Faculty of Medicine

Systems biology

Department of Human and Medical Genetics

Miglė Kazlauskaitė, second year

Master's thesis

5' ir 3' nekoduojančių Uhano uodų viruso 6 sekų nustatymas

Recovery of 5' and 3' UTR sequences of Wuhan Mosquito Virus 6

Supervisor

Dr. Gytis Dudas

Head of the department or clinic

Prof. Dr. Algirdas Utkus

Advisor (if any)

Vilnius, 2026

Student's email migle.kazlauskaite@mf.stud.vu.lt

CONTENTS

LIST OF ABBREVIATIONS

CI – confidence interval.

Ct – cycle threshold.

F – fusion (protein).

GP – glycoprotein.

gp64 – glycoprotein 64.

HA – hemagglutinin.

HE – hemagglutinin-esterase.

HEF – hemagglutinin-esterase-fusion.

hyp1 – hypothetical segment 1.

ICTV – International Committee on Taxonomy of Viruses.

ISAV – infectious salmon anaemia virus.

M – matrix (protein).

MLE – maximum likelihood estimate.

NA – neuraminidase.

NP – nucleoprotein.

NS – non-structural (protein).

ORF – open reading frame.

P3 – polymerase protein 3.

PA – polymerase acidic (protein).

PB1 – polymerase basic protein 1.

PB2 – polymerase basic protein 2.

RACE – rapid amplification of cDNA ends.

RNP – ribonucleoprotein.

SSP – segment specific primer.

TdT – terminal deoxynucleotidyl transferase.

UTR – untranslated region.

vRNA – viral RNA.

WuMV-6 – Wuhan Mosquito Virus 6.

INTRODUCTION

Orthomyxoviruses represent a diverse family of segmented, negative-sense RNA viruses that includes major human and animal pathogens. The *Orthomyxoviridae* family encompasses globally important influenza A and B viruses, which cause seasonal epidemics and pandemics, infectious salmon anaemia virus (ISAV) that devastates aquaculture, and emerging tick-borne pathogens such as Bourbon virus (Kibenge *et al.*, 2004; Sullivan *et al.*, 2010; Kosoy *et al.*, 2015). Recent metagenomic studies have dramatically expanded our understanding of orthomyxovirus diversity, revealing numerous arthropod-associated species worldwide (Shi *et al.*, 2016).

All orthomyxoviruses possess segmented genomes consisting of six to eight single-stranded RNA segments encoding essential viral proteins including RNA polymerase subunits, surface glycoproteins, nucleoprotein, and accessory factors (Bouvier & Palese, 2008; Eisfeld *et al.*, 2015). Each genome segment is flanked by untranslated regions (UTRs) containing highly conserved terminal sequences that form complementary panhandle structures essential for viral replication and transcription (Hsu *et al.*, 1987). In influenza A virus, the terminal 12-13 nucleotides are remarkably conserved across all segments, while ISAV demonstrates variable UTR lengths of 7-48 nucleotides (Desselberger *et al.*, 1980; Kulshreshtha *et al.*, 2010). These UTRs serve critical functions as polymerase recognition sites, regulate mRNA processing, and contain packaging signals for genome assembly (Marsh *et al.*, 2008).

The conserved UTR sequences have enabled powerful research applications including universal primer design for genome amplification, reverse genetics systems for generating recombinant viruses, and mini-genome assays for studying replication mechanisms (Neumann *et al.*, 1999; Lutz *et al.*, 2005; Inoue *et al.*, 2010). However, UTR sequences remain completely uncharacterized for newly discovered arthropod-associated orthomyxoviruses, creating fundamental barriers to functional studies.

Wuhan Mosquito Virus 6 (WuMV-6) exemplifies this knowledge gap as a globally distributed quaranjavirus with an exceptional geographic spread across six continents and a unique eight-segment genome architecture (Dudas & Batson, 2023). The swift worldwide dispersal of WuMV-6 indicates novel evolutionary mechanisms that require examination. Yet the absence of UTR data prevents development of essential molecular tools.

This study aims to characterize the 5′ and 3′ UTRs of WuMV-6 using rapid amplification of cDNA ends (RACE) from Lithuanian *Culex pipiens* mosquito populations. Successful UTR characterization will provide the foundation for developing universal primers and functional assays necessary to understand the molecular biology of this globally important virus group.

The aim of the study:

To determine the 5′ and 3′ untranslated regions of Wuhan Mosquito Virus 6.

The main objectives of the study:

1. To determine the exact 5′ and 3′ UTRs of all eight genome segments of WuMV-6 recovered from Lithuanian *Culex pipiens* mosquito pools.
2. To compare the recovered WuMV-6 UTRs with those of related orthomyxoviruses and assess sequence conservation patterns across viral genera.

LAY DESCRIPTION

Viruses are everywhere, including in mosquitoes. Scientists have recently discovered many new viruses in insects, but very little is known about how these viruses work. It’s unknown whether they might jump to humans or animals. One particularly concerning virus is Wuhan Mosquito Virus 6, which has spread remarkably quickly around the world and is now found on six continents.

WuMV-6 belongs to the same family as influenza, which causes seasonal flu. However, the virus's rapid global spread is puzzling because mosquitoes cannot fly across oceans. This suggests WuMV-6 has special abilities that we don't yet understand.

To study any virus properly, scientists need its complete genetic blueprint. This includes special control regions called "untranslated regions" that work like regulatory hubs. These sequences are essential for developing the tools needed to study the virus, understand how it spreads, and determine if it poses any risks.

This study focused on finding these missing regulatory centres in WuMV-6 samples from Lithuanian mosquitoes using specialized laboratory techniques to decode the virus's complete genetic instructions.

This work provides the foundation scientists need to answer important questions. Could this virus evolve to infect humans? How does it spread so effectively? Understanding WuMV-6 today helps us prepare for potential health threats tomorrow, just as early flu research led to vaccines and treatments that save lives every winter.

LITERATURE REVIEW

1. Orthomyxoviruses: taxonomy, diversity and ecology

1.1 Taxonomy and classification of *Orthomyxoviridae*

The *Orthomyxoviridae* family represents a group of enveloped viruses with segmented, negative-sense RNA genomes (Li *et al.*, 2015; Dudas & Batson, 2023). The term 'negative-sense' (or 'negative-strand') indicates that the viral genomic RNA is complementary to messenger RNA (mRNA) and cannot be directly translated by host ribosomes. Instead, it must first be transcribed into positive-sense RNA by the viral RNA-dependent RNA polymerase. The family name derives from the Greek words *orthos* (meaning 'correct' or 'right') and *myxa* (meaning 'mucus'). This nomenclature reveals their affinity for binding to mucins (mucoproteins) on respiratory tract cells, which was a key characteristic observed when they were first classified (Andrewes *et al.*, 1955). It was established to reflect fundamental differences from the related *Paramyxoviridae* family, such as genome segmentation in orthomyxoviruses versus non-segmented genomes in paramyxoviruses (Lamb & Takeda, 2001; ICTV, 2024). Currently, the International Committee on Taxonomy of Viruses recognises nine genera within the family: *Alphainfluenzavirus*, *Betainfluenzavirus*, *Gammainfluenzavirus*, *Deltainfluenzavirus*, *Isavirus*, *Mykissvirus*, *Quaranjavirus*, *Sardinovirus*, and *Thogotovirus* (ICTV, 2024).

The *Orthomyxoviridae* family was historically comprised primarily of influenza viruses, but recent taxonomic revisions reflect advances in viral discovery and molecular characterization (ICTV, 2024). The four influenza genera include: *Alphainfluenzavirus* (influenza A virus, discovered in 1931), *Betainfluenzavirus* (influenza B virus, discovered in 1936), *Gammainfluenzavirus* (influenza C virus, discovered in 1947), and *Deltainfluenzavirus* (influenza D virus, discovered in 2011), which infect vertebrate hosts (Hause *et al.*, 2013). The genus *Isavirus* contains infectious salmon anaemia virus, originally considered the primary orthomyxovirus of fish, though multiple other fish-infecting orthomyxoviruses have since been discovered across various aquatic species (Rimstad & Mjaaland, 2002; Petrone *et al.*, 2023). The arthropod-associated genera *Thogotovirus* and *Quaranjavirus*, with most thogotoviruses associated with ticks and quaranjaviruses showing diverse arthropod associations, have members capable of infecting vertebrate hosts (Presti *et al.*, 2009; Kosoy *et al.*, 2015).

1.2 Vertebrate-associated orthomyxoviruses

Influenza A and B viruses represent the most globally significant orthomyxoviruses due to their capacity to cause seasonal epidemics and periodic pandemics in human populations. Influenza A viruses exhibit the broadest host range, naturally infecting humans, domestic and wild birds, pigs, horses, marine mammals, and other mammalian species (Webster *et al.*, 1992). The ability of influenza A viruses to undergo antigenic drift through gradual mutation and antigenic shift through genetic reassortment has enabled them to remain pandemic threats throughout human history. Genetic reassortment is the exchange of intact genome segments between different viral strains when they co-infect the same cell. This process creates new viral combinations containing genetic material from both parent viruses. Notable pandemic events include the 1918 H1N1 "Spanish flu", the 1957 H2N2 "Asian flu", the 1968 H3N2 "Hong Kong flu", and the 2009 H1N1 "swine flu" pandemic (Sullivan *et al.*, 2010).

Influenza B viruses have a more restricted host range, primarily circulating in human populations with occasional spillovers from humans into seals (Osterhaus *et al.*, 2000). These viruses do not undergo antigenic shift due to their limited host range, but continue to evolve through antigenic drift, contributing to seasonal influenza epidemics alongside influenza A viruses (Wolff & Veit, 2021). Influenza C viruses infect humans and pigs, typically causing mild respiratory symptoms, particularly in children (Matsuzaki *et al.*, 2006).

Influenza D virus represents a relatively recent addition to the orthomyxovirus taxonomy, first isolated from a pig in Oklahoma in 2011 but subsequently identified as primarily circulating in cattle populations since at least 2003 (Hause *et al.*, 2013). Serological evidence indicates widespread circulation in cattle globally, with seroprevalence rates reaching 80-90% in some regions (Ferguson *et al.*, 2016; Snoeck *et al.*, 2018). Although influenza D virus can infect multiple livestock species, with confirmed infections documented in pigs and cattle and serological evidence of exposure in sheep, goats, and horses, cattle appear to serve as the primary reservoir host (Snoeck *et al.*, 2018).

Infectious salmon anaemia virus represents one of several orthomyxoviruses adapted to aquatic vertebrate hosts. ISAV causes a devastating disease in farmed Atlantic salmon (*Salmo salar*) with mortality rates that can reach 100% during outbreaks (Kibenge *et al.*, 2004). The virus displays unique adaptations to its fish host, including replication at temperatures of 10-15°C and

the ability to infect nucleated erythrocytes. Two major genotypes exist: European and North American strains, which exhibit distinct molecular characteristics and virulence patterns (Falk *et al.*, 1997). The aquatic environment may provide ISAV with transmission routes unavailable to other orthomyxoviruses, potentially influencing its evolutionary trajectory.

1.3 Invertebrate-associated orthomyxoviruses

Thogotoviruses constitute a diverse group of tick-borne orthomyxoviruses with global distribution across Africa, Asia, Europe, Australia, and North America. The prototype Thogoto virus was first isolated from brown (*Rhipicephalus*) ticks in Kenya and has since been detected in various tick species and vertebrate hosts (Sang *et al.*, 2006; Shi *et al.*, 2018). Dhori virus, isolated from Mediterranean (*Hyalomma marginatum)* ticks in India, shares similar biological properties and has been associated with human encephalitis cases (Hao *et al.*, 2022). Both viruses demonstrate the capacity to infect both invertebrate vectors and vertebrate hosts, with documented transmission to cattle, goats, and humans through tick bites (Shi *et al.*, 2018).

Bourbon virus emerged as a human pathogen in 2014 when it was isolated from a fatal case in Kansas, United States. Phylogenetic analysis places Bourbon virus within the *Thogotovirus* genus, closely related to Dhori virus (Kosoy *et al.*, 2015). Subsequent investigations have identified Bourbon virus in lone star (*Amblyomma americanum*) ticks and documented high seroprevalence in white-tailed deer and raccoons (Savage *et al.*, 2017; Bamunuarachchi *et al.*, 2024). The clinical presentation includes fever, thrombocytopenia, and leukopenia, distinguishing it from classical influenza-like illness (Kosoy *et al.*, 2015).

Quaranjaviruses represent arthropod-associated orthomyxoviruses with poorly understood ecology. Most members have not been isolated or sufficiently characterized to determine their host range or transmission cycles. Quaranfil virus, the prototype species, was originally isolated from children with febrile illness in Egypt and subsequently detected in ticks and seabirds across Africa and the Middle East (Taylor *et al.*, 1966; Kemp, Lee & Moore, 1975; Converse & Moussa, 1982; Al-Khalifa, Diab & Khalil, 2007). Johnston Atoll virus, serologically related to Quaranfil virus, was isolated from ticks collected from seabird nests in the central Pacific Ocean (Presti *et al.*, 2009). Another orthomyxovirus, Wellfleet Bay virus, was identified in Cape Cod, Massachusetts, and was linked to recurrent mass die-offs of common eiders (Allison *et al.*, 2015). An arthropod-vectored transmission cycle was suggested for this quaranjavirus.

Metagenomic studies have revolutionized our understanding of invertebrate-associated orthomyxovirus diversity. High-throughput sequencing of arthropod samples has revealed numerous novel orthomyxovirus-like sequences from mosquitoes, midges, flies, and other insects (Shi *et al.*, 2016). These discoveries have expanded the known orthomyxovirus host range beyond traditional vertebrate-tick-mosquito associations to include a vast array of arthropod species. Wuhan Mosquito Virus 6 exemplifies this expanded diversity, representing a globally distributed mosquito-associated orthomyxovirus with rapid evolutionary dynamics (Dudas & Batson, 2023). Thus, the remarkable diversity of arthropod-associated orthomyxoviruses raises questions about potential zoonotic emergence, as demonstrated by the recent discovery of human-pathogenic Bourbon virus.

2. Genome organization of orthomyxoviruses

2.1 General features of orthomyxovirus genomes

All orthomyxoviruses share fundamental structural and genomic characteristics. Their genomes consist of six to eight segments of single-stranded, negative-sense RNA ranging from approximately 10-15 kilobases in total length (Bouvier & Palese, 2008). Each genomic RNA segment exists as a ribonucleoprotein (RNP) complex. It is formed through association of the viral RNA with multiple copies of viral nucleoprotein and the heterotrimeric viral RNA-dependent RNA polymerase complex consisting of PB2, PB1, and PA subunits (Ye *et al.*, 2013).

The segmented nature of orthomyxovirus genomes facilitates genetic reassortment when multiple viral strains co-infect the same cell, a process fundamental to viral evolution and adaptation. Each RNP complex forms rod-shaped, double-helical structures with lengths ranging from 30 to 110 nanometres, correlating directly with the size of the encapsidated RNA segment. These flexible, helical structures enable packaging of diverse segment sizes within a single virion while maintaining structural integrity (Chou *et al.*, 2012; Zheng & Tao, 2013).

The genus-specific variation in segment number reflects distinct evolutionary strategies within the *Orthomyxoviridae* family. Influenza A and B viruses contain eight segments, influenza C and D viruses possess seven segments, infectious salmon anaemia virus maintains eight segments, thogotoviruses typically have six segments, and quaranjaviruses exhibit variable segment numbers (ICTV, 2024). This diversity in genomic organization suggests multiple

independent evolutionary pathways within the family, potentially associated with adaptation to different ecological niches (Allison *et al.*, 2015; Petrone *et al.*, 2023).

2.2 Segment architecture and coding strategies

All orthomyxovirus segments exhibit a conserved architectural organization consisting of 5′ untranslated region, protein-coding sequences, and 3′ untranslated region (Kulshreshtha *et al.*, 2010). The coding region typically contains a major open reading frame that directs expression of one or more viral proteins.

Influenza A and B viruses represent the most extensively characterized orthomyxoviruses, with eight segments encoding PB2, PB1, PA, hemagglutinin (HA), nucleoprotein (NP), neuraminidase (NA), matrix (M), and non-structural (NS) proteins (Eisfeld *et al.*, 2015). The three largest segments encode the heterotrimeric RNA polymerase complex, where PB1 functions as the catalytic subunit containing conserved RNA-dependent RNA polymerase motifs, PB2 provides cap-binding activity for transcription initiation, and PA contributes endonuclease function for cap-snatching (Biswas & Nayak, 1994; Guilligay *et al.*, 2008; Dias *et al.*, 2009). Segments encoding HA and NA proteins determine viral antigenicity and host range specificity, while the M and NS segments utilize alternative splicing to produce multiple proteins from single transcripts (Medina & García-Sastre, 2011; Dubois, Terrier & Rosa-Calatrava, 2014).

Influenza C and D viruses demonstrate alternative genomic strategies with seven segments encoding PB2, PB1, P3, hemagglutinin-esterase-fusion (HEF), NP, M, and NS proteins (Matsuzaki *et al.*, 2006; Skelton & Huber, 2022). The P3 protein is a homologue of PA in influenza A and B. The HEF protein represents a unique evolutionary solution, combining receptor-binding, receptor-destroying, and membrane fusion activities in a single polypeptide that functionally replaces both HA and NA proteins found in influenza A and B viruses (Herrler *et al.*, 1988; Rosenthal *et al.*, 1998).

Infectious salmon anaemia virus maintains eight segments but exhibits distinct protein organization compared to influenza viruses. Segments encode PB2, PB1, PA, NP, fusion (F), hemagglutinin-esterase (HE), NS, and M proteins (Aspehaug *et al.*, 2005). The F protein provides membrane fusion activity analogous to the HA2 subunit of cleaved influenza hemagglutinin, while HE combines receptor-binding and esterase functions, representing a

functional arrangement distinct from influenza virus glycoproteins (Aspehaug *et al.*, 2005, Müller *et al.*, 2010; Cook, Sultana & Lee, 2017).

Thogotoviruses typically possess six segments encoding PB2, PB1, PA, glycoprotein (GP), NP, and M proteins (Weber *et al.*, 1996). Thogotovirus GP do not share significant sequence similarities with the class I fusion proteins (HA, HE, or F) found in other orthomyxovirus genera, indicating that thogotoviruses use a completely different type of membrane fusion protein (Garry & Garry, 2008). Rather, it shows sequence similarity to baculovirus gp64, suggesting independent evolutionary origin and adaptation for arthropod vector transmission (Peng *et al.*, 2017).

Quaranjaviruses exhibit variable segment organization, with members maintaining six to eight genome segments similar to thogotoviruses. However, current evidence suggests that the minimum number of segments is likely seven (NC_052930.1; MT774252.1; MT774251.1; NC_025794.1). Like thogotoviruses, the quaranjavirus GP shows similarities with the gp64 glycoprotein of baculoviruses rather than other orthomyxovirus glycoproteins. The independent acquisition of gp64-like proteins by both quaranjaviruses and thogotoviruses suggests convergent evolutionary adaptation to arthropod host environments, with both genera developing similar molecular solutions to overcome the same host-related challenges (Dr. G. Dudas, pers. comm.).

3. Untranslated regions in orthomyxoviruses

3.1 Definition, orientation, and basic properties

The untranslated regions of orthomyxovirus RNA segments represent critical regulatory elements that flank the protein-coding sequences at both segment termini. In the context of negative-sense RNA viruses, the terminology requires careful distinction between viral RNA (vRNA), messenger RNA, and complementary DNA (cDNA) strands. The orientation and function of UTRs differ fundamentally between the negative-sense vRNA and the positive-sense mRNA/cDNA forms (Kulshreshtha *et al.*, 2010).

