [12] Cudini, J., Roy, S., Houldcroft, C.J., Bryant, J.M., Depledge, D.P., et al., 2019. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. *Proc. Natl. Acad. Sci. U.S.A.* 116 (12), 5693–5698.
- [13] Nguyen, C.C., Siddiquey, M.N.A., Zhang, H., Li, G., Kamil, J.P., 2018. Human Cytomegalovirus tropism modulator UL148 interacts with SEL1L, a cellular factor that governs endoplasmic reticulum-associated degradation of the viral envelope glycoprotein gO. *J. Virol.* 92 (18).
- [14] Wang, E.C.Y., Pjechova, M., Nightingale, K., Vlahava, V.M., Patel, M., et al., 2018. Suppression of costimulation by human cytomegalovirus promotes evasion of cellular immune defenses. *Proc. Natl. Acad. Sci. U.S.A.* 115 (19), 4998–5003.
- [15] Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22 (5), 1208–1222.
- [16] Wei, Y., Cai, Y., Han, X., Han, Z., Zhang, Y., et al., 2023. Genetic diversity and molecular evolution of Seoul virus in Hebei province, China. *Infect. Genet. Evol.* 114, 105503.
- [17] Lurain, N.S., Fox, A.M., Lichy, H.M., Bhorade, S.M., Ware, C.F., et al., 2006. Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. *Virol. J.* 3, 4.
- [18] Sun, Z.R., Ji, Y.H., Ruan, Q., He, R., Ma, Y.P., et al., 2006. Genetic variability of human cytomegalovirus UL132 gene in strains from infected infants. *Microbiol. Immunol.* 50 (10), 773–779.
- [19] Chen, H.P., Jiang, J.K., Chan, C.H., Teo, W.H., Yang, C.Y., et al., 2015. Genetic polymorphisms of the human cytomegalovirus UL144 gene in colorectal cancer and its association with clinical outcome. *J. Gen. Virol.* 96 (12), 3613–3623.