

¹ Turkey Hemorrhagic Enteritis Virus transcriptome profiling

²

³ **Running Title:** Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

⁴ Abraham Quaye^{1*}, Bret Pickett^{*}, Joel S. Griffitts^{*}, Bradford K. Berges^{*}, Brian D. Poole^{†*}

⁵ *Department of Microbiology and Molecular Biology, Brigham Young University

⁶ ¹First-author

⁷ [†] Corresponding Author

⁸ **Corresponding Author Information**

⁹ brian_poole@byu.edu

¹⁰ Department of Microbiology and Molecular Biology,

¹¹ 4007 Life Sciences Building (LSB),

¹² Brigham Young University,

¹³ Provo, Utah

¹⁴

15 **ABSTRACT**

16 **Background:** Hemorrhagic enteritis (HE) is a disease affecting 6-12-week-old turkeys characterized by *im-*
17 *munosuppression (IS)* and bloody diarrhea. This disease is caused by *Turkey Hemorrhagic Enteritis Virus*
18 (*THEV*) of which avirulent strains (*THEV-A*) that do not cause HE but retain the immunosuppressive ability
19 have been isolated. The *THEV-A* Virginia Avirulent Strain (VAS) is still used as a live vaccine despite its
20 immunosuppressive properties. *Our objective is to understand the genetic basis by which VAS induces*
21 *IS.* The transcriptome of *THEV* was studied to set the stage for further experimentation with specific viral
22 genes that may mediate IS.

23 **Methods:** After infecting a turkey B-cell line (MDTC-RP19) with the VAS vaccine strain, samples in tripli-
24 cates were collected at 4-, 12-, 24-, and 72-hours post-infection. Total RNA was subsequently extracted,
25 and poly-A-tailed mRNA sequencing done. After trimming the raw sequencing reads with the FastQC, reads
26 were mapped to the *THEV* genome using Hisat2 and transcripts assembled with StringTie. An in-house
27 script was used to consolidate transcripts from all time-points, generating the final transcriptome. PCR, gel
28 electrophoresis, and Sanger sequencing were used to validate all identified splice junctions.

29 **Results and Conclusions:** A total of **18.1** million reads mapped to *THEV* genome providing good cover-
30 age/depth, leaving no regions unmapped. All predicted genes in the genome were represented. In keeping
31 with all adenoviruses, all transcripts were spliced with either with 5'- or 3'-multi exon UTRs hitherto un-
32 known. *Thirteen* novel exons were identified which were validated by PCR and Sanger sequencing. The
33 splicing patterns strongly suggest that there are *three* main promoters (E1, E3, and major late promoters)
34 driving expression of most of the genes with *two* possible minor promoters driving single genes (ORF7 and
35 ORF8). This RNA-sequencing experiment is the first study of *THEV* gene expression to date. In keeping
36 with other Adenoviruses, almost all *THEV* genes are spliced, and several genes are expressed as one tran-
37 scription unit under a single promoter. This insight into *THEV*'s transcriptome may allow the engineering of
38 the VAS to provide immune protection with less or no associated IS.

39 **INTRODUCTION**

40 Adenoviruses (AdVs) are non-enveloped icosahedral-shaped DNA viruses, causing infection in virtually all
41 vertebrates. Their double-stranded linear DNA genomes range between 26 and 45kb in size, producing a
42 broad repertoire of transcripts via a highly complex alternative splicing pattern (1, 2). The AdV genome is
43 one of the most optimally economized; both the forward and reverse DNA strands harbor protein-coding
44 genes, making it highly gene-dense. There are 16 genes termed “genus-common” that are homologous in
45 all AdVs; these are thought to be inherited from a common ancestor. All other genes are termed “genus-
46 specific”. “Genus-specific” genes tend to be located at the termini of the genome while “genus-common”
47 genes are usually central (1). This pattern is observed in *Adenoviridae*, *Poxviridae*, and *Herpesviridae* (1,
48 3, 4). The family *Adenoviridae* consists of five genera: *Mastadenovirus* (MAdV), *Aviadenovirus*, *Ataden-
49 ovirus*, *Ichtadenovirus*, and *Siadenovirus* (SiAdV) (5, 6). Currently, there are three recognized members
50 of the genus SiAdV: frog adenovirus 1, raptor adenovirus 1, and turkey adenovirus 3 also called turkey
51 hemorrhagic enteritis virus (THEV) (5, 7–10). Members of SiAdV have the smallest genome size (~26 kb)
52 and gene content (~23 genes) of all known AdVs, and many “genus-specific” putative genes of unknown
53 functions have been annotated (see **Figure 1**) (1, 2, 7).