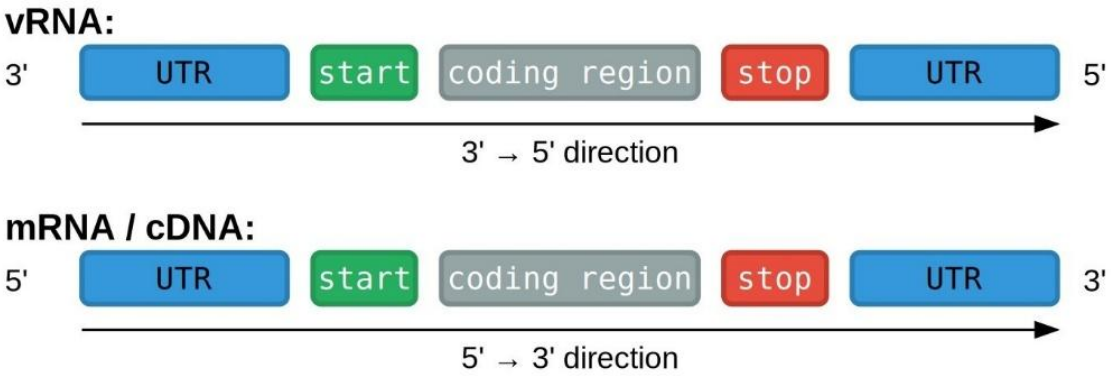

Figure 1. Schematic representation of vRNA and mRNA/cDNA structural organization. Comparison of negative-sense vRNA and positive-sense mRNA/cDNA showing the inverse polarity relationship. UTRs are shown in blue, start codons in green, stop codons in red, and coding regions in grey. Arrows indicate directional orientations. Programmatically generated using Python (v3.12).

The 5′ UTR of negative-sense vRNA contains sequences downstream of the stop codon, while the 3′ UTR of vRNA contains sequences upstream of the translation initiation codon when viewed in the negative-sense orientation. However, because orthomyxovirus genomes consist of negative-sense RNA, the vRNA runs antiparallel to mRNA, such that the 5′ UTR of vRNA becomes the 3′ UTR of the resulting mRNA transcript, and the 3′ UTR of vRNA becomes the 5′ UTR of the mRNA following transcription (Fig. 1) (Kibenge & Godoy, 2016). This inverse relationship between genomic and mRNA orientations necessitates precise nomenclature to avoid confusion in functional studies.

3.2 Sequence characteristics and conservation

The terminal UTR sequences of orthomyxovirus segments demonstrate extraordinary conservation both within and across viral genera, reflecting strong functional constraints on these regulatory elements. Influenza A virus vRNA segments possess highly conserved 12-nucleotide sequences at the 3′ terminus where the open reading frame (ORF) begins (3′-UCGUUUUCGUCC{ORF}-5′) and 13-nucleotide sequences at the 5′ terminus (3′-{ORF}GGAACAAAGAUGA-5′) (Desselberger *et al.*, 1980).

Influenza B virus exhibits similar patterns of conservation with 9 nucleotides at the vRNA 3′ termini and 11 nucleotides at the vRNA 5′ termini remaining common across all eight segments, though the specific sequences differ from those of influenza A virus (Stoeckle *et al.*, 1987). Influenza C virus maintains comparable terminal conservation patterns while displaying

genus-specific sequence motifs that distinguish it from both influenza A and B viruses (Crescenzo-Chaigne, Barbezange & van der Werf, 2008).

Infectious salmon anaemia virus demonstrates unique terminal UTR sequence characteristics among orthomyxoviruses. The vRNA 3′ UTRs vary in length from 7 nucleotides in segment 6 to 48 nucleotides in segment 3, with the terminal 7 nucleotides showing conservation across all eight segments except for specific mutations in segment 5. Notably, ISAV differs from other orthomyxoviruses in having ACU-3′ rather than the typical GCU-3′ motif at the vRNA 3′ terminus of segment 5 (Kulshreshtha *et al.*, 2010).

Thogotoviruses present distinctive terminal UTR sequence patterns that reflect their evolutionary divergence from influenza viruses. Analysis of Thogoto virus vRNA segments reveals that the conserved U at position 3 found in influenza virus vRNA is absent, while an additional U appears at position 8. (Weber *et al.*, 1997).

Comparative analysis of terminal UTR sequence conservation across representative orthomyxoviruses demonstrates both the highly conserved nature of terminal nucleotides and the genus-specific patterns that distinguish different viral lineages (Fig. 2). Beyond the conserved terminal motifs lies the segment-specific region of each UTR, which varies considerably in length and sequence composition between different genome segments. These segment-specific sequences show perfect conservation within segment types across viral strains (Desselberger *et al.*, 1980; Zhao *et al.*, 2014).

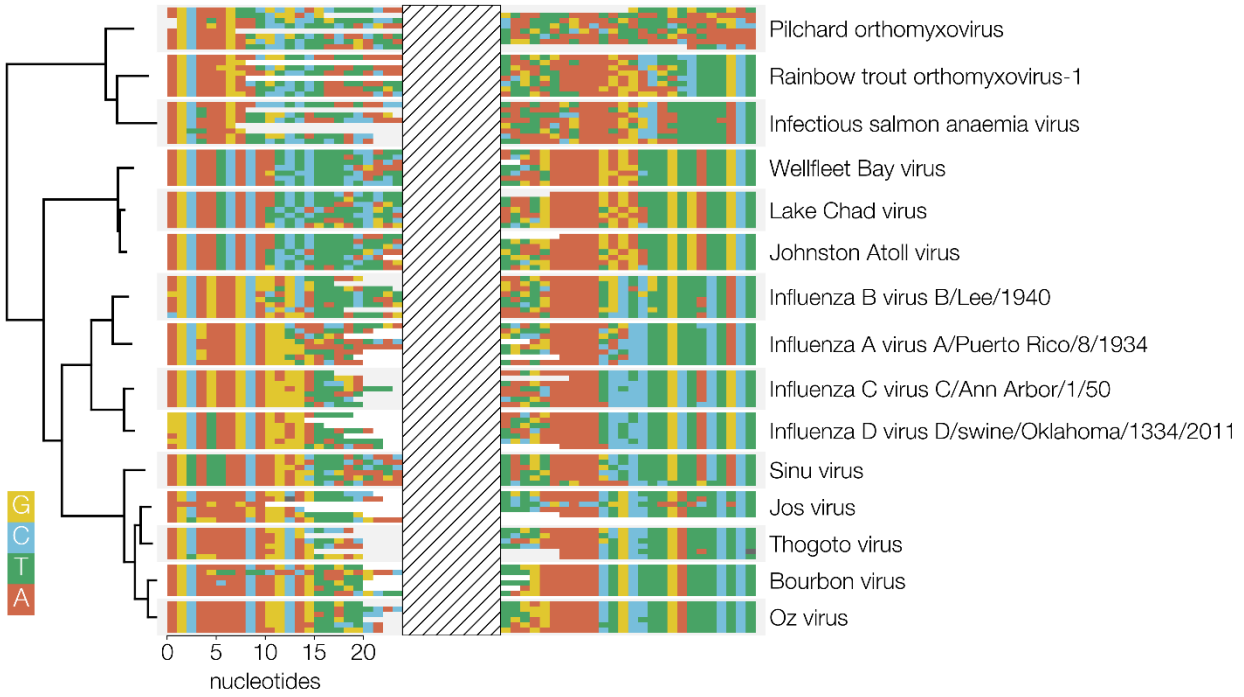

Figure 2. Nucleotide composition of conserved terminal UTR sequences across orthomyxoviruses. Viral ordering is determined by a PB1 amino acid maximum likelihood phylogenetic tree. All known genome segments are shown arranged in the following order from top to bottom: PB1, PB2, PA, NP, HA/HE/HEF, GP/GP64, F, UNKNOWN (the surface protein of Rainbow trout orthomyxovirus-1 labelled as hemagglutinin that bears no recognisable homology to any other protein on public sequence databases), NA, M, NS, VP7. The hatched region represents the coding region. G (yellow) – guanine; C (blue) – cytosine; T (green) – thymine; A (orange) – adenine.

3.3 Structural organization of UTRs

The partially complementary nature of orthomyxovirus terminal UTR sequences enables the formation of distinctive secondary structures that serve as critical regulatory elements. The most extensively characterized of these structures is the panhandle configuration, which results from base-pairing between conserved sequences at the 5′ and 3′ ends of viral RNA segments (Hsu *et al.*, 1987).

In influenza A virus, the panhandle structure consists of a partially double-stranded RNA duplex formed by the terminal 12-13 nucleotides of each segment end. Nuclear magnetic resonance analysis has revealed that this structure adopts an A-form helical geometry on average, with specific bulges occurring at several positions. Between these bulges, Watson-Crick base pairing is maintained, including critical base pairs that contribute to structural stability (Cheong *et al.*, 1996).

Thogotovirus terminal UTR structures demonstrate significant variations from the classical influenza virus panhandle model. The shifted position of conserved nucleotides in Thogoto virus creates a modified panhandle structure with altered base-pairing patterns that affect both the stability and functional properties of the RNA termini. Despite these structural differences, the fundamental requirement for terminal complementarity is preserved, indicating conservation of the basic recognition mechanism while allowing for genus-specific adaptations (Weber *et al.*, 1997).

The structural integrity of terminal regions depends not only on primary sequence conservation but also on the precise positioning of unpaired nucleotides that may serve as recognition elements for viral and cellular proteins or interact with the RNA sequences of other segments. Mutations that disrupt the predicted secondary structure often result in severe defects in RNA synthesis, confirming the functional importance of these structural features (Li & Palese, 1992).

Alternative models for terminal RNA organization have been proposed to explain the complex requirements for transcription and replication. The corkscrew model, which proposes extended base pairing interactions beyond the terminal 12-13 nucleotides to include additional complementary regions further into the segment sequences, suggests additional interactions beyond the simple panhandle. Short-range intramolecular base pairings within each vRNA strand contribute to a more compact and functionally competent promoter conformation (Flick *et al.*, 1996). While the hook model, which suggests that certain unpaired nucleotides in the terminal duplex form protruding loop or hook-like structures, incorporates these unpaired nucleotides as elements that may facilitate polymerase binding and activity (Weber *et al.*, 1997).

3.4 Functional roles of UTRs

Orthomyxovirus UTRs fulfil multiple essential functions in the viral replication cycle, with their roles extending far beyond simple RNA packaging signals. The most fundamental function involves serving as recognition elements for the viral RNA-dependent RNA polymerase, which binds to the panhandle structure formed by the complementary terminal sequences to initiate both transcription and replication (Fodor *et al.*, 1995).

During transcription, the viral polymerase complex recognises the vRNA panhandle structure to begin cap-snatching, a mechanism by which influenza viruses acquire 5′ caps from host cell mRNAs and use the host RNA bound to the cap as a primer for viral mRNA synthesis (Plotch *et al.*, 1981). The terminal sequences provide specific binding sites for individual polymerase subunits. PB1 interacts directly with both 5′ and 3′ ends, while PB2 positions itself to engage in cap-binding and endonuclease activities (Pflug *et al.*, 2014).

In contrast, during replication, the same polymerase complex utilizes the vRNA panhandle structure differently to synthesize full-length anti-genome intermediates, which serve as templates for progeny vRNA production (Fodor *et al.*, 1995). The replication process involves primer-independent initiation at the 3′ terminus of vRNA segments, with the polymerase reading through the entire template to generate full-length complementary copies. These copies then maintain the complementary terminal sequences essential for subsequent rounds of transcription and replication (Fodor, 2013).

The segment-specific regions of UTRs play crucial roles in genome packaging by providing unique identity elements that ensure incorporation of one copy of each segment into progeny virions (Marsh *et al.*, 2008). These sequences, together with adjacent coding regions, constitute bipartite packaging signals that enable the selective assembly of complete genome sets. The specificity of these packaging signals helps explain the remarkable fidelity with which influenza A/B-like viruses assemble complete eight-segment genomes despite the potential for random packaging (Hutchinson *et al.*, 2010).

UTR sequences also regulate polyadenylation of viral mRNAs through a unique stuttering mechanism involving the viral polymerase. The presence of uridine-rich stretches near the 5′ end of vRNA segments causes the polymerase to stutter. It repeatedly copies a short sequence during transcription, resulting in the addition of variable-length poly(A) tails to the developing mRNA. This process differs fundamentally from cellular polyadenylation machinery and represents an autonomous viral mechanism for mRNA maturation (Li & Palese, 1994).

Recent studies have revealed additional regulatory functions for specific UTR motifs in controlling protein expression levels. Analysis of influenza A virus PB1 segment UTRs identified a UAAACU motif in the 3′ UTR that significantly reduces protein expression, suggesting post-transcriptional regulatory mechanisms that fine-tune viral protein synthesis (Ma

et al., 2013). Similar regulatory elements may exist in other segments, contributing to the temporal and quantitative control of viral protein production.

The conservation of UTR sequences under strong purifying selection indicates their critical importance to viral fitness. Evolutionary analysis demonstrates that UTR regions evolve significantly more slowly (2-3 times) than coding sequences, with particularly strong constraints on the terminal nucleotides that form the panhandle structure (Furuse & Oshitani, 2011). This evolutionary pattern confirms the multiple essential functions performed by these regulatory elements and highlights their vulnerability to disruptive mutations.

4. Applied use of orthomyxovirus UTRs

4.1 UTR-based universal primers and full-genome amplification

The conserved terminal sequences of orthomyxovirus UTRs have enabled development of universal primer sets that amplify entire viral genomes without prior sequence knowledge. For Influenza A viruses, this approach exploits the high conservation of 12-13 nucleotides at the cDNA 5′ terminus and 9-12 nucleotides at the cDNA 3′ terminus across all segments within each genus. Universal primers targeting the conserved termini (in cDNA sense) 5′ (5′-AGCAAAAGCAGG-3′, corresponding to vRNA 3′ terminus) and 3′ (5′-AGTAGAAACAAGG-3′, corresponding to vRNA 5′ terminus) can simultaneously amplify all eight genome segments in a single reaction (Inoue *et al.*, 2010). Similarly, influenza B universal primers exploit the 9-nucleotide 5′ (5′-AGCAGAAGC-3′) and 10-nucleotide 3′ (5′-TGTTTCTACT-3′ or 5′-TGTTACTACT-3′) conserved termini (cDNA sense genome), with some variation at position 6 from the 3′ terminus (Zhou *et al.*, 2014).

The high conservation of these terminal sequences also makes them attractive targets for broad-spectrum PCR assays that detect multiple strains within a genus while maintaining high specificity (Muradrasoli *et al.*, 2010). Real-time PCR targeting conserved UTR regions enables rapid influenza diagnosis and subtyping, offering advantages over culture-based methods in speed and sensitivity (Sakurai *et al.*, 2011).

Additionally, the combination of universal primer approaches with modern sequencing technologies has revolutionized viral surveillance by enabling rapid whole-genome sequencing from clinical specimens with minimal viral loads. The integration with next-generation

sequencing platforms allows direct library preparation without cloning steps, facilitating real-time surveillance during outbreaks as demonstrated during the 2009 H1N1 pandemic (Zhou *et al.*, 2009; Licheri *et al.*, 2025). These methods are critical for informing vaccine strain selection and monitoring the emergence of antiviral resistance (Zhou *et al.*, 2014).

Future applications could focus on newly discovered orthomyxoviruses, particularly arthropod-associated species identified through metagenomic surveys. Characterization of UTR sequences from newly discovered viruses like Wuhan Mosquito Virus 6 will be essential for developing molecular tools to study their biology and assess zoonotic potential, as the continued expansion of orthomyxovirus diversity through metagenomic discovery emphasizes the need for systematic UTR characterization efforts.

4.2 Reverse-genetics systems

Reverse genetics systems, which allow researchers to generate infectious viruses from cloned cDNA copies of viral genomes, depend absolutely on authentic viral UTR sequences to reconstitute functional ribonucleoprotein complexes and generate recombinant viruses. The first successful plasmid-based system established by Neumann *et al.* (1999) for influenza A virus. It required precise incorporation of 5′ and 3′ UTR sequences into RNA polymerase I-driven transcription cassettes, as the authentic terminal nucleotides form the panhandle promoter structure essential for viral polymerase recognition.

The eight-plasmid system developed by Hoffmann *et al.* (2000) simplified virus rescue using bidirectional plasmids containing viral cDNA flanked by authentic UTRs under RNA polymerase I control for vRNA synthesis and RNA polymerase II control for protein expression. This approach has become standard for influenza A, B, and C viruses (Hoffmann *et al.*, 2002; Crescenzo-Chaigne & van der Werf, 2007). For infectious salmon anaemia virus, specialized vectors utilizing Atlantic salmon internal transcribed spacer region 1 promoters were necessary because mammalian promoters function poorly in fish cells, and ISAV UTRs show variable lengths (7-48 nucleotides at 3′ termini) compared to influenza viruses (Toro-Ascuy *et al.*, 2015).

4.3 Mini-genome and reporter assays

Mini-genome assays utilize authentic viral UTRs flanking reporter genes to study orthomyxovirus polymerase activity without requiring infectious virus production. The first

influenza system employed firefly luciferase in the negative sense, flanked by nucleoprotein segment 5′ and 3′ UTRs, expressed from RNA polymerase I promoters to generate authentic viral RNA-like transcripts (Lutz *et al.*, 2005). These systems enable investigation of viral replication mechanisms, antiviral screening, and host factor studies.

Gaussia luciferase-based systems offer significant advantages, including 50-fold higher luminescent activity, heat tolerance, and secretion enabling supernatant sampling without cell lysis (Zhu *et al.*, 2011). For influenza B virus, larger UTR sequences (103 nucleotides for 5′ UTR, 53 nucleotides for 3′ UTR) required novel cloning strategies using overlapping primers containing UTR sequences (Sharma *et al.*, 2022). High-throughput screening applications have proven valuable. Bourbon virus mini-genome systems identified dihydroorotate dehydrogenase inhibitors as potent antivirals effective at nanomolar concentrations, demonstrating broad applicability of these systems across orthomyxovirus genera (Yamaoka *et al.*, 2022).

5. WuMV-6 and other mosquito-associated orthomyxoviruses

5.1 Discovery and spread of WuMV-6

Following the initial metagenomic discovery of diverse quaranjaviruses in multiple arthropod species, Wuhan Mosquito Virus 6 emerged as a particularly notable example of viral global dispersal. WuMV-6 was first identified in 2013 from mosquito samples collected in Wuhan, China (Li *et al.*, 2015). The virus exhibits an extraordinary geographic distribution spanning six continents, with detection of closely related genomes in China, California, Missouri, Australia, Greece, Cambodia, Trinidad and Tobago, Puerto Rico, Tunisia, Madagascar (called the “global lineage”), and Sweden, Lithuania, France, the Netherlands, United Kingdom and Belgium (the “Baltic/Northern European” lineage) within a short timeframe (Dudas & Batson, 2023; G. Dudas, pers. comm.). This swift global establishment suggests either highly efficient dispersal mechanisms or human-mediated transportation of infected mosquito vectors (Dudas & Batson, 2023).

Phylogenetic analysis reveals that most globally distributed WuMV-6 variants share a common ancestor from approximately 60 years ago, with segments from different continents showing common ancestry within the past 20 years (Dudas & Batson, 2023). The virus

demonstrates persistent circulation in *Culex* mosquito populations across diverse climatic zones (Birnberg *et al.*, 2020; Abel *et al.*, 2023; Mohamed Ali *et al.*, 2023; Bennouna *et al.*, 2023).

5.2 Genomic features of WuMV-6

The most distinctive feature of WuMV-6 is its eight-segment genome architecture, representing the largest confirmed segment complement among quaranjaviruses (Fig. 3). Beyond the seven segments typical of quaranjaviruses (PB2, PB1, PA, gp64, NP, M (hypothetical 1), VP7 (hypothetical 3)), WuMV-6 possesses one additional hypothetical segment (hypothetical 2), whose function remains unknown (Batson *et al.*, 2021; Dr. A. Karp, pers. comm.). The acquisition of an additional segment may represent evolutionary innovation enabling novel regulatory mechanisms or host adaptation strategies. Alternatively, this pattern may reflect segment loss in earlier discovered quaranjaviruses (Dr. G. Dudas, pers. comm.).

