Elucidating the Transcriptome of Turkey Hemorrhagic Enteritis Virus

Running Title: Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

Abraham Quaye[1*], Bret Pickett[*], Joel S. Griffitts[*], Bradford K. Berges[*], Brian D. Poole[†*]

[*]Department of Microbiology and Molecular Biology, Brigham Young University

[1]First-author

[†] Corresponding Author

Corresponding Author Information

brian_poole@byu.edu

Department of Microbiology and Molecular Biology,

4007 Life Sciences Building (LSB),

Brigham Young University,

Provo, Utah

ABSTRACT

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

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

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

INTRODUCTION

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

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

Myriads of studies have elucidated the AdV transcriptome in fine detail (17, 18). However, a large pre-

ponderance of studies focus on MAdVs - specifically human AdVs - thus, most of the current knowledge regarding AdV gene expression and replication is based on MAdV studies, which is generalized for all other AdVs (6, 19). MAdV genes are transcribed in a temporal manner; therefore, genes are categorized into five early transcription units (E1A, E1B, E2, E3, and E4), two intermediate (IM) units (pIX and IVa2), and one major late unit (MLTU), which generates five families of late mRNAs (L1-L5). An additional gene (UXP or U exon) is located on the reverse strand. The early genes encode non-structural proteins such as enzymes or host cell modulating proteins, primarily involved in DNA replication or providing the necessary intracellular niche for optimal replication while late genes encode structural proteins. The immediate early gene E1A is expressed first, followed by the the delayed early genes, E1B, E2, E3 and E4. Then the intermediate early genes, IVa2 and pIX are expressed followed by the late genes (6, 17, 18). MAdV makes an extensive use of alternative RNA splicing to produce a very complex array of mRNAs; all but pIX mRNA undergo at least one splicing event. The MLTU produces over 20 distinct splice variants all of which contain three non-coding exons at the 5'-end (collectively known as the tripartite leader, TPL) (17, 18). There is also an alternate 5' three non-coding exons present in varying amounts on a subset of MLTU mRNAs (known as the x-, y- and z-leaders). Lastly, there is the i-leader exon, which is infrequently included between the second and third TPL exons, and codes for the i-leader protein (20). Thus, the MLTU produces a complex repertoire of mRNA with diverse 5'-UTRs spliced onto different 3' coding exons which are grouped into five different 3'-end classes (L1-L5). Each transcription unit (TU) contains its own promoter driving the expression of all the array of mRNA transcripts produced via alternative splicing of the genes encoded in the unit(6, 17, 18). Almost all AdV mRNAs are generated by the excision of one or more introns and most of these introns are located in the 5' or 3' UTRs of pre-mRNA. Thus the viral introns scarcely interrupt the open reading frames (ORFs) (1, 18).

High throughput sequencing methods have facilitated the discovery of many novel transcribed regions and splicing isoforms. It is also a very powerful tool to study alternative splicing under different conditions at an unparalleled depth (18, 21). In this paper, a paired-end deep sequencing experiment was performed to characterize for the first time, the transcriptome of THEV (VAS vaccine strain) during different phases of the infection, yielding the first THEV splicing map. Our paired-end sequencing allowed for reading **149** bp long high quality (mean Phred Score of 36) sequences from each end of cDNA fragments, which were mapped to the genome of THEV. The generated data from our paired-end sequencing experiment should thus be reliable.

RESULTS

Overview of sequencing data and analysis pipeline outputs

A previous study by Zeinab *et al* showed that almost all THEV transcripts were detectable beginning at 4 hours (22). Therefore, infected MDTC-RP19 cells were harvested at 4-, 12-, 24-, and 72-hours post-infection(h.p.i) to ensure an amply wide time window to sample all transcripts. Our paired-end RNA sequencing (RNA-seq) experiment yielded an average of **107.1** million total reads of **149**bp in length per time-point, which were simultaneously mapped to both the virus (THEV) and host (*M.gallopavo*) genomes using the `Hisat2` (23) alignment program. A total of **18.1** million reads from all time-points mapped to the virus genome; this provided good coverage/depth, leaving no regions unmapped. The mapped reads 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 1**, **Figure 2a**). From the mapped reads, we identified an overall total of **2,457** unique THEV splice junctions from all time-points, with splice junctions from the later time-points being supported by significantly more sequence reads than earlier time-points. For example all the **13** unique junctions at 4 h.p.i had less than 10 reads supporting each one, averaging a mere **2.8** reads/junction. Conversely, the **2374** unique junctions at 72 h.p.i averaged **898.4** reads/junction, some junctions having coverage as high as **322,677** reads. The substantial increases in splice junctions and mapping reads to the THEV genome over time denotes an active infection, and correlates with our qPCR assay quantifying the total number of viral genome copies over time (**Figure 2b**). Using `StringTie` (23), an assembler of RNA-seq alignments into potential transcripts, the mapped reads for each time-point were assembled into transcripts using the genomic location of the predicted THEV ORFs as a guide. In the final consolidated transcriptome of THEV, a composite of all unredundant transcripts from all time points, we counted a total of **28** transcripts all of which are novel. Although some exons in some transcripts match the predicted ORFs exactly, most of our identified exons are longer, spanning multiple predicted ORFs (**Figure 3**). The complete list of unique splice junctions mapped to THEV's genome has been submitted to the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under **accession no. XXXXXX**.

Changes in THEV splicing profile over time

AdV gene expression occurs under exquisite temporal control, supervised by designated promoters for each transcription unit (TU). Each promoter typically produce one or few pre-mRNAs that undergo alternative splicing to yield the manifold repertoire of complex transcripts characteristic of AdVs. To evaluate the activity of each promoter over time, Firstly, `Ballgown` - a program for statistical analysis of assembled transcriptomes, was used to estimate and normalize expression levels of all transcripts in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) units over time. Very few unique splice junctions,

reads, and transcripts were counted at 4 h.p.i; hence, this time point was excluded in this analysis. This analysis revealed DBP - from the E2 region - to be the most significantly expressed at 12 h.p.i, whereas MLP region transcripts were most abundantly expressed at 24 and 72 h.p.i as expected (**Figure 4a**). Also, our analysis of the FPKM values of transcripts by region showed a similar pattern: the E2 region as the most significantly overexpressed at 12 h.p.i, after which the MLP region assumes predominance (**Figure 4b**).

Secondly, we estimated relative abundances of all splice junctions for each time point, counting as significantly expressed only those with coverage of at least 1% of the total splice junction reads counted at the given time point. At 12 h.p.i, **20** junctions meet the 1% threshold, and were predominantly early region (E1, E2, and E3) junctions, albeit a significant proportion are MLP-derived (**Table 2a**). Junctions from the E2 region were the most abundant at this time point, constituting **26.6**% of all junctions. The two most abundant junctions at 12 h.p.i were maintained at 24 h.p.i also as the most significantly expressed. However, the MLP-derived junctions were unsurprisingly the most preponderant overall at this time point, accounting for **45.7**% of all the junction reads counted (**Table 2b**). At 72 h.p.i, the trend of increased activity of the MLP continued as expected; at this time, the MLP-derived junctions were not only the most abundant overall - accounting for **67.4**% of all junctions reads - but also contained the most significantly expressed individual junctions (**Table 2c**. Also see **Supplementary Tables 1a-c**; **Figure 4c**). When we limited this analysis to only junctions in the final transcriptome, the relative abundances of the junctions for each region over time was substantially similar to the pattern seen with all the junctions included (**Figure 4d**).

Furthermore, we analyzed splice donor and acceptor site nucleotide usage over time to investigate any peculiarities that THEV may show, generally or over the course of the infection. We found that most splice donor-acceptor sequences were unsurprisingly the canonical GU-AG nucleotides.