54 Virulent strains (THEV-V) and avirulent strains (THEV-A) of THEV are serologically indistinguishable, infect-
55 ing turkeys, chickens, and pheasants. They infect via the fecal-oral route, and the THEV-V cause different
56 clinical diseases in these birds (2, 11). In turkeys, the THEV-V cause hemorrhagic enteritis (HE), a debil-
57 itating acute disease affecting predominantly 6-12-week-old turkeys characterized by immunosuppression
58 (IS), weight loss, intestinal lesions leading to bloody diarrhea, splenomegaly, and up to 80% mortality (11–
59 13). HE is the most economically significant disease caused by any strain of THEV (11). While the current
60 vaccine strain (a THEV-A isolated from a pheasant, Virginia Avirulent Strain [VAS]) have proven effective
61 at preventing HE in young turkey pouls, it still retains the immunosuppressive ability. Thus, vaccinated
62 birds are rendered more susceptible to opportunistic infections and death than unvaccinated cohorts lead-
63 ing to substantial economic losses (11, 14–16). The induced IS also interferes with vaccination schemes
64 for other infections of turkeys (11, 14). To eliminate this immunosuppressive side-effect of the vaccine, a
65 thorough investigation of the culprit viral factors (genes) mediating this phenomenon is essential. However,
66 the transcriptome (splicing and gene expression patterns) of THEV has not been characterized, making
67 the investigation of specific viral genes for possible roles in causing IS impractical. A well-characterized
68 transcriptome of THEV is required to set the stage for experimentation with specific viral genes that may
69 mediate IS.

70 Myriads of studies have elucidated the AdV transcriptome in fine detail (17, 18). However, a large pre-
71 ponderance of studies focus on MAdVs - specifically human AdVs - thus, most of the current knowledge
72 regarding AdV gene expression and replication is based on MAdV studies, which is generalized for all other
73 AdVs (6, 19). MAdV genes are transcribed in a temporal manner; therefore, genes are categorized into five
74 early transcription units (E1A, E1B, E2, E3, and E4), two intermediate (IM) units (pIX and IVa2), and one
75 major late unit (MLTU), which generates five families of late mRNAs (L1-L5). An additional gene (UXP or U
76 exon) is located on the reverse strand. The early genes encode non-structural proteins such as enzymes or
77 host cell modulating proteins, primarily involved in DNA replication or providing the necessary intracellular
78 niche for optimal replication while late genes encode structural proteins. The immediate early gene E1A
79 is expressed first, followed by the the delayed early genes, E1B, E2, E3 and E4. Then the intermediate
80 early genes, IVa2 and pIX are expressed followed by the late genes (6, 17, 18). MAdV makes an extensive
81 use of alternative RNA splicing to produce a very complex array of mRNAs; all but pIX mRNA undergo
82 at least one splicing event. The MLTU produces over 20 distinct splice variants all of which contain three
83 non-coding exons at the 5'-end (collectively known as the tripartite leader, TPL) (17, 18). There is also
84 an alternate 5' three non-coding exons present in varying amounts on a subset of MLTU mRNAs (known
85 as the x-, y- and z-leaders). Lastly, there is the i-leader exon, which is infrequently included between the
86 second and third TPL exons, and codes for the i-leader protein (20). Thus, the MLTU produces a complex
87 repertoire of mRNA with diverse 5'-UTRs, spliced onto different 3' coding exons grouped into five different
88 3'-end classes (L1-L5). Each transcription unit contains its own promoter that drives the expression of all
89 the array of mRNA transcripts produced via alternative splicing of the genes encoded in the unit(6, 17, 18).

90 Almost all AdV mRNAs are generated by the excision of one or more introns and most of these introns are
91 located in the 5' or 3' UTRs of pre-mRNA. Thus the viral introns do only in a few cases interrupt the open
92 reading frames (ORFs) (1, 18). The development of high throughput sequencing methods has facilitated
93 the discovery of many novel transcribed regions and splicing isoforms. It is also a very powerful tool to study
94 alternative splicing under different conditions at an unparalleled depth (18, 21). In this paper, a paired-end
95 deep sequencing experiment was performed to characterize for the first time, the transcriptome of THEV
96 (VAS strain) during different phases of the infection, yielding a complete THEV splicing map. Our paired-
97 end sequencing allowed for reading **149** bp long high quality (mean Phred Score of 36) sequences from
98 each end of cDNA fragments, which were mapped to the genome of THEV. The generated data from our
99 paired-end sequencing experiment should thus be reliable.

