FINAL THESIS

TIAR SARA

2024/2025

HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

SZENT ISTVÁN CAMPUS

INSTITUTE OF GENETICS AND BIOTECHNOLOGY

MASTER OF SCIENCE IN AGRICULTURAL BIOTECHNOLOGY

INVESTIGATION OF VIROMES OF CROPS AND WEEDS USING HIGH THROUGHPUT SEQUENCING BASED METAGENOMICS APPROACHES

Insider consultant : Dr. Éva Várallyay,

Scientific advisor

Insider consultant's

Institute/department: Institute for Plant Protection, MATE

Insider consultant: Dr. Zsuzsanna Galbács

Research fellow,

Insider consultant's

Institute/department: Institute for Plant Protection, MATE

Created by: TIAR SARA

GÖDÖLLŐ

2024/2025

CONTENTS

CONTENTS

INTRODUCTION AND WORK OBJECTIVES

Plants and their environment are geared to a generalized within natural ecosystems, where plants interact with their environment through diverse mechanisms, including plant-ecosystem dynamics, macro- and micro-species interactions, symbiosis, and allelopathy.

These ecosystems face threats from a vast array of pests, including approximately 10,000 insect species, 1,800 weed species, 1,500 plant disease species, 1,500 plant virus species, and various rodent species, all of which impact plant health. Through long-term natural selection, these heterotrophic organisms, which obtain nutrients and reproduce by harming plants, achieve a dynamic equilibrium.

To counter these threats, plant protection practices have evolved through three distinct phases: an initial stage based on natural farming, followed by a period dominated by pesticide use, and finally, the current integrated approach. This integrated strategy incorporates agricultural, physical, biological, and chemical methods, alongside plant quarantine and integrated pest management (IPM) techniques, aiming to provide comprehensive plant protection (Zheng & Xu, 2023).

Algeria's agricultural landscape heavily relies on vegetable crops, which dominate 70% of its cultivable land. The favorable climate and soil conditions of the Sahel and coastal regions have led to widespread plasticulture, unfortunately also creating an environment conducive to the proliferation of diverse diseases and pests. (Sabri et al., 2022)

In Algeria, little is known about the specific plant protection practices employed by farmers. However, studies indicate that pesticide use is prevalent, with a significant majority (86%) of farmers relying on regular pesticide applications for crop protection. A smaller proportion (14%) resort to pesticides only occasionally, typically in response to severe infestations by bio-aggressors(Bachouche et al., 2024)

This highlights the need for a balanced approach, integrating time-honored strategic methods with innovative, contemporary plant protection techniques, to mitigate the risks associated with intensive cultivation practices and ensure sustainable agricultural productivity.

Recent advances in pesticide application technology involve plant protection UAVs, intelligent sprayers, and spraying robots. Additionally, techniques like metagenomics and metabarcoding are culture-independent methodologies that offer detailed descriptions of the composition, abundance, and structure of thousands of genomes within a community. **Metabarcoding** amplifies known gene regions such as 16S, ITS, or 18S, which are highly variable in length and

GC nucleotide content. **Metagenomics** involves sequencing all the DNA in a sample to provide a comprehensive view, both techniques can be used to detect plant pathogen evolution. (Adeleke et al., 2022)

Other methods include customized cultivation of resistant plants using genetic variations and transgenic or gene-editing methodologies. Furthermore, seed enhancement technologies like pesticide seed dressing, coating, and priming enhance seeds and seedlings against early-season pest(Zheng & Xu, 2023)

High throughput sequencing (HTS) has revolutionized virus detection and discovery by allowing for the untargeted characterization of whole viromes. HTS methods have been developed to look for virus-like sequences without the bias of only detecting known viruses. It allows a generic approach to virus identification and can deliver species, or even strain-specific results. (Olmos et al., 2018)

HTS has multiple applications in plant virus diagnostics, such as identifying causes of viral diseases in economically important crops, screening for specific viruses when their presence is suspected, detecting asymptomatic or cryptic viruses, and discovering novel viruses. HTS offers advantages over traditional methods like ELISA or PCR, which target only a few viral species and require prior knowledge of the virus (Hasiów-Jaroszewska, Boezen, & Zwart, 2021)

E. canadensis is an invasive weed, which can also act as a feeding source of several different insects. Consequently, if infected the viruses can easily transmitted to the neighboring crops, but at the moment we don’t know which viruses can infect this weed?

Our work objectives consist of:

Utilization of HTS to sequence and analyze viral genetic material from environmental samples of *E. canadensis*, allowing for the comprehensive characterization of viral communities

Employment of advanced bioinformatics tools to process and analyze HTS data, identifying known and novel viruses present in this plant species.

Characterization of the identified viruses in terms of their genetic diversity, structure, and potential impact on plant health.

LITERATURE REVIEW

1. The importance of weed management

Over the past century, the global population has quadrupled, rising from 1.8 billion people in 1915 to a milestone of 8 billion people as of November 15th, 2022, according to the most recent estimate by the UN. Projections suggest that the global population is expected to reach around 9.7 billion by 2050(Jiménez-Arias, Morales-Sierra, Borges, Herrera, & Luis, 2022).The agriculture sector must prioritize higher yields and increased food grain production. As the world's population continues to grow, per capita arable land decreases, underscoring the need to enhance food production efficiency both in terms of time and space to guarantee food and nutritional security. (Laxman Navi, 2024)

Weeds are defined as unwanted plants that compete with crops for essential resources such as nutrients, light, and water. Their presence can lead to substantial yield losses across various crops. Not only do weeds compete directly with crops, but they can also serve as hosts for pests and pathogens, exacerbating the challenges faced by farmers in maintaining crop health and maximizing yield potential (Balas, Pargi, Lakhani, Mehta, & Bambhaniya, 2023)

The significance of effective weed management cannot be understated. A robust weed control strategy can significantly enhance crop yield; for example, timely hand weeding is reported to improve yields by up to 70% compared to non-weeded plots(Saqib et al., 2015) **.**Furthermore, the competitive advantage that crops hold over weeds can be crucial for achieving the necessary output levels to safeguard against food insecurity as the population expands. It is estimated that unchecked weed growth could lead to substantial crop yield losses globally, underscoring the importance of implementing comprehensive weed management plans (Farooq, Flower, Jabran, Wahid, & Siddique, 2011).

2. Weed management strategies

To Achieving this goal different strategies can be used

2.1 Strategies for Conservation Agriculture (CA)

conservation agriculture is basically a sustainable farming system defined as "a methodology for resource saving agricultural crop production, emphasizing on enhancement of natural and biological processes above and below ground"(FAO, 2019)**.**This context define the frame of the management methods what can be used. According to (Laxman Navi, 2024)the possibilities are the following:

- **Preventive Weed Management** involves using weed-free crop seeds and ensuring that machinery, irrigation water (through screens), and livestock do not transfer weed seeds between areas.
- **Cultural Practices** include techniques such as maintaining ground cover through mulch, using competitive crop varieties, managing crop residues, applying cover crops, intercropping, sequential cropping, and utilizing allelopathy. Proper management of water and nutrients is also essential.
- **Chemical Weed Management** refers to the application of herbicides to control weed populations.
- **Integrated Weed Management (IWM)** combines multiple strategies, including cultural, preventive, and chemical methods, to effectively manage weeds.
- **Robotic Weed Control** employs technologies like machine vision, artificial intelligence (AI), and precision application equipment to detect and eliminate weeds efficiently(Gerhards et al., 2022)

2.2 Nonconventional Weed Management Strategies for Modern Agriculture

(Bajwa, Mahajan, & Chauhan, 2015)mentioned that to combat evolving weed challenges exacerbated by intensive agriculture and climate change, a shift towards sustainable weed management is crucial. They emphasis on potential nonconventional strategies, including improvements in tillage regimes, harvest weed seed control, seed predation, allelopathy, biotechnological advancements (like herbicide-tolerant crops), and thermal techniques. These approaches aim to mitigate herbicide resistance, environmental pollution, and biodiversity decline, offering a path towards more effective and ecologically sound weed management.

2.3 New Strategies and Technologies in Weed Management

Contemporary weed management is evolving towards precision agriculture, traditional methods while still relevant, are being augmented by emerging technologies like, utilizing GPS

and remote sensing for site-specific interventions, and genomic approaches, which aim to develop weed or herbicide-resistant crops. Robotic weed control, employing machine vision and AI, enables targeted weed removal, minimizing chemical use. These technologies, while promising, require addressing challenges like cost and complexity for widespread adoption. Integrated weed management, combining these advanced tools with traditional methods, offers a sustainable path forward. (Shaikh, Rasool, & Lone, 2022)

3. Weeds as virus Reservoirs

Weeds are reservoirs of plant viruses, playing a crucial role in viral epidemiology by potentially spreading viruses to cultivated plants and leading to epidemics. (Hasiów-Jaroszewska et al., 2021)

Weeds act as reservoirs for plant viruses, harboring pathogens that can infect crops and reduce yield quality (Byron, Treadwell, & Dittmar, 2019). These viruses can spread through insect vectors, making weed management crucial for preventing crop diseases. Identifying and removing weeds that host viruses can help mitigate economic losses in agriculture. Understanding the diversity of viruses in weeds is essential for developing effective management strategies. (Kusumavathi et al., 2025)

To understand virus diversity in wild plants to gain insights into long-term coexistence between hosts and viruses and to evaluate disease Weeds constitute potential reservoirs for viruses that may spread into cultivated plants, leading to epidemics or the emergence of novel viruses. This "crossover" is influenced by factors such as agricultural practices, international trade, and human-induced changes in plant biodiversity. emergence risk suitable metagenomic approaches are needed. Knowledge of virus diversity, prevalence, and dynamics in wild plant populations is relevant to understanding virus epidemiology and emergence in crops.

4. *E. canadensis*

E. canadensis, also known as horseweed, is an annual Plant, Its taxonomic classification places it in the Kingdom Plantae, Family Asteraceae, and Genus Erigeron/Conyza. This species is widely distributed across various regions, including Asia, North America, and Europe, thriving in disturbed habitats and agricultural fields (Figure01).

Figure 01: *E. canadensis* (Flora, 2018)

4.1 Biology and Ecology of *E. canadensis*

E. canadensis, is an annual plant, exhibits fluctuating population dynamics driven by seed input, germination, and survival. High seed input, dispersed by wind, can range from 12,500 seeds m-2 near the seed source to 126 seeds m-2 further away. Germination primarily occurs in spring and autumn, triggered by rainfall and moisture. Optimal temperatures range from 18 to 23°. In Europe, *E. canadensis* typically flowers from May to September and Seedling survival rates are generally high but winter mortality, mainly due to frost-heaving, significantly reduces rosette numbers. Winter survival is closely correlated with rosette size, with smaller rosettes experiencing higher mortality. Surviving rosettes bolt in spring, and flower and seed production are positively correlated with plant height, though reproductive effort decreases with increasing height. At maturity, seed production is proportional to plant height, but reproductive effort is inversely related to plant height. Despite high winter mortality, the abundant seed production of survivors ensures the continuation of the *E. canadensis* population. These dynamics highlight the adaptive strategy of *E. canadensis* as a colonizing species in disturbed areas (Regehr & Bazzaz, 1979)

E. canadensis exhibits high genetic diversity by analyzing 312 individuals in both its native (USA) and invasive (China) ranges, suggesting multiple introductions and a lack of significant genetic bottlenecks during invasion. Despite being a self-pollinating species, evidence of admixture points to a higher-than-expected outcrossing rate, potentially contributing to its adaptability. Analysis of population structure reveals distinct genetic clusters corresponding to

native and invasive regions, with further subdivision within the invasive range, likely shaped by historical population structure and subsequent gene flow. These findings highlight the complex interplay of factors, including multiple introductions, admixture, and gene flow, that have contributed to the successful invasion of *E. canadensis* in Eastern China. (Bhattacharya et al., 2022)

4.2 Distribution and habitat

E. canadensis, thrives in disturbed habitats and is frequently found in agricultural fields and their margins. This species typically germinates in the summer and autumn, overwinters as a rosette, and bolts in the spring. It is characterized by its high seed production and wind-dispersed seeds, which contribute to its widespread distribution. *E. canadensis* can survive cold climates and tolerate brackish water, making it highly adaptable to various environments. Its invasiveness can lead to local extirpations of native plant species and changes in lake ecosystems, affecting both abiotic and biotic conditions. (Bhattacharya et al., 2022)and (Garibaldi et al., 2011).

