University of Szeged
Albert Szent-Györgyi Medical School
Doctoral School of Multidisciplinary Medical Sciences

Exploring the Transcriptome Profile of Caviid Gammaherpesvirus 1 using Long-Read Sequencing

PhD Thesis Booklet

Ákos Dörmő

Supervisors: Dr. habil. Dóra Tombácz and prof. Dr. Zsolt Boldogkői

Szeged,
2026

Publication directly related to the subject of the thesis

1. Torma Gábor ; Dörmő Ákos*; Fülöp Ádám ; Tombácz Dóra ; Mizik Máté ; Pretory Amanda M. ; Lee See-Chi ; Toth Zsolt** ;Boldogkői Zsolt
Long-read transcriptomics of caviid gammaherpesvirus 1: compiling a comprehensive RNA atlas
MSYSTEMS 10: 3 Paper: e0167824 , 19 p. (2025)
Folyóirat szakterülete: Scopus - Ecology, Evolution, Behavior and Systematics SJR indikátor: D1

List of abbreviations

ncRNA	Non-coding RNA
EBV	Eppstein-Barr Virus
KSHV	Kaposi's Sarcoma-associated HerpesVirus
ORF	Open Reading Frame
CaGHV-1	Caviid GammaHerpesVirus-1
MHV-68	Murine gammaHerpesVirus 68
ONT	Oxford Nanopore Technologies
TSS	Transcription Start Site
TES	Transcription End Site
RTA	Repilication and Transcription Activator
raRNA	Replication origin-Associated RNA

Introduction

The viruses are considered to be the most frequently occurring[1] obligate intracellular microorganisms on Earth that do not reproduce by cell division and possess a non-cellular organization[2]. The herpesviruses are also included in this category, for which approximately one hundred thirty distinct viral species have already been identified and have been classified into three families, namely the *Orthoherpesviridae,* the *Alloherpesviridae,* and the *Malacoherpesviridae*[3–5].

The members of the *Orthoherpesviridae* family are characterized by the possession of a large, enveloped, linear double-stranded DNA genome, the genome size of which may reach approximately 125-241 kb and which contains 70-170 genes[3,5–9]. Three subfamilies are included, namely the *Alphaherpesvirinae*, the *Betaherpesvirinae* and the *Gammaherpesvirinae*[3,8,10]. Among these, nine herpesviruses are currently recognized as human pathogens. Approximately 90% of humanity is

infected with some type of herpesvirus. A lytic and a latent cycle are characteristic of all herpesviruses[8,10,11].

Viruses belonging to the *Gammaherpesvirinae* subfamily are characterized by the establishment of lifelong latent infection[12] The development of latency, as well as the investigation of the effects of non-coding RNAs (ncRNAs) on lytic and latent infection, has been regarded as a dynamically and rapidly expanding research field in which the persistence of gammaherpesviruses in the human population is examined. Included here are two oncogenic human pathogenic gammaherpesviruses: *Epstein–Barr virus* (EBV, or HHV four), belonging to the lymphocryptoviruses, and *Kaposi's sarcoma-associated herpesvirus* (KSHV, or HHV eight), belonging to the *rhadinoviruses*[13–15]. More than 200,000 new cancer cases are attributed to them worldwide; therefore, they are classified as first-class carcinogenic viruses by the WHO[16–18]. Their transmission is primarily mediated via saliva and occurs orally[17].

KSHV is characterized as an oncogenic gammaherpesvirus that has been associated with three

distinct neoplastic diseases[19]. A large double-stranded DNA genome of approximately 165 kb is possessed, comprising ~90 open reading frames (ORFs), and lifelong infection is established following infection of B lymphocytes. Similarly to other herpesviruses, a lytic and a latent life cycle are exhibited by KSHV[20–24]. Currently, no appropriate animal model against the virus is available, nor is a perfect model for human viruses available; only a single aspect of the disease can be represented. Therefore, the establishment of an inexpensive and easily manipulable model is considered important in order to facilitate the understanding of pathogenesis and to enable the development of antiviral therapies[25].

Caviid gammaherpesvirus 1 (CaGHV-1), formerly designated as guinea pig herpes-like virus (GPHLV), was isolated from strain two guinea pigs in which spontaneous acute lymphoblastic leukemia had developed[26–28] Over time, host tropism was successfully broadened; however, the emergence of *Murine gammaherpesvirus 68* (MHV-68, *Murid gammaherpesvirus* 4) resulted in its replacement, and it became established as the new model

organism for gammaherpesvirus research[26,29]. Subsequently, the genome of CaGHV-1 was elucidated and its phylogenetic affiliation was determined by Stanfield and colleagues.