100 **RESULTS**

101 **Overview of sequencing data and analysis pipeline outputs**

102 A previous study by Zeinab *et al* showed that almost all THEV transcripts were detectable beginning at
103 4 hours (22). Therefore, infected MDTC-RP19 cells were harvested at 4-, 12-, 24-, and 72-hours post-
104 infection(h.p.i) to ensure an amply wide time window to sample all transcripts. Our paired-end RNA se-
105 quencing (RNA-Seq) experiment yielded an average of **119.1** million total reads of **149**bp in length per
106 time-point. Using the HISAT2 alignment program, a total of **18.1** million reads from all time-points mapped
107 to the virus genome; this provided good coverage/depth, leaving no regions unmapped. The mapped reads
108 to the virus genome increased substantially from **432** reads at 4 h.p.i to **16.9** million reads at 72 h.p.i (**Table**
109 **1, Figure 2**). From the mapped reads, we identified an overall total of **2,867** THEV splice junctions from all
110 time-points, with splice junctions from the later time-points being supported by significantly more sequence
111 reads than earlier time-points. The substantial increases in splice junctions and mapping reads to the
112 THEV genome from earlier time-points to later time-points corresponds to the progression of the infection
113 and correlates with our qPCR assay quantifying the total number of viral genome copies over time (**Figure**
114 **3**). Using STRINGTIE, an assembler of RNA-Seq alignments into potential transcripts, the mapped reads for
115 each time-point were assembled into transcripts using the location of the predicted ORFs within the viral
116 genome. The transcripts were consolidated to generate the final transcriptome of THEV (Maybe put the
117 figure of the final transcriptome here? XXXXXXXXXXXXXXXXX).

118 The complete list of splice junctions mapped to THEV's genome has been submitted to the National Center
119 for Biotechnology Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under **acces-**
120 **sion no. XXXXXX.**

121 **MATERIALS AND METHODS**

122 **Cell culture and THEV Infection**

123 The Turkey B-cell line (MDTC-RP19, ATCC CRL-8135) was grown as suspension cultures in 1:1 complete
124 Leibovitz's L-15/McCoy's 5A medium with 10% fetal bovine serum (FBS), 20% chicken serum (ChS), 5%
125 tryptose phosphate broth (TPB), and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Strepto-
126 mycin), at 41°C in a humidified atmosphere with 5% CO₂. When infected, the cells were maintained in
127 1:1 serum-reduced Leibovitz's L15/McCoy's 5A media (SRLM) with 2.5% FBS, 5% ChS, 1.2% TPB, and
128 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Streptomycin). A commercially available HE
129 vaccine was purchased from Hygieia Biological Labs as a source of THEV-A (VAS strain). The stock virus
130 was titrated using an in-house qPCR assay with titer expressed as genome copy number(GCN)/mL, similar
131 to Mahshoub *et al*(23) with modifications. Cells were infected at a multiplicity of infection (MOI) of 100
132 GCN/cell and samples in triplicates were harvested at 4-, 12-, 24-, and 72-h.p.i for RNA-Seq. A second
133 infection was done but samples in triplicates were harvested at 12-, 24-, 36-, 48-, and 72-h.p.i for PCR
134 validation of novel splice sites.

135 **RNA extraction and Sequencing**

136 Total RNA was extracted from infected cells using Thermofishers' RNAqueous™-4PCR Total RNA Isolation
137 Kit (#AM1914) as per manufacturer's instructions. An agarose gel electrophoresis was performed to check
138 RNA integrity. The RNA quantity and purity was initially assessed using nanodrop, and RNA was used only
139 if the A260/A280 ratio was 2.0 ± 0.05 and the A260/A230 ratio was >2 and <2.2. Extracted total RNA sam-
140 ples were sent to LC Sciences, Houston TX for poly-A-tailed mRNA sequencing where RNA integrity was
141 checked with Agilent Technologies 2100 Bioanalyzer High Sensitivity DNA Chip and poly(A) RNA-
142 Seq library was prepared following Illumina's TruSeq-stranded-mRNA sample preparation protocol.
143 Paired-end sequencing was performed on Illumina's NovaSeq 6000 sequencing system.