Early Region 1 (E1) transcripts. This region in MAdVs is the first transcribed after successful entry of the viral DNA into the host cell nucleus, albeit at low levels (18). The host transcription machinery solely mediates the transcription of this region. After their translation, the E1 proteins in concert with a myriad of host transcription factors activate the other viral promoters (6). Only two ORFs (ORF1 [sialidase] and Hyd) are predicted in this region; however, we discovered **four** novel transcripts in this region, which collectively contain **3** unique splice junctions (**Figure 5**). Most of the encoded proteins of the novel transcripts are distinct from the predicted ORFs, although they all have the potential to encode the Hyd protein as the 3'-most coding sequence (CDS) if the first start codon (SSC) is skipped. The 5'-most CDS of TRXPT_1 is multi-exonic, producing a 17.9 kilodalton (kDa) protein of 160 residues [amino acids (aa)]. The CDS begins in the first exon, starting at position 211, spans the second exon, and terminates in the third exon

at position 2312. From the 5'-most SSC, TRXPT_2 encodes the largest protein in this region —— a 64.3 kDa, 580 aa protein with the same SSC as TRXPT_1 (position 211). This CDS spans almost the entire predicted ORF1 and Hyd, coming short in two regards: it is spliced from 1655 to 1964 (ORF1's C-terminus, including the stop codon), and it's stop codon (STC; position 2312) is 13 bp short of the Hyd STC. However, it has an SSC 102 bp upstream and in-frame with ORF1's predicted SSC. The N-terminus of TRXPT_2 CDS therefore, shares substantial protein sequence similarity with ORF1 but Hyd and TRXPT_2 are not in-frame; hence no protein sequence similarity. TRXPT_3 is almost identical to TRXPT_1, except for the lack of TRXPT_1's second exon. From our analysis, TRXPT_3 and TRXPT_4 seem to have transcription start sites (TSS) downstream of the TSS of TRXPT_1 and TRXPT_2; however, given that studies in MAdVs show that E1 mRNAs share a common 5' and 3' positions, only differing from each other regarding the introns (18), it is likely that TRXPT_3 and TRXPT_4 are truncated, and the TSS just like the transcription termination site (TTS) are identical for all E1 transcripts. Regardless of the TSS considered for TRXPT_3, the coding potential remains unaffected. Its 5'-most CDS, beginning at 1965 and sharing the same STC as TRXPT_1 and TRXPT_2, produces a 13.1 kDa, 115 residue protein. This CDS (ORF4) was predicted in an earlier study (24) but was excluded in later studies (1, 12), but our data suggests it is a bona fide ORF. The coding potential of TRXPT_4 is affected by the TSS considered; if we consider its unmodified TSS, then its coding potential is the same as TRXPT_3 (ORF4 as the first CDS and Hyd as second CDS if the first SSC is skipped). However, if we assume that TRXPT_4 shares the same TSS as TRXPT_1, then the 5'-most CDS is a distinct multi-exonic 15.9 kDa, 143 aa protein with the same SSC as TRXPT_1 and TRXPT_2 but with a unique STC. All splice junctions of the transcripts in this region (except the junction for TRXPT_4) have been validated by cloning and Sanger sequencing of cDNA (**Figure 5b**; **supplementary PCR methods**). Finally, during our validation of TRXPT_2, ORF1 was present on the agarose gel and Sanger sequencing results as a bona fide transcript (**supplementary PCR methods**). This is corroborated by 3' Rapid Amplification of cDNA Ends (3'RACE) experiment, which shows transcripts spanning the entire ORF1 and Hyd ORFs without any splicing. The 5'-most CDS of this transcript would encode ORF1. Given that the SSC of the predicted ORF1 is in-frame but downstream of TRXPT_2 SSC, it suggests that the predicted ORF1 CDS is truncated; it shares the same SSC as TRXPT_2, but has a unique STC, albeit it has the same TTS.

Early Region 2 (E2) and Intermediate Region (IM) transcripts. The E2 TU expressed on the reverse strand, is subdivided into E2A and E2B and encodes three classical AdV proteins: pTP and Ad-pol (E2B proteins), and DBP (E2A protein) [####]. Unlike MAdV where two promoters (E2-early and E2-late) were discovered [###], we discovered only a single promoter from which both E2A and E2B transcription is

initiated. However, similar to MAdVs, E2A and E2B transcripts have distinct TTSs; the E2B transcripts sharing the TTS of the IVa2 transcript of the IM region [####] (**Figure 6**). ~~Analysis of the coding potential of the transcripts; include the variant of DBP from 3'RACE data~~.

The IM region is a single-transcript TU, encoding a single classical protein, IVa2. The promoter expressing this single transcript (TRXPT_5) is embedded in E2B region and shares a TTS with E2B transcripts (17, 18). TRXPT_5 is a two-exon transcript located on the reverse strand spliced at 3447-3615. The first exon is an untranslated region (UTR), except the last 2 nucleotides, which connect with the first nucleotide of second exon to form the 5'-most SSC. This first SSC is 4 codons upstream and in-frame of the predicted IVa2 SSC. Regardless of the SSC considered, the encoded protein (IVa2) is largely unaffected. Except for the four extra residues at the N-terminus (considering the 5'-most SSC), the entire protein sequence is identical.

The splice junction of ~~TRXPT_5 add trxpts from E2~~ were confirmed by cloning and Sanger sequencing of cDNA (**supplementary PCR methods**).

Early Region 3 (E3) transcripts.

Early Region 4 (E4) transcripts. This transcription unit (TU) is the found at the tail-end (3'-end) of the genome, on the reverse strand. Based on nucleotide position, ORF7 and ORF8 were predicted in this region (1); however, as ORF7 is neither on the same strand as ORF8 nor transcribed from a promoter in the E4 region, only ORF8 can legitimately be classified as a transcript in this TU. This is corroborated by our RNA-seq data, as only one transcript was identified in this region on the reverse strand (**Figure 8**). The transcript (TRXPT_28) spans 25192-26247 and is spliced at 25701-26055; hence, a two-exon transcript. The second exon fully matches the predicted ORF8 with 12 extra base pairs at the 3'-end; however, the encoded protein is an exact match. There is a SSC in the first exon at position 26246 (second nucleotide of the transcript). The encoded protein from this SSC is in-frame with the SSC of ORF8 in the second exon; hence, the C-terminus of this longer protein (26.4 kDa, 229 aa) would be the identical to the predicted ORF8 protein. The splice junction of TRXPT_28 was validated by cloning and Sanger sequencing of cDNA (**supplementary PCR methods**).

Major Late Promoter Region (MLP) transcripts.

DISCUSSION/CONCLUSIONS

In the original study where the ORFs of THEV were predicted, ORF4 was predicted in the E1 region spanning the Hyd gene. However, later studies predicted and preferred Hyd instead of ORF4; hence, the current prediction map. However, this study shows that while both Hyd and ORF4 may be both expressed, ORF4 is most likely the bona fide gene. For fig2a: There is a dramatic increase of mean coverage/depth from **2.42** at 4 h.p.i to **95,042** at 72 h.p.i, strongly demonstrating an active infection. Unexpectedly, the pileup of reads seems consistently skewed over similar regions of the genome. We could speculate that the temporal gene expression regulation of THEV is different from MAdVs or this could simply mean that the infection was not well synchronized. However, the relative proportions over these similar regions shows some variation over time. For fig2b: titer reaching a plateau at 120 h.p.i, probably due to high cell death

MATERIALS AND METHODS

Cell culture and THEV Infection

The Turkey B-cell line (MDTC-RP19, ATCC CRL-8135) was grown as suspension cultures in 1:1 complete Leibovitz's L-15/McCoy's 5A medium with 10% fetal bovine serum (FBS), 20% chicken serum (ChS), 5% tryptose phosphate broth (TPB), and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Streptomycin), at 41°C in a humidified atmosphere with 5% CO_2. Infected cells were maintained in 1:1 serum-reduced Leibovitz's L15/McCoy's 5A media (SRLM) with 2.5% FBS, 5% ChS, 1.2% TPB, and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Streptomycin). A commercially available HE vaccine was purchased from Hygieia Biological Labs as a source of THEV-A (VAS strain). The stock virus was titrated using an in-house qPCR assay with titer expressed as genome copy number(GCN)/mL, similar to Mahshoub *et al*(25) with modifications. Cells were infected at a multiplicity of infection (MOI) of 100 GCN/cell and samples in triplicates were harvested at 4-, 12-, 24-, and 72-h.p.i for RNA-seq. The infection was repeated but samples in triplicates were harvested at 12-, 24-, 36-, 48-, and 72-h.p.i for PCR validation of novel splice sites.