Following genome sequencing and bioinformatic processing of the data, phylogenetic analyses demonstrated that the virus belongs to the genus *Rhadinovirus* within the *Gammaherpesvirinae* subfamily. After de novo genome assembly, a genome size of 103,374 bp was determined, 75 ORFs were identified, and a GC content of 35,45% was reported. Of the 75 ORF, sequence homology with the human pathogenic KSHV was exhibited by 63 ORFs. Furthermore, it was demonstrated that phylogenetically CaGHV-1 is more closely related to KSHV than MHV-68 therefore, further investigation of the molecular mechanisms of CaGHV-1 as a relevant model organism is justified[26].

RNA sequencing is currently applied routinely. A highly important role is fulfilled in transcriptomics; however, short-read sequencing technologies are not suitably applied for transcriptome assembly, splice isoform

identification, and novel gene annotation due to read-length limitations[30]. In order to overcome these biases and errors, long-read sequencing technologies, such as Oxford Nanopore Technologies (ONT), have been introduced. Sequenced reads exceeding 30 kb in length are generated[30–32]. This sequencing technology has also been applied in viral transcriptome investigations[33]. Novel transcriptomic complexities have been revealed, and transcript isoforms have been identified, including splice variants and transcriptional overlaps[34,35].

Aims

The transcriptome of *Caviid gammaherpesvirus 1* (CaGHV-1) has not previously been characterized; only its genome has been characterized; therefore, the aim of the present study was the mapping of the transcriptome, as well as the identification of viral gene regulatory regions, thereby supporting *Caviid gammaherpesvirus 1* as a potential model organism for the investigation of human gammaherpesviruses.

The objectives were as follows:

1. Native RNA and direct cDNA sequencing of the CaGHV-1 virus were to be performed on a long-read sequencing platform.
2. Transcription start sites (TSS) and transcription end sites (TES) were to be identified, and the elements regulating viral gene expression were to be determined at nucleotide-level resolution.
3. The most frequently occurring (canonical) transcripts, antisense RNAs, non-coding RNAs,

complex transcripts, and transcriptional overlaps were to be annotated.

4. The transcriptional activity of the replication and transcription activator (RTA) encoded by the ORF50 gene was to be examined, and it was to be determined whether the CaGHV-1 ORF50 promoter can be activated by RTA proteins derived from other gammaherpesviruses.

Materials and methods

Fibroblast 104C1 cells derived from guinea pig embryonic tissue were infected with Caviid gammaherpesvirus 1 (CaGHV-1), and the infection was terminated at eight different time points. Following total RNA isolation and poly(A) selection, for the native RNA sequencing library, all replicates of each time point were mixed in equal proportions and were sequenced on the PromethION P2 Solo device of Oxford Nanopore Technologies (ONT). In the case of the direct cDNA libraries, the samples were assigned unique barcodes for their separation during sequencing; therefore, a protocol modification was applied at the library preparation step, since the manufacturer did not provide a barcode step for the direct cDNA sequencing protocol. Subsequently, the replicates of the first two early time points were sequenced on the MinION platform of ONT, whereas the late time points were sequenced on the PromethION P2 Solo in order to avoid the potential barcode hopping phenomenon.

The coding sequence of the CaGHV-1 ORF50 protein was cloned into an expression vector. The N-terminally 3xFLAG-tagged CaGHV-1 RTA was transfected into HEK239T cells. Subsequently, the previously generated ORF50 promoter fragments were cloned into luciferase reporter vectors, and the reporter plasmid was transfected together with the RTA expression plasmid.

The raw reads obtained from sequencing were mapped to the reference genome using the minimap2 program. For the annotation of transcripts derived from the direct cDNA data, the LoRTIA software was used. For the quantification of transcripts obtained from direct RNA sequencing, the NAGATA software was applied, and for visualization the IGV program was used.