4.3 Nutritional Values and uses of *E. canadensis*

Erigeron species, including E. Canadensis, have been found to possess diverse phytochemistry, pharmacological activities, and traditional uses. Phytochemical studies of *E. Canadensis* reveal the presence of saponins, diterpenoids, terpenoids, glycosides, tannins, anthraquinones, steroids, and flavonoids. These compounds contribute to a range of pharmacological activities, including antimicrobial, antioxidant, anti-inflammatory, and anticancer effects (Sharma, Verma, Jha, Singh, & Kumar, 2014). Traditionally, *E. Canadensis* species have been used for various medicinal purposes, such as treating wounds, pain, inflammation, fever, and microbial infections. *E. canadensis* has been specifically used as a tonic for diarrhea, diabetes, and hemorrhages, as well as for treating urinary and respiratory infections. The plant has also been employed in traditional medicine to address granuloma annulare, sore throats, and for medicinal baths. These uses align with the observed pharmacological properties, highlighting the potential of Erigeron species as a source of medicinal agents (Al-Snafi, 2017)

5. Plant viruses

Plant pathogens, are a diverse group of microorganisms, induce a variety of diseases in plants, leading to substantial economic losses in agricultural sectors (Sun, Xiao, & Xue, 2023). Following fungi, viruses are recognized as the second most common type of plant pathogen.

These minute parasites, are composed of nucleic acid, either RNA or DNA, encapsulated within a protein coat (Kovalskaya & Hammond, 2014; Manjunatha et al., 2022), and (Tang et al., 2023). Characteristically, viruses are measured in nanometers and upon infecting plants, they can cause a spectrum of diseases and significant crop damage.

5.1 viral genome reproduction, transmission strategies

Viruses depend on host cells for replication, exploiting cellular machinery for genome replication and expression (Tennant, Fermin, & Foster, 2018).Viral genomes, which include dsDNA, ssDNA, dsRNA, or ssRNA, function as mRNAs or templates. RNA viruses employ unique pathways such as RNA-dependent RNA synthesis or reverse transcription. Regulation involves trans-acting proteins and structural elements. DNA viruses utilize host or viral enzymes for mRNA transcription and DNA replication, employing mechanisms like rolling circle and bidirectional replication. Transmission is achieved through generating abundant genome copies, packaging them into virions, and infecting new hosts (Figure02).

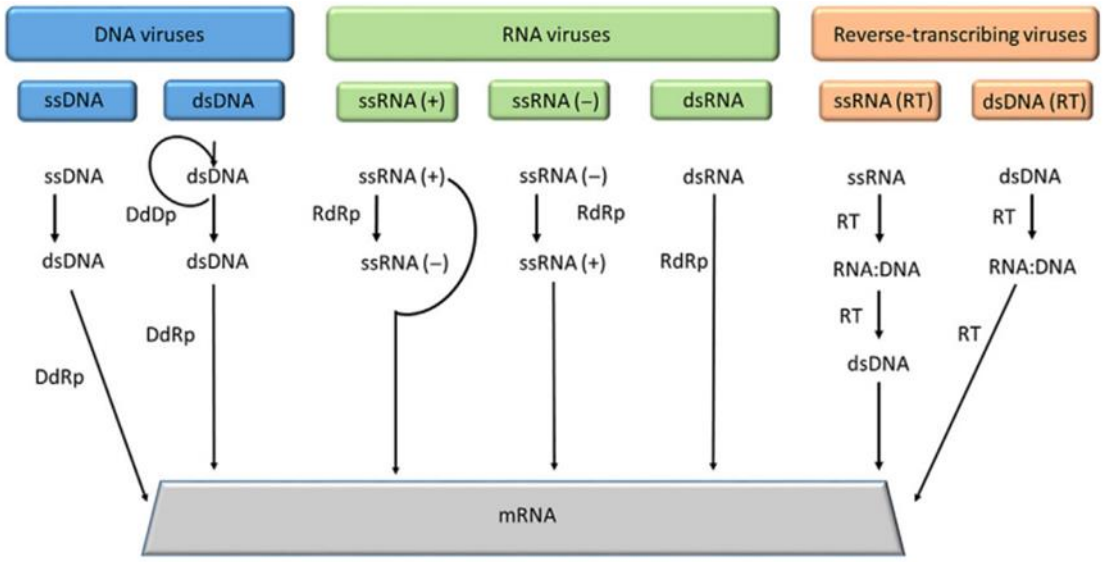


Figure02: Summary of replication and transcription modes of different classes of viruses. DdDp, DNA-dependent DNA polymerase; DdRp, DNA-dependent RNA polymerase; RdRP, RNA- dependent RNA polymerase; RT, reverse transcriptase. The ssRNA (+) can serve as the template for translation and does not undergo any modification prior to translation (Tennant et al., 2018).

5.2 Host-pathogen interaction investigation

Pathogens must interact with the host genome to survive and propagate, a process that involves several critical steps during pathogenesis. Initially, the pathogen adheres to the host cell surface, often using specific receptors to facilitate entry. Once inside, the pathogen evades

the host's immune response by various mechanisms, such as altering its surface proteins or secreting factors that inhibit immune signaling(Ponce-Cusi, Bravo, Paez, Pinto, & Pilco-Ferreto, 2024). The pathogen then manipulates the host cell's machinery to replicate its genetic material, often integrating its genome into the host's DNA. This integration can disrupt normal cellular functions and promote the production of viral proteins necessary for the pathogen's survival and spread (Team, 2025). Finally, the pathogen exits the host cell, often causing cell damage or death, and spreads to new cells or hosts, continuing the cycle of infection (Ponce-Cusi et al., 2024)

5.3 Description about biotechnology methods used to set resistance against viruses

Metagenomic methods, unlike traditional methods (ELISA, PCR), use high-throughput sequencing (HTS) for untargeted virus detection revolutionizing plant virology by enabling the untargeted characterization of whole viromes and revealing the vast diversity of viruses, particularly in wild plants (Kyrychenko, Shcherbatenko, & Kovalenko, 2021)

Metagenomic approaches could base on different starting materials for the High-throughput sequencing total RNA or DNA, virus-like particles (VLPs), double-stranded RNAs (dsRNAs), and small interfering viral RNAs could also serve as a starting point of the HTS and are essential tools for identifying novel viruses and characterizing their ecological roles.

Metagenomic methods could be based on:

•**Total RNA or DNA** sequencing: identifying viral sequences from symptomatic and asymptomatic plants.

•**Virus-like particles (VLPs)** analysis: differential centrifugation to obtain a VLP fraction, followed by nucleic acid isolation, amplification, sequencing, and bioinformatic analysis.

•**Double-stranded RNAs (dsRNAs) analysis**: dsRNAs are often formed by RNA viruses as a viral genome or replicative form.

•**Small interfering RNAs (microRNAs) sequencing**: specific double-stranded RNA molecules that regulate mRNA translation and are synthesized in response to RNA/DNA viruses

5.4 Viral Diseases Affecting *E. canadensis*

E. canadensis is vulnerable to a range of viral and mycoplasma diseases, impacting its health and reproductive success. Notably, tomato viruses have been identified in infected plants in mixed infections of tomato spotted wilt virus (TSWV) and Impatiens necrotic spot virus (INSV) have been observed in weeds near tobacco fields, causing symptoms like stunted growth,

chlorosis, and necrosis(Grbelja, Erić, & Jeknić, 1988).Additionally, *E. canadensis* is susceptible to aster yellows, a mycoplasma disease transmitted by the aster leafhopper, Macrosteles fascifrons (Regehr & Bazzaz, 1979).The geographical spread of aster yellows is influenced by summer wind patterns, with southerly winds and cool, damp weather favoring vector transport and disease development. Symptoms include chlorotic branches, stunted growth, and necrosis, with early or severe infections potentially causing plant death before flowering, and later infections leading to shoot apex necrosis. The severity of aster yellows is visually classified, and it significantly reduces seed production, directly affecting the plant's reproductive capacity. Effective management of these diseases involves monitoring and controlling vector populations, practicing crop rotation, and removing infected plants to limit viral and mycoplasma spread.(Timilsina et al., 2020),(Grbelja et al., 1988)and(Martínez-Ochoa, Langston, Mullis, & Flanders, 2003), Table 01 below group different viruses that host *E. canadensis*

Table01: Viruses Found in *E. canadensis* Based on NCBI Records

Virus Name	Accession	Genome Type	Length (bp)
Chilli veinal mottle virus isolate 24YNZT52	PQ847521.1	Linear RNA	861
Chilli veinal mottle virus isolate 24YNZT34	PQ847520.1	Linear RNA	861
Chilli veinal mottle virus isolate 24YNZT76	PQ847519.1	Linear RNA	861
Ageratum yellow vein Singapore alphasatellite isolate YN-2023	OR509735.1	Circular DNA	1,357
Tobacco curly shoot betasatellite isolate YN-2023	OR509734.1	Circular DNA	1,339
Tobacco leaf curl Yunnan virus isolate YN-2023	OR509733.1	Circular DNA	2,75
Tomato brown rugose fruit virus isolate Jo-Conc	OP009335.1	Linear RNA	872
Malvastrum yellow vein betasatellite isolate MYVbeta/YN251	OK120267.1	Circular DNA	1,341
Conyza yellow vein virus isolate CYVV/YN251	OK120266.1	Circular DNA	2,733
Malvastrum yellow vein betasatellite isolate MYVbeta/YN250	OK120265.1	Circular DNA	1,341
Conyza yellow vein virus isolate CYVV/YN250	OK120264.1	Circular DNA	2,733

6. Virus Diagnostics Methods

Accurate and timely virus diagnostics are critical for effective disease control and agricultural productivity. Plant virus detection relies on both conventional and innovative methods

6.1 Standard diagnostic methods

Traditional standard methods, used for virus diagnostics are enzyme-linked immunosorbent assay (ELISA), DNA amplification-based assays like PCR and Reverse transcription (RT-PCR. Their reliability and accuracy are different and they could detect the presence or absence of a particular pathogen.

6.1.1 Detection by ELISA (enzyme-linked immuno-sorbent assay)

ELISA is a convenient and sensitive serological method used to detect and quantify viral agents in plant tissues. It relies on the specific interaction between antibodies and viral proteins, mainly the capsid protein. Several types of ELISA exist, including direct, indirect, sandwich, and competitive, offering varying levels of sensitivity and specificity. Despite its dependence on antibody quality and potential for false results, ELISA remains a world standard for virus detection because of its ease of implementation, sensitivity, and adaptability (Zhang et al., 2024),(Figure03).

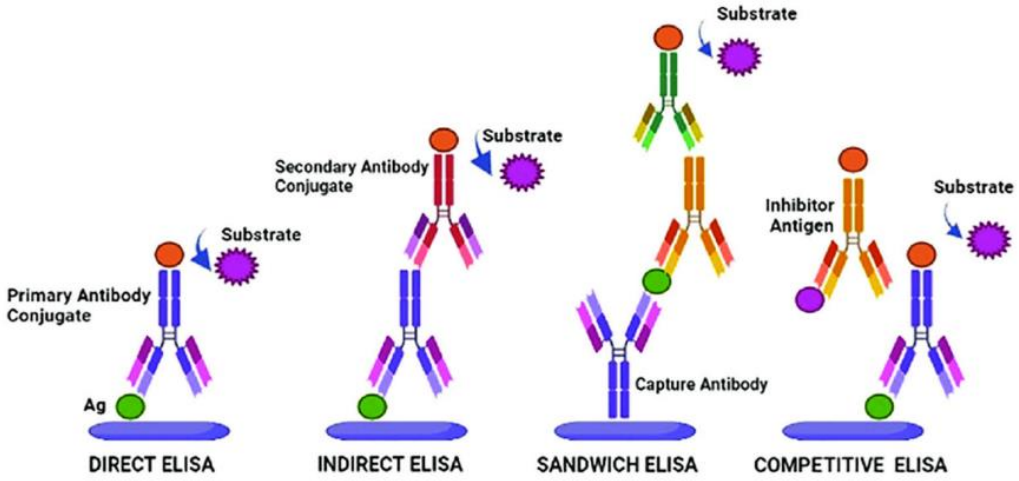


Figure03: ELISA types and their working principles. (Khan et al., 2023)

6.1.2 Detection by Polymerase Chain Reaction methods (PCR)

The polymerase chain reaction (PCR) was developed in 1983 by Dr. Kary Mullis while working for Cetus Corporation. In 1993, he received the Nobel Prize in Chemistry for this important contribution that revolutionized molecular biology (Mullis & Faloona, 1987).The technique can be used to amplify DNA sequences from any type of organism. It has been adapted over the years to allow amplification of RNA samples, as well as quantification of the amount of DNA or RNA in a sample (Suchman, 2011).