RNA extraction and Sequencing

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

Validation of Novel Splice Junctions

All splice junctions identified in this work are novel except one predicted splice site each for pTP and DBP, which were corroborated in our work. However, these predicted splice junctions have not been experimentally validated hitherto, and we identified additional novel exons, giving a more complete picture of the transcripts. ~~The novel splice junctions after consolidating all transcripts with `StringTie` which we validated~~

by PCR and Sanger Sequencing are shown in Table###1. We designed primers that crossed a range of novel exon–exon boundaries for each specific transcript in a transcription unit paired with their respective universal primers (supplementary, PCR methods). Each forward primer contained a KpnI restriction site and reverse primers, an XbaI site. After first-strand cDNA synthesis with SuperScript™ III First-Strand Synthesis System (ThermoFisher SCIENTIFIC), these primers were used in a targeted PCR experiment, the PCR products were analysed on Agarose gels, cloned by traditional restriction enzyme method and Sanger sequenced to validate these splice junctions at the sequence level.

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

Computational Analysis of RNA Sequencing Data: Mapping and Transcript characterization

Analysis of our sequence reads were analyzed following a well established protocol described by Pertea *et al* (23), using `SNAKEMAKE 7.24.0` to drive the pipeline. Briefly, sequencing reads were trimmed with the `FastQC - version 0.11.9` (26) program to achieve an overall `Mean Sequence Quality (Phred Score)` of 36. Trimmed reads were mapped to the complete sequence of avirulent turkey hemorrhagic enteritis 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` (23) with default settings without relying on known splice sites. The generated `BAM` files from each infection time-point were filtered for reads mapping to the THEV genome and fed into `StringTie - version 2.2.1` (23) using a `gff3` file from NCBI containing the predicted ORFs of THEV as a guide. A custom script was used to consolidate all transcripts from all time-points without redundancy, generating the final transcriptome of THEV.

TRXPT_2 and ORF1 are isoforms

SCRIPTS AND SUPPLEMENTARY MATERIALS

DATA AVAILABILITY

CODE AVAILABILITY

All the code/scripts written for analysis of the data is available on github (linkXXXXX)

ACKNOWLEDGMENTS

LC Sciences - RNA sequencing was done here Eton Bioscience, Inc, San Diego, CA - All Sanger sequencing validations was done here

REFERENCES

1. Davison A, Benko M, Harrach B. 2003. Genetic content and evolution of adenoviruses. The Journal of general virology 84:2895–908.

2. Harrach B. 2008. Adenoviruses: General features, p. 1–9. *In* Mahy, BWJ, Van Regenmortel, MHV (eds.), Encyclopedia of virology (third edition). Book Section. Academic Press, Oxford.

3. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: Toward defining the minimum essential poxvirus genome. Journal of virology 77:7590–7600.

4. McGeoch D, Davison AJ. 1999. Chapter 17 - the molecular evolutionary history of the herpesviruses, p. 441–465. *In* Domingo, E, Webster, R, Holland, J (eds.), Origin and evolution of viruses. Book Section. Academic Press, London.

5. Harrach B, Benko M, Both GW, Brown M, Davison AJ, Echavarría M, Hess M, Jones M, Kajon A, Lehmkuhl HD, Mautner V, Mittal S, Wadell G. 2011. Family adenoviridae. Virus Taxonomy: 9th Report of the International Committee on Taxonomy of Viruses 125–141.

6. Guimet D, Hearing P. 2016. 3 - adenovirus replication, p. 59–84. *In* Curiel, DT (ed.), Adenoviral vectors for gene therapy (second edition). Book Section. Academic Press, San Diego.

7. Kovács ER, Benkő M. 2011. Complete sequence of raptor adenovirus 1 confirms the characteristic genome organization of siadenoviruses. Infection, Genetics and Evolution 11:1058–1065.

8. Davison AJ, Wright KM, Harrach B. 2000. DNA sequence of frog adenovirus. J Gen Virol 81:2431–2439.

9. Kovács ER, Jánoska M, Dán Á, Harrach B, Benkő M. 2010. Recognition and partial genome characterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample originating from parus major, a great tit. Journal of Virological Methods 163:262–268.

10. Katoh H, Ohya K, Kubo M, Murata K, Yanai T, Fukushi H. 2009. A novel budgerigar-adenovirus belonging to group II avian adenovirus of siadenovirus. Virus Research 144:294–297.

11. Beach NM. 2006. Characterization of avirulent turkey hemorrhagic enteritis virus: A study of the molecular basis for variation in virulence and the occurrence of persistent infection. Thesis.

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. J Gen Virol 90:1978–85.

13. Gross WB, Moore WE. 1967. Hemorrhagic enteritis of turkeys. Avian Dis 11:296–307.

14. Rautenschlein S, Sharma JM. 2000. Immunopathogenesis of haemorrhagic enteritis virus (HEV) in turkeys. Dev Comp Immunol 24:237–46.

15. Larsen CT, Domermuth CH, Sponenberg DP, Gross WB. 1985. Colibacillosis of turkeys exacerbated by hemorrhagic enteritis virus. Laboratory studies. Avian Dis 29:729–32.

16. Dhama K, Gowthaman V, Karthik K, Tiwari R, Sachan S, Kumar MA, Palanivelu M, Malik YS, Singh RK, Munir M. 2017. Haemorrhagic enteritis of turkeys – current knowledge. Veterinary Quarterly 37:31–42.

17. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. 2020. Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. Communications Biology 3:124.

18. Zhao H, Chen M, Pettersson U. 2014. A new look at adenovirus splicing. Virology 456-457:329–341.

19. Wolfrum N, Greber UF. 2013. Adenovirus signalling in entry. Cell Microbiol 15:53–62.

20. Falvey E, Ziff E. 1983. Sequence arrangement and protein coding capacity of the adenovirus type 2 "i" leader. Journal of Virology 45:185–191.

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.

22. Aboezz Z, Mahsoub H, El-Bagoury G, Pierson F. 2019. In vitro growth kinetics and gene expression analysis of the turkey adenovirus 3, a siadenovirus. Virus Research 263:47–54.

23. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. Nature Protocols 11:1650–1667.

24. Pitcovski J, Mualem M, Rei-Koren Z, Krispel S, Shmueli E, Peretz Y, Gutter B, Gallili GE, Michael A, Goldberg D. 1998. The complete DNA sequence and genome organization of the avian adenovirus, hemorrhagic enteritis virus. Virology 249:307–315.

25. Mahsoub 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 vaccines. Journal of Virological Methods 239:42–49.

26. 2015. FastQC.