144 **Computational Analysis of RNA Sequencing Data: Mapping and Transcript characterization**

145 Analysis of our sequence reads were analyzed following a well established protocol described by Pertea
146 *et al* (24), using SNAKEMAKE 7.24.0 to drive the pipeline. Briefly, sequencing reads were trimmed with the
147 FastQC - version 0.11.9 (25) program to achieve an overall Mean Sequence Quality (Phred Score)
148 of 36. Trimmed reads were mapped to the complete sequence of avirulent turkey hemorrhagic enteritis

149 virus strain Virginia (<https://www.ncbi.nlm.nih.gov/nuccore/AY849321.1/>) and *Meleagris gallopavo* (<https://www.ncbi.nlm.nih.gov/genome/?term=Meleagris+gallopavo>) using Hisat2 – version 2.2.1 (24) with de-
150 fault settings without relying on known splice sites. The generated BAM files from each infection time-point
151 were filtered for reads mapping to the THEV genome and fed into StringTie – version 2.2.1 (24) us-
152 ing a gff3 file from NCBI containing the predicted ORFs of THEV as a guide. A custom script was used
153 to consolidate all transcripts from all time-points without redundancy, generating the final transcriptome of
154 THEV.
155

156 **Validation of Novel Splice Junctions**

157 All splice junctions identified in this work are novel except one predicted splice site each for pTP and DBP,
158 which were corroborated in our work. However, these predicted splice junctions have not been experimen-
159 tally validated hitherto, and we identified additional novel splice junctions beside the predicted junctions,
160 giving a more complete picture of the transcripts.

161 The novel splice junctions after consolidating all transcripts with StringTie which we validated by PCR and
162 Sanger Sequencing are shown in Table####1. We designed primers that crossed a range of novel exon–
163 exon boundaries for each specific transcript in a transcription unit with their respective universal primers
164 (supplementary PCR methods). Each forward primer contained a KpnI restriction site and reverse primers,
165 an XbaI site. After first-strand cDNA synthesis with SuperScript™ III First-Strand Synthesis System (Ther-
166 moFisher SCIENTIFIC), these primers were used in a targeted PCR experiment, the PCR products were
167 analysed on Agarose gels, cloned by traditional restriction enzyme method and Sanger sequenced to vali-
168 date these splice junctions at the sequence level.

169 **3' Rapid Amplification of cDNA Ends (3'RACE)**

¹⁷⁰ **DISCUSSION/CONCLUSIONS**

¹⁷¹ **SCRIPTS AND SUPPLEMENTARY MATERIALS**

¹⁷² **DATA AVAILABILITY**

¹⁷³ **CODE AVAILABILITY**

¹⁷⁴ All the code/scripts written for analysis of the data is available on github ([linkXXXXXX](#))