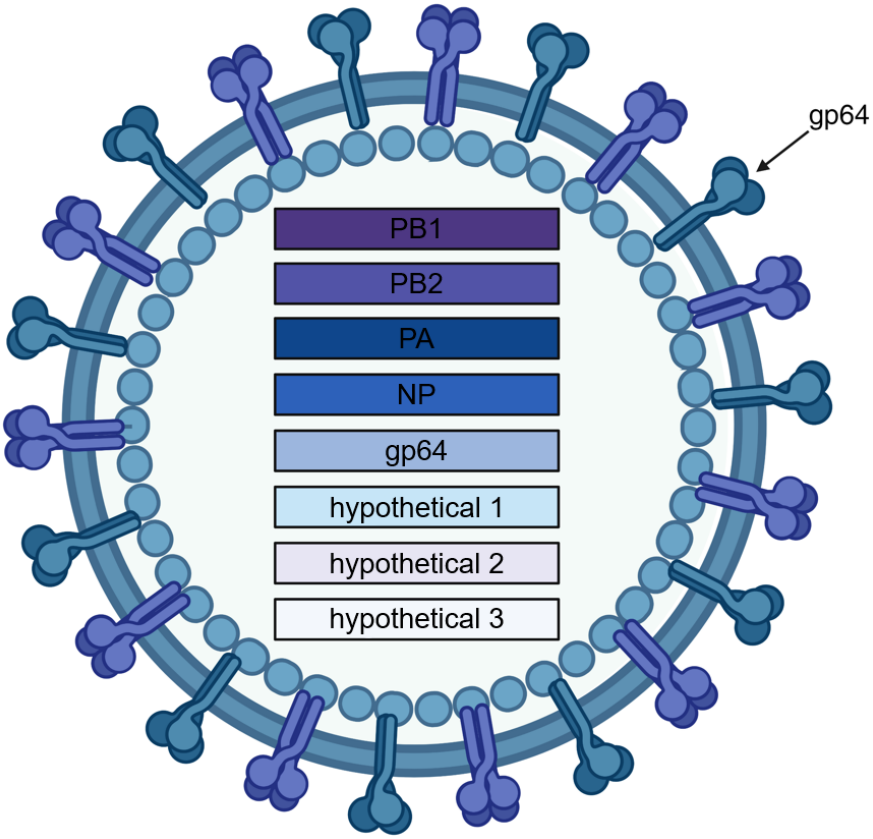

Figure 3. Hypothetical genome organization of WuMV-6 showing eight-segment architecture. Schematic diagram showing the hypothetical virion structure with eight genome segments (labelled rectangular bars). Surface glycoprotein spikes (gp64, arrow) are embedded in the viral envelope. Created using BioRender (https://biorender.com).

Co-occurrence analysis across multiple metagenomic datasets enabled the reconstruction of complete segment repertoires, revealing that hypothetical segments show similar length and predicted transmembrane domains despite sequence divergence. This conservation of structural features suggests functional importance despite their unknown roles. Phylogenetic analysis indicates that the entire eight-segment clade, including related viruses like Guadeloupe mosquito quaranja-like virus 1, Ūsinis virus and Astopletus virus, *Halyomorpha halys* orthomyxo-like virus 1, and Wuhan mosquito virus 4, likely shares this expanded genomic architecture (Batson *et al.*, 2021; Papa *et al.*, 2023; Liu *et al.*, 2025).

The surface glycoprotein gp64 of the clade to which WuMV-6 belongs exhibits faster non-synonymous evolution rates that significantly exceed those observed in other orthomyxovirus proteins. This evolutionary pattern typically characterizes viral proteins under immune selection pressure from vertebrate hosts (Smith *et al.*, 2004; Dudas & Batson, 2023). The observation of antigenic evolution in a purportedly mosquito-specific virus raises intriguing possibilities about unknown vertebrate host involvement or previously unrecognised immune-like responses in arthropod hosts (Dudas & Batson, 2023). The rapid surface protein evolution, combined with the virus's global distribution pattern exceeding typical mosquito flight ranges, suggests potential involvement of migratory vertebrate hosts.

5.3 UTR knowledge gaps in WuMV-6 and other mosquito orthomyxoviruses

Despite comprehensive characterization of coding sequences, the 5′ and 3′ untranslated regions of WuMV-6 and related quaranjaviruses remain completely uncharacterized. This represents perhaps the most significant knowledge gap limiting functional studies of these globally important viruses.

Limited attempts to characterize quaranjavirus UTRs using RACE approaches have achieved only partial success, recovering 3′ UTR sequences from select segments (surface glycoprotein and NP) while failing to obtain 5′ UTRs or complete terminal characterizations (Cholleti *et al.*, 2018).

The distinction between active viral infections and expression from endogenous viral elements (EVEs) represents an additional complication for quaranjavirus studies. Arthropod genomes contain numerous quaranjavirus-derived sequences (piRNA) that can be transcriptionally

active, potentially confounding viral detection efforts (Brait *et al.*, 2024). Evidence suggests that EVEs arise from viral mRNA rather than genomic viral RNA during integration. This is indicated by the preferential incorporation of highly expressed viral genes and the absence of sequences from poorly expressed viral proteins (Whitfield *et al.*, 2017). Authentic UTR sequences would provide definitive markers for distinguishing exogenous viral infections from endogenous element expression, a distinction critical for understanding viral ecology and epidemiology.

The absence of UTR sequence data creates a fundamental barrier to functional investigation of WuMV-6 biology. Universal primer design requires knowledge of conserved terminal motifs that remain unknown for quaranjaviruses. Reverse genetics systems depend absolutely on UTR sequences for generating functional ribonucleoprotein complexes (Neumann *et al.*, 1999; Toro-Ascuy *et al.*, 2015). The inability to develop these tools severely constrains investigation of WuMV-6 replication mechanisms, host range determinants, and pathogenic potential.

METHODS

1. Materials

Mosquito samples collected from 25 sampling localities distributed across Lithuania between July and September 2023 were used in this study. RNA from 3310 *Culex pipiens* adults was previously extracted at LSMU Veterinary Faculty and stored as pools of 1-50 mosquitoes. A total of 86 pools were provided for this study. The pools were collected by Dr. Arnoldas Pautienius and his team (Table 1, APPENDICES). The kits employed in this study are presented in Table 1.1, the reagents in Table 1.2, the primers, probes and adapters in Table 1.3, and the equipment in Table 1.4. All kits and their reagents were stored at room temperature, whereas samples were stored at -80 °C and reaction components were stored at -20 °C at the Life Sciences Center, Biotechnology Institute (LSC BTI), in the laboratory of the Eukaryote Gene Engineering.

Table 1.1 Kits used in this study.

Kit	Catalogue No.	Manufacturer	Purpose
SensiFAST Probe No-ROX One-Step Kit	BIO-76005	Bioline	RT-qPCR
Maxima H Minus Double-Stranded cDNA Synthesis Kit	K2562	Thermo Fisher Scientific	Synthesis of ds cDNA
DNA Clean & Concentrator-5	D4014	Zymo Research	Purification of cDNA
GeneJET Gel Extraction Kit	K0692	Thermo Fisher Scientific	Gel extraction

Table 1.2 Reagents used in this study.

Reagent	Catalogue No.	Manufacturer	Purpose
Klenow Fragment, exo-	EP0421	Thermo Fisher Scientific	A-tailing
T4 DNA Ligase (5 U/μL)	EL0011		Adapter ligation
DreamTaq Hot Start Green PCR Master Mix (2X)	K9021		RACE
Dynabeads Streptavidin for Target Enrichment	65605D	Invitrogen	Purification of linear PCR product

Continuation of Table 1.2			
UltraPure Ethidium Bromide, 10 mg/mL	15585011	Invitrogen	Gel electrophoresis
TAE Buffer (Tris-acetate-EDTA) (50X)	B49	Thermo Fisher Scientific	
GeneRuler Low Range DNA Ladder, ready-to-use	SM0383		
GeneRuler DNA Ladder Mix	SM0333		

Table 1.3 Primers, probes, and adapters used in this study.

ID	Sequence (5' → 3')	Manufacturer
hyp1_fwd	CCTCACCAGGAAAGAGAAGT	NorthSpeed Biotech
hyp1_rev	ATCATGCCGCATGACAAC	
hyp1_probe	FAM-TTCAC+CGAACCAGTTA+CAAA+GA-IBQ	
NP_fwd	TAAGTTGGCGCACCACAA	
NP_rev	GTTCACAGGTTCATGAGGGAA	
NP_probe	HEX-AGAATCCTCACGGAGCTGGAAGTTGC-IBQ	
Adapter	CGTAGCGTACGTAGCTAGCGTGCAGCACTTATGTCGACTCTTCT	
A1	CGTAGCGTACGTAGCTAGCGT	
A2	GCAGCACTTATGTCGACTCTTCT	

Table 1.4 Equipment used in this study.

Device	Manufacturer
Fisherbrand Mini-Centrifuge	Fisher Scientific
Fisherbrand Wizard Infrared Vortex Mixer	
Centrifuge 5415 R	Eppendorf
Qubit 4 Fluorometer	Thermo Fisher Scientific
VeritiPro Thermal Cycler	
QuantStudio 5 Real-Time PCR System for Human Identification	
Standart Power Pack P25	Biometra
GelDoc Go Imaging System	BIO-RAD

2. Methods

2.1 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR was performed using the SensiFAST Probe No-ROX One-Step Kit (Bioline). Diagnostic primer and probe combinations used for the detection of the WuMV-6 hypothetical segment 1 (hyp1) and WuMV-6 NP segment are presented in Table 2.1. Upon thawing, all the components except Reverse transcriptase and RiboSafe RNase Inhibitor were mixed by vortexing, briefly centrifuged, and kept on ice. Each 12,5 µL reaction was prepared in a separate PCR tube by adding the components in the order and volumes indicated in Table 2.2. Each reaction was thoroughly mixed by pipetting, briefly centrifuged, and incubated in the QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific). The RT-qPCR cycling conditions are presented in Table 2.3.

Table 2.1 Primer and probe combinations used for RT-qPCR detection.

ID	Sequence (5' → 3')	PCR product size, bp	Purpose
hyp1_fwd	CCTCACCAGGAAAGAGAAGT	128	Amplification of WuMV-6 hyp1 fragment
hyp1_rev	ATCATGCCGCATGACAAC		
hyp1_probe	FAM-TTCAC+CGAACCAGTTA+CAAA+GA-IBQ		
NP_fwd	TAAGTTGGCGCACCACAA	106	Amplification of WuMV-6 NP fragment
NP_rev	GTTCACAGGTTCATGAGGGAA		
NP_probe	HEX-AGAATCCTCACGGAGCTGGAAGTTGC-IBQ		

Table 2.2 RT-qPCR reaction mixture composition.

Reaction component	Stock conc.	Final conc.	Volume (µL)
Nuclease-free water			0,15
2x SensiFAST Probe No-ROX One-Step Mix	2X	1X	6,25
hyp1_fwd	10 µM	500 nM	0,625
hyp1_rev	10 µM	500 nM	0,625
NP_fwd	10 µM	500 nM	0,625
NP_rev	10 µM	500 nM	0,625
hyp1_probe	6 µM	180 nM	0,375
NP_probe	6 µM	180 nM	0,375
Reverse transcriptase			0,13

Continuation of Table 2.2			
RiboSafe RNase Inhibitor			0,22
RNA			2,5
Final volume			**12,5**

Table 2.3 RT-qPCR cycling conditions.

Step	Stage	Cycles	Temperature, °C	Time
Reverse transcription	1	1	45	10 min
RT inactivation/initial denaturation	2	1	95	10 min
Amplification	3	40	95	15 sec
			60	45 sec

2.2 Maximum likelihood prevalence estimation

Individual-level prevalence of WuMV-6 was estimated from pooled sampling data using maximum likelihood estimation with a negative binomial distribution model. The model assumes uniform prevalence across all pools, with no spatial or temporal variation in infection rates. For each pool of size n with detection outcome d (positive = 1, negative = 0), the probability of observing the outcome was calculated as:

$$P(\text{positive} \mid n, p) = F(n; 1, p) \text{ for positive pools,}$$

$$P(\text{negative} \mid n, p) = 1 - F(n; 1, p) \text{ for negative pools,}$$

where $F(n; 1, p)$ is the cumulative distribution function of the negative binomial distribution with parameters $r = 1$ and probability p. Here, r = 1 represents the shape parameter (number of failures before the first success), and the function calculates the probability that at least one infected individual is present in a pool of size n given individual prevalence p.

The log-likelihood function was calculated as:

$$\log L(p) = \Sigma_i \log P(d_i \mid n_i, p),$$

where d_i and n_i represent the detection outcome and pool size for pool i, respectively. Natural logarithms (base e) were used throughout the calculations.

The prevalence estimate was determined by finding the value of p that maximized the log-likelihood function across all pools. Confidence intervals were calculated using the likelihood ratio test, where the 95% confidence interval includes all values of p where $\log L(p) > \log L(\hat{p}) - 1{,}92$, with $\hat{p}$ being the maximum likelihood estimate.

2.3 cDNA synthesis

cDNA synthesis was performed exclusively on RNA samples that tested positive for WuMV-6 by RT-qPCR.

First Strand cDNA synthesis was performed using the Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific). Upon thawing, the kit components were mixed by vortexing, briefly centrifuged, and kept on ice. Reagents were added to a sterile, RNase-free tube on ice in the following order: 1 µL of random hexamer primers and 13 µL of template RNA to a total volume of 14 µL. The mixture was gently mixed, briefly centrifuged, and incubated at 65 °C for 5 min, then chilled on ice and briefly centrifuged again. Subsequently, 5 µL of 4X First Strand Reaction Mix and 1 µL of First Strand Enzyme Mix were added. The mixture was gently mixed and centrifuged. Incubation was carried out in the VeritiPro Thermal Cycler (Thermo Fisher Scientific) for 10 min at 25 °C, followed by 30 min at 50 °C. The reaction was terminated by heating at 85 °C for 5 min, after which the tube was placed on ice.

Second Strand cDNA synthesis was performed immediately following first strand synthesis. The following components were pipetted directly into the first strand reaction tube on ice in the indicated order: 55 µL of nuclease-free water, 20 µL of 5X Second Strand Reaction Mix, and 5 µL of Second Strand Enzyme Mix, bringing the total reaction volume to 100 µL. The mixture was gently mixed and briefly centrifuged, then incubated in the VeritiPro Thermal Cycler (Thermo Fisher Scientific) at 16 °C for 60 min. The reaction was stopped by adding 6 µL of 0,5 M EDTA (pH 8,0) and mixing gently. The reaction mixture was kept on ice until further processing.

To remove residual RNA, 10 µL of RNase I (100 U) was added to the second strand synthesis reaction tube and incubated for 5 min at room temperature. The reaction mixture was then either proceeded to cDNA purification or stored at -20 °C.

2.4 cDNA purification

cDNA purification was performed using the DNA Clean & Concentrator-5 (Zymo Research) kit. Following RNase I treatment, DNA Binding Buffer was added to the reaction mixture in a 5:1 (buffer:sample) ratio and mixed thoroughly by vortexing. The mixture was transferred to a Zymo-Spin IC Column placed in a Collection Tube and centrifuged at 12,000 × g for 30 sec in the Centrifuge 5415 R (Eppendorf); this centrifugation speed and duration apply to all subsequent steps unless otherwise stated. The flow-through was discarded. The column was washed by adding 200 µL of DNA Wash Buffer and centrifuging; this wash step was repeated once. To ensure complete removal of the DNA Wash Buffer, the column was centrifuged for 2 min. To elute the purified ds cDNA, 20 µL of nuclease-free water was added directly to the column matrix and incubated at room temperature for 1 min. The column was then transferred to a 1,5 mL microcentrifuge tube and centrifuged. To improve yield, elution was repeated a second time. cDNA concentration was determined using a Qubit 4 Fluorometer (Thermo Fisher Scientific). The purified cDNA was then either proceeded to A-tailing or stored at -20 °C.

2.5 A-tailing

A 3'-dA mononucleotide extension was added to the purified ds cDNA by incubation with Klenow Fragment, exo- (Thermo Fisher Scientific) polymerase. Upon thawing, all the components were mixed by vortexing, briefly centrifuged, and kept on ice. Each 20 µL reaction was prepared in a separate PCR tube by adding the components in the order and volumes indicated in Table 2.4. Each reaction was thoroughly mixed by pipetting, briefly centrifuged, and incubated in the VeritiPro Thermal Cycler (Thermo Fisher Scientific) at 37 °C for 45 min, followed by enzyme inactivation at 75 °C for 15 min. The A-tailed cDNA was subsequently purified as described in Section 2.4 (cDNA purification).

Table 2.4 A-tailing reaction mixture composition.

Reaction component	Stock conc.	Final conc.	Volume (µL)
Purified ds cDNA			16,8
10X reaction buffer	10X	1X	2
dATP	10 mM	0,5 mM	1
Klenow Fragment, exo-	5 U	1 U	0,2
Final volume			**20**

2.6 Adapter ligation

Adapter ligation was performed using a 3' T-overhang compatible adapter, the sequence of which is presented in Table 1.4. Upon thawing, all the components except T4 DNA ligase (Thermo Fisher Scientific) were mixed by vortexing, briefly centrifuged, and kept on ice. Each 30 µL reaction was prepared in a separate PCR tube by adding the components in the order and volumes indicated in Table 2.5. Each reaction was thoroughly mixed by pipetting, briefly centrifuged, and incubated in the VeritiPro Thermal Cycler (Thermo Fisher Scientific) at 22 °C overnight, followed by enzyme inactivation at 65 °C for 10 min. The adapter-ligated cDNA was subsequently purified as described in Section 2.4 (cDNA purification).

Table 2.5 Adapter ligation reaction mixture composition.

Reaction component	Stock conc.	Final conc.	Volume (µL)
A-tailed purified cDNA			20
Adapter	33,33 µM	3,33 µM	3
T4 DNA ligase buffer	10X	1X	3
T4 DNA ligase	5 U	15 U	3
Nuclease-free water			1
Final volume			**30**

2.7 Rapid amplification of cDNA ends

RACE was performed using the DreamTaq Hot Start Green PCR Master Mix (2X) (Thermo Fisher Scientific). The segment-specific and adapter-complementary primer combinations used are presented in Table 2.6. Upon thawing, all the components were mixed, briefly centrifuged, and kept on ice. Each 25 µL reaction was prepared in a separate PCR tube by adding the components in the order and volumes indicated in Table 2.7. Each reaction was thoroughly mixed by vortexing, briefly centrifuged, and incubated in the VeritiPro Thermal Cycler (Thermo Fisher Scientific). The PCR cycling conditions are presented in Table 2.8.

Table 2.6 Segments-specific and adapter-complementary primer combinations used for RACE.

ID	Sequence (5' → 3')	PCR product size, bp	T_a, °C	Purpose
hyp1_fwd	CCTCACCAGGAAAGAGAAGT	~795	61,4	hyp1 3' RACE
A1	CGTAGCGTACGTAGCTAGCGT			
A1	CGTAGCGTACGTAGCTAGCGT	~297	59,9	hyp1 5' RACE
hyp1_rev	ATCATGCCGCATGACAAC			
NP_fwd	TAAGTTGGCGCACCACAA	~328	61,7	NP 3' RACE
A1	CGTAGCGTACGTAGCTAGCGT			
A1	CGTAGCGTACGTAGCTAGCGT	~1727	61,8	NP 5' RACE
NP_rev	GTTCACAGGTTCATGAGGGAA			

Table 2.7 RACE reaction mixture components.