TABLES AND FIGURES

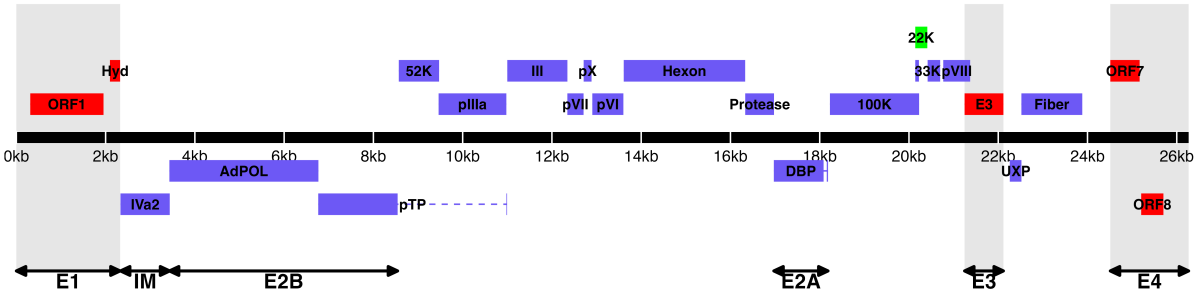


Figure 1. ***Genomic map of THEV avirulent strain***. The central horizontal line represents the double-stranded DNA marked at 5kb intervals as white line breaks. Blocks represent viral genes. Blocks above the DNA line are transcribed rightward, those below are transcribed leftward. pTP, DBP and 33K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

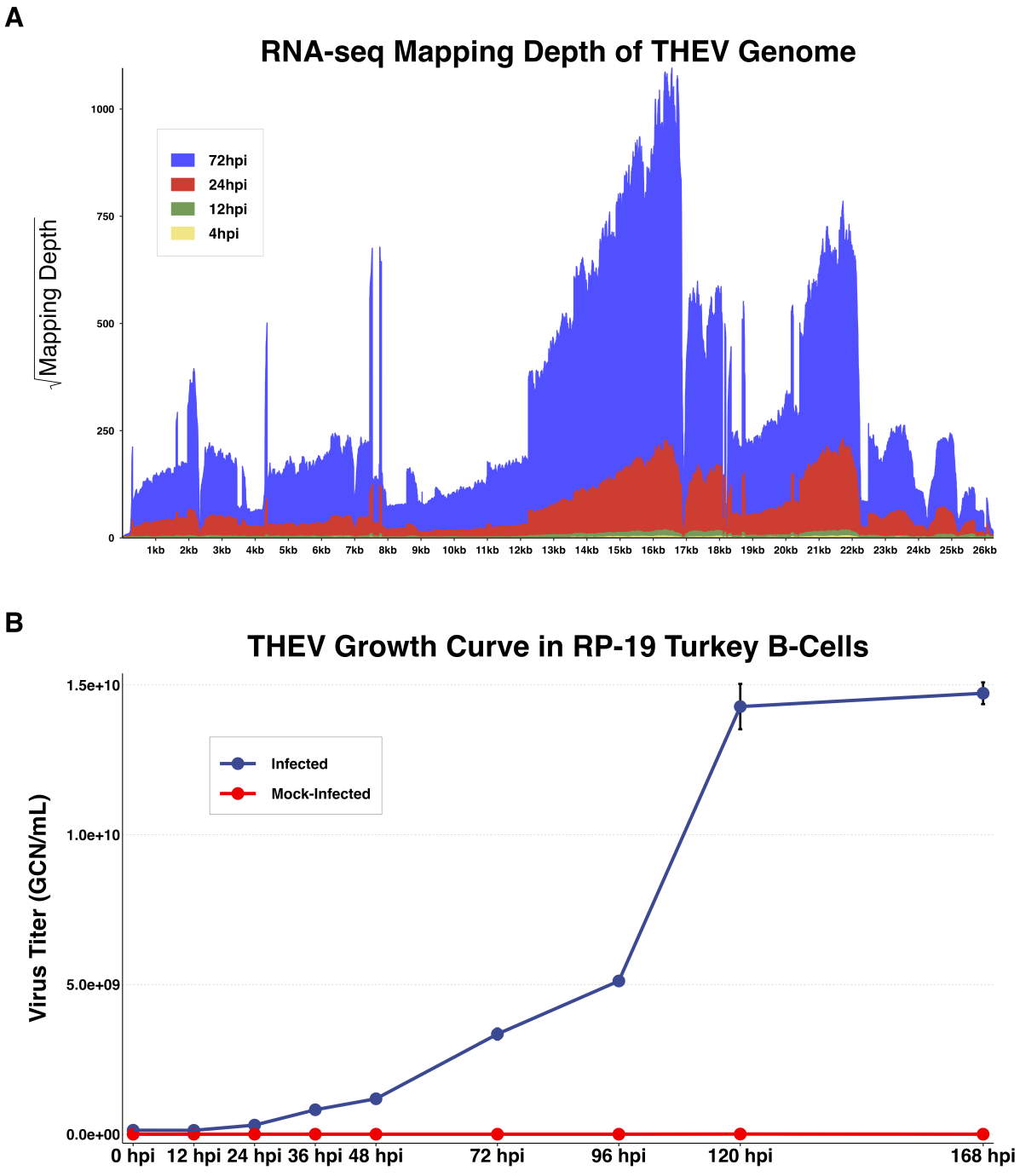
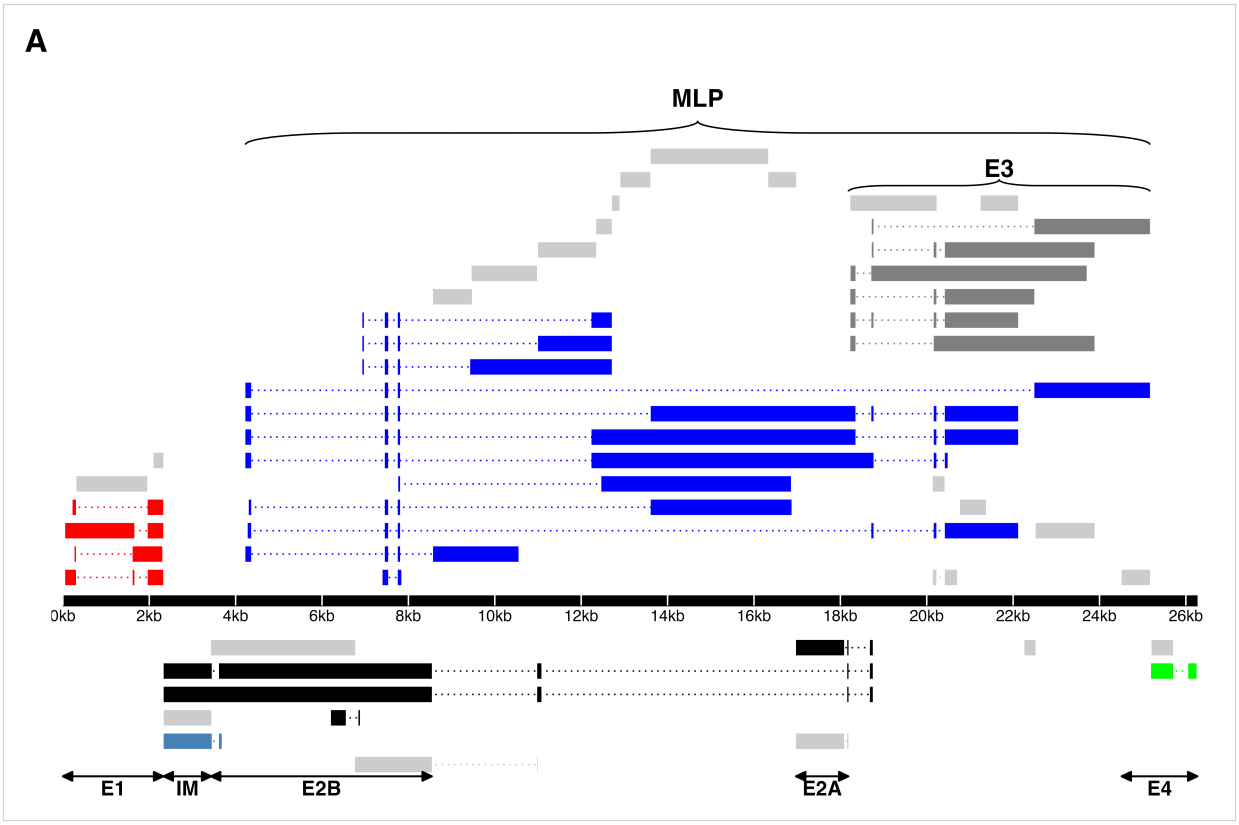


Figure 2: Increasing levels of THEV over time. a) ***Per base coverage of sequence reads mapping to THEV genome by time point***. The pileup of mRNA reads mapping to THEV genome at the base-pair level for each indicated time point. **b)** ***One-step growth of THEV (VAS vaccine strain) in MDTC-RP19 cell line***. After infecting cells at an MOI of 100 GCN/cell, triplicates of harvested infected cells were quantified with an in-house qPCR assay measuring the total copies of THEV genome. There is no discernible

increase in virus titer up 12 h.p.i, after which there is a steady increase in virus titer is measured. The virus titer expands exponentially beginning from 48 h.p.i, increasing by orders of magnitude before reaching a plateau at 120 h.p.i. GCN: genome copy number.