PCR amplifies target DNA through a cycling process involving denaturation, annealing, and extension. The reaction mixture includes dNTPs, primers, PCR buffer, specific polymerase, template DNA, and water. Additionally, chemicals like MgCl2 or DMSO are used to enhance amplification efficiency. Gradient PCR is employed to optimize each parameter within a single reaction. (Chauhan, Pandey, & Jain, 2021),(Figure04).

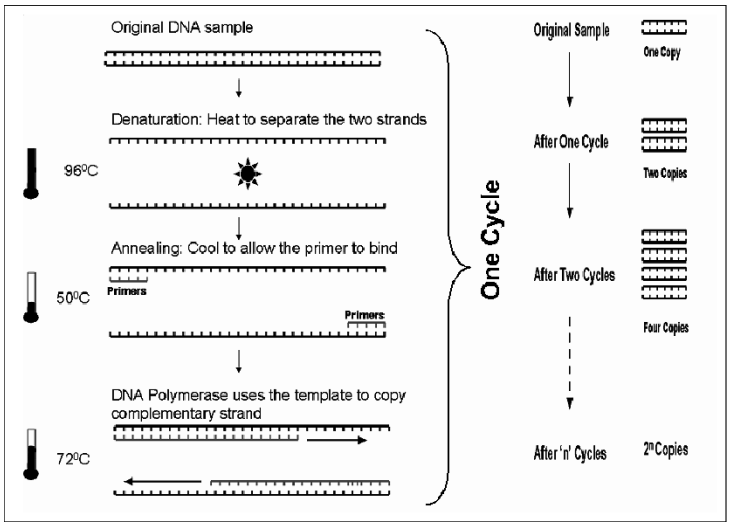


Figure 04: Schematic Diagram of the Polymerase Chain Reaction (PCR) (Pabla & Pabla, 2008)

6.1.3 Reverse transcription PCR (RT-PCR)

RNA viruses cannot be directly diagnosed by PCR. The first step in their diagnosis is the cDNA preparation using a reverse transcriptase enzyme. This cDNA can be used further in the RT-PCR reaction as a template(Sinha, John, Singh, & Johri, 2022).

It is a highly sensitive and versatile method used for detecting and quantifying RNA viruses. The process involves first converting viral RNA into complementary DNA (cDNA) using reverse transcriptase, followed by PCR amplification of the cDNA. RT-PCR is valuable because it can detect infectious agents at the earliest stages of infection, increasing the effectiveness of measures to combat plant diseases. Several modifications of RT-PCR have been developed, including quantitative PCR (qPCR), nested PCR, and multiplex PCR, each offering unique advantages in terms of sensitivity, quantification, and throughput. For example, qPCR is more sensitive in detecting small concentrations of target viruses, while significantly reducing detection time compared to other PCR methods.

6.1.4 Loop-mediated isothermal amplification (LAMP)

LAMP is valued for its simplicity, high specificity, and sensitivity, enabling rapid amplification of target sequences without needing expensive equipment

(Jang et al., 2022), LAMP uses multiple primers targeting different areas within the nucleic acid sequence, including internal primers, external primers, and loop primers, to initiate DNA synthesis with unique enzymes

The reaction displays the synthesis of new strands continuously in a cyclical manner, resulting in an exponential increase in the specific target sequence(Jeong, Lee, Ko, Ko, & Seo, 2022). LAMP can be performed using a crude extract from an infected plant, reducing preparation time and making it suitable for on-site detection (Panno et al., 2020

LAMP can amplify DNA at detectable levels within 30-60 minutes and can achieve detection limits that are comparable to, if not better than, those of PCR(Qin et al., 2021)

6.1.5 Micro and macroarray techniques

Array techniques are based on hybridization methods and they are used for the simultaneous detection of multiple plant viral pathogens (Sinha et al., 2022).The method is based on the use of virus-specific oligonucleotides that bind to a membrane or glass support. After the total RNA is converted to cDNA and amplified by PCR using pathogen-specific primers labeled with markers suitable for detecting molecules, the amplified and labeled products are applied to the array and DNA hybridization is performed. Microarrays can simultaneously analyze many molecules increasing the effectiveness of pathogen screening. Microarrays have been improved to detect hundreds of plant viruses, including genus-specific oligoprobes.

6.2 Metagenomics diagnostic methods

Metagenomics provides detailed genetic information, improving virus detection, characterization, and surveillance, and facilitating the transition from virus discovery to the development of specific detection methods like PCR or LAMP

6.2.1 Next-generation (high-throughput) sequencing (HTS) techniques

HTS offer high performance, sensitivity, and scalability for detecting plant viruses (Kanapiya et al., 2024)HTS technologies allow for the rapid, high-throughput sequencing of DNA, providing detailed information about the genetic composition of viruses in a timely and accurate manner. HTS overcomes the limitations of traditional methods, providing more accurate and complete analyses.

It has revolutionized plant virus diagnostics (Maree, Fox, Al Rwahnih, Boonham, & Candresse, 2018). HTS enables the detection of all viruses present in a plant, including those still unknown, without requiring any previous knowledge of viral sequences. HTS can be used for relative quantification based on the number of reads for the same sequence

Third-generation sequencing is based on sequencing single molecules in real-time without the need for clonal amplification, thus shortening DNA preparation time and giving long reads of

several kilobases(Bleidorn, 2016; Goodwin, McPherson, & McCombie, 2016)**.**HTS is the most powerful technique for multiplex detection as it can identify and discover an unlimited number of viruses and virus variants within a plant.

MATERIALS AND METHODS

1. Origin of the samples and their preparation for RNAseq

Leaf samples of *E. canadensis* were collected from crop fields in Tusnád, Romania, during June, July, and August of 2016. 10 individuals were sampled each month. Total nucleic acid was extracted from the leaf samples, DN-ase treated, then pooled according to the sample collection time, generating three pools. These three pools were unified as one pool, representing RNA of all of the sampled plant and sent for ribodepleted RNA sequencing. The sequencing was done using Illumina platform by NOVOGEN as a service. The library was assigned by a unique [TB_4_1] to track paired-end sequencing data. Processed library was stored in project-specific folders. Raw FASTQ files underwent trimming and quality control using CLC Genomic Workbench, retaining paired 150nt reads for downstream analysis. This standardized workflow ensured reliable detection and characterization of viral communities in *E. canadensis* via RNAseq.

2. Bioinformatic analysis of ribodepleted RNAseq

The result of the sequencing was the fastq files, containing the sequences and the quality of the sequencing. The fastq files were analyzed

For bioinformatics analysis we used CLC Genomic workbench. After quality control and trimming of the Illumina sequencing Reads, RNA sequences were quantified. De novo assembly of the reads generated contigs, which were annotated using four specialized databases: fungi viruses, land plants viruses, Slovenian weeds viruses, and invertebrate host viruses. To identify potential viral and viroid sequences, a BLAST search was performed against the NCBI database, with an E-value threshold used to filter for high-confidence hits.

A comprehensive list of potential viral and viroid matches was generated by aligning contigs to the NCBI reference genome database, with matches ranked by E-value to indicate probability. To ensure high confidence, only hits with an E-value of 0 were prioritized for further analysis. Consensus sequences were created from these contigs and aligned to their respective NCBI reference genomes. The final list of identified pathogens was organized by E-value and supported by the number of corresponding contigs. To validate the bioinformatics findings, RT-PCR was employed as an independent method, targeting specific viral sequences detected through this rigorous workflow. Using map reads to reference algorithm of the CLC the reads of both libraries were mapped to the reference genome visualized and counted.

- Improved Workflow: RNA -Seq Analysis

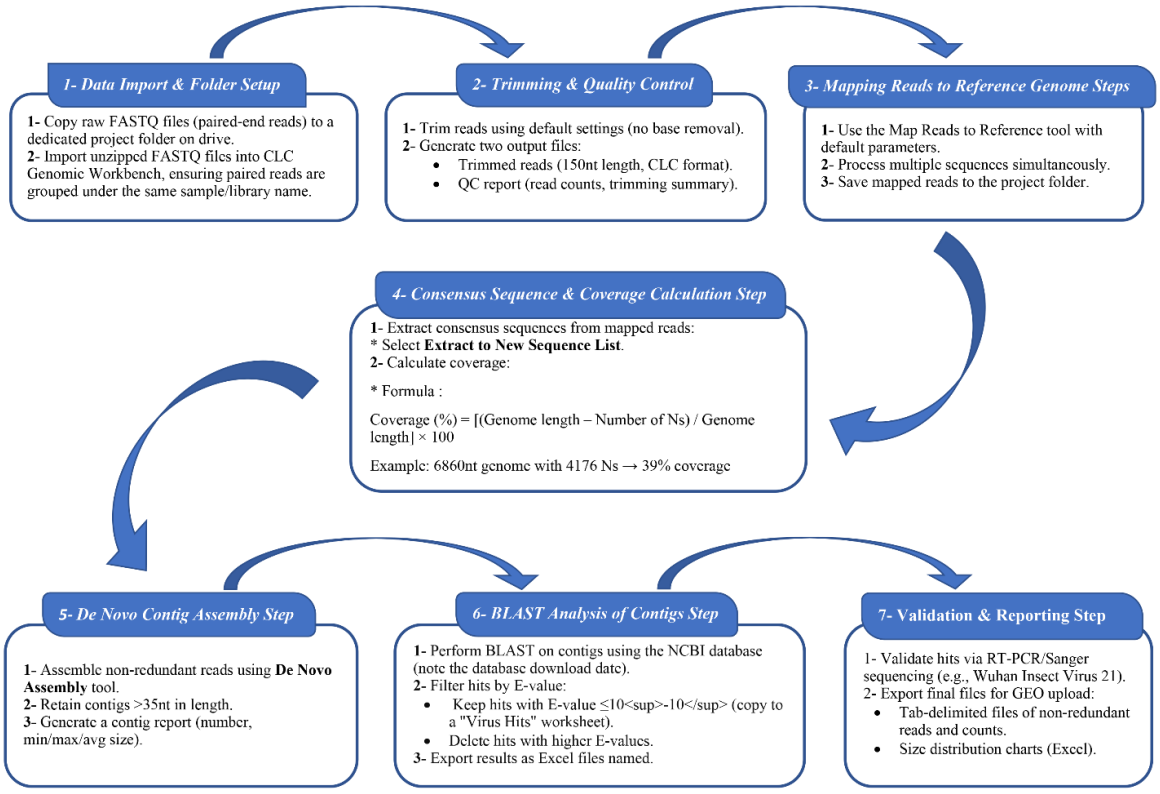


Figure05: Bioinformatic pipeline of the RNA- seq analysis **(Original,2025)**

3. cDNA Quality control

My research process began with checking the quality of the synthesized cDNA with a PCR reaction, amplifying a part of the host actin mRNA, using the "Q5 DNA polymerase". For the test I used 10 times diluted cDNA and primers specific to the host actin mRNA. As a positive control cDNA samples, ZS741/2 (prepared from Grapevine sample) and ZS237/3 (prepared from *E. canadensis* sample), while as a negative control water were used, (see Table 02).

Table:02: Different reaction components of the cDNA (quality control) PCR mixture.

Reaction mixture	**Quantity per sample**
MQ (Sterile water)	37 µl
5x Q5 Reaction Buffer	10 µl
Primer **F** (**Vv actin 601 s**)	2.5 µl
Primer **R** (**Vv actin 1200** as)	2.5 µl
10 mM dNTP	1 µl
Q5 DNA Polymerase	0.5 µl
Template (**Zs237/3**) / (**Zs237/2**)10x RT	0.5 µl

3.2.RT-PCR validation of the HTS detected viruses.

3.2.1. Primer design

To validate the presence of viruses detected by HTS RT-PCR To validate the presence of the viruses which were detected by HTS RT-PCR was employed. Primer pairs targeting specific viral sequences were designed. For the primer design the HTS detected contig and the sequenced genomes of the particular virus was used. We tried to find a position where these sequences were the same and the amplified part was 1737bp long. We tried to avoid repetition of the nucleotides and designed C or G to the 3' end of the primer. Based on these the following primers shown in Table 03 were designed and Figure 06 illustrates it.