Reaction component	Stock conc.	Final conc.	Volume (µL)
DreamTaq Hot Start Green PCR Master Mix (2X)	2X	1X	12,5
Forward primer	10 µM	0,5 µM	1,25
Reverse primer	10 µM	0,5 µM	1,25
Adapter-ligated purified cDNA			0,25
Nuclease-free water			9,75
Final volume			**25**

Table 2.8 PCR cycling conditions.

Step	Stage	Cycles	Temperature, °C	Time
Initial denaturation	1	1	95	1 min
Denaturation	2	40	95	30 sec
Annealing	3	40	T_a	1 sec
Extension	4	40	72	1 min
Final extension	5	1	72	5 min

T_a for each primer pair is presented in Table 2.6.

2.8 Linear polymerase chain reaction (PCR)

Linear PCR was performed as an alternative to RACE using biotinylated segment-specific primers to enable subsequent purification and enrichment of target sequences. The segment-specific primers used were identical to those presented in Table 2.6, except for the biotin attached at the 5′ end of each primer. Linear PCR was conducted using a single biotinylated

segment-specific primer per reaction, without the corresponding adapter primer. Upon thawing, all components were mixed, briefly centrifuged, and kept on ice. Each 25 µL reaction was prepared in a separate PCR tube by adding the components in the order and volumes indicated in Table 2.9. Each reaction was thoroughly mixed by vortexing, briefly centrifuged, and incubated in the VeritiPro Thermal Cycler (Thermo Fisher Scientific). The PCR cycling conditions are presented in Table 2.10.

Table 2.9 Linear PCR reaction mixture components.

Reaction component	Stock conc.	Final conc.	Volume (µL)
DreamTaq Hot Start Green PCR Master Mix (2X)	2X	1X	12,5
Biotinylated segment-specific primer	10 µM	0,5 µM	1,25
Adapter-ligated purified cDNA			0,25
Nuclease-free water			11
Final volume			**25**

Table 2.10 Linear PCR cycling conditions.

Step	Stage	Cycles	Temperature, °C	Time
Initial denaturation	1	1	95	1 min
Denaturation	2	375	95	30 sec
Annealing	3	375	T_a	30 sec
Extension	4	375	72	1 min
Final extension	5	1	72	5 min

T_a for single-use segment-specific primers remained the same as specified for the corresponding primer pairs in Table 2.6.

2.9 Streptavidin magnetic bead purification

Purification of biotinylated linear PCR products was performed using Dynabeads Streptavidin for Target Enrichment (Invitrogen). Binding and washing buffers were prepared according to manufacturer specifications. The 2X B&W Buffer contained 10 mM Tris-HCl (pH 7,4), 1 mM EDTA, and 2 M NaCl. The 1X B&W Buffer was prepared by diluting 2X B&W Buffer with equal volume of nuclease-free water (1:1 ratio).

The beads were resuspended completely by vortexing and placed on a roller mixer for 20 min. Following resuspension, 12,5 µL of bead suspension was transferred to a new

microcentrifuge tube for washing to remove excess streptavidin. The tube was placed on a magnet for 1 min and the supernatant was carefully discarded. An equal volume (12,5 µL) of 1X B&W Buffer was added and beads were resuspended by vortexing. The tube was placed on a magnet for 1 min and the supernatant was carefully discarded. This washing step was repeated twice for a total of three washes.

For immobilization of biotinylated products, pre-washed beads were resuspended in 25 µL of 2X B&W Buffer to achieve a final concentration of 5 µg/µL. An equal volume (25 µL) of biotinylated linear PCR product was added and the mixture was incubated for 15 min at room temperature using gentle rotation. Biotinylated DNA-coated beads were separated using a magnet for 1 min and the supernatant was carefully discarded. The coated beads were washed 3 times with 50 µL of 1X B&W Buffer, with magnetic separation for 1 min and supernatant removal between each wash. For elution, beads were resuspended in 50 mM Tris-HCl and heated at 95 °C for 5 min. The supernatant was collected rapidly on a magnet while still hot and immediately processed for first strand cDNA synthesis only, following the procedure described in Section 2.3 (cDNA synthesis) but omitting the second strand synthesis step. PCR amplification was then performed as described in Section 2.7 (RACE), with annealing time modified to 30 sec instead of 1 sec.

2.10 Agarose gel electrophoresis

Agarose gel electrophoresis was performed to visualize and analyse RACE and linear PCR products. For shorter WuMV-6 products, 1,5% agarose gels were prepared by dissolving 1,5 g agarose in 100 mL of 1X TAE buffer, while 1% agarose gels were prepared using 1 g agarose in 100 mL of 1X TAE buffer for longer WuMV-6 product visualization. The mixture was heated until completely dissolved, cooled to 60 °C, and 10 µL of ethidium bromide was added before pouring into the gel casting system where it was allowed to solidify for 20 min. The solidified gel was placed in an electrophoresis tank connected to a Standart Power Pack P25 power supply (Biometra). PCR products and DNA size standards were loaded into gel wells. Electrophoresis was conducted to separate DNA fragments according to size. Following electrophoresis, gels were visualized using a GelDoc Go Imaging System (BIO-RAD). The resulting images were used for analysis of DNA fragment size and identification of target bands for excision.

2.11 Gel excision

Following agarose gel electrophoresis, DNA bands of expected size were visualized under UV illumination and excised using a sterile scalpel blade. Target bands were identified by comparison with molecular weight markers and cut as close to the DNA as possible to minimize excess agarose. Each excised gel slice was placed in a sterile 1,5 mL microcentrifuge tube and weighed to determine gel volume for subsequent purification. Care was taken to avoid cross-contamination between samples by cleaning the scalpel blade in 100% ethanol between each excision. DNA extraction from excised gel slices was performed immediately using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) following manufacturer's instructions.

2.12 Sequencing

All gel-excised samples were advanced to sequencing. Sample concentration was determined using a Qubit 4 Fluorometer (Thermo Fisher Scientific). Sequencing was performed using a third-generation Oxford Nanopore MinION sequencer (Oxford Nanopore Technologies) at the MB SeqVision sequencing facility.

2.13 Sequence data analysis

Sequencing reads obtained from Oxford Nanopore MinION sequencing were analysed to identify WuMV-6 genomic segments. Assembled contigs were mapped to the WuMV-6 reference genome using minimap2 (v2.24) with assembly-to-reference alignment parameters (-ax asm5) (Li, 2018). Mapped sequences were processed using SAMtools (v1.15) to extract contigs aligning to the reference genome (Danecek *et al.*, 2021). Adapter sequences (Table 1.3, METHODS) were removed using Cutadapt (v4.1) with default parameters (Martin, 2011). WuMV-6 gene segments were manually classified into hyp1 and NP categories based on homology to known WuMV-6 gene regions.

2.14 Phylogenetic analysis

Generated sequences were analysed together with WuMV-6 genome data collected by Linas Gasparavičius from publicly available sources (China National GeneBank DataBase, CNGBdb; Science Data Bank, SciDB; National Center for Biotechnology Information, NCBI). To integrate new sequences with the existing dataset, multiple sequence alignment was performed

using MAFFT (G-INS-i mode) (v7.490) (Katoh *et al.*, 2005). Phylogenetic tree construction was performed using PhyML (v3.3) with the Hasegawa-Kishino-Yano nucleotide substitution model with gamma-distributed evolutionary rates using four rate categories (HKY+Gamma(4)) (Hasegawa, Kishino & Yano, 1985; Yang, 1994). This model employs the maximum likelihood method for tree construction. The tree was rooted with the branch leading to Baltic/Northern European diversity used as the tree root (Dudas & Batson, 2023). Final phylogenetic tree visualization was performed using matplotlib (v3.10.9) and baltic (v0.3.0) (Hunter, 2007).

2.15 UTR conservation analysis

For comparative analysis of UTR conservation across orthomyxoviruses, non-WuMV-6 sequences were obtained from NCBI Virus database. Multiple sequence alignments of 5′ and 3′ UTR sequences were performed using MAFFT (v7.490) in G-INS-i mode (Katoh *et al.*, 2005). Phylogenetic tree construction was performed using PhyML (v3.3) with the LG substitution matrix (Le & Gascuel, 2008). The resulting phylogenetic tree and UTR alignments were visualized using matplotlib (v3.10.9) and baltic (v0.3.0) libraries (Hunter, 2007).

RESULTS

1. RT-qPCR detection of WuMV-6 in mosquito pools

qPCR screening was performed on all 86 samples to identify those positive for WuMV-6, as only confirmed positive samples were eligible for subsequent RACE analysis. Of these, 36 (~42%) tested positive and 50 (~58%) tested negative. Of the 36 samples that tested positive by at least one target, 27 were positive for both the hyp1 and NP segments, while 5 were positive exclusively for hyp1 and 4 exclusively for NP (Table 1, APPENDICES). Positive samples exhibited cycle threshold (Ct) values ranging from 16,88 to 29,43 (mean 23,78 ± 4,03 SD) for the hyp1 and from 18,94 to 29,53 (mean 25,06 ± 3,53 SD) for the NP segment (Fig.4; Fig. 5). Negative samples returned Ct values above the pre-defined detection threshold of 29,5 or showed no amplification. A subset of samples produced Ct values proximal to the detection threshold (Ct 29-30), representing the zone of analytical uncertainty inherent to qPCR-based detection at low template concentrations. These borderline results were classified according to the pre-defined cutoff of Ct 29,5, which was chosen as a conservative threshold to ensure high confidence in positive detection while maintaining assay sensitivity.

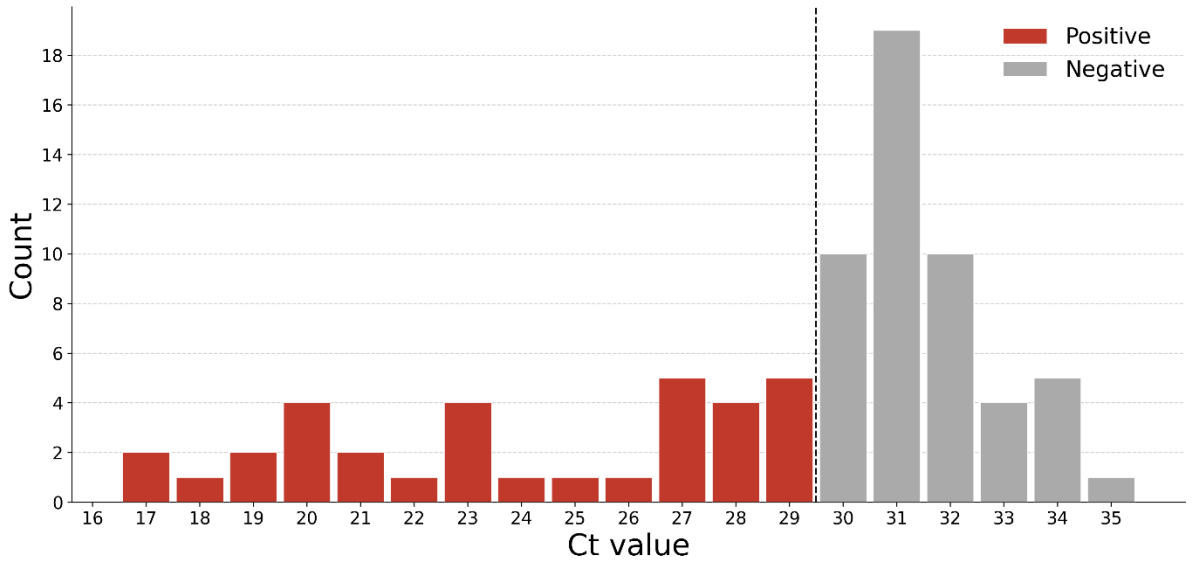

Figure 4. Distribution of Ct values for WuMV-6 hyp1. Frequency histogram of observed Ct values coloured by detection outcome based on a pre-defined threshold (Ct 29,5 – dashed vertical line). Values at or below the threshold are classified as positive (red), values above as negative (grey). n = 82. Samples that did not yield a

detectable Ct value within 40 cycles (undetermined; n = 4) are not shown but were treated as negative for the purposes of this analysis. Programmatically generated using Python (v3.12) with matplotlib (v3.10.9).

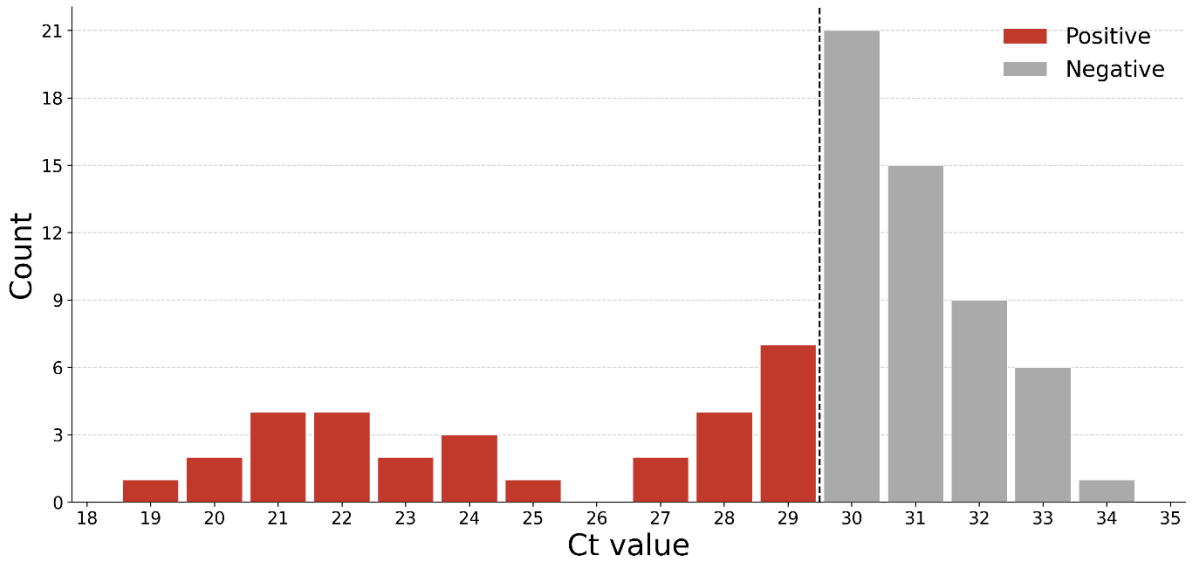

Figure 5. Distribution of Ct values for WuMV-6 NP segment. Frequency histogram of observed Ct values coloured by detection outcome based on a pre-defined threshold (Ct 29,5 – dashed vertical line). Values at or below the threshold are classified as positive (red), values above as negative (grey). n = 82. Samples that did not yield a detectable Ct value within 40 cycles (undetermined; n = 4) are not shown but were treated as negative for the purposes of this analysis. Programmatically generated using Python (v3.12) with matplotlib (v3.10.9).

Ct values of all positive samples were compared between hyp1 (n = 32, mean 23,78 ± 4,03 SD) and NP (n = 31, mean 25,06 ± 3,53 SD). No statistically significant difference was detected in unpaired analysis (Mann-Whitney U test, p = 0,12). However, paired analysis of samples with valid results for both targets (n = 27) revealed a significant difference (Wilcoxon signed-rank test, p = 0,0005), with hyp1 yielding mean Ct values 1,12 cycles lower than NP.

2. Geographic distribution and individual prevalence of WuMV-6 across Lithuania

Having established the detection status of all 86 pools by qPCR, the geographic provenance of positive and negative results was examined to assess the spatial distribution of WuMV-6 circulation across the sampled region. Of the 25 localities sampled, 12 (48%) yielded at least one WuMV-6-positive pool, indicating a geographically widespread presence of the virus across the country (Fig. 6).

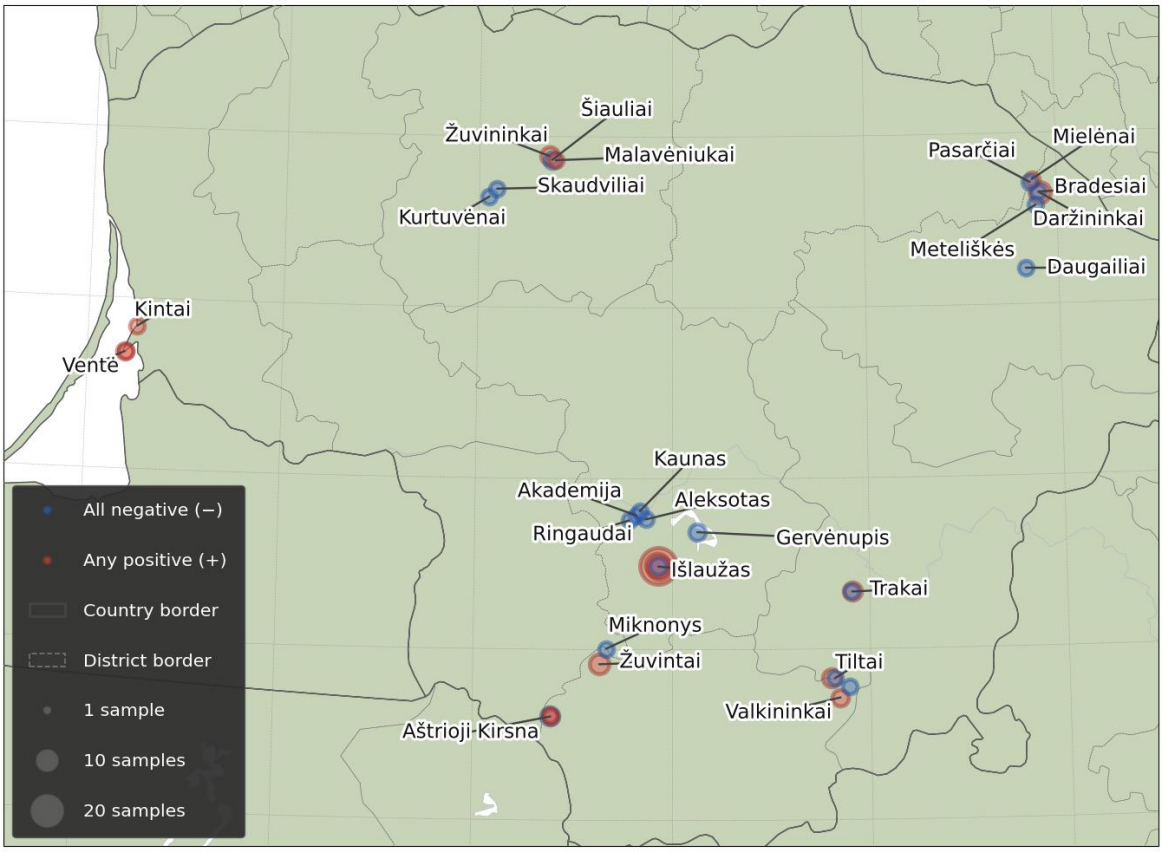

Figure 6. Geographic distribution of WuMV-6 detection in *C. pipiens* mosquito pools collected across Lithuania in 2023. Each circle represents a unique sampling coordinate; circle size is proportional to the total number of pools collected at that location. Red circles indicate sites where at least one pool tested positive for WuMV-6; blue circles indicate sites where all pools tested negative. Administrative district boundaries are shown as dashed lines; country borders as solid lines. Programmatically generated using Python (v3.12) with cartopy (v0.23).

The highest number of positive pools was recorded at Išlaužas (16 of 37 pools; 43,2%), which also represented the site of greatest sampling intensity in the study. Among sites with multiple pools, Ventė, a coastal locality in western Lithuania, exhibited the highest positivity rate, with all three pools testing positive (100%). Similarly, all pools from Kintai, Malavėniukai and Mielėnai were positive, though each site contributed only a single pool. Žuvininkai (66,7%), Daržininkai (60%), Žuvintai (50%), Trakai (50%), and Tiltai (50%) also showed substantial positivity rates. While Aštrioji Kirsna and Valkininkai returned lower positivity (40% and 33%, respectively). Thirteen localities, including Kaunas, Aleksotas, Akademija, Ringaudai, Gervėnupis, Šiauliai, and several northeastern sites, yielded exclusively negative results across all

pools tested (Fig. 6). However, the limited number of pools per site at many of these locations warrants caution in interpreting negative results as indicative of absent virus circulation.