A
MLP
E3
0kb
2kb
4kb
6kb
8kb
10kb
12kb
14kb
16kb
18kb
20kb
22kb
24kb
26kb
E1
IM
E2B
E2A
E4

362

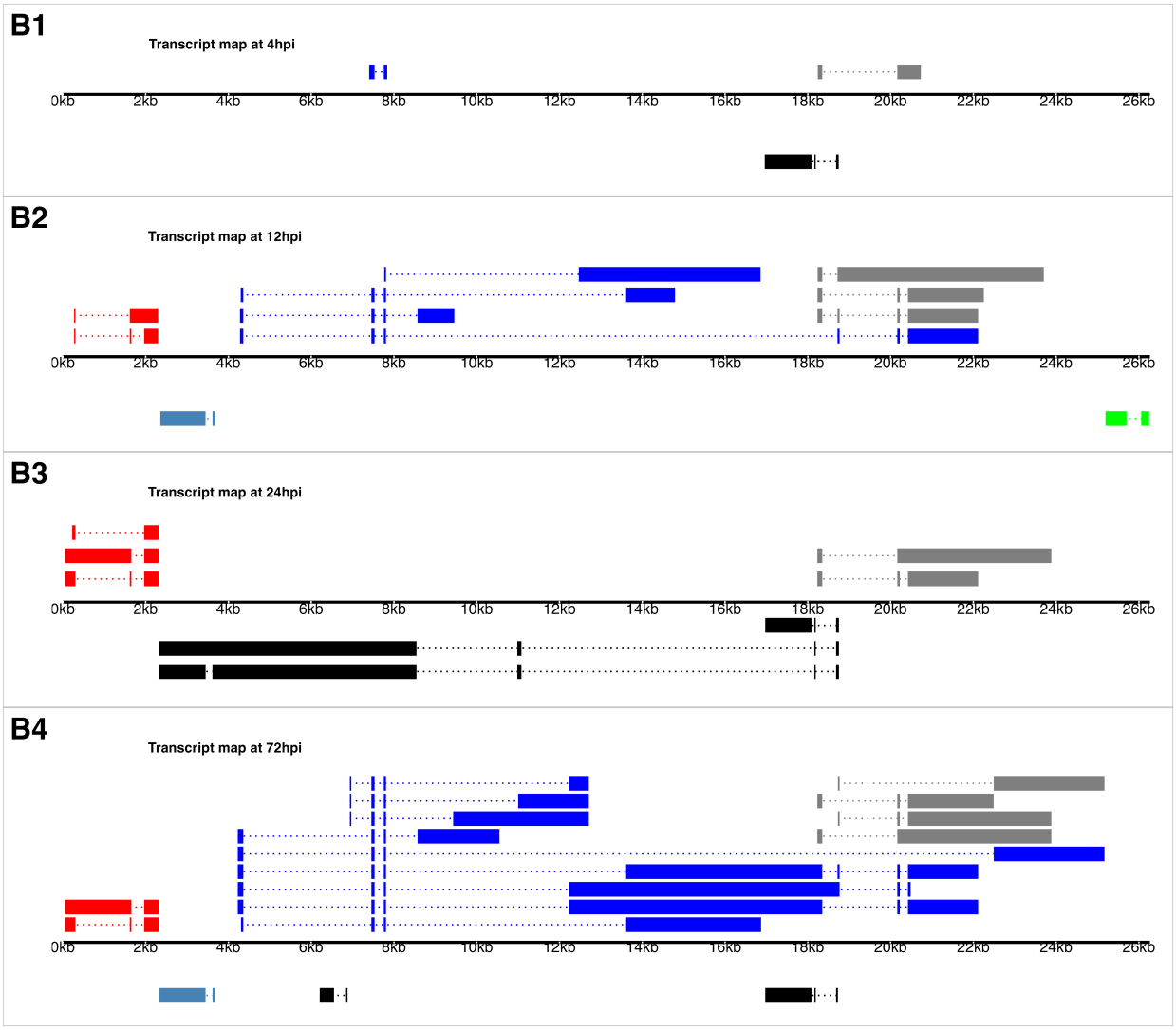


Figure 3. a) Full transcriptome of THEV. THEV transcripts assembled from all time points by `StringTie` are unified forming this final transcriptome (splicing map). Transcripts belonging to the same transcription unit (TU) are located in close proximity on the genome and are color coded and labeled in this figure as such. The organization of TUs in the THEV genome is unsurprisingly similar to MAdVs; however, the MAdV genome shows significantly more transcripts. The TUs are color coded: E1 transcripts - red, E2 - black, E3 - dark grey, E4 - green, MLP - blue. Predicted ORFs are also indicated here, colored light grey. **b) THEV transcripts identified at given time points**. Transcripts are color coded as explained in **a**.