Table:03: specific primers for Wuhan insect virus 21

Virus	Primer name	Primer sequence (5'-3')	Function of the amplified region	Genome used as a reference	Reference
Wuhan insect virus 21	WIV21_163F WIV21_1668R	CGTGTGAGCAGCCGCTCAGCG GCACAGAATATGATGATATGTG	hypothetical protein 1	NC_033481	Shi,M.,et al,2016
	WIV21_97F WIV21_1834R	CAAGTGCAAGACATCCTGCACC CACCGACCGTGGAATGATATTC	hypothetical protein	NC_033491	

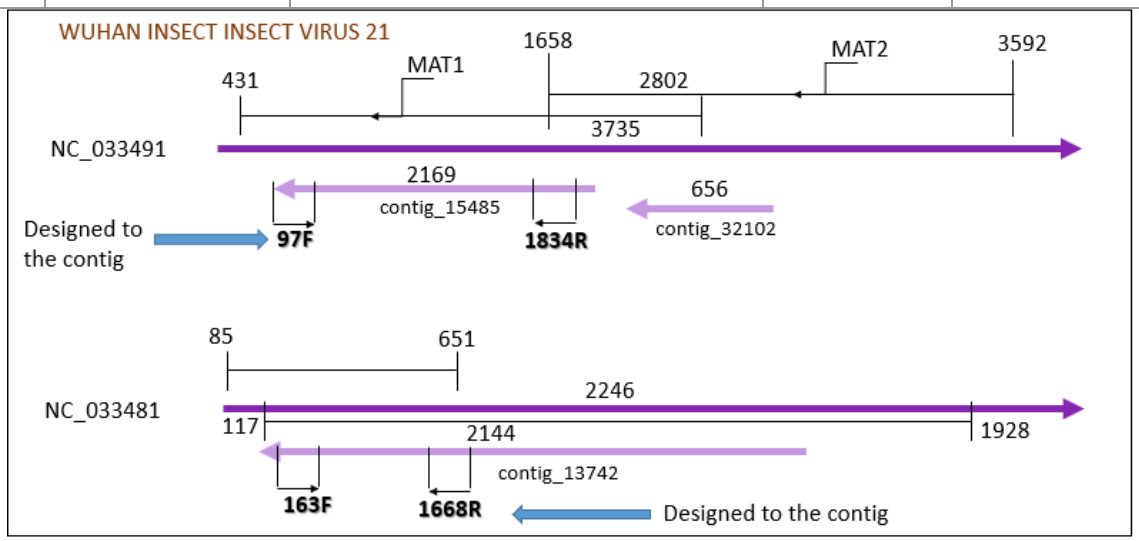


Figure06: BLAST Alignment and Primer-Target Sequence Conservation (Original,2025)

3.2.2. Optimization of the annealing temperature

To find the best amplification we tested different DNA Polymerases for the viral product amplification: Q5(Biolabs, 2025), and Phire Hot Start II DNA Polymerase

(ThermoFisherScientific, 2019) As these polymerases are different in their optimal annealing temperature and also in their temperature optimum for extension, the conditions for the reaction, including the optimization of the annealing temperature of the primers were done.

Q5 DNA Polymerase(Biolabs, 2025), a high-fidelity enzyme known for its accuracy and efficiency, and it allows a very high temperature for the annealing.

Phire Hot Start II DNA Polymerase (ThermoFisherScientific, 2019)is a very fast acting, polymerase which has a very high amplification rate, but allows only lower annealing temperature. Both enzymes have a very unique hot-start mechanism that prevents non-specific primer annealing at lower temperatures, significantly reducing background noise and improving the overall yield of specific PCR products.

To optimize the annealing temperature conditions detailed below were used: Table 04 and 05, and later Table 08,09 and 10 for Q5 and Table 06 and 07 for Phire polymerase.

Table:04: Component of the reaction mixture used for the amplification using Q5 DNA polymerase

Reactions Mixture	**Quantity per sample**
MQ (Sterile water)	37 µl
5x Q5 Reaction Buffer	10 µl
WIV21 Primer (**97F/ 1834R**)	2.5 µl
WIV 21 Primer (**163F / 1668R**)	2.5 µl
10 mM dNTP	1 µl
Q5 DNA Polymerase	0.5 µl
Template (**Zs237/3**) / (**Zs237/2**)10x RT	1.5 µl

Table:05: PCR amplification program when Q5 enzyme was used as a DNA polymerase

Stage	**Temperature °C**	**Time**	**Number of Cycle**
Initial Denaturation	98	30 seconds	1
Denaturation	98	10 seconds	40X
Annealing	58-66	30 seconds	
Extension	72	1 minutes	
Final extension	72	2 minutes	1
Hold	4	∞	-

Table:06: Component of the reaction mixture used for the amplification using Phire DNA polymerase

Component	**Quantity per 1x sample**	**Quantity per 2x sample**
MQ (Sterile water)	33 µl	66 µl
5x phire Buffer	10 µl	20 µl
WIV21_163F	2.5 µl	2.5 µl
WIV21_1668R	2.5 µl	2.5 µl
10 mM dNTP	1 µl	2 µl
Phire Polymerase	1 µl	2 µl
Template **Zs237/310**x RT	1 µl+49 ul	

Table07: Gradient PCR using Phire hot start DNA polymerase

Stage	**Temperature °C**	**Time**	**Number of Cycle**
Initial Denaturation	98	30 seconds	1
Denaturation	98	10 seconds	40X
Annealing	55 - 65	30 seconds	
Extension	72	1 minutes	
Final extension	72	2 minutes	1
Hold	4	∞	-

Table08: Gradient PCR using Q5 DNA polymerase of three different samples collected on three different dates

Component	**Quantity per 1x sample**	**Quantity per 15x sample**
MQ (Sterile water)	9 µl	135 µl
5x Q5 Buffer	3 µl	45 µl
WIV21_97F	1 µl	1.5 µl
WIV21_1834R	1 µl	15 µl
10 mM dNTP	0.3 µl	4.5 µl
Q5 Polymerase	0.2 µl	3 µl
Templates **Zs808/7** **Zs808/8** **Zs808/10** x RT	0.5 µl+ 14,5ul Mix	

Table09: PCR Test using Q5 DNA polymerase of the big pool and the three different samples collected on three different dates at Tusnád station

Component	**Quantity per 1x sample**	**Quantity per 8x sample**
MQ (Sterile water)	9 µl	72 µl
5x Q5 Buffer	3 µl	24 µl
WIV21_97F	1 µl	8 µl
WIV21_1834R	1 µl	8 µl
10 mM dNTP	0.3 µl	2.4 µl
Q5 Polymerase	0.2 µl	1,6 µl
Templates **Zs800/4 (Big pool)** **Zs803/10** **Zs803/11** **Zs803/12** x RT	0.5 µl+ 14,5ul Mix	

Table10: Gradient PCR using Q5 DNA polymerase

Stage	**Temperature °C**	**Time**	**Number of Cycle**
Initial Denaturation	98	30 seconds	1
Denaturation	98	10 seconds	40X
Annealing	62.5	30 seconds	
Extension	72	2 minutes	
Final extension	72	3 minutes	1
Hold	4	∞	-

3.3. Investigation of the results of the RT-PCR reaction (Agarose gel electrophoresis)

To test the results of the amplification of target DNA sequences, agarose gel electrophoresis was employed. This technique separates DNA fragments based on size, allowing for the identification of PCR products with the targeted size. Separating the PCR product, along with a DNA ladder as a size reference, and applying an electric field, the DNA fragments migrate through the agarose gel matrix. The presence of distinct bands corresponding to the expected size of the virus-specific amplicons confirms successful PCR amplification.

3.3.1. Agarose Gel Preparation

To prepare the **1.2%** agarose gel, 3.6 g of agarose powder was weighed and added to 300 mL of 1X TBE buffer. The agarose was solubilised and melted by heat in a microwave oven in short bursts, with gentle shaking between each heating cycle, until a clear, homogeneous solution was obtained. Ethidium bromide (EtBr), a fluorescent dye that intercalates into the DNA double helix, was added to the aliquot of the molten agarose solution at a final concentration of 0.5 µg/mL. The mixture was then poured into a gel casting tray containing a comb to create wells for sample loading. The gel was allowed to solidify at room temperature for approximately 20 minutes.

3.3.2. Setting Up the Electrophoresis Gel

The electrophoresis setup is identical for both the Q5 Polymerase and Phire Hot Start DNA Polymerase samples. The solidified agarose gel was carefully transferred to an electrophoresis

tank and submerged in 1X TBE buffer. 1.2% agarose gel was used to separate the approximately **1737/1505** base pair (bp) PCR product generated from the cDNA.

3.3.3. Gel Electrophoresis

5vl DNA molecular weight marker (DNA ladder 100bp + of Thermo Scientific) was loaded into the first well of the gel, followed by 10 µL of each PCR product (amplified at different annealing temperatures). Before loading 5 µL of DNA loading dye containing bromophenol blue as a tracking dye was added to each sample. Electrophoresis was performed at a constant voltage of 132-135 Volt allowing for the separation of DNA fragments based on their size, till the bromophenol blue dye reached the end of the gel (approximately 30-35 minutes).

Following electrophoresis, the gel was visualized under ultraviolet (UV) light to detect the ethidium bromide (EtBr)-stained DNA bands. A Bio-Rad Chemi Doc MP Imaging System was used to capture a digital image of the illuminated gel.

3.3.4. Purification of PCR fragments from agarose gel

To purify the PCR product of interest, a NucleoSpin Gel & PCR Clean-up kit (Macherey and Nagel) was employed. This purification step is essential for subsequent downstream applications, such as DNA sequencing or cloning. The gel slice containing the target DNA band was excised from the gel, using a sterile scalpel and transferred to a pre-weighed microcentrifuge tube. The gel was melted using NT1 buffer (200 ul was used for 100m g gel) at 50 C for 10 minutes. To maximise the melting after 2-3 minutes the solution was mixed by vortexing it. After dissolving the gel, the liquid containing the DNA was carefully pipetted into a NucleoSpin column sitting inside a 2 mL collection tube. The tube was spun in a centrifuge at high speed (11,000 × g) for 30 seconds to stick the DNA onto the column's filter. The leftover liquid was thrown away. Next, the column was washed twice with a special cleaning solution (Wash Buffer NT3, made by mixing ethanol with a concentrate) to remove any impurities—each wash involved spinning the column again at the same speed for 30 seconds. After washing, the column was spun one last time for a full minute to get rid of any leftover ethanol, which could interfere with later experiments.

To collect the purified DNA, 15–30 µL of a mild, slightly alkaline solution (Elution Buffer NE) was added to the column. After letting it sit for a minute at room temperature, another quick spin in the centrifuge released the DNA into a clean tube. For bigger DNA pieces (over 1,000 bp), we boosted recovery by warming the elution buffer to 70°C, letting it sit on the column for

5 minutes, and repeating the elution step once more. This process gave us clean, high-quality DNA ready for experiments like sequencing or cloning.

3.3.5. Cloning (Ligation and Transformation)

To clone the PCR product, a pJET1.2/blunt cloning vector was utilized. This vector, provided by Thermo Scientific, is capable of accepting DNA inserts ranging from 6 to 10 kilobase pairs (kb). A ligation reaction was set up in a 1.5 mL microcentrifuge tube, combining 7.5 µL of 2X reaction buffer, 0.75 µL of pJET1.2/blunt vector, 0.5 µL of T4 DNA ligase, 1 µL of nuclease-free water, and 5 µL of the purified PCR product. The ligation reaction was incubated at room temperature for 10 minutes, resulting in the formation of recombinant plasmids. The ligation mixture was then ready for transformation into competent cel

A vial containing 200 ul of competent E. coli DH5α cells was thawed on ice for 10-15 minutes. 100 µL of thawed competent cells were transferred to a sterile 15 ml polystirol tube, followed by the addition of 5 µL of the ligation mixture. The cells were incubated on ice for -20 minutes to facilitate DNA uptake. Heat shock transformation was performed by incubating the cells at 42°C for 30 seconds, followed by immediate transfer to ice to halt the heat shock process. 500 µL of SOC medium was added to each tube, and the cells were incubated at 37°C with shaking for 40 minutes to allow for recovery and expression of antibiotic resistance genes. Subsequently, 250 µL of the transformed cells were spread onto LB agar plates containing ampicillin. The plates were incubated at 37°C overnight to allow for the growth of transformed colonies.