To estimate the individual-level infection rate underlying the pool-based detection data, a maximum likelihood approach was applied across the full set of 86 pools. The resulting prevalence estimate for WuMV-6 in *C. pipiens* was 1,45% (95% confidence interval (CI): 1,03% - 1,98%), indicating that approximately 1 in 69 individual mosquitoes carried the virus at the time of collection (Fig. 7). This estimate is in close agreement with an independently derived winter estimate of approximately 1,5% from the same year (Dr. G. Dudas, pers. comm.), lending additional confidence to the robustness of the result.

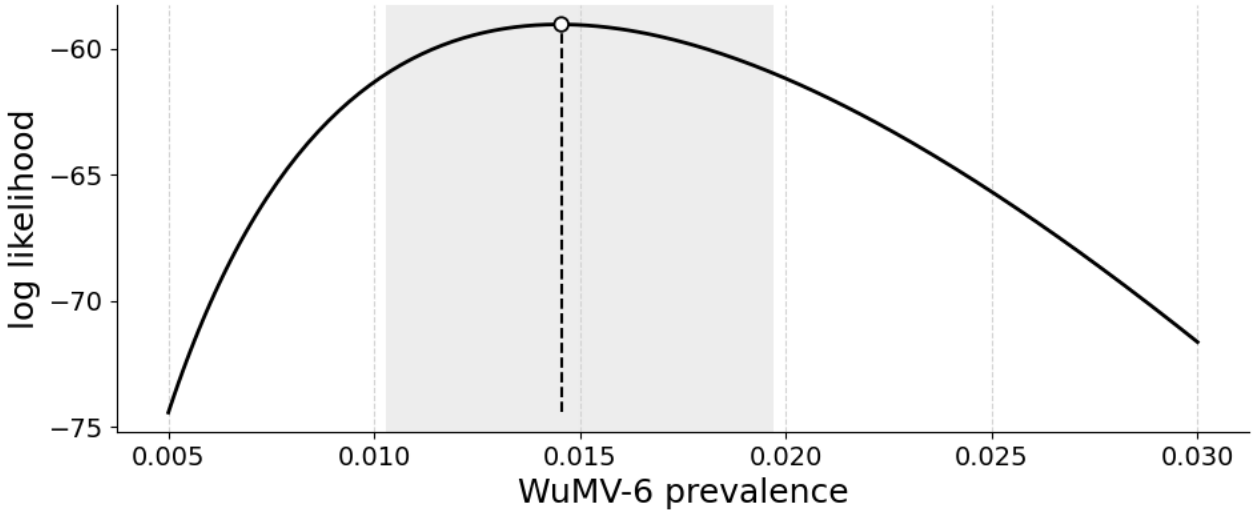

Figure 7. Maximum likelihood profile estimate of WuMV-6 prevalence in *C. pipiens* mosquitoes sampled across Lithuania in 2023. The dashed vertical line and open circle indicate the maximum likelihood estimate (MLE = 1,45%). The shaded region represents the 95% CI (1,03% - 1,98%).

3. Amplification of WuMV-6 UTRs

All 36 qPCR-confirmed positive samples were advanced to RACE analysis to characterise the UTRs of the WuMV-6 genome. A critical prerequisite for RACE is the ligation of a known adapter sequence to the RNA termini, against which a universal adapter primer can subsequently be used in combination with segment-specific primers (SSPs). To ensure specificity, the adapter was required to be non-homologous to both the *C. pipiens* host genome and the WuMV-6 viral genome. The adapter sequence was constructed by combining two primer sequences: A1 (5'-CGTAGCGTACGTAGCTAGCGT-3'), containing 15 bases with homology to

the mosquito genome, and A2 (5'-GCAGCACTTATGTCGACTCTTCT-3'), containing 16 bases of host genome homology. These represented the optimal variants identified following extensive sequence homology screening, with the combined adapter exhibiting the least overall homology to both the host and viral genomes among all candidates evaluated.

Four SSPs were used against the WuMV-6 genome, two targeting the hyp1 and two targeting the NP segment, covering both the 3' and 5' UTRs of each segment (Table 1.3, METHODS). These SSPs were identical in sequence to the qPCR primers designed by Dr. Philippe Selhorst (Institute of Tropical Medicine in Antwerp) and used for initial screening. Each adapter primer was evaluated in combination with all four SSPs. Agarose gel electrophoresis of representative RACE products from five positive samples (IS1-2, IS1-3, IS1-4, IS2-6, IS3-5) using the hyp1 5' RACE combination is shown in Figure 8 (Table 2.6, METHODS). Comparable amplification patterns were observed across the remaining SSP combinations (data not shown). A2 in combination with SSPs failed to produce any detectable amplification product and was therefore deemed unsuitable for further analysis. A1 in combination with SSPs produced amplification products in the expected size range (~297 bp). However, non-specific bands were also observed, attributable to the partial homology of A1 with the *C. pipiens* genome. Despite this, A1 was retained for subsequent RACE analysis, as it consistently yielded products of the expected size alongside the non-specific background. Bands of the expected size were therefore excised from the gel and submitted for Nanopore sequencing to confirm product identity. Excised bands are indicated by red rectangles in the corresponding gel image (Fig. 8).

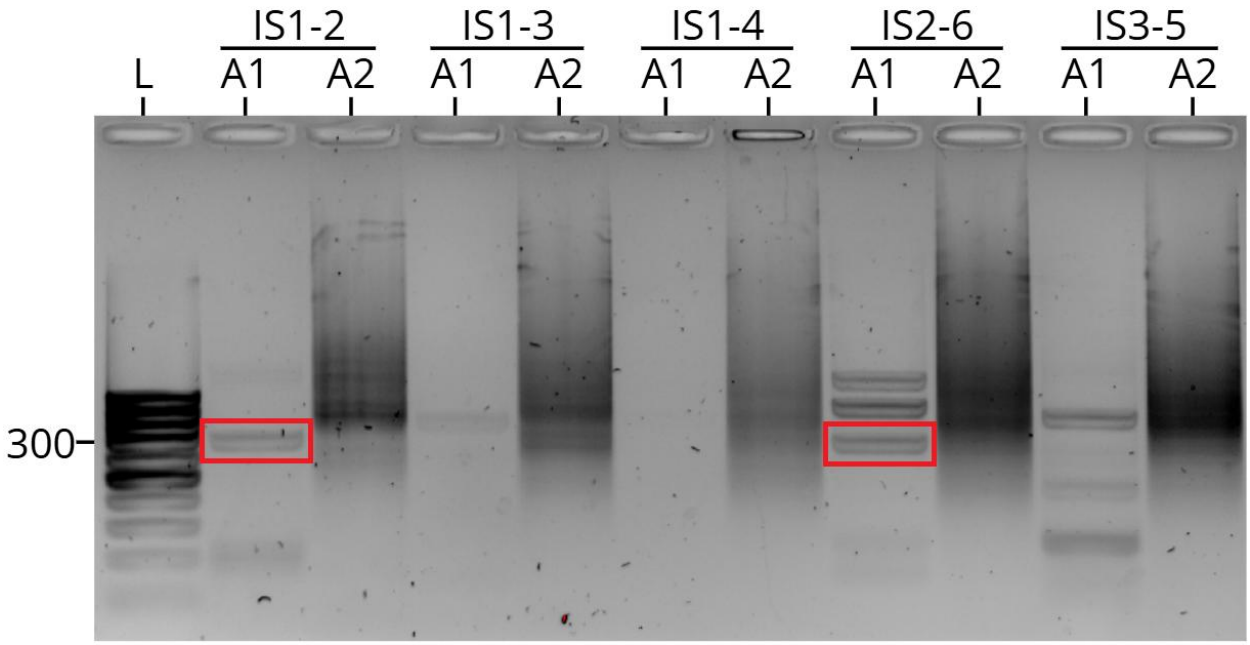

Figure 8. Agarose gel electrophoresis (1,5% agarose) of 5' RACE products for the WuMV-6 hyp1 segment. L – GeneRuler Low Range DNA Ladder (Thermo Fisher Scientific); A1 – adapter primer 1; A2 – adapter primer 2. Red rectangles indicate excised bands submitted for Nanopore sequencing.

To improve product yield and specificity over the initial RACE analysis, a modified linear PCR-based RACE protocol was employed. In this approach, biotinylated variants of the same SSP sequences used in the preceding RACE analysis were used for linear PCR amplification, with biotin incorporated at the 5' end of each primer. The resulting single-stranded biotinylated products were subsequently purified using streptavidin-coated magnetic beads, which selectively retain biotin-bearing molecules. The intention was to eliminate non-specific amplification products lacking the biotin tag. First-strand cDNA synthesis was then performed on the bead-purified product so the same SSP and adapter primer A1 combinations as in the initial RACE analysis could be used for subsequent PCR amplification.

Agarose gel electrophoresis of representative RACE products from eight positive samples (IS1-2, IS1-3, IS1-4, IS2-6, IS3-5, IS3-6, IS4-4, IS4-8) across all four RACE combinations is shown in Figure 9. hyp1 3' RACE (expected product ~795 bp) yielded no detectable amplification products. hyp1 5' RACE (expected ~297 bp) and NP 3' RACE (expected ~328 bp) produced amplification products, with notably improved yield compared to the initial RACE analysis. However, non-specific amplification was not reduced. Both high molecular weight smearing and discrete bands of unexpected size were observed alongside bands of the

expected size. NP 5' RACE (expected ~1727 bp) yielded no detectable products. As in the initial RACE analysis, bands of the expected size were excised from the gel and submitted for sequencing. The excised bands are marked by red rectangles in the gel image (Fig. 9).

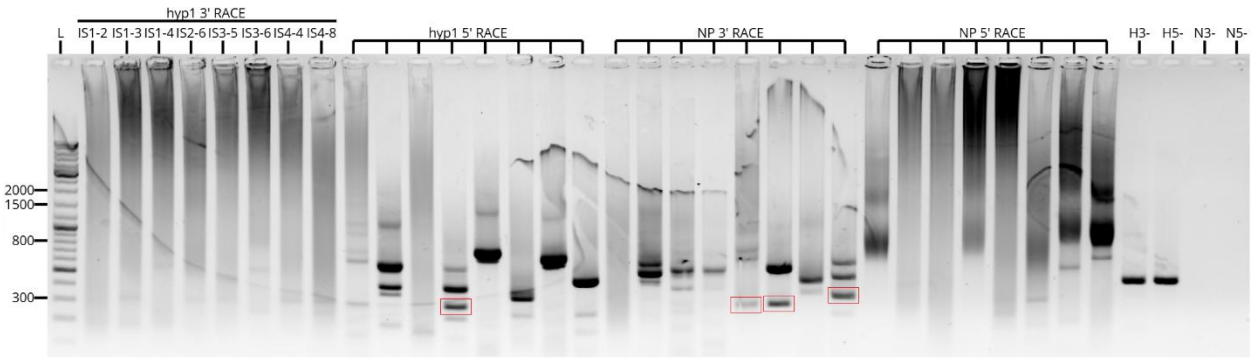

Figure 9. Agarose gel electrophoresis (1% agarose) of linear PCR-based RACE products for the WuMV-6 hyp1 and NP segments. L – GeneRuler DNA Ladder Mix (Thermo Fisher Scientific); H3-, H5-, N3-, N5-, no-template controls for hyp1 3', hyp1 5', NP 3' and NP 5' RACE combinations, respectively. Red rectangles indicate excised bands submitted for Nanopore sequencing.

Critically, amplification was detected in the no-template controls for the hyp1 3' and hyp1 5' combinations (H3- and H5-, respectively), indicating that the non-specific products observed in these reactions are not exclusively derived from the sample template and may reflect primer self-amplification or low-level contamination. No amplification was observed in the NP no-template controls (N3-, N5-) (Fig. 9).

4. Recovery and phylogenetic placement of WuMV-6 sequences

Gel excision and sequencing of RACE products of the expected size was undertaken to identify genuine WuMV-6-derived sequences spanning the terminal UTRs, and to compare any recovered UTR sequences with those of other orthomyxoviruses. A total of 57 excised RACE products were submitted for Nanopore sequencing. Despite many samples yielding especially low (<1 ng/µL) or undetectable DNA concentrations, all 57 were sequenced successfully (Table 2, APPENDICES). This demonstrates the sensitivity of Nanopore sequencing even at minimal template input.

However, mapping of the resulting sequences to the WuMV-6 reference genome revealed that only 4 of the 57 sequenced products showed homology to WuMV-6: samples 37_N3, 59_H5, IS4-10_H3 and IS4-11_H3 (Table 2, APPENDICES). The remaining sequences were

either homologous to the *C. pipiens* genome or were bacterial sequences. Given that only four sequences with WuMV-6 homology were recovered, and that sequencing depth across these sequences was insufficient to support reliable consensus calling, the data were insufficient to support robust UTR analysis alone. Nevertheless, the four recovered sequences were incorporated into nucleotide-based phylogenetic analysis of the WuMV-6 hyp1 and NP segments alongside sequences compiled in a previous study (Kazlauskaitė, 2024). It was done to assess the phylogenetic placement of the newly obtained Lithuanian WuMV-6 sequences within the global WuMV-6 diversity.

The four WuMV-6 sequences recovered through RACE and subsequent Nanopore sequencing were assigned standardised identifiers following the scheme WuMV-6/LT-<county code>-<municipality code>-<sample number>/<date> and incorporated into phylogenetic analysis, extending that performed in a previous study (Kazlauskaitė, 2024). Two separate phylogenetic trees were constructed, one for the hyp1 and one for the NP segment. Each sequence was included only in the tree corresponding to its respective segment.

In the hyp1 tree, WuMV-6/LT/KL-KLP-10-1/2024-09-07 clustered most closely with sequences from Sweden, while WuMV-6/LT/KN-PRN-16/2023-08-08 and WuMV-6/LT/KN-PRN-17/2023-08-08 were most closely related to sequences from Belgium (Fig. 10). In the NP tree, WuMV-6/LT/KL-KLP-8-1/2024-09-07 clustered closest to sequences from both Sweden and Belgium (Fig. 11). Despite these within-clade differences, all Lithuanian sequences fell within the Baltic/Northern European lineage of WuMV-6, comprising sequences from Sweden, Belgium, the Netherlands and Great Britain, and were distantly related to the global lineage.

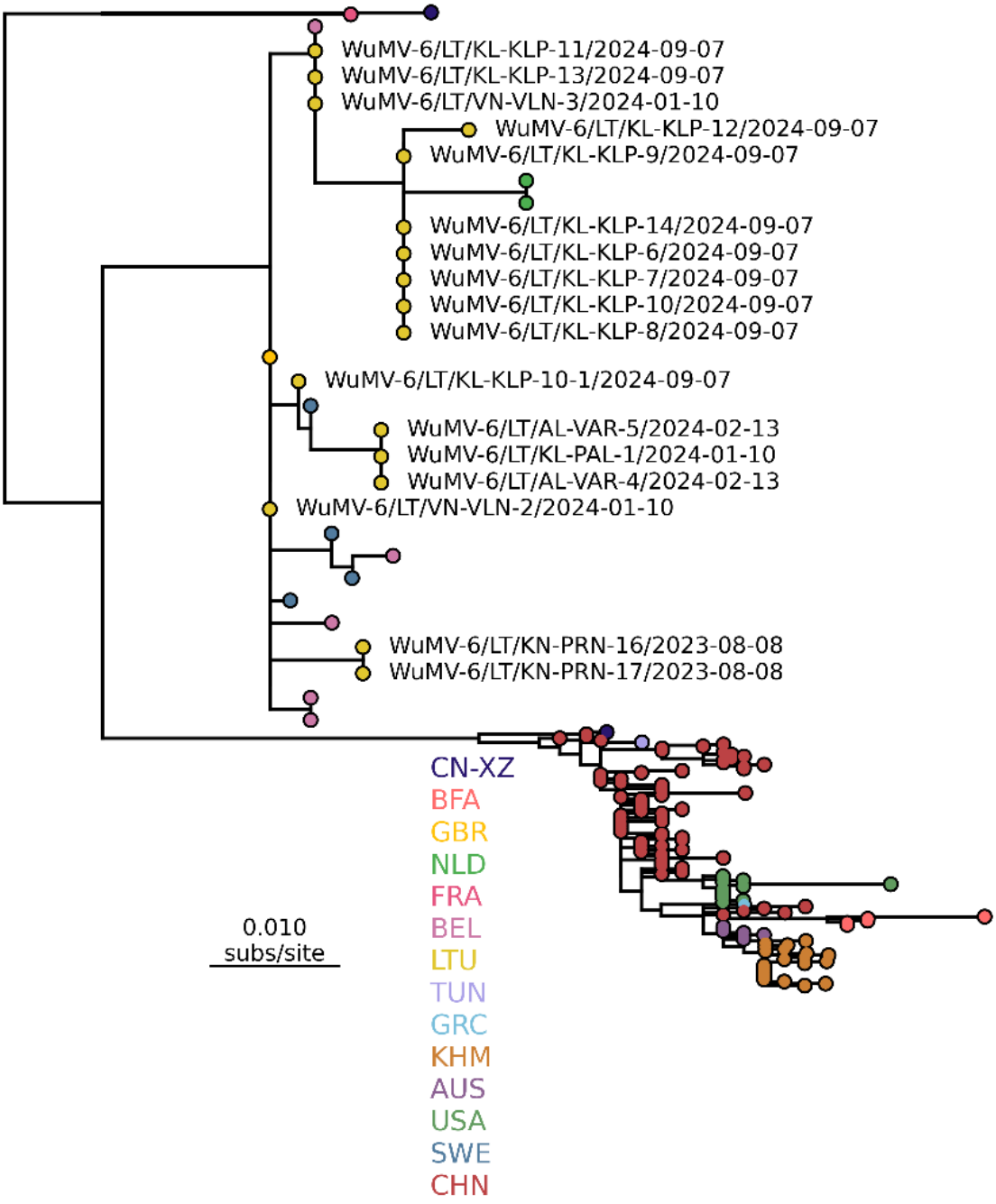

Figure 10. Maximum likelihood phylogenetic tree of WuMV-6 hyp1 sequences. CN-XZ – Tibet, BFA – Burkina Faso, GBR – Great Britain, NLD – Netherlands, FRA – France, BEL – Belgium, LTU – Lithuania, TUN – Tunisia, GRC – Greece, KHM – Cambodia, AUS – Australia, USA – United States of America, SWE – Sweden, CHN – China.