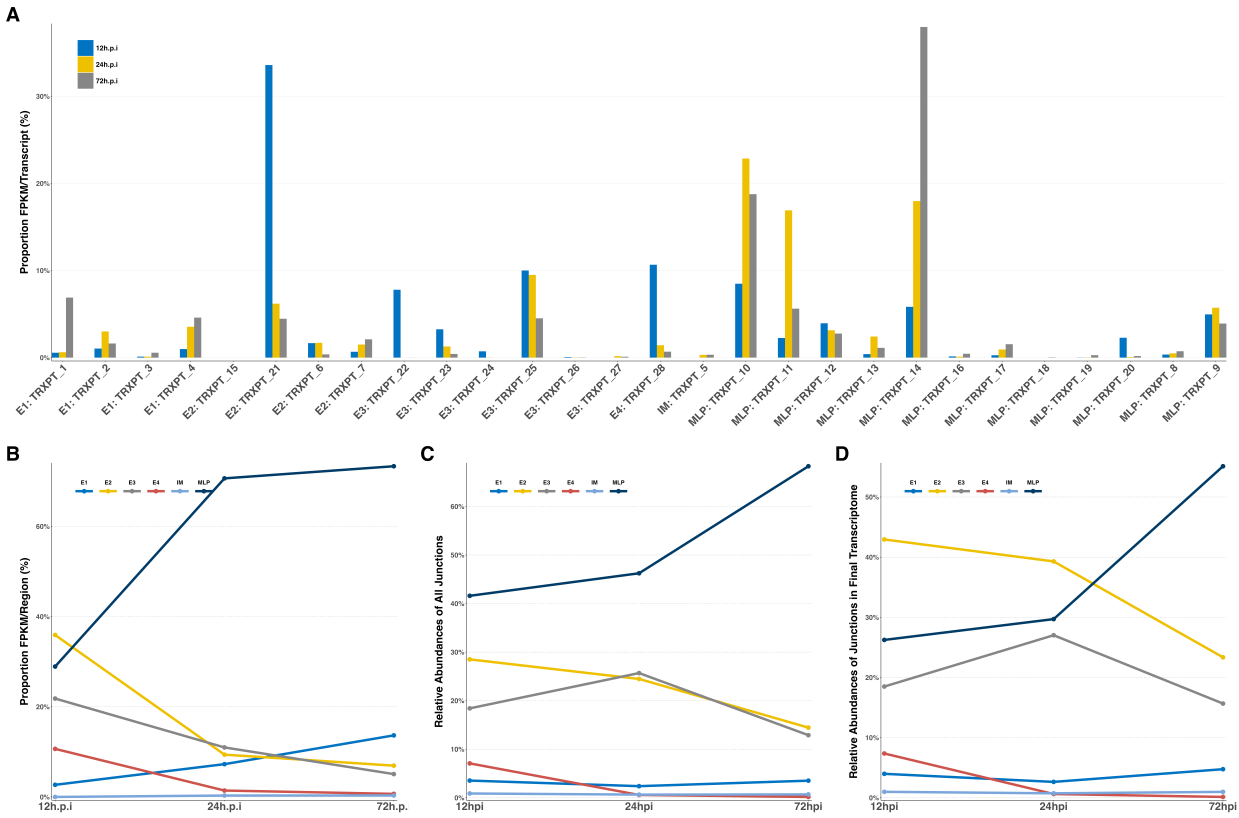
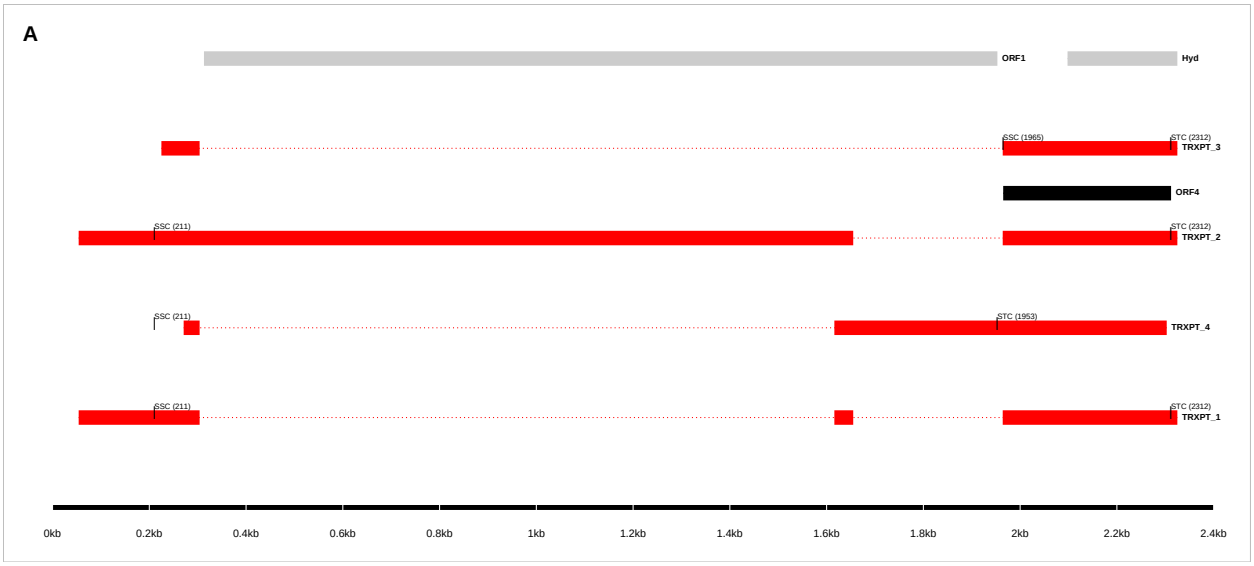


Figure 4: Changes in splicing and expression profile of THEV over time. **a)** Expression levels of transcripts over time.**b)** Expression levels of transcripts by region over time.**c)** Relative abundances of all splice junctions over time. **d)** Relative abundances of junctions in transcriptome.

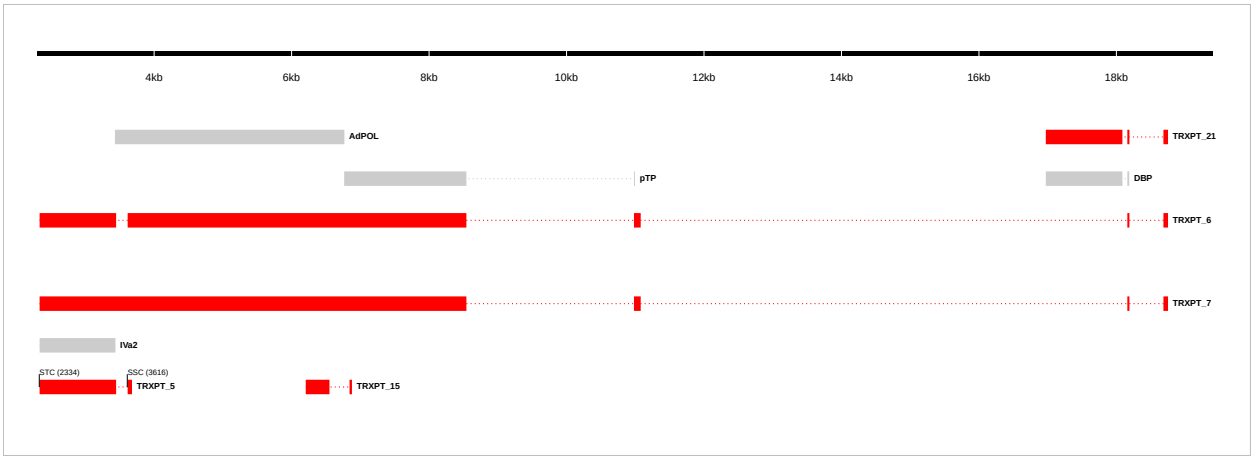


B

Transcript ID	Splice Junction				Strand	Junction Reads				Junction Status
	Start	End	Intron Length	Splice Donor−Acceptor		4h.p.i	12h.p.i	24h.p.i	72h.p.i	
TRXPT_1, TRXPT_4	304	1616	1313bp	GT−AG	+	0	9	1019	25041	Validated
TRXPT_3	304	1964	1661bp	GT−AG	+	0	2	168	1588*	Validated
TRXPT_2, TRXPT_1	1655	1964	310bp	GT−AG	+	0	9	1395	38491	Validated

*Not validated for TRXPT_4

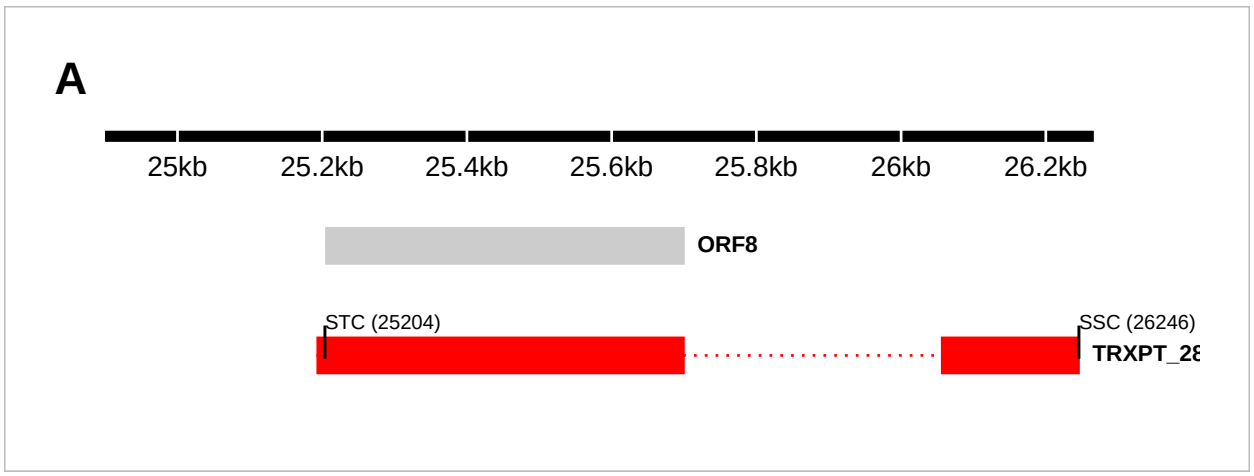
Figure 5: E1 region transcripts. a) The splice map of the E1 transcription unit. Exons are depicted as boxes connected by introns (dotted lines). Transcripts from RNA-seq data are colored red, predicted ORFs are colored grey, and the previously annotated ORF4 is colored black. Each transcript or ORF is labelled with its name to the right. The start codon (SSC) and stop codon (STC) of the 5'-most CDS of each transcript is indicated with the nucleotide position in brackets. The region of the virus is depicted at the bottom as a black line with labels of the nucleotide positions for reference **b)** The sequence reads covering the splice junctions are indicated with information about their validation status using cloning and Sanger sequencing.