3.3.6. Plasmid Purification for Sanger sequencing

3.3.6.1. liquid culture

Two hundred fifty microliters of ampicillin were added to a 250 mL Erlenmeyer flask containing LB medium. Three millilitres of this antibiotic-supplemented LB medium was then transferred to individual inoculation flasks. 3 colonies from each transformation plate were picked using sterile toothpicks and streaked onto fresh LB agar plates containing ampicillin. The toothpicks were also used to inoculate the 3 mL LB medium cultures. These cultures were incubated overnight at 37°C with shaking to allow for bacterial growth and plasmid amplification

3.3.6.2. Plasmid Purification

To isolate plasmid DNA, a NucleoSpin Plasmid kit (Macherey and Nagel) was utilized. 1.5 ml of the overnight culture was harvested by centrifugation at 8000 rpm for 3 minutes. The supernatant was discarded, and the cell pellet was resuspended in 250 µL of Buffer A1. Lysis Buffer A2 (250 µL) was added to the cell suspension, and the mixture was gently inverted six times to ensure complete lysis. After a 5-minute incubation at room temperature, 300 µL of Precipitation Buffer A3 was added, and the mixture was inverted gently to mix. The lysate was centrifuged at 11,000 x g for 5 minutes to precipitate cellular debris. The supernatant containing the plasmid DNA was transferred to a spin column and centrifuged at 11,000 x g for 1 minute. The flow-through was discarded, and the column was washed with 600 µL of Buffer A4 to remove impurities. The column was centrifuged at 11,000 x g for 1 minute and then after removing the flow through for a further 2 minutes to dry the silica membrane. Finally, plasmid DNA was eluted from the column with 50 µL of Buffer AE, after 1 minute room temperature incubation, followed by a 1-minute centrifugation at 11,000 x g. The purified plasmid DNA was stored at -20°C for subsequent analysis.

3.3.6.3. Testing the success of the ligation (Plasmid digestion with restriction enzymes)

To test, if the cloning process was successful, the purified plasmid DNA was subjected to restriction enzyme digestion. A reaction mixture containing 2 µL of 10X Tango Yellow Buffer, 0.1 µL of XhoI restriction enzyme, 0.1 µL of XbaI restriction enzyme, and 5,8 µL of nuclease-free water was prepared. 8 µL of this digestion mixture was added to 2 µL of the purified plasmid DNA, and the reaction was incubated at 37°C for 1 hour. The digested DNA fragments were analyzed by agarose gel electrophoresis to visualize the expected band pattern. Plasmids that exhibited the correct restriction pattern were selected and sent for Sanger sequencing to confirm the presence and orientation of the inserted DNA fragment.

4. Analysis of the sequences

The result of the Sanger sequencing arrived as a chromatogram, which could be analyzed using Chromas 2.6.5. This software is a user-friendly tool for visualizing and editing DNA sequence chromatograms. The software allows for the extraction of sequences in FASTA format, which can be subsequently analyzed using other bioinformatics tools. In the case of this study, Chromas 2.6.5 was used to assess the quality of the sequences.

To analyze the genetic sequences from Sanger sequencing, we first cleaned up the raw data using Chromas. This step involved trimming off poor-quality sections at the ends of reads, fixing unclear or ambiguous bases, and removing any "N" nucleotides to improve accuracy. Next, we stitched together overlapping reads to build complete sequences and manually refined them using ApE (Davis, 2011). This included fixing mismatches, labeling important features (like genes), merging shorter sequence fragments, and ensuring the forward and reverse strands of the cloned product Zs808I10I3 aligned perfectly without gaps.

To identify what viruses these sequences belonged to, we compared them to the NCBI's global genetic database and specialized viral reference databases using BLASTn(Altschul, Gish, Miller, Myers, & Lipman, 1990). For cloned PCR products linked to Wuhan Insect Virus 21, we adjusted the search by selecting the **"**somewhat similar sequences**"** option in BLASTn, which helped find close relatives of the virus. Matches were ranked based on how well they aligned (alignment scores), how much of the sequence they covered (query coverage), and their relevance to known viruses.

To double-check our findings, we reran the analysis using the reverse complements of the sequences and compared results across multiple databases. This cross-validation step ensured no mistakes crept in and confirmed the reliability of our identifications.

5. Phylogenetic analysis of viruses detected by HTS

To reconstruct the phylogenetic relationships of viruses identified via HTS, aligned viral sequences (conserved regions and whole genomes) were imported into Geneious Prime (v2023.1.1). Sequences were aligned using the built-in MUSCLE alignment tool, followed by manual curation to remove gaps and poorly aligned regions. The neighbor-joining model using the Jukes-Kantor algorithm was employed for tree construction. Node support was assessed using 1,000 bootstrap replicates, with nodes retaining ≥70% bootstrap values considered statistically robust. Genetic distances were scaled accordingly. Final trees were exported in JPEG format for visualization (Dotmatics, 2023)

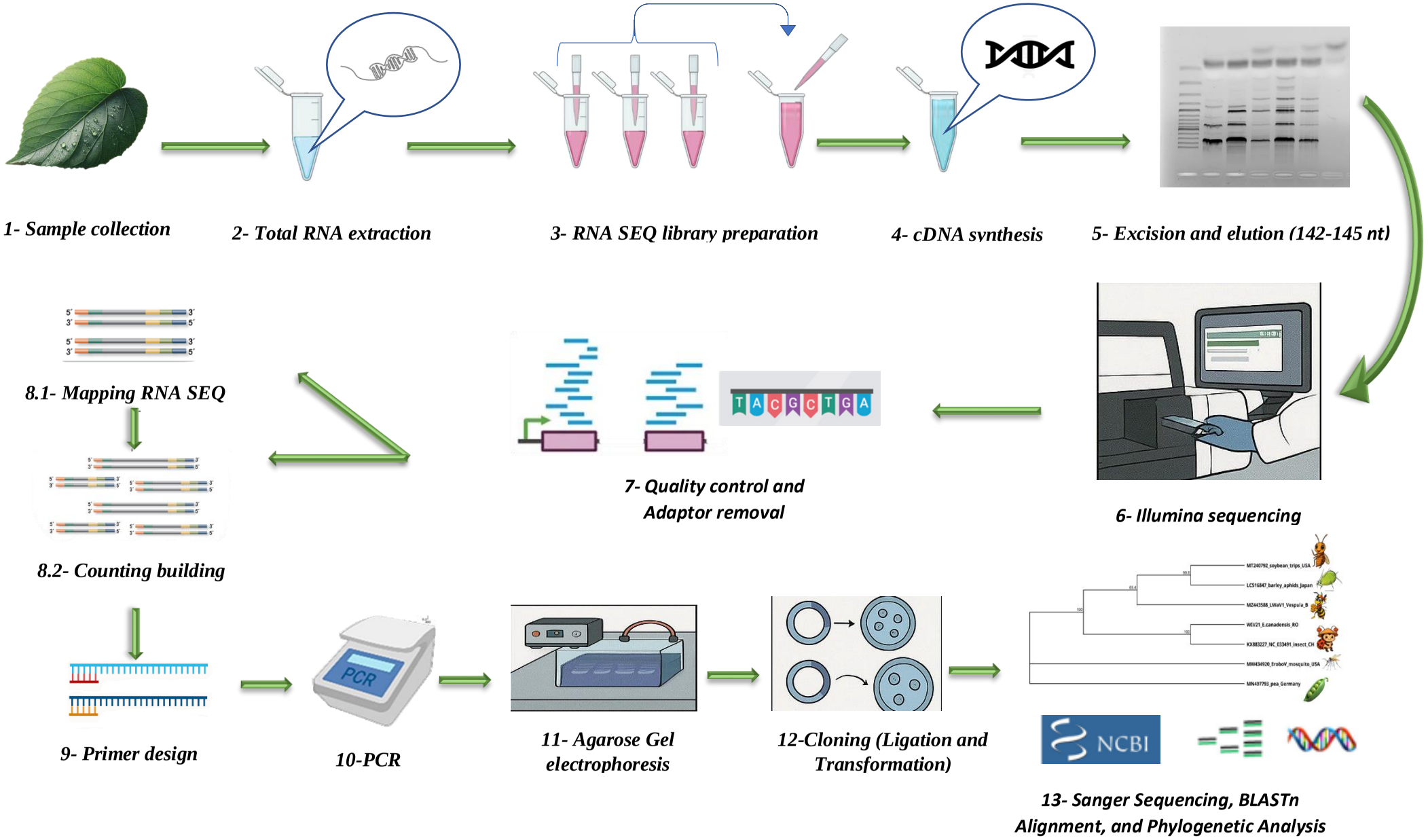


Figure 07: Comprehensive Methodology for Viral Sequence Analysis and Phylogenetic Reconstruction (Original,2025)

RESULTS AND DISCUSSION

The study started in 2016 with the sampling, RNA extraction and continued by HTS of ribodepleted RNA from weed plant samples ordered as a service from NOVOGEN. My contribution to this research project commenced at the stage of data analysis from the point where all above was ready and I started to analyze the resulting sequence data. Utilizing the CLC Genomic Workbench, I performed a comprehensive analysis, including BLAST searches against reference genomes of different sequence databases including Land Plant- infecting Viruses, Fungi Viruses, Invertebrate Host Viruses and Slovenian Weeds Viruses, to identify potential viral infections within the weed samples.

1. Results of the initial statistics

Initial RNA sequencing analysis revealed 20,313,050 total reads, of which 500,000,94 reads were retained after quality trimming and 109,401 reads discarded due to low quality or adapter contamination. Following this, de novo assembly of the trimmed reads generated 56,260 contigs, with lengths ranging from 300 bp (minimum) to 15,821 bp (maximum) and an average length of 859 bp.

To characterize the viromes and identify potential viral pathogens, the contigs were matched to the NCBI reference genome database, resulting in a list of viral pathogens potentially occurring in the sample, indicating their probability as an E-value. I sorted the obtained results based on the corresponding E-value, and then continued the further tests with the most probable pathogens, i.e. those with an E-value of 0. To check all of the SNP's present in the sequenced viral variant I created consensus sequences from the mapped reads which were separately aligned to the reference genome of the given virus downloaded from the NCBI database.

The analysis resulted a list of viral hits exhibiting low read counts while others demonstrated high read abundance. A curated list of viral candidates was subsequently generated.

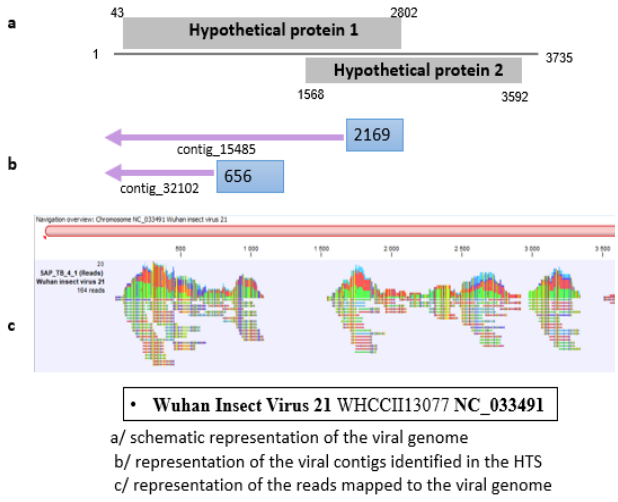
To quantify the extent of information about the presenting viral sequence, coverage of the viral genome: the percentage of RNA-seq reads aligning to each viral hit was calculated table 11 below illustrates the results.