Results

Based on the obtained results, 162 canonical transcription start sites (TSSs) were identified. In addition, 92 potential TATA boxes, 18 putative CAAT boxes, and 5 possible GC boxes were annotated. Promoter elements were detected, the majority of which contained a TATA box sequence harboring the TATTWAA motif, which had previously been demonstrated in Kaposi's sarcoma-associated herpesvirus (KSHV). An extremely high transcriptional activity was indicated, as more than 1.5 million reads confirmed the TSS of the PAN non-coding RNA.

A total of 140 canonical transcription end sites (TESs) were identified, of which 131 were associated with polyadenylation signals (PASs). Based on the transcription start and end sites, 278 canonical transcripts were ultimately identified. In order to explore the splicing landscape of the viral transcriptome, 56 introns were identified on the basis of the direct RNA data, which were also confirmed by the direct cDNA sequencing data. Introns were also identified within the coding regions of

two genes, ORF50 and ORF57, the structures of which were conserved and corresponded to those of their homologous genes in KSHV. Furthermore, 54 monocistronic, 60 multigenic, and 108 complex transcripts were identified, indicating a high degree of transcriptional complexity, primarily due to overlapping transcripts and alternative splicing events.

.

The non-coding RNAs included intergenic transcripts (PAN), antisense transcripts, and presumably several long 5′ UTR mRNA variants. A total of 44 antisense RNAs were annotated, of which 26 contained TATA boxes, and among these, 15 harbored the TATTWAA motif characteristic of late gene promoters of KSHV. Among the non-coding RNAs, replication origin-associated RNAs (raRNAs) represented a distinct category. Several RNAs were identified whose promoter regions were located directly adjacent to OriLyt-R, and long complex RNAs were also detected in the region that completely overlapped the lytic origin. A high degree of transcriptional complexity was exhibited by the viral

genome, as numerous convergent, divergent, and tandem transcriptional overlaps characterized the genome. Similarly extensive overlaps were observed in KSHV.

During the analyses, it was found that the most abundant mRNAs of the CaGHV-1 ORF50 gene were initiated from two adjacent transcription start sites and contained four exons. The results also demonstrated that the replication and transcription activator (RTA) efficiently induced the CaGHV-1 ORF promoter in both cell lines, whereas the RTAs of other gammaherpesviruses exhibited differential promoter activation in the two cell lines. In summary, functional similarity was exhibited by RTA to its homologs in gammaherpesviruses with respect to the ability to induce its own gene promoter.

Discussion

In this study, data derived from ONT direct RNA and direct cDNA sequencing, as well as the bioinformatic software LoRTIA and NAGATA, were applied for the mapping of the transcriptome of Caviid gammaherpesvirus 1 (CaGHV-1). A large number of canonical monocistronic, multigenic, and complex transcripts were identified. Intergenic (PAN) and non-coding RNAs, RNAs containing truncated ORFs, antisense RNAs, and replication origin-associated RNAs were also included.

In numerous TSSs, the TATTWAA motif was detected, which has previously been demonstrated in other beta- and gammaherpesviruses[36]. A large proportion of TESs contained the AAUAAA consensus sequence characteristic of eukaryotic mRNAs, which has also been demonstrated in other herpesvirus subfamilies.

raRNAs were identified in both the OriLyt-L and OriLyt-R regions of CaGHV-1. Among these, transcripts directly originating from the replication origin were detected.

raRNAs overlapping the Ori region were also identified. An interesting observation was that the TATTWAA consensus sequence required for late gene transcription was present in the OriLyt-R region, indicating interactions between DNA replication and transcriptional regulation. Numerous introns were detected within the coding UTR regions of mRNAs, as well as within non-coding RNAs. A splicing pattern similar to that observed in Kaposi's sarcoma-associated herpesvirus (KSHV) and Murine gammaherpesvirus 68 (MHV-68) was observed in spliced transcripts, such as ORF50, which is responsible for the transition from latency to the lytic cycle, and ORF64[37–39]. n addition, a high degree of isoform diversity was identified in ORF73 encoding LANA, the expression of which plays one of the most important roles following infection in the establishment and maintenance of latency[40]. Numerous multigenic and complex transcripts were identified, with large portions of genes overlapping across the genome, further confirming its complexity.