¹⁷⁵ **ACKNOWLEDGMENTS**

¹⁷⁶ LC Sciences Eton Bioscience, Inc, San Diego, CA

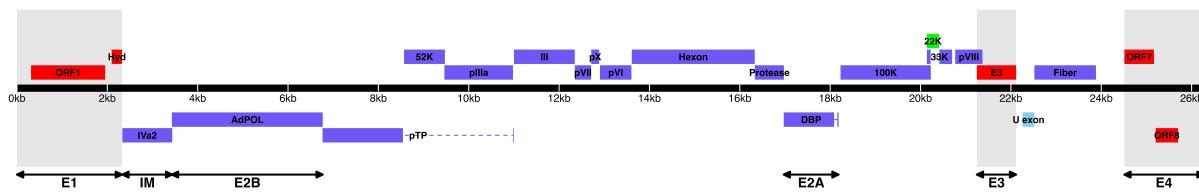
177 REFERENCES

- 178 1. Davison A, Benko M, Harrach B. 2003. Genetic content and evolution of adenoviruses. *The Journal*
179 of general virology
- 180 2. Harrach B. 2008. Adenoviruses: General features, p. 1–9. *In* Mahy, BWJ, Van Regenmortel, MHV
181 (eds.), *Encyclopedia of virology* (third edition). Book Section. Academic Press, Oxford.
- 182 3. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: Toward
183 defining the minimum essential poxvirus genome. *Journal of virology* 77:7590–7600.
- 184 4. McGeoch D, Davison AJ. 1999. Chapter 17 - the molecular evolutionary history of the herpesviruses,
185 p. 441–465. *In* Domingo, E, Webster, R, Holland, J (eds.), *Origin and evolution of viruses*. Book
Section. Academic Press, London.
- 186 5. Harrach B, Benko M, Both GW, Brown M, Davison AJ, Echavarría M, Hess M, Jones M, Kajon A,
187 Lehmkühl HD, Mautner V, Mittal S, Wadell G. 2011. Family adenoviridae. *Virus Taxonomy: 9th*
Report of the International Committee on Taxonomy of Viruses 125–141.
- 188 6. Guimet D, Hearing P. 2016. 3 - adenovirus replication, p. 59–84. *In* Curiel, DT (ed.), *Adenoviral*
189 vectors for gene therapy (second edition). Book Section. Academic Press, San Diego.
- 190 7. Kovács ER, Benkő M. 2011. Complete sequence of raptor adenovirus 1 confirms the characteristic
191 genome organization of siadenoviruses. *Infection, Genetics and Evolution* 11:1058–1065.
- 192 8. Davison AJ, Wright KM, Harrach B. 2000. DNA sequence of frog adenovirus. *J Gen Virol* 81:2431–
193 2439.
- 194 9. Kovács ER, Jánoska M, Dán Á, Harrach B, Benkő M. 2010. Recognition and partial genome char-
acterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample
195 originating from parus major, a great tit. *Journal of Virological Methods* 163:262–268.
- 196 10. Katoh H, Ohya K, Kubo M, Murata K, Yanai T, Fukushi H. 2009. A novel budgerigar-adenovirus
197 belonging to group II avian adenovirus of siadenovirus. *Virus Research* 144:294–297.
- 198 11. Beach NM. 2006. Characterization of avirulent turkey hemorrhagic enteritis virus: A study of the
199 molecular basis for variation in virulence and the occurrence of persistent infection. Thesis.

- 200 12. Beach NM, Duncan RB, Larsen CT, Meng XJ, Sriranganathan N, Pierson FW. 2009. Comparison of
12 turkey hemorrhagic enteritis virus isolates allows prediction of genetic factors affecting virulence.
201 J Gen Virol 90:1978–85.
- 202 13. Gross WB, Moore WE. 1967. Hemorrhagic enteritis of turkeys. Avian Dis 11:296–307.
- 203
- 204 14. Rautenschlein S, Sharma JM. 2000. Immunopathogenesis of haemorrhagic enteritis virus (HEV) in
205 turkeys. Dev Comp Immunol 24:237–46.
- 206 15. Larsen CT, Domermuth CH, Sponenberg DP, Gross WB. 1985. Colibacillosis of turkeys exacerbated
207 by hemorrhagic enteritis virus. Laboratory studies. Avian Dis 29:729–32.
- 208 16. Dhama K, Gowthaman V, Karthik K, Tiwari R, Sachan S, Kumar MA, Palanivelu M, Malik YS, Singh
209 RK, Munir M. 2017. Haemorrhagic enteritis of turkeys – current knowledge. Veterinary Quarterly
37:31–42.
- 210 17. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. 2020. Deep splicing plasticity
211 of the human adenovirus type 5 transcriptome drives virus evolution. Communications Biology 3:124.
- 212 18. Zhao H, Chen M, Pettersson U. 2014. A new look at adenovirus splicing. Virology 456-457:329–341.
- 213
- 214 19. Wolfrum N, Greber UF. 2013. Adenovirus signalling in entry. Cell Microbiol 15:53–62.
- 215
- 216 20. Falvey E, Ziff E. 1983. Sequence arrangement and protein coding capacity of the adenovirus type 2
217 "i" leader. Journal of Virology 45:185–191.