Table 11: Bioinformatic analysis, Using CLC and NCBI Blast obtained for the RNA-seq library based on the lowest E-value (Original,2025)

Libraries	Virus abbreviation	GenBank ID	Number of contigs	Number of mapped reads redundant reads	Coverage of the viral genome (%)
LAND PLANTS VIRUSES 12\05\2023	WHCCII13077	NC_033491	2	164	80%
	WHCCII13077	NC_033481	1	145	84%
	WHAV2	NC_028382	3	107	90%
	WHAV2	NC_028386	2	697	97%
	WHAV2	NC_028383	3	103	99%
	WHAV2	NC_028387	2	93	89%
	CMV	NC_001440	1	13	47%
	AgLV	NC_022127	1	27	28%
	BVF(BBV-3X)	NC_029303	1	15	5%
SLOVENIAN WEEDS VIRUSES 2023	SnIV-1	OL472060	2	74	81%
	SnIV-1	OL472062	2	52	89%
	CMV	OL472039	1	13	27,57%
	EnOLV82	OL472300	1	25	38%
FUNGI VIRUSES 12/05/2023-27/02/2024	AaMV1	NC_030747	1	51	83%

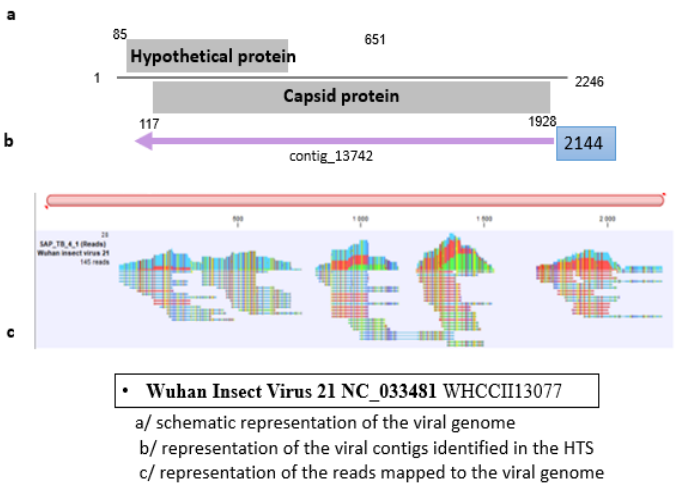
To be able to reliable investigate further the initial bioinformatics analysis, I explored more into identifying the viral hits. This involved research at NCBI databases to gather information on the origin, host range, and genomic characteristics of each potential virus. It aims to track the origins of the identified viral sequences and provide a more comprehensive understanding of their potential impact. A comprehensive table below is compiled to summarize details on the RNA type (linear or circular), the natural host of the virus, its geographical origin, and the orientation of its genome.

In order to visualize accurately our data, we mapped the contig sequences to the known sequence of viruses to see how entirely their genome is covered by RNA. The figures from (08 to 13) below represents and illustrates all virus hits.



• **Wuhan Insect Virus 21** WHCCII13077 **NC_033491**

a/ schematic representation of the viral genome
b/ representation of the viral contigs identified in the HTS
c/ representation of the reads mapped to the viral genome



• **Wuhan Insect Virus 21 NC_033481** WHCCII13077

a/ schematic representation of the viral genome
b/ representation of the viral contigs identified in the HTS
c/ representation of the reads mapped to the viral genome

Figure08: Analysis of Viral Genome Whan Insect Virus21(NC_033491-NC_033481) and HTS Data

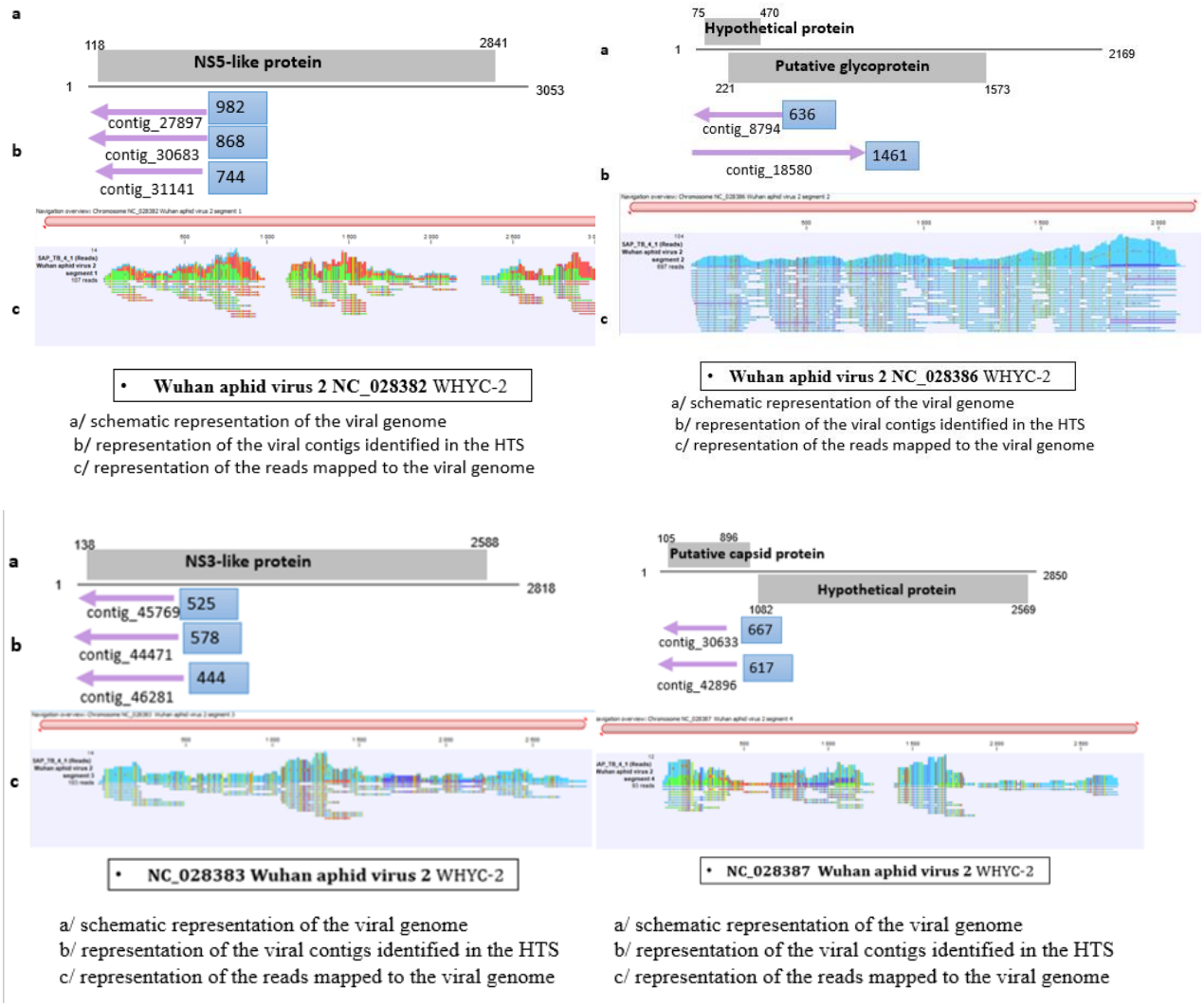


Figure09: Analysis of Viral Genome Whan Aphid Virus 2 (NC_028382-NC_028386-NC_028383- NC_028387) and HTS Data

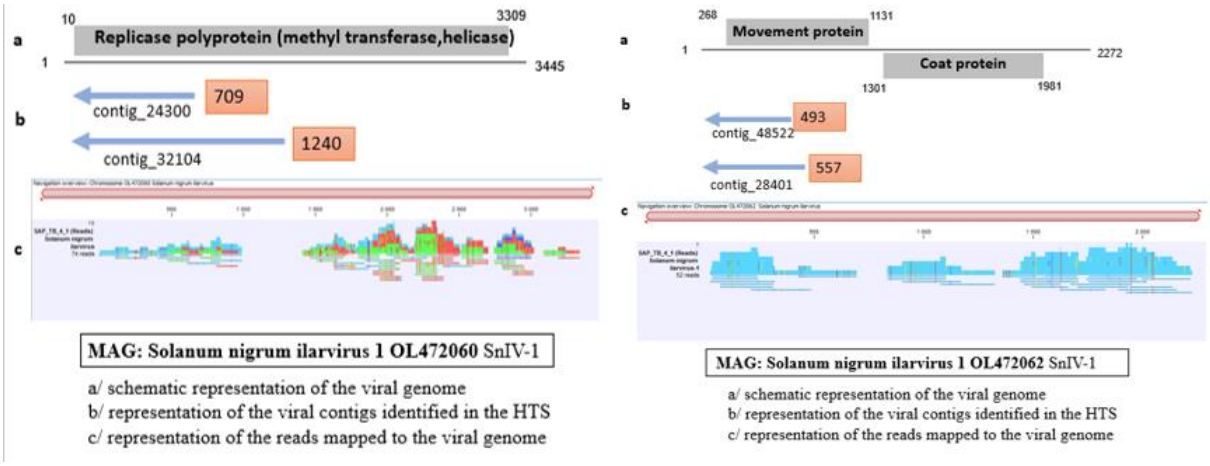


Figure10: Analysis of Viral Genome Solanum nigrum ilarvirus 1 (OL472060- OL472062) and HTS Data

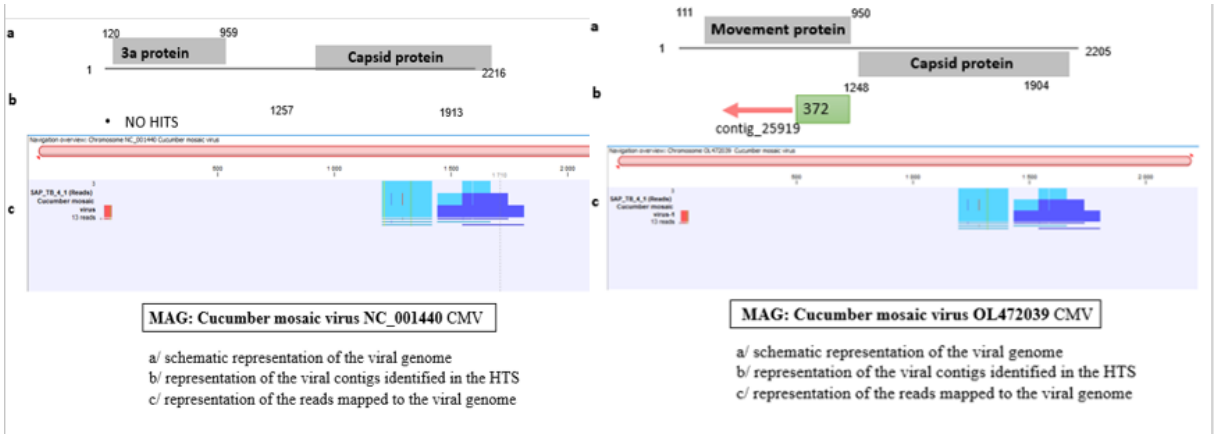


Figure11: Analysis of Viral Genome Cucumber mosaic virus (NC_001440- OL472039) and HTS Data

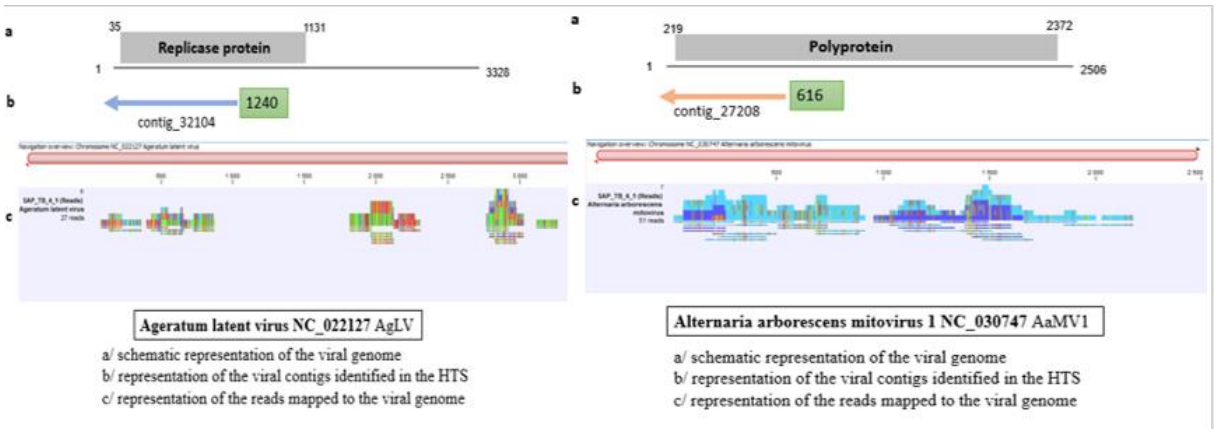


Figure12: Analysis of Viral Genome Ageratum latent virus (NC_022127) Alternaria arborescens mitovirus 1 (NC_030747) and HTS Data

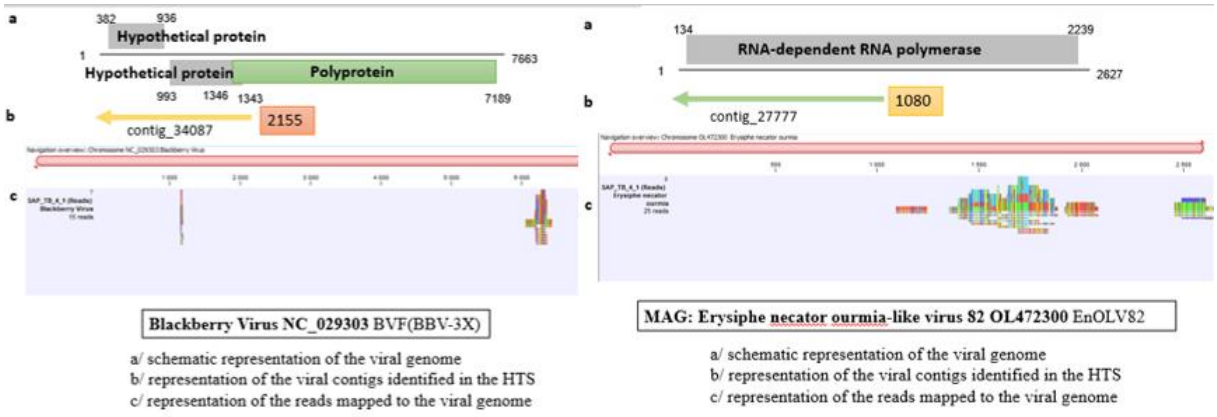


Figure13: Analysis of Viral Genome Blackberry Virus (NC_029303) Erysiphe necator ourmia-like virus 82 (OL472300) and HTS Data

2. Checking the quality of the cDNA of *E. canadensis* samples

To validate the presence of these two viruses RT-PCR has been carried out. For RT-PCR reactions cDNA has to be used as a template. cDNAs originating from the investigated sample had been available in the lab, therefore I started my validation work by checking the quality of

the cDNA templates for virus screening. To do this a PCR reaction detecting the actin mRNA in the sample was carried out using Vvactine primers with grapevine and *E. canadensis* samples as mentioned in the protocol above. Actin is a highly conserved mRNA, the primers designed for grapevine could amplify actin from a wide range of plants. However, we did not know if these primers would amplify actin from *E. canadensis* as well and if they do at which annealing temperature. As a positive control we used grapevine cDNA.