The CaGHV-1 ORF50 promoter was activated by the CaGHV-1 RTA, and it was also induced by RTAs derived

from other gammaherpesviruses, indicating functional similarity in transcriptional activation. The transcriptional mechanisms characteristic of gammaherpesviruses, mediated by viral proteins as well as non-coding RNAs, are evolutionarily conserved and play key roles in the establishment of the biological characteristics and processes of gammaherpesviruses. An example is the PAN non-coding RNA, which is highly abundant during the lytic cycle of KSHV and plays an important role in the regulation of viral gene expression, as has also been demonstrated in other gammaherpesviruses, such as *Rhesus rhadinovirus* (RRV) and *Equid herpesvirus 2* (EHV-2)[41–43]. It is of particular importance that although PAN is absent from MHV-68, it is encoded by CaGHV-1, thereby providing an opportunity for the first investigation of its biological significance and pathogenesis in a potential model organism.

The elucidation of the genes and gene regulatory regions of CaGHV-1, together with its tumorigenic capacity, would provide an excellent opportunity for its application

as a prospective model organism in the investigation of the pathogenicity of human gammaherpesvirus infections.

References

1. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**, 578–585 (2021).

2. Takács, M. A vírusok kémiai összetétele és szerkezete. in *Orvosi Virológia* 3 (Medicina Könyvkiadó Zrt., 2022).

3. Dotto-Maurel, A., Arzul, I., Morga, B. & Chevignon, G. Herpesviruses: overview of systematics, genomic complexity and life cycle. *Virol J* **22**, 155 (2025).

4. Fu, M., Deng, R., Wang, J. & Wang, X. Whole-genome phylogenetic analysis of herpesviruses. *Acta Virol* **52**, 31–40 (2008).

5. Gatherer, D. *et al.* ICTV Virus Taxonomy Profile: Herpesviridae 2021. *J Gen Virol* **102**, 001673 (2021).

6. Davison, A. J. *et al.* The order Herpesvirales. *Arch Virol* **154**, 171–177 (2009).

7. Connolly, S. A., Jardetzky, T. S. & Longnecker, R. The structural basis of herpesvirus entry. *Nat Rev Microbiol* **19**, 110–121 (2021).

8. Gruffat, H., Marchione, R. & Manet, E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front. Microbiol.* **7**, (2016).

9. Morissette, G. & Flamand, L. Herpesviruses and chromosomal integration. *J Virol* **84**, 12100–12109 (2010).

10. Carneiro, V. C. de S., Pereira, J. G. & de Paula, V. S. Family Herpesviridae and neuroinfections: current status and research in progress. *Mem Inst Oswaldo Cruz* **117**, e220200 (2022).

11. Houldcroft, C. J. Human Herpesvirus Sequencing in the Genomic Era: The Growing Ranks of the Herpetic Legion. *Pathogens* **8**, 186 (2019).

12. Sorel, O. & Dewals, B. G. The Critical Role of Genome Maintenance Proteins in Immune Evasion During Gammaherpesvirus Latency. *Front. Microbiol.* **9**, 3315 (2019).

13. Weed, D. J. & Damania, B. Pathogenesis of Human Gammaherpesviruses: Recent Advances. *Curr Clin Micro Rpt* **6**, 166–174 (2019).

14. Jha, H., Banerjee, S. & Robertson, E. The Role of Gammaherpesviruses in Cancer Pathogenesis. *Pathogens* **5**, 18 (2016).

15. Blake, N. Immune evasion by gammaherpesvirus genome maintenance proteins. *Journal of General Virology* **91**, 829–846 (2010).

16. Münz, C. Human γ-Herpesvirus Infection, Tumorigenesis, and Immune Control in Mice with Reconstituted Human Immune System Components. *Front. Immunol.* **9**, 238 (2018).

17. Böni, M., Rieble, L. & Münz, C. Co-Infection of the Epstein–Barr Virus and the Kaposi Sarcoma-Associated Herpesvirus. *Viruses* **14**, 2709 (2022).

18. Guzha, B. T. *et al.* The impact of DNA tumor viruses in low-to-middle income countries (LMICS): A literature review. *Tumour Virus Res* **18**, 200289 (2024).

19. Sin, S.-H. *et al.* The complete Kaposi sarcoma-associated herpesvirus genome induces early-onset, metastatic angiosarcoma in transgenic mice. *Cell Host Microbe* **32**, 755-767.e4 (2024).

20. Wen, K. W. & Damania, B. Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis. *Cancer Lett* **289**, 140–150 (2010).

21. Yu, C. J. & Damania, B. Molecular Mechanisms of Kaposi Sarcoma-Associated Herpesvirus (HHV8)-Related Lymphomagenesis. *Cancers (Basel)* **16**, 3693 (2024).