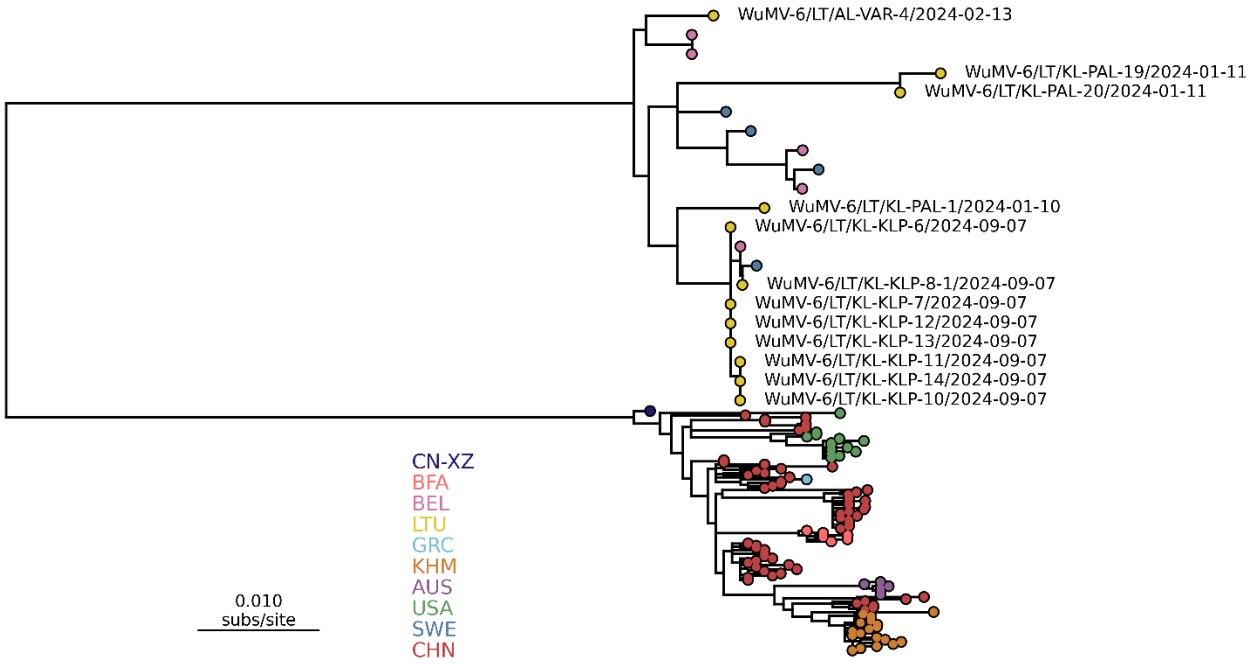

Figure 11. Maximum likelihood phylogenetic tree of WuMV-6 NP segment sequences. CN-XZ – Tibet, BFA – Burkina Faso, BEL – Belgium, LTU – Lithuania, GRC – Greece, KHM – Cambodia, AUS – Australia, USA – United States of America, SWE – Sweden, CHN – China.

In addition to phylogenetic placement, despite the limited number of recovered sequences and low sequencing depth, an attempt was made to examine the terminal UTR regions of WuMV-6 in comparison with those of other orthomyxoviruses. Two figures were compiled, one depicting the conserved terminal nucleotide positions across orthomyxoviruses and one depicting the full available UTR sequences. Viral ordering in both was determined by a PB1 amino acid maximum likelihood phylogenetic tree, placing WuMV-6 among other quaranjavirids, clustering most closely with Wellfleet Bay virus and Lake Chad virus (Fig. 12; Fig. 1, APPENDICES).

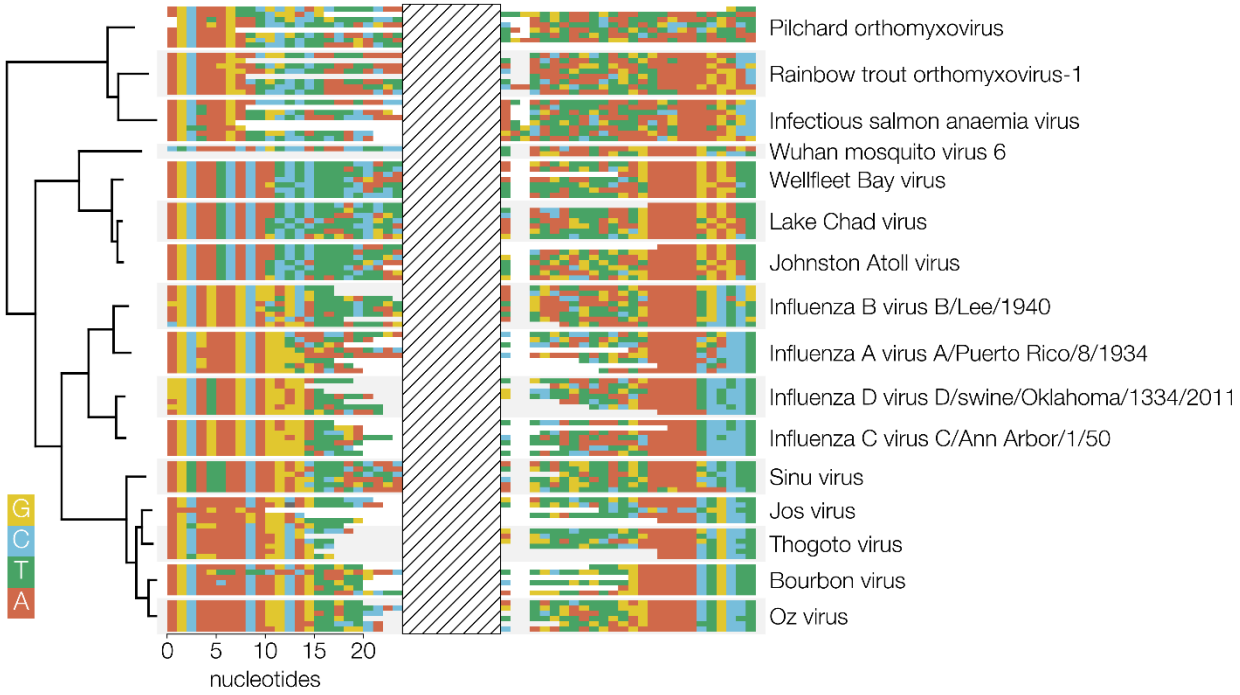

Figure 12. Nucleotide composition of conserved terminal UTR sequences across orthomyxoviruses including WuMV-6. Viral ordering is determined by a PB1 amino acid maximum likelihood phylogenetic tree. All known genome segments are shown (except for WuMV-6) arranged in the following order from top to bottom: PB1, PB2, PA, NP, HA/HE/HEF, GP/GP64, F, UNKNOWN (the surface protein of Rainbow trout orthomyxovirus-1 labelled as hemagglutinin that bears no recognisable homology to any other protein on public sequence databases), NA, M/hyp1, NS, VP7. The hatched region represents the coding region. G (yellow) – guanine; C (blue) – cytosine; T (green) – thymine; A (orange) – adenine.

Examination of the conserved terminal UTR sequences revealed notable differences between WuMV-6 and other orthomyxoviruses. The 5' conserved terminal sequence of WuMV-6 appears markedly different from those of other orthomyxoviruses, notably lacking the AG dinucleotide that typically initiates the 5' UTR across the family. In contrast, the 3' conserved terminal sequence of WuMV-6 showed greater similarity to other quaranjavirids, even retaining the CT dinucleotide ending characteristic of orthomyxovirus 3' UTRs.

Broader examination of the full UTR sequences revealed that beyond the conserved terminal positions, WuMV-6 appears to exhibit additional nucleotide sequence extending past the expected 3' UTR boundary (Fig. 1, APPENDICES). Although, the recovered WuMV-6 UTR sequences were comparable in length to other orthomyxoviruses. The apparent 5' UTR spanning approximately 38 nucleotides and the apparent 3' UTR approximately 153 nucleotides, the latter

being slightly longer than in several other members of the family. It must be noted that all observations described above are based on only four sequences recovered from the entire sample set of 86 pools. Furthermore, the expected RACE product sizes were calculated assuming approximately 20 nucleotides of UTR sequence. However, the actual UTR lengths of WuMV-6 appear considerably longer. Meaning that the true specific RACE products would have migrated at a larger size than anticipated and were therefore not selected for sequencing.

DISCUSSION

The present study aimed to extend the surveillance of WuMV-6 in Lithuanian *C. pipiens* mosquito populations and, for the first time, to characterise the terminal UTR sequences of the WuMV-6 genome through RACE analysis. To the best of our knowledge, targeted RACE-based UTR characterisation of WuMV-6 has not previously been attempted, making this the first such effort for this virus. While the detection of WuMV-6 was successful and yielded results consistent with previous findings, the UTR characterisation was substantially limited by technical challenges that are discussed below.

The consistently lower Ct values observed for the hyp1 relative to NP segment in paired analysis are unlikely to reflect differences in PCR amplification efficiency, given that the amplified fragments for both targets were of similar length (Table 2.1, METHODS). The more probable explanation is that hyp1 RNA is present at higher abundance in infected mosquitoes. Given that total RNA was used as input, the measured signal reflects the overall RNA pool of the sample, meaning that genuine differences in segment abundance would be captured directly. This abundance difference likely arises from a combination of higher genomic copy number and potentially higher transcriptional activity of the hyp1 relative to NP segment. Transcriptional differences being the more probable explanation. Differential transcription of individual genome segments has been described in segmented RNA viruses, where polymerase activity and promoter strength can vary between segments, resulting in unequal RNA accumulation (Widjaja *et al.*, 2012). Whether this reflects a genuine biological feature of WuMV-6 replication or a technical artefact warrants further investigation.

WuMV-6 was detected across 12 of 25 sampling localities throughout Lithuania, spanning coastal wetland habitats in the west, agricultural lowlands in the central and southern regions, and forested areas in the north and northeast. This broad geographic and ecological distribution is consistent with and extends the earlier finding that WuMV-6 circulates across three distinct Lithuanian regions (Kazlauskaitė, 2024). This is also consistent with the characterisation of WuMV-6 as a geographically widespread virus with high genetic variability and rapid migration potential. Previous studies have shown that WuMV-6 strains found hundreds to thousands of kilometres apart can be closely genetically related, suggesting efficient dispersal across large geographic scales (Dudas & Batson, 2023). The particularly high positivity rates at western coastal

sites are notable given that prior surveillance identified these habitats, including the Palanga area, as supporting exceptionally dense *C. pipiens* populations, conditions that may facilitate sustained virus amplification (Kazlauskaitė, 2024). Taken together, the spatial distribution of positive detections supports the conclusion that WuMV-6 is broadly established in *C. pipiens* populations throughout Lithuania rather than confined to a single focal area. The maximum likelihood prevalence estimate of 1,45% (95% CI: 1,03 - 1,98%) mirrors an independently derived winter estimate of approximately 1,5% from the same year (Dr. G. Dudas, pers. comm.). Meaning that individual-level WuMV-6 prevalence in Lithuanian *C. pipiens* remains consistent across seasons. Together, these findings indicate that while WuMV-6 is broadly established across Lithuanian *C. pipiens* populations, individual-level prevalence remains relatively low.

The RACE analysis encountered substantial technical challenges that warrant detailed discussion. The high molecular weight smearing observed in the linear PCR-based RACE products likely reflects partial or complete degradation of DNA fragments during the high-temperature (95°C) elution step of the bead purification procedure. Elevated temperatures accelerate DNA strand breakage disproportionately affecting longer molecules. Consistent with this, a clear size-dependent pattern emerged. The two combinations yielding no detectable products, hyp1 3' RACE (expected ~795 bp) and NP 5' RACE (expected ~1727 bp), had considerably larger expected product sizes than the two that produced amplification, hyp1 5' RACE (expected ~297 bp) and NP 3' RACE (expected ~328 bp). The distinct non-specific bands are most likely attributable to incomplete purification. While streptavidin exhibits extremely high affinity for biotin, the beads themselves, along with the biotin-binding proteins attached to them, can cause non-specific binding of host-derived DNA, which is subsequently carried through to the PCR step.

Retrospective examination of the adapter primer A1 sequence revealed internal repetition that was initially overlooked during adapter design. Dot plot analysis of the A1 sequence against itself demonstrated that two hexameric sequences, 5'-CGTAGC-3' (positions 1-6 and 10-15) and 5'-TAGCGT-3' (positions 3-8 and 16-21), each occur twice within the primer, along with reverse complement matches, rendering it capable of folding back on itself and self-priming (Fig. 14). This provides a mechanistic basis for the spurious amplification products observed in the no-

template controls and highlights the importance of screening adapter sequences for internal repeats during RACE design, in addition to host and viral genome homology.

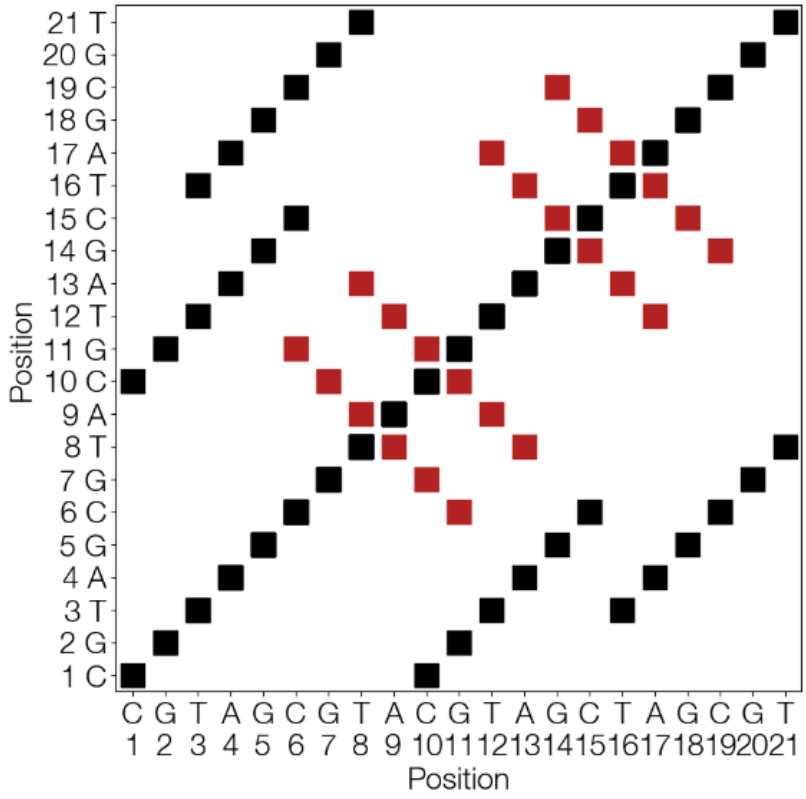

Figure 14. Dot plot of the adapter primer A1 sequence aligned against itself, revealing internal sequence repetition. Diagonal lines off the main axis indicate regions of self-complementarity within the primer. Black squares represent forward sequence matches, while red squares represent reverse complement matches. Programmatically generated using Python (v3.12) with matplotlib (v3.10.9).

The presence of bacterial-derived sequences among the sequenced products is consistent with the use of total RNA as input material, which would include nucleic acids from mosquito-associated gut microbiota alongside viral RNA (Abel *et al.*, 2023). Under these conditions, the adapter and segment-specific primers may have non-specifically primed bacterial templates present in the extract, resulting in amplification of non-target sequences. More broadly, the low proportion of genuine WuMV-6-derived products reflects the cumulative effect of non-specific amplification throughout the RACE workflow. Which is a consequence of the approach being functionally analogous to metatranscriptomic methods in its indiscriminate capture of all RNA species present in the sample (Batson *et al.*, 2021). An additional limitation specific to UTR characterisation is that random hexamers employed for cDNA synthesis may fail to prime

efficiently at the very termini of RNA molecules, particularly where secondary structures are present. Meaning potentially resulting in loss of terminal UTR sequence prior to amplification.

RACE methodology encompasses various approaches for UTR sequence recovery. Typically, it employs distinct strategies for 5′ and 3′ ends. For 3′ RACE, the original method involves reverse transcription of mRNA using an oligo-dT adapter primer that contains both a poly(dT) sequence complementary to the mRNA poly(A) tail and a unique adapter sequence at its 5′ end. Following cDNA synthesis, PCR amplification is performed using a gene-specific primer that anneals to a known internal sequence and an adapter primer that targets the unique adapter sequence, thus capturing 3′ UTR sequences between the gene-specific primer site and the poly(A) tail. For 5′ RACE, first-strand cDNA synthesis is primed using a gene-specific antisense oligonucleotide to maximize extension to the 5′ end of the target mRNA. Following purification to remove unincorporated nucleotides and primers, terminal deoxynucleotidyl transferase (TdT) is used to add homopolymeric tails (typically dC) to the 3′ ends of the cDNA. The tailed cDNA is then amplified by PCR using a second gene-specific primer nested within the first and an abridged anchor primer complementary to the homopolymeric tail, enabling capture of 5′ UTR sequences (Frohman *et al.*, 1988). The approach employed in this study utilized a unified methodology for both termini: A-tailing of double-stranded cDNA followed by adapter ligation and PCR amplification. This differs from standard RACE protocols that employ terminus-specific strategies and require different primer sets for 5′ versus 3′ amplification. WuMV-6 genomic RNA, like other orthomyxoviruses, lacks poly(A) tails, making poly(A)-dependent 3′ RACE methods unsuitable for these targets. Additionally, TdT-based 5′ RACE would likely have increased non-specific amplification in the total RNA samples used in this study, as homopolymeric sequences promote non-specific priming in complex templates (Frohman, 1994).

Despite these limitations, all four recovered sequences clustered within the Baltic/Northern European lineage of WuMV-6, consistent with and extending the finding of the previous study (Kazlauskaitė, 2024). The inclusion of the Netherlands and Great Britain in the Baltic/Northern European lineage reflects the accumulation of additional sequences in the dataset since the previous analysis. Which further supports the existence of a distinct European WuMV-6 clade with limited connectivity to the global lineage (Dudas & Batson, 2023). These phylogenetic findings reinforce the conclusion that WuMV-6 detected in Lithuania originates from active

infections rather than endogenous viral elements. The consistent clustering of both hyp1 and NP segment within the same lineage makes independent integration of both segments into the host genome as endogenous viral elements highly unlikely (Kazlauskaitė, 2024).

The comparative UTR analysis, while necessarily preliminary, yielded observations of potential biological interest. The retention of the CT dinucleotide at the 3' terminus, characteristic of orthomyxovirus 3' UTRs, suggests that this terminal feature is conserved even in highly divergent orthomyxovirus lineages (Desselberger *et al.*, 1980). The apparent absence of the AG dinucleotide typically initiating the 5' UTR must be interpreted with caution. Given the low sequencing depth of the recovered sequences, it cannot be excluded that this reflects incomplete or artefactual sequence rather than a genuine biological feature of the WuMV-6 5' UTR. Similarly, the additional nucleotide sequence observed extending beyond the expected 3' UTR boundary most likely reflects artefactual sequence captured beyond the true RNA terminus during RACE rather than a genuine biological feature. Additionally, the expected RACE product sizes were calculated assuming approximately 20 nucleotides of UTR sequence, whereas the actual UTR lengths appear to be considerably longer. Approximately 38 nucleotides for the 5' UTR and 153 nucleotides for the 3' UTR. Consequently, the true specific RACE products would have migrated at larger sizes than anticipated and were therefore not selected during gel excision, further limiting confidence in the recovered sequences.

Future work should prioritise accurate characterisation of the WuMV-6 terminal UTR sequences, as these are essential for the development of molecular tools that would enable detailed functional studies of this virus. Should RACE be attempted again, several methodological improvements should be considered. Firstly, virus-specific rather than random hexamers for cDNA synthesis to restrict reverse transcription to viral sequences. Secondly, rigorous screening of adapter sequences for internal repeats and host genome homology. And finally, accurate estimation of expected product sizes informed by the UTR length data obtained in the present study, which indicates that products considerably larger than initially anticipated should be targeted during gel excision. Alternative RNA-based approaches, such as direct RNA sequencing or targeted enrichment methods, may circumvent the non-specificity limitations encountered here and should be considered as complementary strategies. Once UTR sequences are reliably established, they would enable not only the design of universal primers for full genome amplification, as has been

achieved for other orthomyxoviruses (Inoue *et al.*, 2010; Zhou *et al.*, 2009; Zhou *et al.*, 2014), but also the development of reverse genetics systems (Neumann *et al.*, 1999; Hoffmann *et al.*, 2000; Toro-Ascuy *et al.*, 2015) and mini-genome assays (Zhu *et al.*, 2011; Lutz *et al.*, 2005) for WuMV-6. Which would open the door to functional characterisation of viral proteins, replication mechanisms, and host interactions that currently remain inaccessible.