Gradient PCR is an adaptation of conventional PCR that facilitates the optimization of the PCR process. The annealing temperature must be set high enough to minimize non-specific products while maintaining the yield of the target product. By setting a range of annealing temperatures, the optimal temperature of the annealing step of the PCR reaction, which can vary according to the used polymerase, can be identified.

The actin test, using "sense and antisense" actin primers, confirmed the successful synthesis of cDNA by producing amplified products. Positive controls validated the reliability of our process.

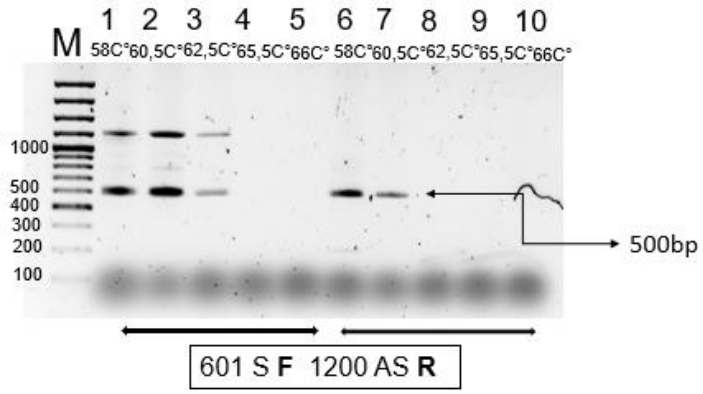


Figure14: Gel Electrophoresis Results for Actin Test and Primer Check (NC_033491-NC_033481) – (M) marker

The samples were checked using Vvactine primers pairs (601S F- 1200 AS R) from previous grapevine (lanes 1 to 5) and our sample template, *E. canadensis* (lanes 6 to 10) for both viruses the expected size is 500bp which suggests Positive Results, figure14 above Represents the gel electrophoresis results showing the actin test.

3. RT PCR validation of the presence of WIV21

3.1. Primer design

The primer design process for Wuhan Insect Virus 21 targeted conserved genomic regions to ensure specificity and amplification efficiency. Using bioinformatics tools (e.g., NCBI Primer-BLAST, Geneious Prime), primers were designed to flank a ~500 bp region of the viral genome, avoiding secondary structures and regions of high GC content. The forward and reverse primers were validated *in silico* for specificity against the NCBI nucleotide database to minimize cross-reactivity with non-target sequences. Melting temperatures (Tm) were optimized to ensure compatibility with standardized PCR protocols. Experimental validation included testing primer pairs on cDNA synthesized from *E. canadensis* samples, followed by gel electrophoresis and Sanger sequencing to confirm amplification of the intended viral sequence.

3.2. RT PCR validation of the presence of WIV21(First experiment)

3.2.1. Optimization of the annealing temperature for the virus specific primers

A gradient PCR using Q5 DNA polymerase using *E. canadensis* cDNA as a template did not yield the expected 507 bp product. Subsequent PCR amplifications were performed using three different cDNA templates (*E. canadensis* PCR, VS6/8 produced by RT, and Zs 747/8 produced using Maxima). These experiments confirmed a low percentage of viral presence, as indicated by the absence of distinct bands on the gel Presented below in figure 15.

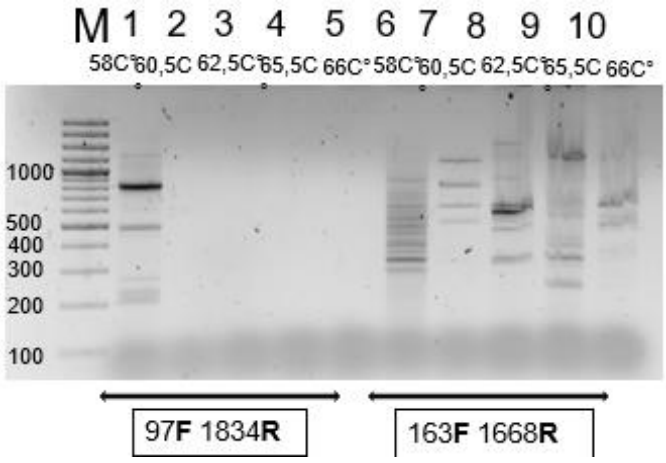


Figure15: PCR Gradient with the use of Q5 Polymerase enzyme to obtain the optimization annealing temperature of *E. canadensis* sample for Wuhan insect viruses (NC_033491-NC_033481).

The virus-specific product using specific primers to detect WHCCII13077 should yield a product around 1737 bp when using primer pairs (97F - 1834R), and 1505 bp is expected when using primer pairs (163F - 1668R). The gradient-PCR was performed with the Q5 polymerase enzyme, showing a negative reaction result, indicating that the sample may not contain the virus or, if present, contains it at a low concentration. Using the Q5 enzyme was not very successful, as we had some amplifications, but they were not at the correct size. Therefore, we opted to use the Phire Hot Start II DNA Polymerase,(Figure 16).

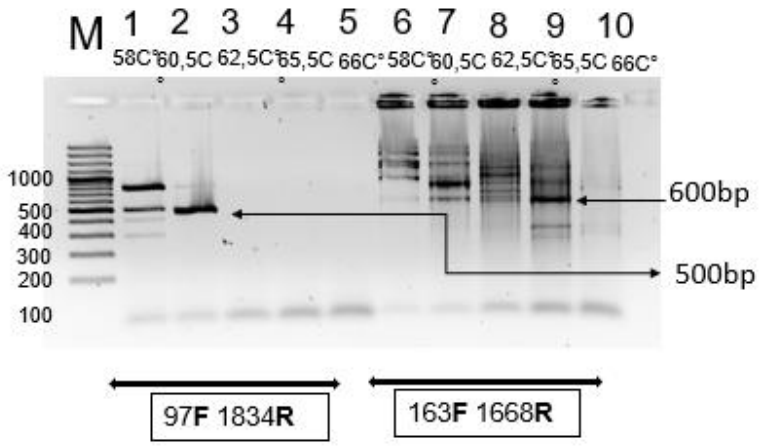


Figure16: PCR Gradient with the use of Phire hot start DNA polymerase enzyme to obtain the optimization annealing temperature of *E. canadensis* sample for Wuhan insect viruses (NC_033491 -033481).

Samples 1 to 5 were handled with the use of primer pairs (97F - 1834R) which shows a band around 500 bp. Samples 6 to 10 were handled with the use of primer pairs (163F - 1668R) which shows a band around 600bp. These bands illustrated positive results.

3.2.2. Purification, Cloning and analyzing of PCR fragments Results

To check if the amplified part could be viral origin we decided to determine the sequence of the amplified product. To do this, the products have been purified from the gel and cloned into pJET1.2 vector demonstrated in figure 17.

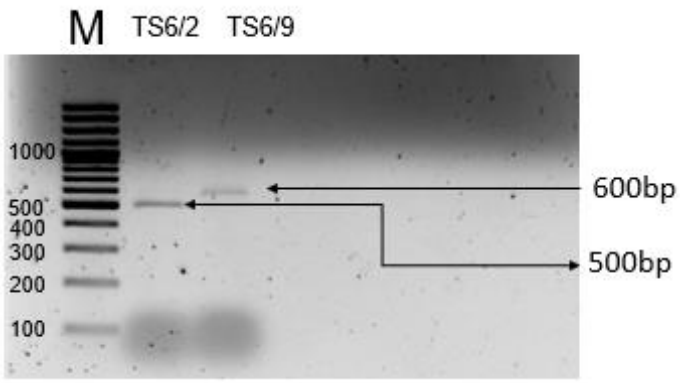


Figure 17: DNA purification results

Following this, the purified products was ligated, and transformed into the vector. The result of the cloning was checked on three resulted and purified plasmids by cleavage with XhoI and Xbal restriction enzyme digestion. Results were:

3 clones have been produced from sample TS6/2 (TS6/2 1, TS6/2 2, TS6/2 3) and 3 clones have been produced from sample TS9/2(TS9/2 1, TS9/2 2, TS9/2 3) as the figure 18 below indicates. it.

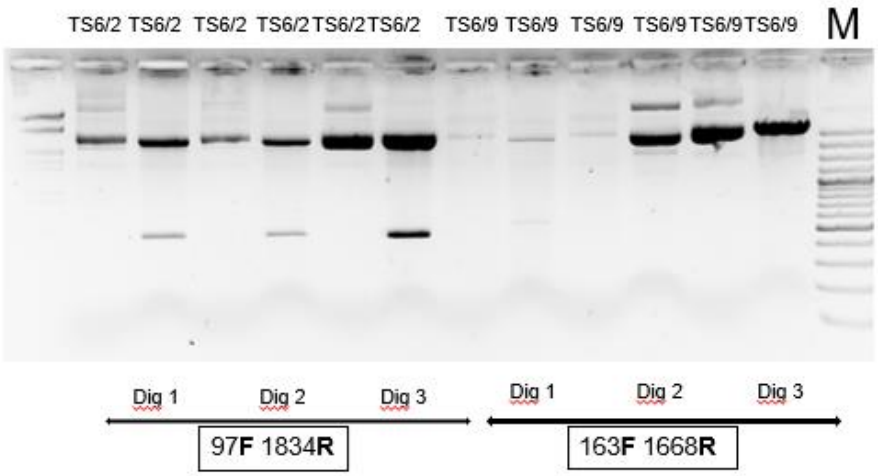


Figure 18: DNA cloning and digestion Results (dig: digestion, Ø of dig: not digested)

To proceed with bioinformatics analysis, we sent the cloned samples to be Sanger sequenced. The resulting sequences were visualized using Chromas 2.6.6 software. The sequence was analyzed in order to identify its possible origin from Wuhan insect viruses (NC_033491-NC_033481), figure 19 shows part of the obtained long sequence .

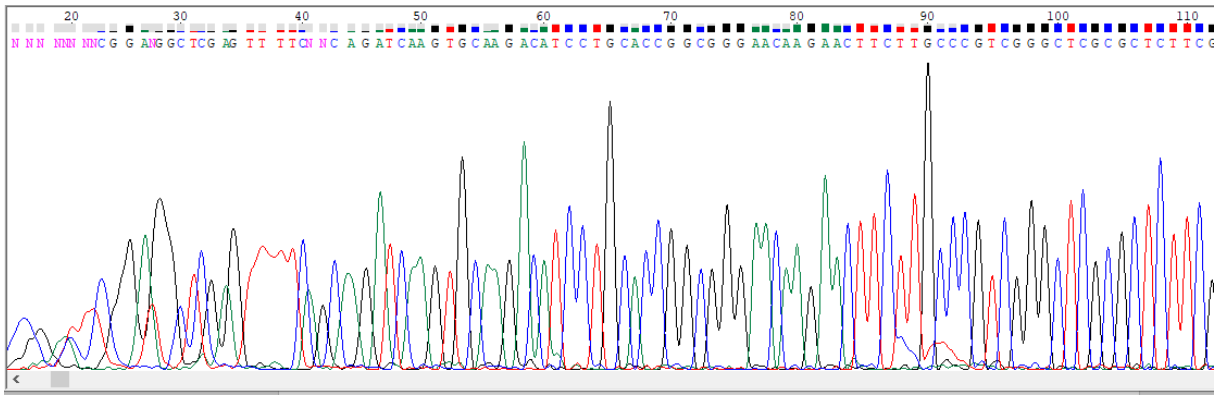


Figure19: chromatogram of received sequence opened in Chromas 2.6.6

Based on the chromatogram from the Sanger sequencing visualized, the sequence was subjected to bioinformatics analysis. This involved aligning the obtained sequence and checking for significant matches or similarities. The analysis concludes that the sequence may originate from the plant genome.