Transcript ID	Splice Junction				Strand	region	Junction Reads				Junction Status
	Start	End	Splice Donor–Acceptor	Intron Length			4h.p.i	12h.p.i	24h.p.i	72h.p.i	
TRXPT_5, TRXPT_7	3447	3615	GT–AG	169bp	–	IM, E2	1	5	720	13422	Validated
TRXPT_6, TRXPT_7	11079	18159	GT–AG	7081bp	–	E2	0	2	0	0	Validated
TRXPT_21	18087	18159	GT–AG	73bp	–	E2	9	103	0	0	Validated
TRXPT_21, TRXPT_6, TRXPT_7	18189	18684	CT–AC, GT–AG	496bp	–	E2	0	111	18794	156037	Validated
TRXPT_6, TRXPT_7	8543	10981	GT–AG	2439bp	–	E2	0	0	298	850	Validated
TRXPT_15	6551	6843	GT–GC	293bp	–	E2	0	0	0	6	Unvalidated*

*Truncated transcript

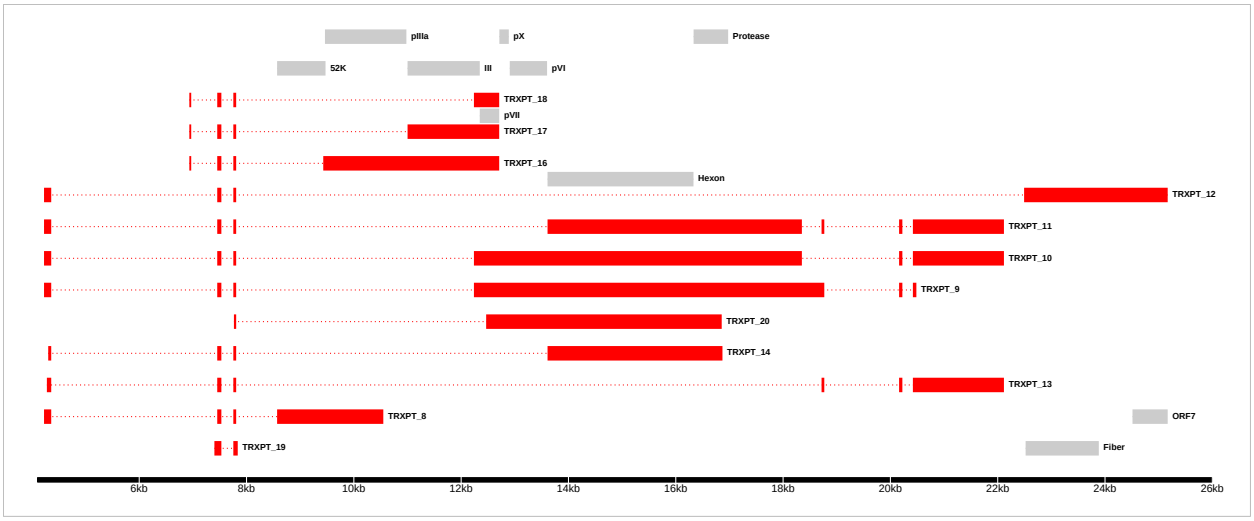
Figure 6: E2 and IM region transcripts. a) The splice map of the E1 and IM transcription units. Exons are depicted as boxes connected by introns (dotted lines). Red transcripts are generated from RNA-seq data and predicted ORFs are colored grey. Each transcript or ORF is labelled with its name to the right. The start codon (SSC) and stop codon (STC) of the 5'-most CDS of each transcript is indicated with the nucleotide position in brackets. The region of the virus is depicted at the bottom as a black line with labels of the nucleotide positions for reference **b)** The sequence reads covering the splice junctions are indicated with information about their validation status using cloning and Sanger sequencing.



B

Transcript ID	Splice Junction Start	Splice Junction End	Splice Junction Donor–Acceptor	Intron Length	Strand	Junction Reads 4h.p.i	Junction Reads 12h.p.i	Junction Reads 24h.p.i	Junction Reads 72h.p.i	Junction Status
TRXPT_28	25701	26055	GT–AG	355bp	–	1	37	602	1,894	Validated

Figure 8: E4 region transcripts. a) The splice map of the E4 transcription unit. Exons are depicted as boxes connected by introns (dotted lines). The transcript from RNA-seq data is colored red and the predicted ORF, grey. The transcript and ORF are labelled with their names to the right. The start codon (SSC) and stop codon (STC) of the 5'-most CDS is indicated with the nucleotide position in brackets. The region of the virus is depicted at the bottom as a black line with labels of the nucleotide positions for reference **b)** The sequence reads covering the splice junction are indicated.



Transcript ID	Splice Junction				Strand	Junction Reads				Junction Status
	Start	End	Splice Donor–Acceptor	Intron Length		4h.p.i	12h.p.i	24h.p.i	72h.p.i	
TRXPT_13, TRXPT_8, TRXPT_14, TRXPT_10, TRXPT_12, TRXPT_11, TRXPT_9	4360	7454	GT–AG	3095bp	+	1	32	7128	179607	Validated
TRXPT_19, TRXPT_13, TRXPT_8, TRXPT_14, TRXPT_10, TRXPT_16, TRXPT_12, TRXPT_17, TRXPT_18, TRXPT_11, TRXPT_9	7531	7754	GT–AG	224bp	+	4	58	12319	322677	Validated
TRXPT_8	7807	8570	GT–AG	764bp	+	0	5	707	0	Validated
TRXPT_16	7807	11001	GT–AG	3195bp	+	0	1	226	10254	Validated
TRXPT_18, TRXPT_11, TRXPT_9	7807	12238	GT–AG	4432bp	+	0	7	2446	0	Validated
TRXPT_20	7807	12466	GT–AG	4660bp	+	0	3	437	0	Validated
TRXPT_14, TRXPT_10	7807	13610	GT–AG	5804bp	+	0	18	2553	0	Validated
TRXPT_13	7807	18717	GT–AG	10911bp	+	0	2	1391	0	Validated
TRXPT_12	7807	22492	GT–AG	14686bp	+	1	6	1642	21891	Validated
TRXPT_25, TRXPT_24, TRXPT_10	18350	18717	GT–AG	368bp	+	4	21	3930	35490	Validated
TRXPT_23, TRXPT_22, TRXPT_11	18350	20162	GT–AG	1813bp	+	3	18	6619	38841	Validated
TRXPT_26, TRXPT_24, TRXPT_13, TRXPT_9, TRXPT_10	18768	20162	GT–AG	1395bp	+	2	21	5207	45062	Validated
TRXPT_26, TRXPT_22, TRXPT_24, TRXPT_13, TRXPT_11, TRXPT_9, TRXPT_10	20223	20419	GT–AG	197bp	+	3	33	10583	93238	Validated
TRXPT_16, TRXPT_17, TRXPT_18	6969	7454	GT–AG	486bp	+	0	0	143	13482	Validated
TRXPT_17	7807	9430	GT–AG	1624bp	+	0	0	67	0	Validated