CONCLUSIONS

1. Complete determination of the exact 5′ and 3′ untranslated regions for all eight genome segments of Wuhan Mosquito Virus 6 was not achieved, with only 4 authentic viral sequences recovered from 57 samples processed, representing only two genome segments (hypothetical segment 1 and nucleoprotein segment). Non-specific primer binding is considered the primary limiting factor, with insufficient sequencing depth further preventing comprehensive characterization.
2. Comparative analysis of Wuhan Mosquito Virus 6 untranslated regions with related orthomyxoviruses could not be completed due to insufficient sequence data from only 4 recovered sequences of poor quality, though preliminary observations suggested the absence of the typical AG dinucleotide at the 5′ untranslated region terminus and retention of the CT dinucleotide at the 3′ terminus.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Gytis Dudas for the opportunity to be a part of this project, for his guidance throughout, and for his trust in my abilities. I am grateful to Monika Repšytė and Dr. Milda Norkienė for their guidance in laboratory work, and to Joris Balčiūnas for his support in experimental design. I thank Philippe Selhorst for the creation of the qPCR primers and Linas Gasparavičius for compiling the available WuMV-6 genome data.

I would also like to thank Aušrinė Žvirblytė for being an engaged and dedicated student and for her invaluable work in the laboratory. I am grateful to Matas Mikulskis and Bushran Noor Mohammed Iqbal for stepping in and being an extra pair of hands when it was needed most. Emre Mert Asar and Zeynep Özlem Cinar deserve a special thanks for always being there with a reassuring second opinion and for their moral support throughout this work.

Finally, I am grateful to Dr. Rasa Petraitytė-Burneikienė for the opportunity to work in the Department of Eukaryote Gene Engineering, and to the entire department team for their warm welcome and collegial atmosphere.

Claude (Anthropic, 2026) was used to assist with language editing, grammar, and syntax proofreading of the manuscript.

REFERENCES

1. Abel, S.M., Hong, Z., Williams, D., Ireri, S., Brown, M.Q., Su, T., Hung, K.Y., Henke, J.A., Barton, J.P. and Le Roch, K.G. (2023). Small RNA sequencing of field *Culex* mosquitoes identifies patterns of viral infection and the mosquito immune response. *Scientific Reports*, 13(1), 10598.
2. Al-Khalifa, M. S., Diab, F. M., & Khalil, G. M. (2007). Man-threatening viruses isolated from ticks in Saudi Arabia. *Saudi Medical Journal*, 28(12), 1864-1867.
3. Allison, A.B., Ballard, J.R., Tesh, R.B., Brown, J.D., Ruder, M.G., Keel, M.K., Munk, B.A., Mickley, R.M., Gibbs, S.E.J. and Travassos da Rosa, A.P.A. (2015). Cyclic avian mass mortality in the northeastern United States is associated with a novel orthomyxovirus. *Journal of Virology*, 89(13), 6822-6832.
4. Andrewes, C.H., Bang, F.B. and Burnet, F.M. (1955). A short description of the *Myxovirus* group (influenza and related viruses). Virology, 1(2), 176-184.
5. Anthropic (2026). *Claude* (claude.ai) [Large language model]. Available at: https://claude.ai (Accessed: 15 May 2026).
6. Aspehaug, V., Mikalsen, A.B., Snow, M., Biering, E. and Villoing, S. (2005). Characterization of the infectious salmon anemia virus fusion protein. *Journal of Virology*, 79(12), 7676-7687.
7. Bamunuarachchi, G., Najera, F., Aziati, I.D., Palmer, J.L., Towner, J.S., Replogle, A.J., Bosco-Lauth, A., Root, J.J. and Balasuriya, U.B.R. (2024). Serosurveillance and the first detection of Bourbon virus RNA in a wildlife host. *bioRxiv*.
8. Batson, J., Dudas, G., Haas-Stapleton, E., Kistler, A.L., Li, L.M., Logan, P., Ratnasiri, K. and Retallack, H. (2021). Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. *eLife*, 10, e68353.
9. Bennouna, A., Tantely, M.L., Raharinosy, V., Andriamandimby, S.F., Bigot, T., Chrétien, D., Jacquemet, E., Volant, S., Temmam, S., Dussart, P. and Lacoste, V. (2023). Comprehensive characterization of viral diversity of female mosquitoes in Madagascar. *Viruses*, 15(9), 1852.
10. Birnberg, L., Temmam, S., Aranda, C., Correa-Fiz, F., Talavera, S., Bigot, T., Eloit, M. and Busquets, N. (2020). Viromics on honey-baited FTA cards as a new tool for the detection of circulating viruses in mosquitoes. *Viruses*, 12(3), 274.
11. Biswas, S. K., & Nayak, D. P. (1994). Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *Journal of Virology*, 68(3), 1819-1826.
12. Bouvier, N.M. and Palese, P. (2008). The biology of influenza viruses. *Vaccine*, 26, D49-D53.
13. Cheong, H.K., Cheong, C. and Choi, B.S. (1996). Secondary structure of the panhandle RNA of influenza virus A studied by NMR spectroscopy. *Nucleic Acids Research*, 24(22), 4197-4201.
14. Cholleti, H., Hayer, J., Mulandane, F.C., Falk, K., Fafetine, J., Berg, M. and Blomström, A.L. (2018). Viral metagenomics reveals the presence of highly divergent quaranjavirus in Rhipicephalus ticks from Mozambique. *Infection Ecology & Epidemiology*, 8(1), 1556140.
15. Converse, J. D., & Moussa, M. I. (1982). Quaranfil virus from *Hyalomma dromedarii* (*Acari: Ixodoidea*) collected in Kuwait, Iraq and Yemen. *Journal of Medical Entomology*, 19(2), 209-210.
16. Cook, J.D., Sultana, A. and Lee, J.E. (2017). Structure of the infectious salmon anemia virus receptor complex illustrates a unique binding strategy for attachment. *Proceedings of the National Academy of Sciences*, 114(14), E2929-E2936.
17. Crescenzo-Chaigne, B. and van der Werf, S. (2007). Rescue of influenza C virus from recombinant DNA. *Journal of Virology*, 81(20), 11282-11289.
18. Crescenzo-Chaigne, B., Barbezange, C. and van der Werf, S. (2008). Non-coding extremities of the seven influenza virus type C vRNA segments: Effect on transcription and replication by the type C and A polymerase complexes. *Virology Journal*, 5, 132.
19. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008.
20. Desselberger, U., Racaniello, V.R., Zazra, J.J. and Palese, P. (1980). The 3′ and 5′-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene*, 8(3), 315-328.
21. Dias, A., Bouvier, D., Crépin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Ruigrok, R. W. H., & Cusack, S. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458(7240), 914-918.

22. Dubois, J., Terrier, O. and Rosa-Calatrava, M. (2014). Influenza viruses and mRNA splicing: Doing more with less. *mBio*, 5(3), e00070-14.
23. Dudas, G. and Batson, J. (2023). Accumulated metagenomic studies reveal recent migration, whole-genome evolution, and undiscovered diversity of orthomyxoviruses. *Journal of Virology*, 97(10), e01056-23.
24. Eisfeld, A.J., Neumann, G. and Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. *Nature Reviews Microbiology*, 13(1), 28-41.
25. Falk, K., Namork, E., Rimstad, E., Mjaaland, S. and Dannevig, B.H. (1997). Characterization of infectious salmon anemia virus, an orthomyxo-like virus isolated from Atlantic salmon (*Salmo salar L.*). *Journal of Virology*, 71(12), 9016-9023.
26. Ferguson, L., Olivier, A. K., Genova, S., Epperson, W. B., Smith, D. R., Schneider, L., Barton, K., McCuan, K., Webby, R. J., & Wan, X. F. (2016). Pathogenesis of influenza D virus in cattle. *Journal of Virology*, 90(12), 5636-5642.
27. Flick, R., Neumann, G., Hoffmann, E., Neumeier, E. and Hobom, G. (1996). Promoter elements in the influenza vRNA terminal structure. *RNA*, 2(10), 1046-1057.
28. Fodor, E. (2013). The RNA polymerase of influenza A virus: mechanisms of viral transcription and replication. *Acta Virologica*, 57(2), 113-122.
29. Fodor, E., Pritlove, D.C. and Brownlee, G.G. (1995). Characterization of the RNA-fork model of virion RNA in the initiation of transcription in influenza A virus. *Journal of Virology*, 69(7), 4012-4019.
30. Frohman, M. A. (1994). On beyond classic RACE (rapid amplification of cDNA ends). *Genome Research*, 4(1), S40-S58.
31. Frohman, M. A., Dush, M. K., & Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), 8998-9002.
32. Furuse, Y. and Oshitani, H. (2011). Evolution of the influenza A virus untranslated regions. *Infection, Genetics and Evolution*, 11(5), 1150-1154.
33. Garry, C. E., & Garry, R. F. (2008). Proteomics computational analyses suggest that baculovirus GP64 superfamily proteins are class III penetrenes. *Virology Journal*, 5, 28.
34. Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W. H., & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology*, 15(5), 500-506.
35. Hao, S., Yang, D., Wang, S., Li, C., Du, M., Zhao, X., Duan, Q. and Zhang, Y. (2022). Eight years of research advances in Bourbon virus, a tick-borne thogotovirus. *Zoonoses*, 2(1), 1-11.
36. Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160-174.
37. Hause, B.M., Ducatez, M., Collin, E.A., Ran, Z., Liu, R., Sheng, Z., Chen, A., Hinckley, S., Searby, R., Spradbrow, P.B., Zhang, Y., Webby, R.J., Simonson, R.R. and Li, F. (2013). Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *Proceedings of the National Academy of Sciences*, 110(43), E4126-E4134.
38. Herrler, G., Dürkop, I., Becht, H., & Klenk, H. D. (1988). The glycoprotein of influenza C virus is the haemagglutinin, esterase and fusion factor. *Journal of General Virology*, 69(4), 839-846.
39. Hoffmann, E., Krauss, S., Perez, D., Webby, R. and Webster, R.G. (2002). Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine*, 20(25-26), 3165-3170.
40. Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. and Webster, R.G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences*, 97(11), 6108-6113.
41. Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M. and Palese, P. (1987). Genomic RNAs of influenza viruses are held in a circular conformation in virions and infected cells by a terminal panhandle. *Proceedings of the National Academy of Sciences*, 84(22), 8140-8144.
42. Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
43. Hutchinson, E.C., von Kirchbach, J.C., Gog, J.R. and Digard, P. (2010). Genome packaging in influenza A virus. *Journal of General Virology*, 91(2), 313-328.
44. ICTV Online Report, 2011. Virus taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Academic Press. (Accessed: 4 December 2025)
45. ICTV Online Report, 2024. Orthomyxoviridae. (Accessed: 4 December 2025)

46. Inoue, E., Wang, X., Osawa, Y. and Maeda, J. (2010). Full genomic amplification and subtyping of influenza A virus using a single set of universal primers. *Microbiology and Immunology*, 54(8), 389-392.
47. Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, 33(2), 511-518.
48. Kazlauskaitė, M. (2024). *Uhano uodų viruso 6 paieškos Lietuvoje* [Bachelor's thesis]. Vilnius University.
49. Kemp, G. E., Lee, V. H., & Moore, D. L. (1975). Isolation of Nyamanini and Quaranfil viruses from *Argas* (*Persicargas*) *arboreus* ticks in Nigeria. *Journal of Medical Entomology*, 12(5), 535-537.
50. Kibenge, F.S., Munir, K., Kibenge, M.J., Joseph, T. and Moneke, E. (2004). Infectious salmon anemia virus: causative agent, pathogenesis and immunity. *Animal Health Research Reviews*, 5(1), 65-78.
51. Kibenge, F.S.B. and Godoy, M. (2016). Orthomyxoviruses of fish. *Veterinary Research*, 47(1), 110.
52. Kosoy, O.I., Lambert, A.J., Hawkinson, D.J., Pastula, D.M., Goldsmith, C.S., Hunt, D.C., Staples, J.E. and the CDC Bourbon virus working group (2015). Novel thogotovirus associated with febrile illness and death, United States, 2014. *Emerging Infectious Diseases*, 21(5), 760-764.
53. Kulshreshtha, V., Kibenge, M., Salonius, K., Simard, N., Riveroll, A. and Kibenge, F. (2010). Identification of the 3′ and 5′ terminal sequences of the eight RNA genome segments of European and North American genotypes of infectious salmon anaemia virus (an orthomyxovirus) and evidence for quasispecies based on the non-coding sequences of transcripts. *Virology Journal*, 7(1), 338.
54. Lamb, R.A. and Takeda, M. (2001). Death by influenza virus protein. *Nature Medicine*, 7(12), 1286-1288.
55. Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307-1320.
56. Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C. and Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife*, 4, e05378.
57. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100.
58. Li, X. and Palese, P. (1992). Mutational analysis of the promoter required for influenza virus virion RNA synthesis. *Journal of Virology*, 66(7), 4331-4338.
59. Li, X. and Palese, P. (1994). Characterization of the polyadenylation signal of influenza virus RNA. *Journal of Virology*, 68(2), 1245-1249.
60. Licheri, S., Smith, H., Manso, C.F., Amato, R., Aleman, N., Haldenby, S., Kwiatkowski, D., Tyson, J., Baker, D., Bryer, A., Robinson, D.G. and Wruck, W. (2025). Nanopore sequencing for real-time influenza surveillance. *Nature Microbiology*, 1-15.
61. Liu, X., Kong, J., Shan, Y., Yang, Z., Miao, J., Pan, Y., Luo, T., Shi, Z., Wang, Y., Gou, Q., Yang, C., Li, H., Li, C., Li, S., Zhang, X., Sun, Y., Holmes, E. C., Guo, D., & Shi, M. (2025). SegFinder: an automated tool for identifying complete RNA virus genome segments through co-occurrence in multiple sequenced samples. *Briefings in Bioinformatics*, 26(4), bbaf358.
62. Lutz, A., Dyall, J., Olivo, P.D. and Pekosz, A. (2005). Virus-inducible reporter genes as a tool for detecting and quantifying influenza A virus replication. *Journal of Virological Methods*, 126(1-2), 13-20.
63. Ma, J., Shen, H., Liu, Q., Bawa, B., Qi, T., Duff, M., Mu, J., Chen, J., Qin, C., Wang, M., Shu, Y., Jiang, L. and Gao, G.F. (2013). Influenza A virus nucleoprotein C-terminal domain is involved in the recruitment of transcriptional machinery. *Journal of Virology*, 87(15), 8302-8310.
64. Marsh, G.A., Hatami, R. and Palese, P. (2008). Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *Journal of Virology*, 82(11), 5413-5422.
65. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10-12.
66. Matsuzaki, Y., Sugawara, K., Takashita, E., Muraki, Y., Hongo, S., Katsushima, N., Mizuta, K. and Nishimura, H. (2006). Genetic diversity of influenza C virus in Japan during the past 10 years. *Journal of Clinical Microbiology*, 44(7), 2522-2528.
67. Müller, K.E., Subramaniam, S., Gregory, L., Stæhr, K., Cunningham, A.A., Lauridsen, J.C., Salman, M.D. and Evensen, Ø. (2010). Phylogenetic analysis of infectious salmon anemia virus isolates from Norway, Faroe Islands and Scotland. *Diseases of Aquatic Organisms*, 90(1), 1-12.
68. Muradrasoli, S., Mohamed, N., Belák, S., Czifra, G., Herrmann, B., Berencsi, G. and Blomberg, J. (2010). Broadly targeted triplex real-time PCR detection of influenza A, B and C viruses based on the nucleoprotein gene and a novel MegaBeacon probe strategy. *Journal of Virological Methods*, 163(2), 313-322.
69. Nadja, B., Thomas, H., Côme, M., Antoni, E., Serafin, G. and Sebastian, L. (2024). A tale of caution: How endogenous viral elements affect virus discovery in transcriptomic data. *Virus Evolution*, 10(1), vead088.

70. Neumann, G., Watanabe, T., Ito, H., Watanabe, S., Goto, H., Gao, P., Hughes, M., Perez, D.R., Donis, R., Hoffmann, E., Hobom, G. and Kawaoka, Y. (1999). Generation of influenza A viruses entirely from cloned cDNAs. *Proceedings of the National Academy of Sciences*, 96(16), 9345-9350.
71. Osterhaus, A. D., Rimmelzwaan, G. F., Martina, B. E., Bestebroer, T. M., & Fouchier, R. A. (2000). Influenza B virus in seals. *Science*, 288(5468), 1051-1053.
72. Papa, G., Abbà, S., Galetto, L., Parise, C., Marzachì, C., & Negri, I. (2023). Distribution and prevalence of viral genomes in Italian populations of the invasive brown marmorated stink bug *Halyomorpha halys*. *Journal of Invertebrate Pathology*, 107977.
73. Peng, R., Zhang, S., Cui, Y., Shi, Y., Gao, G.F. and Qi, J. (2017). Structures of human-infecting Thogotovirus fusogens support a common ancestor with insect baculovirus. *Proceedings of the National Academy of Sciences*, 114(42), E8905-E8912.
74. Petrone, M. E., Parkinson, N., Turnbull, J. D., Mordecai, G. J., Geoghegan, J. L., & Holmes, E. C. (2023). Evidence for an ancient aquatic origin of the RNA viral order Articulavirales. *Proceedings of the National Academy of Sciences*, 120(33), e2310529120.
75. Petrone, M.E., Kelly, C.M., Brierley, L., Shi, M., Blasdell, K.R., Lauber, C. and Holmes, E.C. (2023). Evidence for an ancient aquatic origin of the RNA viral superfamily Articulavirales. *Proceedings of the National Academy of Sciences*, 120(35), e2310529120.
76. Pflug, A., Guilligay, D., Reich, S. and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, 516(7531), 355-360.
77. Plotch, S.J., Bouloy, M., Ulmanen, I. and Krug, R.M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, 23(3), 847-858.
78. Presti, R.M., Zhao, G., Beatty, W.L., Mihindukulasuriya, K.A., da Rosa, A.P., Popov, V.L., Tesh, R.B., Virgin, H.W. and Wang, D. (2009). Quaranfil, Johnston Atoll, and Lake Chad viruses are novel members of the family *Orthomyxoviridae*. *Journal of Virology*, 83(22), 11599-11606.
79. Rimstad, E. and Mjaaland, S. (2002). Infectious salmon anaemia virus. *APMIS*, 110(4), 273-282.
80. Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C. H., Meier-Ewert, H., Skehel, J. J., & Wiley, D. C. (1998). Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature*, 396(6706), 92-96.
81. Sakurai, A., Nomura, N., Nanba, R., Sinkai, T., Iwaki, T., Obuchi, M., Sato, H., Shoham, D., Takemae, N., Tsunekuni, R., Saito, T. and Shibasaki, F. (2011). Rapid typing of influenza viruses using super high-speed quantitative real-time PCR. *Journal of Virological Methods*, 178(1-2), 65-70.
82. Sang, R., Onyango, C., Gachoya, J., Mabinda, E., Konongoi, S., Ofula, V., Dunster, L., Okello, F., Macleod, E., Crabtree, M. and Miller, B. (2006). Tickborne arbovirus surveillance in market livestock, Nairobi, Kenya. *Emerging Infectious Diseases*, 12(7), 1074-1080.
83. Savage, H.M., Burkhalter, K.L., Godsey Jr, M.S., Panella, N.A., Ashley, D.C., Nicholson, W.L., Lambert, A.J. and the CDC Bourbon virus investigation team (2017). Bourbon virus in field-collected ticks, Missouri, USA. *Emerging Infectious Diseases*, 23(12), 2017-2022.
84. Sharma, A., Kumar, A., Aggarwal, M., Nagpal, R. and Vashishtha, M. (2022). A comprehensive roadmap towards the generation of an influenza B reporter assay using a single DNA polymerase-based cloning of the reporter RNA construct. *Frontiers in Microbiology*, 13, 868367.
85. Shi, J., Hu, Z., Deng, F. and Shen, S. (2018). Tick-borne viruses. *Virologica Sinica*, 33(5), 289-304.
86. Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C. and Zhang, Y.Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*, 540(7634), 539-543.
87. Skelton, R.M. and Huber, V.C. (2022). Comparing influenza virus biology for understanding influenza D virus. *Viruses*, 14(5), 1036.
88. Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D., & Fouchier, R. A. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682), 371-376.
89. Snoeck, C.J., Oliva, J., Pauly, M., Losch, S., Wildschütz, F., Muller, C.P., Hübschen, J.M. and Ducatez, M.F. (2018). Influenza D virus circulation in cattle and swine, Luxembourg, 2012-2016. *Emerging Infectious Diseases*, 24(7), 1388-1389.
90. Stoeckle, M.Y., Shaw, M.W. and Choppin, P.W. (1987). Segment-specific and common nucleotide sequences in the noncoding regions of influenza B virus genome RNAs. *Proceedings of the National Academy of Sciences*, 84(9), 2703-2707.