22. Ganem, D. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. *J Clin Invest* **120**, 939–949 (2010).

23. Broussard, G. & Damania, B. Regulation of KSHV Latency and Lytic Reactivation. *Viruses* **12**, 1034 (2020).

24. Prazsák, I. *et al.* KSHV 3.0: a state-of-the-art annotation of the Kaposi's sarcoma-associated herpesvirus transcriptome using cross-platform sequencing. *mSystems* **9**, e0100723 (2024).

25. Dittmer, D. P., Damania, B. & Sin, S.-H. Animal models of tumorigenic herpesviruses--an update. *Curr Opin Virol* **14**, 145–150 (2015).

26. Stanfield, B. A., Ruiz, E., Chouljenko, V. N. & Kousoulas, K. G. Guinea pig herpes like virus is a gamma herpesvirus. *Virus Genes* **60**, 148–158 (2024).

27. Hsiung, G. D. & Kaplow, L. S. Herpeslike virus isolated from spontaneously degenerated tissue culture derived from leukemia-susceptible guinea pigs. *J Virol* **3**, 355–357 (1969).

28. Hsiung, G. D., Kaplow, L. S. & Booss, J. Herpesvirus infection of guinea pigs. I. Isolation, characterization and pathogenicity. *Am J Epidemiol* **93**, 298–307 (1971).

29. Rhim, J. S. Malignant transformation of rat embryo cells by a herpesvirus isolated from L2C guinea pig leukemia. *Virology* **82**, 100–110 (1977).

30. Oikonomopoulos, S. *et al.* Methodologies for Transcript Profiling Using Long-Read Technologies. *Front Genet* **11**, 606 (2020).

31. Wongsurawat, T., Jenjaroenpun, P., Wanchai, V. & Nookaew, I. Native RNA or cDNA Sequencing for Transcriptomic Analysis: A Case Study on Saccharomyces cerevisiae. *Front Bioeng Biotechnol* **10**, 842299 (2022).

32. Udaondo, Z. *et al.* Comparative Analysis of PacBio and Oxford Nanopore Sequencing Technologies for Transcriptomic Landscape Identification of Penaeus monodon. *Life (Basel)* **11**, 862 (2021).

33. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research. *Trends in Microbiology* **27**, 578–592 (2019).

34. Moldován, N. *et al.* Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep* **8**, 8604 (2018).

35. Shekhar, R., McMahon, S., Tibbetts, S. A., Flemington, E. K. & Renne, R. Cross-Species Insights Into Gamma Herpesvirus Transcriptomes: Long-Read and Multi-Omics Perspectives. *J Med Virol* **98**, e70802 (2026).

36. Dremel, S. E. & Didychuk, A. L. Better late than never: A unique strategy for late gene transcription in the beta- and gammaherpesviruses. *Semin Cell Dev Biol* **146**, 57–69 (2023).

37. O'Grady, T. *et al.* Genome-wide Transcript Structure Resolution Reveals Abundant Alternate Isoform Usage from Murine Gammaherpesvirus 68. *Cell Rep* **27**, 3988-4002.e5 (2019).

38. Purushothaman, P., Uppal, T. & Verma, S. C. Molecular biology of KSHV lytic reactivation. *Viruses* **7**, 116–153 (2015).

39. Shekhar, R. *et al.* High-density resolution of the Kaposi's sarcoma associated herpesvirus transcriptome identifies novel transcript isoforms generated by long-range transcription and alternative splicing. *Nucleic Acids Res* **52**, 7720–7739 (2024).

40. Allen, R. D., Dickerson, S. & Speck, S. H. Identification of spliced gammaherpesvirus 68 LANA and v-cyclin transcripts and analysis of their expression in vivo during latent infection. *J Virol* **80**, 2055–2062 (2006).

41. Tycowski, K. T., Shu, M.-D., Borah, S., Shi, M. & Steitz, J. A. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep* **2**, 26–32 (2012).

42. Borah, S., Darricarrère, N., Darnell, A., Myoung, J. & Steitz, J. A. A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog* **7**, e1002300 (2011).

43. Rossetto, C. C., Tarrant-Elorza, M., Verma, S., Purushothaman, P. & Pari, G. S. Regulation of viral and cellular gene expression by Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA. *J Virol* **87**, 5540–5553 (2013).