Figure 9: MLP region transcripts. a)

Table 1: Table 1: Overview of sequencing results

Metric	4h.p.i	12h.p.i	24h.p.i	72h.p.i	Total
Total reads	1.17e+08	7.63e+07	1.20e+08	1.15e+08	4.28e+08
Mapped (Host)	1.04e+08 (89.06%)	6.79e+07 (89.0393%)	1.06e+08 (88.2719%)	8.38e+07 (72.9802%)	3.62e+08
Mapped (THEV)	4.32e+02 (0.0004%)	6.70e+03 (0.0088%)	1.18e+06 (0.9841%)	1.69e+07 (14.6904%)	1.81e+07
Mean Per Base Coverage/Depth	2.42	37.71	6,666.96	95,041.7	101,749
Total unique splice junctions	13	37	236	2374	2,457
Junction coverage Total (at least 1 read)	37	605	115075	2132806	2.25e+06
Junction coverage Mean reads	2.8	16.4	487.6	898.4	351.3
Junction coverage (at least 10 reads)	0	13	132	1791	1,936
Junction coverage (at least 100 reads)	0	1	53	805	859
Junction coverage (at least 1000 reads)	0	0	18	168	186

Table 2: Table 2a: Most abundant splice junctions at 12h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice Acceptor-Donor	Region	Reads	Intron Length	Reads_Percentage
12hpi	-	18,189	18,684	GT-AG	T-A	E2	111	495 bp	111 (18.3%)
12hpi	-	18,087	18,159	GT-AG	T-A	E2	103	72 bp	103 (17%)
12hpi	+	7,531	7,754	GT-AG	T-A	MLP	58	223 bp	58 (9.6%)
12hpi	-	25,701	26,055	GT-AG	T-A	E4	37	354 bp	37 (6.1%)
12hpi	+	20,223	20,419	GT-AG	T-A	E3	33	196 bp	33 (5.5%)
12hpi	+	4,360	7,454	GT-AG	T-A	MLP	32	3,094 bp	32 (5.3%)
12hpi	-	18,751	20,668	GT-AG	T-A	E2	23	1,917 bp	23 (3.8%)
12hpi	+	18,350	18,717	GT-AG	T-A	E3	21	367 bp	21 (3.5%)
12hpi	+	18,768	20,162	GT-AG	T-A	E3	21	1,394 bp	21 (3.5%)
12hpi	-	21,735	22,264	GT-AG	T-A		20	529 bp	20 (3.3%)
12hpi	+	7,807	13,610	GT-AG	T-A	MLP	18	5,803 bp	18 (3%)
12hpi	+	18,350	20,162	GT-AG	T-A	E3	18	1,812 bp	18 (3%)
12hpi	-	18,751	21,682	GT-AG	T-A	E2	12	2,931 bp	12 (2%)
12hpi	+	24,602	25,524	CT-AC	T-A		11	922 bp	11 (1.8%)
12hpi	+	304	1,616	GT-AG	T-A	E1	9	1,312 bp	9 (1.5%)
12hpi	+	1,655	1,964	GT-AG	T-A	E1	9	309 bp	9 (1.5%)
12hpi	-	22,540	24,500	GT-AG	T-A		9	1,960 bp	9 (1.5%)
12hpi	-	18,087	18,163	GT-AG	T-A	E2	8	76 bp	8 (1.3%)

Timepoint	Strand	Start	End	Splice_Site	Splice Acceptor-Donor	Region	Reads	Intron Length	Reads_Percentage
12hpi	+	7,807	12,238	GT-AG	T-A	MLP	7	4,431 bp	7 (1.2%)
12hpi	+	7,807	22,492	GT-AG	T-A	MLP	6	14,685 bp	6 (1%)

Table 3: Table 2b: Most abundant splice junctions at 24h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice Acceptor-Donor	Region	Reads	Intron Length	Reads_Percentage
24hpi	-	18,087	18,159	GT-AG	T-A	E2	18,826	72 bp	18,826 (16.4%)
24hpi	-	18,189	18,684	GT-AG	T-A	E2	18,794	495 bp	18,794 (16.3%)
24hpi	+	7,531	7,754	GT-AG	T-A	MLP	12,319	223 bp	12,319 (10.7%)
24hpi	+	20,223	20,419	GT-AG	T-A	E3	10,583	196 bp	10,583 (9.2%)
24hpi	+	4,360	7,454	GT-AG	T-A	MLP	7,128	3,094 bp	7,128 (6.2%)
24hpi	+	18,350	20,162	GT-AG	T-A	E3	6,619	1,812 bp	6,619 (5.8%)
24hpi	+	18,768	20,162	GT-AG	T-A	E3	5,207	1,394 bp	5,207 (4.5%)
24hpi	-	18,751	20,668	GT-AG	T-A	E2	4,123	1,917 bp	4,123 (3.6%)
24hpi	+	18,350	18,717	GT-AG	T-A	E3	3,930	367 bp	3,930 (3.4%)
24hpi	+	7,807	13,610	GT-AG	T-A	MLP	2,553	5,803 bp	2,553 (2.2%)
24hpi	+	7,807	12,238	GT-AG	T-A	MLP	2,446	4,431 bp	2,446 (2.1%)
24hpi	+	7,807	22,492	GT-AG	T-A	MLP	1,642	14,685 bp	1,642 (1.4%)
24hpi	+	1,655	1,964	GT-AG	T-A	E1	1,395	309 bp	1,395 (1.2%)
24hpi	+	7,807	18,717	GT-AG	T-A	MLP	1,391	10,910 bp	1,391 (1.2%)
24hpi	-	18,751	21,128	GT-AG	T-A	E2	1,124	2,377 bp	1,124 (1%)
24hpi	+	20,223	20,894	GT-AG	T-A	E3	1,208	671 bp	1,208 (1%)

Table 4: Table 2c: Most abundant splice junctions at 72h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice Acceptor-Donor	Region	Reads	Intron Length	Reads_Percentage
72hpi	+	7,531	7,754	GT-AG	T-A	MLP	322,677	223 bp	322,677 (15.1%)
72hpi	+	4,360	7,454	GT-AG	T-A	MLP	179,607	3,094 bp	179,607 (8.4%)
72hpi	-	18,087	18,159	GT-AG	T-A	E2	161,488	72 bp	161,488 (7.6%)
72hpi	-	18,189	18,684	GT-AG	T-A	E2	156,037	495 bp	156,037 (7.3%)
72hpi	+	20,223	20,419	GT-AG	T-A	E3	93,238	196 bp	93,238 (4.4%)
72hpi	+	7,807	13,610	GT-AG	T-A	MLP	81,434	5,803 bp	81,434 (3.8%)
72hpi	+	7,807	12,238	GT-AG	T-A	MLP	77,617	4,431 bp	77,617 (3.6%)
72hpi	+	18,768	20,162	GT-AG	T-A	E3	45,062	1,394 bp	45,062 (2.1%)
72hpi	+	1,655	1,964	GT-AG	T-A	E1	38,491	309 bp	38,491 (1.8%)
72hpi	+	18,350	20,162	GT-AG	T-A	E3	38,841	1,812 bp	38,841 (1.8%)
72hpi	+	18,350	18,717	GT-AG	T-A	E3	35,490	367 bp	35,490 (1.7%)
72hpi	-	18,751	20,668	GT-AG	T-A	E2	28,454	1,917 bp	28,454 (1.3%)
72hpi	+	304	1,616	GT-AG	T-A	E1	25,041	1,312 bp	25,041 (1.2%)
72hpi	+	7,807	12,904	GT-AG	T-A	MLP	21,948	5,097 bp	21,948 (1%)
72hpi	+	7,807	22,492	GT-AG	T-A	MLP	21,891	14,685 bp	21,891 (1%)