- 218 21. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W,
Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. 2012. Landscape of transcription in human cells. *Nature* 489:101–108.
- 219
- 220 22. Aboeza Z, Mabsoub H, El-Bagoury G, Pierson F. 2019. In vitro growth kinetics and gene expression
221 analysis of the turkey adenovirus 3, a siadenovirus. *Virus Research* 263:47–54.
- 222 23. Mabsoub HM, Evans NP, Beach NM, Yuan L, Zimmerman K, Pierson FW. 2017. Real-time PCR-based infectivity assay for the titration of turkey hemorrhagic enteritis virus, an adenovirus, in live
223 vaccines. *Journal of Virological Methods* 239:42–49.
- 224 24. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of
225 RNA-seq experiments with HISAT, StringTie and ballgown. *Nature Protocols* 11:1650–1667.
- 226 25. 2015. FastQC.
- 227

228 **TABLES AND FIGURES**



229

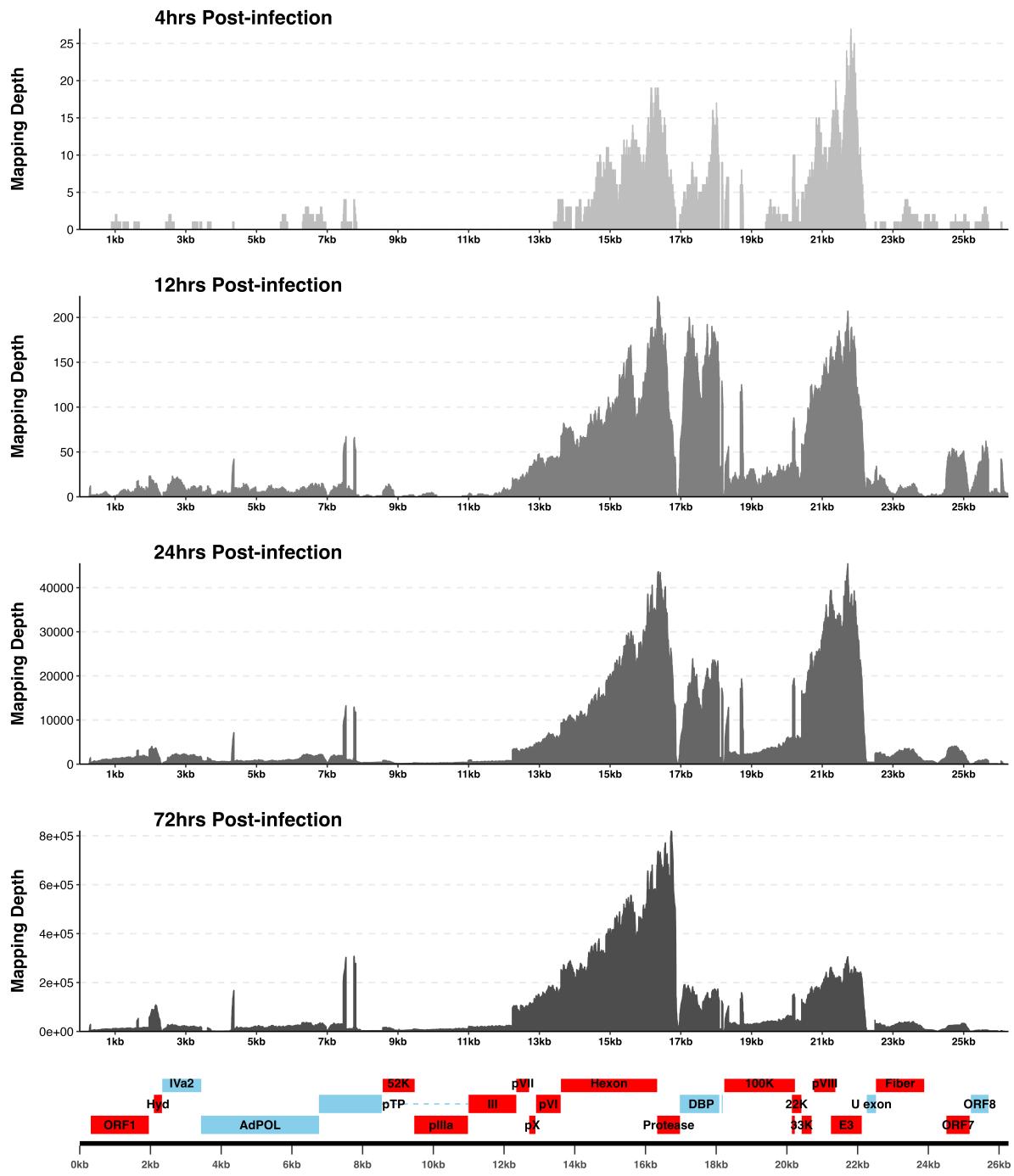
230 **Figure 1. Genomic map of THEV avirulent strain.** The central horizontal line represents the double-
 231 stranded DNA marked at 5kb intervals as white line breaks. Blocks represent viral genes. Blocks above
 232 the DNA line are transcribed rightward, those below are transcribed leftward. pTP, DBP and 33K predicted
 233 to be spliced are shown as having tails. Shaded regions indicate regions containing “genus-specific” genes
 234 (colored red). Genes colored in blue are “genus-common”. Gene colored in light green is conserved in
 235 all but Adenoviruses. The UXP (light blue) is an incomplete gene present in almost all AdVs. Regions
 236 comprising the different transcription units are labelled at the bottom (E1, E2A, E2B, E3, E4, and IM); the
 237 unlabeled regions comprise the MLTU.