91. Sullivan, S.J., Jacobson, R.M., Dowdle, W.R. and Poland, G.A. (2010). 2009 H1N1 influenza. *Mayo Clinic Proceedings*, 85(1), 64-76.
92. Taylor, R. M., Hurlbut, H. S., Work, T. H., Kingston, J. R., & Hoogstraal, H. (1966). Arboviruses isolated from ARGAS ticks in Egypt: Quaranfil, Chenuda, and Nyamanini. *American Journal of Tropical Medicine and Hygiene*, 15(1), 76-86.
93. Toro-Ascuy, D., Tambley, C., Beltrán, C., Mascayano, C., Sandoval, N., Olivares, E., Medina, R.A., Spencer, E. and Cortez-San Martín, M. (2015). Development of a reverse genetic system for infectious salmon anemia virus: rescue of recombinant fluorescent virus by using salmon internal transcribed spacer region 1 as a novel promoter. *Applied and Environmental Microbiology*, 81(4), 1210-1224.
94. Tulauskaitė, A. (2025). Uhano uodų viruso 6 (WuMV-6) vertikalios pernašos tyrimas *Culex pipiens* uoduose [Bachelor's thesis]. Vilnius University.
95. Weber, F., Bridgen, A., Fazakerley, J.K., Streitenfeld, H., Kessler, N., Randall, R.E. and Elliott, R.M. (2002). Bunyamwera bunyavirus nonstructural protein NSs counteracts the induction of alpha/beta interferon. *Journal of Virology*, 76(15), 7949-7955.
96. Weber, F., Haller, O. and Kochs, G. (1996). Nucleoprotein viral RNA and mRNA of Thogoto virus: a novel cap-stealing mechanism in tick-borne orthomyxoviruses? *Journal of Virology*, 70(12), 8361-8367.
97. Weber, F., Kochs, G., Gruber, S. and Haller, O. (1997). A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology*, 230(1), 10-16.
98. Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M. and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiological Reviews*, 56(1), 152-179.
99. Whitfield, Z. J., Dolan, P. T., Kunitomi, M., Tassetto, M., Seetin, M. G., Oh, S., Heiner, C., Paxinos, E., & Andino, R. (2017). The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Current Biology*, 27(22), P3511-3519.E7.
100. Widjaja, I., de Vries, E., Rottier, P. J. M., & de Haan, C. A. M. (2012). Competition between influenza A virus genome segments. *PLOS ONE*, 7(10), e47529.
101. Wolff, T. and Veit, M. (2021). Influenza B, C and D Viruses (*Orthomyxoviridae*). In *Encyclopedia of Virology* (4th ed., pp.561-574). Academic Press.
102. Yamaoka, S., Weisend, C.M., Swenson, V.A. and Ebihara, H. (2022). Development of accelerated high-throughput antiviral screening systems for emerging orthomyxoviruses. *Antiviral Research*, 200, 105291.
103. Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-314.
104. Ye, Q., Guu, T.S., Mata, D.A., Kuo, R.L., Smith, B., Krug, R.M. and Tao, Y.J. (2013). Biochemical and structural evidence in support of a coherent model for the formation of the double-helical influenza A virus ribonucleoprotein. *mBio*, 4(1), e00467-12.
105. Zhao, L., Peng, Y., Zhou, K., Cao, M., Wang, J., Wang, X., Jiang, T. and Deng, T. (2014). New insights into the nonconserved noncoding region of the subtype-determinant hemagglutinin and neuraminidase segments of influenza A viruses. *Journal of Virology*, 88(21), 11789-11803.
106. Zheng, W. and Tao, Y.J. (2013). Structure and assembly of the influenza A virus ribonucleoprotein complex. *FEBS Letters*, 587(8), 1206-1214.
107. Zhou, B., Donnelly, M.E., Scholes, D.T., St. George, K., Hatta, M., Kawaoka, Y. and Wentworth, D.E. (2009). Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza A viruses. *Journal of Virology*, 83(19), 10309-10313.
108. Zhou, B., Lin, X., Wang, W., Halpin, R.A., Bera, J., Stockwell, T.B., Barr, I., Wentworth, D.E. and the Influenza Genome Sequencing Project (2014). Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics. *Journal of Clinical Microbiology*, 52(4), 1330-1337.
109. Zhu, W., Li, J., Zou, R., Li, X., Li, Z., Wang, J., Shao, H., Wu, B., Lin, T., Qin, A. and Ye, J. (2011). A reporter system for assaying influenza virus RNP functionality based on secreted Gaussia luciferase activity. *Virology Journal*, 8, 29.

SUMMARY

Vilnius University

Systems biology

Miglė Kazlauskaitė

Recovery of 5' and 3' UTR sequences of Wuhan Mosquito Virus 6

Orthomyxoviruses represent a diverse family of segmented RNA viruses including major human and animal pathogens such as influenza viruses. Recent metagenomic studies have revealed numerous arthropod-associated orthomyxoviruses, including Wuhan Mosquito Virus 6 (WuMV-6), which has achieved exceptional global distribution across six continents. Understanding viral biology requires characterization of untranslated regions (UTRs) that serve critical regulatory functions. Yet UTR sequences remain completely unknown for WuMV-6 and related quaranjaviruses, preventing development of essential molecular tools.

This study aimed to characterize the 5′ and 3′ UTRs of all eight WuMV-6 genome segments using samples from Lithuanian *Culex pipiens* mosquito populations. Reverse transcription quantitative polymerase chain reaction screening identified WuMV-6-positive pools. Those were later subjected to rapid amplification of complementary DNA ends (RACE) followed by Oxford Nanopore sequencing. WuMV-6 was detected in 42% of mosquito pools across 12 of 25 sampling localities with an estimated individual prevalence of 1,45%, confirming widespread circulation throughout Lithuania. Phylogenetic analysis revealed that recovered sequences clustered within the Baltic/Northern European lineage, distinct from the global lineage. However, RACE analysis encountered significant technical challenges including non-specific amplification and low recovery rates. Only four authentic WuMV-6 sequences were recovered from 57 samples, representing fragments from hypothetical segment 1 and nucleoprotein segments with insufficient sequencing depth for reliable analysis.

Complete UTR characterization was not achieved due to methodological limitations. The study demonstrates the need for improved approaches to characterize UTRs in newly discovered arthropod-associated orthomyxoviruses, as current techniques prove insufficient for these viral lineages.

SUMMARY IN LITHUANIAN

Vilniaus universitetas

Sistemų biologija

Miglė Kazlauskaitė

5' ir 3' nekoduojančių Uhano uodų viruso 6 sekų nustatymas

Ortomiksovirusai yra įvairių segmentuotų RNR virusų šeima, kuriai priklauso tokie žymūs žmonių ir gyvūnų patogenai kaip gripo virusai. Pastaraisiais metais metagenominiai tyrimai atskleidė daugybę su nariuotakojais susijusių ortomiksovirusų, tarp jų ir Uhano uodų virusą 6 (WuMV-6), pasižymintį itin plačiu paplitimu tarp šešių žemynų. Siekiant suprasti virusų biologiją, būtina nustatyti jų nekoduojančias sekas (UTR), atliekančias svarbias reguliacines funkcijas. Tačiau WuMV-6 ir jam giminingų karandža virusų UTR sekos iki šiol nėra žinomos, todėl nėra galimybės kurti esminių molekulinių tyrimų įrankių.

Šio tyrimo metu buvo siekiama nustatyti visų aštuonių WuMV-6 genomo segmentų 5′ ir 3′ UTR sekas, pasitelkus Lietuvoje surinktų *Culex pipiens* uodų mėginius. WuMV-6 teigiami mėginiai buvo identifikuoti taikant atvirkštinės transkripcijos kiekybinę polimerazės grandininę reakciją. Vėliau šie mėginiai buvo analizuojami taikant komplementarios DNR galų greito amplifikavimo metodą (RACE) bei Oxford Nanopore sekoskaitą. WuMV-6 buvo nustatytas 42 % uodų mėginių 12-oje iš 25 tirtų vietovių, o apskaičiuotas individualus viruso paplitimas siekė 1,45 %, patvirtindamas platų viruso cirkuliavimą Lietuvoje. Filogenetinė analizė atskleidė, kad gautos WuMV-6 sekos priklauso Baltijos/Šiaurės Europos linijai ir yra genetiškai nutolusios nuo pasaulinės linijos. Vis dėlto RACE analizės metu kilo reikšmingų techninių sunkumų, įskaitant nespecifinę amplifikaciją ir žemą sekų atkūrimo dažnį. Iš 57 mėginių buvo gautos tik keturios autentiškos WuMV-6 sekos, atitinkančios hipotetinio segmento 1 ir nukleoproteino segmento fragmentus, kurių sekoskaitos gylis buvo nepakankamas patikimai analizei atlikti.

Dėl metodologinių apribojimų UTR sekos nebuvo patikimai nustatytos. Tyrimas atskleidė būtinybę tobulinti metodus, skirtus naujai atrastų su nariuotakojais susijusių ortomiksovirusų UTR sekų nustatymui, kadangi šiuo metu taikomi metodai pasirodė nepakankami šioms virusų atmainoms tirti.

APPENDICES

Table 1 Mosquito pool sampling and WuMV-6 detection information for hyp1 and NP segment fragments.

Pool ID	Location	Coordinates	Collection date	No. of individuals	hyp1	NP
KA1-1	Kaunas	54.9054008, 23.8377843	7/10/2023	24	-	-
AL1-1	Aleksotas	54.8813590, 23.8685700	7/21/2023	15	-	-
AK1-1	Akademija	54.8928523, 23.8198148		6	-	-
RI1-1	Ringaudai	54.8781741, 23.7838481		25	-	-
IS1-1	Išlaužas	54.7449822, 23.9299795	7/23/2023	50	-	-
IS1-2				50	+	+
IS1-3				50	+	+
IS1-4				50	+	+
IS1-5				50	-	-
IS2-1				50	-	+
IS2-2		54.7471715, 23.9268681		50	-	-
IS2-3				50	-	-
IS2-4				50	-	-
IS2-5				50	-	-
IS2-6				42	+	+
TR2	Trakai	54.6669592, 24.9090820	7/24/2023	31	-	-
TI3	Tiltai	54.4138864, 24.8189635		25	-	-
TI2-1		54.416287,24.803705		40	-	+
TI2-2				26	+	+
TI1		54.4159982, 24.8058109		26	-	-
VA1-1	Valkininkai	54.3575228, 24.8447050		48	+	+
VA1-2				48	-	-
VA2-1		54.3880168, 24.8925127		33	-	-
TR1-1	Trakai	54.6670650, 24.9114152		50	-	-
TR1-2				40	+	+
TR1-3				42	+	-
SK1-1	Skaudviliai	55.8458145, 23.0912682	7/26/2023	12	-	-
SI1-1	Šiauliai	55.9318808, 23.3771229		50	-	-
SI1-2				43	-	-
KU3-1	Kurtuvėnai	55.8224393, 23.0521414		6	-	-
ZU1-1	Žuvininkai	55.9427239, 23.3644357		50	-	-
ZU1-2				47	-	+
ZU1-3				46	+	+
MA1-1	Malavėniukai	55.9308640, 23.3958437		50	+	+
DA1-1	Daržininkai	55.824019,25.910332	7/28/2023	50	-	-

Continuation of Table 1						
DA1-2	Daržininkai	55.824019,25.910332	7/28/2023	50	-	-
DA1-3				40	+	+
DA1-4				34	+	+
MA1-2	Malavėniukai	55.9308640, 23.3958437		50	+	-
MIE1-1	Mielėnai	55.8633054, 25.8723025	7/29/2023	20	+	+
GE1-1	Gervėnupis	54.8439793, 24.1293142	7/30/2023	8	-	-
GE2-1				39	-	-
MIK1-1	Miknonys	54.5022806, 23.6690525	8/2/2023	22	-	-
ASK4-1	Aštrioji Kirsna	54.3010774, 23.3936755		35	-	+
ASK3-1		54.3068150, 23.3914053		14	+	+
ZUV1-1	Žuvintai	54.4569480, 23.6353616		50	+	+
ZUV1-4				50	+	+
ZUV1-6				50	-	-
ZUV1-7				27	-	-
ASK6-1	Aštrioji Kirsna	54.3052039, 23.3898879		40	-	-
ASK6-2				40	-	-
ASK6-3				30	-	-
IS3-1	Išlaužas	54.7399378, 23.9354287	8/8/2023	50	-	-
IS3-2				50	+	-
IS3-3				50	-	-
IS3-4				50	-	-
IS3-5				50	+	+
IS3-6				50	+	+
IS3-7				17	-	-
IS4-1		54.7449782, 23.9299664		50	-	-
IS4-2				50	-	-
IS4-3				50	+	-
IS4-4				50	+	+
IS4-5				50	-	-
IS4-6				50	-	-
IS4-7				50	+	-
IS4-8				50	+	+
IS4-9				50	+	+
IS4-10				50	+	+
IS4-11				50	+	+
IS4-12				50	-	-
IS4-13				50	-	-
IS4-14				50	-	-
IS4-15				49	-	-
IS5-1		54.7472474, 23.9268688		40	-	-

Continuation of Table 1						
IS5-2	Išlaužas	54.7472474, 23.9268688	8/8/2023	40	+	+
MET1-1	Meteliškės	55.788173, 25.888037	8/10/2023	15	-	-
BRA1-1	Bradesiai	55.825879,25.905057	8/11/2023	6	-	-
IS7-10	Išlaužas	54.7449147, 23.9299520		30	-	-
IS7-11				29	-	-
KI1-1	Kintai	55.4179383, 21.2506653	8/14/2023	8	+	+
VEN1-1	Ventė	55.3417415, 21.1894509		25	+	+
VEN1-2				29	+	+
VEN2-1		55.3418010, 21.1946738		19	+	+
DAU1-1	Daugailiai	55.6037471, 25.8300282	9/1/2023	14	-	-
PAS1-1	Pasarčiai	55.8551722, 25.8609718		15	-	-
Total mosquitoes tested					**3310**	

Positive and negative detections are denoted by (+) and (-), respectively. hyp1 – WuMV-6 hypothetical segment 1; NP – WuMV-6 nucleoprotein segment. Detection of WuMV-6 hyp1 alone is indicated in yellow, WuMV-6 NP segment alone is indicated in orange, and simultaneous detection of both WuMV-6 segments is indicated in green.

Table 2 DNA concentrations of samples selected for sequencing.

Sample ID	Concentration ng/µL
37_H5	undetectable
59_H5	undetectable
37_N3	0,15
28_H3	0,46
31_H3	0,13
66_H3	undetectable
37_N5	undetectable
28_H3_2	3,06
31_H3_2	3,10
66_H3_2	0,36
IS1-2_H3	0,14
IS1-3_H3	1,42
IS2-6_H3	0,14
IS3-5_H3	0,72
IS3-6_H3	0,83
IS4-4_H3	0,46
IS4-10_H3	0,15
IS4-11_H3	undetectable
IS1-3_N3	undetectable
IS1-4_N3	undetectable
IS3-6_N3	undetectable
IS4-8_N3	undetectable

Continuation of Table 2	
IS4-11_H5	0,25
IS3-5_N3	0,38
IS3-5_H5	6,40
IS3-6_N3	0,79
IS4-8_N3_2	undetectable
IS1-3_H5	6,74
IS4-4_H5	5,50
IS4-8_N3	1,16
IS1-3_H5_2	1,11
IS4-8_H5	4,94
IS2-6_H5	2,94
IS2-6_H5_2	1,79
IS2-6_H5_3	undetectable
IS4-7_H5	0,21
IS4-9_H5	0,35
IS4-11_H5	0,15
IS4-7_N5	0,21
VEN1-1_N3	0,12
IS4-10_H3_2	0,48
IS4-11_H3_2	0,55
VEN1-1_H5	0,52
VA1-1_H5	0,26
ZUV1-1_H5	0,12
ASK3-1_N3	0,25
TRI1-3_N3	undetectable
TRI2-2_N3	0,73
VA1-1_N3-2	0,23
TRI2-2_H5	undetectable
VA1-1_N3	3,04
ZU1-3_N3	0,23
TI2-2_N5	undetectable
TI2-2_H5	0,90
ZU1-3_N3_2	0,17
MA1-1_H5	16,6
DA1-4_H5	12,4

H3 – 3' RACE product of hyp1; H5 – 5' RACE product of hyp1; N3 – 3' RACE product of NP; N5 – 5' RACE product of NP. Additional WuMV-6-positive samples (28, 31, 37, 59, 66) were obtained from a previous study (Tulauskaitė, 2025).

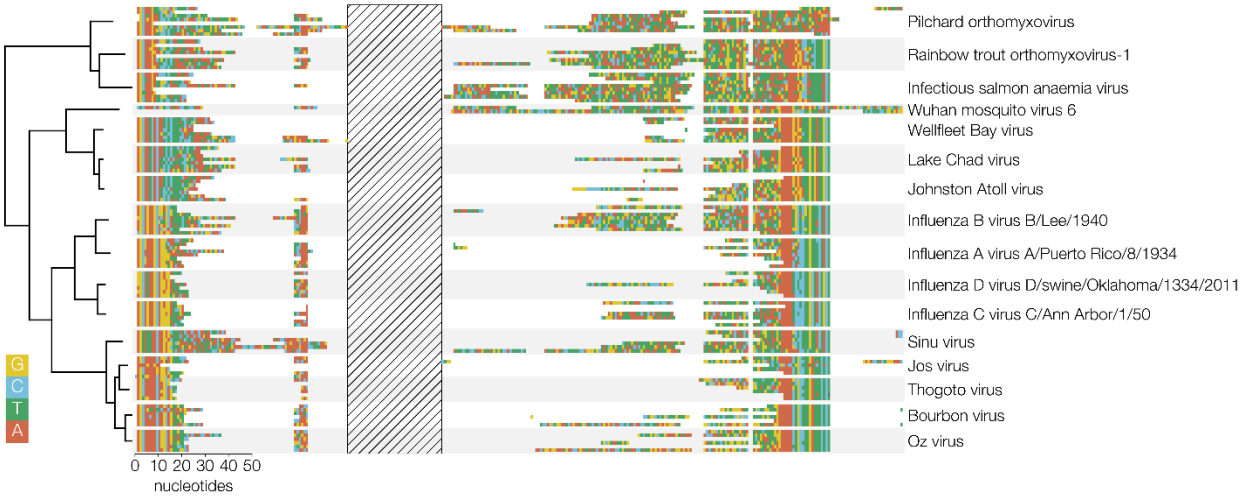

Figure 1. Nucleotide composition of entire UTR sequences across orthomyxoviruses including WuMV-6. Viral ordering is determined by a PB1 amino acid maximum likelihood phylogenetic tree. All known genome segments are shown (except for WuMV-6) arranged in the following order from top to bottom: PB1, PB2, PA, NP, HA/HE/HEF, GP/GP64, F, UNKNOWN (the surface protein of Rainbow trout orthomyxovirus-1 labelled as hemagglutinin that bears no recognisable homology to any other protein on public sequence databases), NA, M/hyp1, NS, VP7. The hatched region represents the coding region. G (yellow) – guanine; C (blue) – cytosine; T (green) – thymine; A (orange) – adenine.