Table 1
Summary of sequence alignment metrics

Metric	4h.p.i	12h.p.i	24h.p.i	72h.p.i
Total reads	1.34e+08	8.26e+07	1.32e+08	1.28e+08
Mapped (Host)	1.04e+08 (77.796%)	6.79e+07 (82.2476%)	1.06e+08 (80.319%)	8.38e+07 (65.54%)
Mapped (THEV)	4.32e+02 (0.0003%)	6.70e+03 (0.0081%)	1.18e+06 (0.8954%)	1.69e+07 (13.61%)

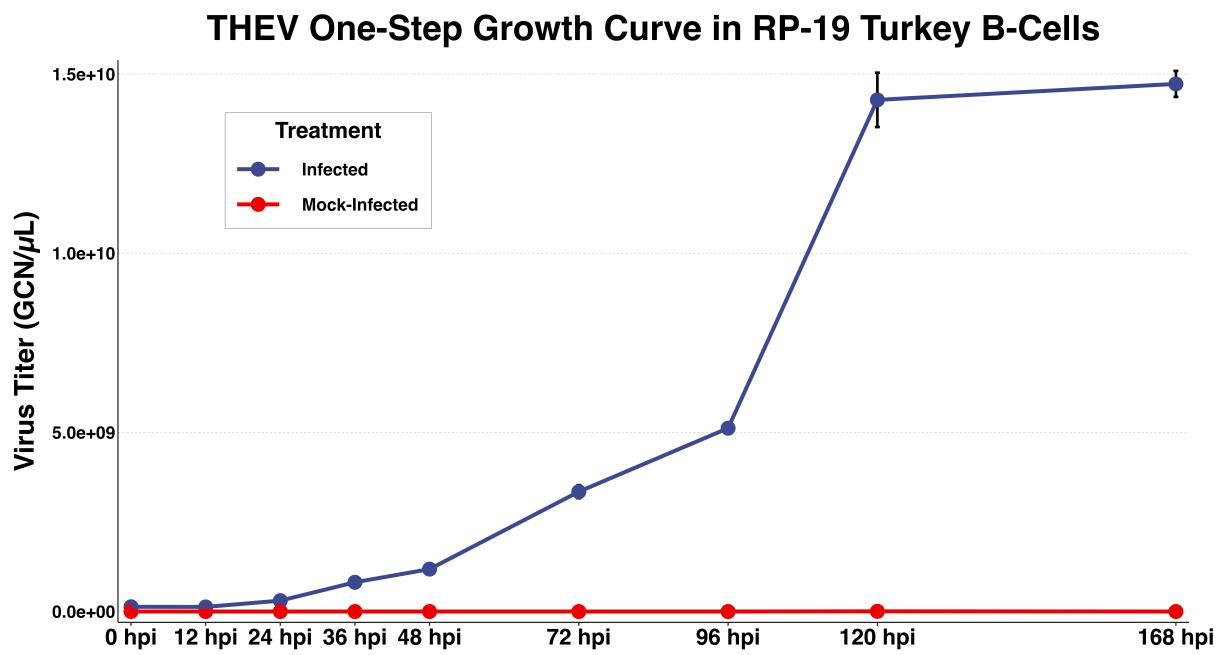
Splice junctions	13	42	247	2,565
Junction coverage >= 1 read	37	605	115,152	2,132,905
Junction coverage >= 10 reads	0	13	132	1,792
Junction coverage >= 100 reads	0	1	54	805
Junction coverage >= 1000 reads	0	0	18	168

RNA-seq Mapping Depth of THEV Genome



238

Figure 2. Sequence reads mapping to THEV genome by time-point. Description #####



240

241 **Figure 3. One-step growth of THEV.** Description###. GCN: genome copy number.