3.3.RT PCR validation of the presence of WIV21(Second experiment)

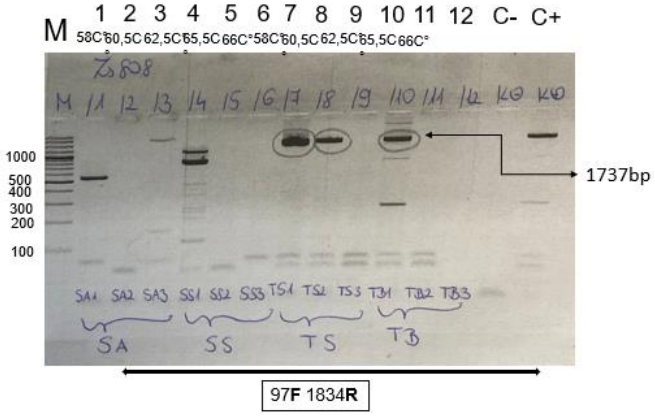


Figure 20: PCR Gradient with the use of Q5 Polymerase enzyme to check the optimizing annealing temperature of E. canadensis sample for Wuhan insect viruses 21 (NC_033491).

To investigate thoroughly my supervisor attempted another gradient PCR using the Q5 enzyme with Wuhan insect virus 21 primer pairs 97F and 1834R. This time, we obtained a good product at the correct size and position1737bp, (Figure20).

I then proceeded to check cDNA for the presence of the virus, as illustrated in figure 21 below.

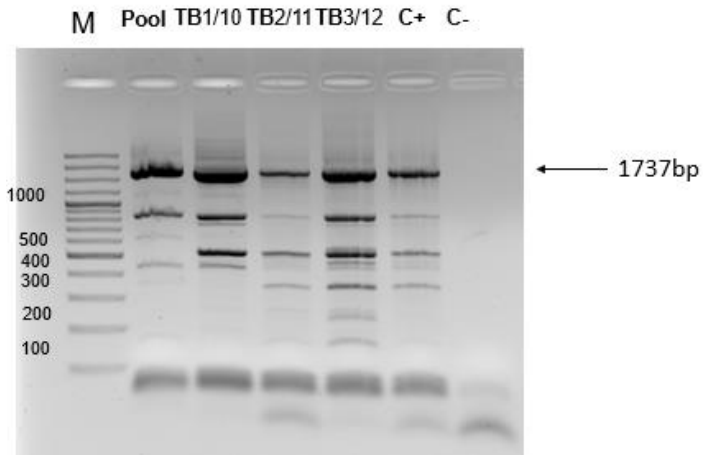


Figure21: PCR Gradient with the use of Q5 Polymerase enzyme to obtain the optimization annealing temperature of *E. canadensis* sample for Wuhan insect viruses 21 (NC_033491-NC_033481).

Utilizing specific primers designed to detect Wuhan Insect Virus 21(WHCCII13077), a gradient PCR was conducted to amplify a target product of 1737 base pairs. The primer pair pairs (97F - 1834R). was employed in conjunction with Q5 polymerase enzyme and four distinct cDNA templates. These templates included cDNA derived from *E. canadensis* samples collected on 3 separate months at the Tusnád station, as well as a "Big Pool" cDNA sample. Additionally, a positive control mix and a negative control mix, using molecular-grade water, were included in the experiment.

The results of the gradient PCR demonstrated a positive reaction across the cDNA samples, confirming the presence of Wuhan Insect Virus 21 within the *E. canadensis* specimens. Following this, the cloned samples were subjected to Sanger sequencing. The generated sequences were visualized and analyzed using Chromas 2.6.6 software. Subsequent sequence comparison aimed to determine their potential origin by aligning them with reference sequences of Wuhan insect viruses (NCBI accessions: NC_033491-NC_033481). As shown in figure22 chromatogram of received sequence Forward and Reverse sequence.

Figure22: chromatogram of received sequence opened in Chromas 2.6.6

4. Phylogenetic Analysis of WIV21 in *E. canadensis*

The phylogenetic tree results illustrate that Wuhan Insect Virus 21 (WIV21) has been described from various countries, including the United States of America, Japan, Belgium, China, and Germany, and from different origins such as soybean, pea, and barley. Multiple alignments in GenBank Revealed that there are two additional viruses under different names: MZ443588_LWaV1_Vespula vulgaris_BG and MW434920_EroboV_mosquito_USA. The Pairwise sequence comparison of these two variants showed that they are 87.5% identical, (Table 11).

Table 12: Percentage Similarity Matrix of Genetic Sequences Across Various Species and Locations

	KX883227 NC_033491 insect_CH	LC516847 mixture	MN497793 pea_Germany	MT240792 soybean	MW434920 EroboV mosquito_USA	MZ443588 LWaV1_Vespula	WIV21 E.canadensis_RO
KX883227 NC_033491_insect_CH		81,734	81,501	81,443	81,453	82,141	83,43
LC516847 mixture	81,734		85,573	94,415	85,523	88,54	82,616
MN497793 pea_Germany	81,501	85,573		85,748	94,826	87,784	80,64
MT240792 soybean	81,443	94,415	85,748		85,64	87,9	81,686
MW434920 EroboV_mosquito_USA	81,453	85,523	94,826	85,64		87,5	81,628
MZ443588 LWaV1_Vespula	82,141	88,54	87,784	87,9	87,5		82,442
WIV21 E.canadensis_RO	83,43	82,616	80,64	81,686	81,628	82,442	

The phylogenetic tree shows that several viruses closely related to WIV21, though named differently, are present across different branches and are associated with various arthropod hosts. Our sequence of the WIV21 *E. canadensis* variant (WIV21_*E. canadensis*_RO) clusters with these isolates, with its closest relative being the variant sequenced in China from insect samples (KX883227_NC_033491_insect_CH), indicating a close genetic relationship despite naming differences., (Figure23).

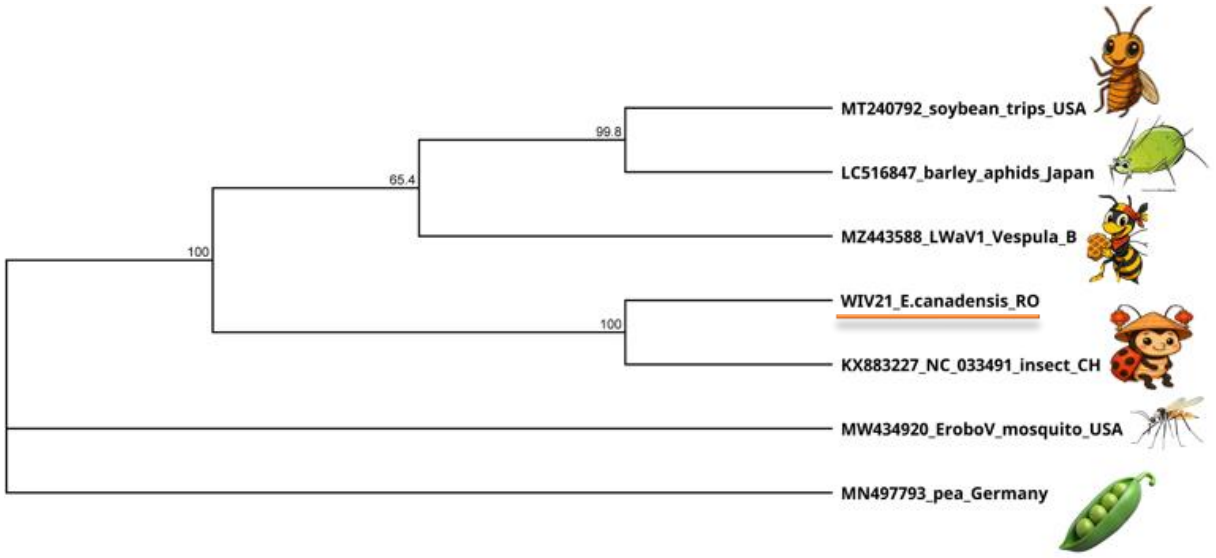


Figure 23: Phylogenic Tree of Wuhan insect virus 21 in *E. canadensis*

CONCLUSION

In conclusion, *E. canadensis*, a ubiquitous plant exhibiting both medicinal and ecological significance, presents a complex case study in weed management due to its rapid colonization of disturbed habitats and increasing herbicide resistance. However, its notable susceptibility to viral infections, as revealed through recent high-throughput sequencing (HTS) analyses, offers a critical avenue for investigation. This study aimed to comprehensively characterize the viral communities (viromes) within 30 plant samples of *E. canadensis* through molecular detection via RT-PCR, thereby validating HTS findings and exploring diverse viral sources, including the targeted Wuhan Insect Virus 21 and Wuhan Aphid 2.

To achieve this, novel primers were designed based on RNA reads generated during bioinformatic analysis, utilizing CLC Workbench and NCBI BLAST databases, enabling specific amplification of the virus. Subsequent purification, cloning, and Sanger sequencing of PCR products were subjected to rigorous bioinformatic analyses using Chromas and BLASTn. The gradient PCR was used to optimize the annealing temperature of the PCR. The results demonstrated a positive reaction across the cDNA samples, confirming the presence of Wuhan Insect Virus 21 within the *E. canadensis* specimens. This was further validated by Sanger sequencing and sequence alignment with reference Wuhan insect viruses.

WIV21's phylogenetic placement links it to globally dispersed crop-infecting isolates (soybean, pea, barley) across five countries, with 87.5% identity to insect-associated viruses (LWaV1_Vespula, EreboV_mosquito). Despite genetic parallels, WIV21 forms a distinct lineage (bootstrap 100), closely related to an insect-derived Chinese variant (KX883227), implicating unresolved host-jumping potential. These findings highlight WIV21's ecological divergence from arthropod viruses, urging re-evaluation of its taxonomy and transmission mechanisms.

In conclusion, WIV21's phylogenetic placement as a plant-associated virus highlights its role in shaping plant-virus interactions in non-crop angiosperms. Its distinct transmission ecology underscores the diversity of plant virus dissemination, emphasizing the need for continued molecular surveillance of viral reservoirs in *E. canadensis* within agroecosystems.

SUMMARY

Plant viruses represent major biotic threats to global agriculture, inflicting substantial economic damage and undermining efforts to achieve sustainable farming systems, The analysis of viromes—complex assemblages of viruses coexisting in ecological niches—aims to unravel their genetic diversity, host interactions, and functional dynamics, in agroecosystems

This research employs high-throughput sequencing (HTS) and advanced bioinformatics tools to analyze viral genetic material from environmental samples, comprehensively characterizing viral communities—including their genetic diversity, structural features, and potential impacts on plant health.

RNA was extracted from E. canadensis plants sampled at Tusnád, Romania. These samples were subjected to high-throughput sequencing (HTS) on the Illumina platform. Bioinformatics analysis was performed using CLC Genomic Workbench, contigs Hits were prioritized based on alignment confidence (E-value), with E = 0. Consensus sequences of these viruses were generated from contigs and aligned to their respective NCBI reference genomes for validation.

To confirm HTS-detected viruses, RT-PCR primers were designed using HTS-derived contigs and reference viral genomes. PCR amplicons were electrophoresed on agarose gels, and target bands were purified. Purified fragments were ligated into cloning vectors, transformed into competent cells, and subjected to plasmid purification. Sanger sequencing of cloned inserts produced chromatograms, which were analyzed to verify the presence of viral sequences identified via HTS.

The analysis revealed viral candidates with varying read abundances, ranging from low to high, leading to a curated list of targets. Among these, Wuhan Insect Virus 21 was prioritized for validation. Virus-specific primers amplified products of ~500 bp and ~600 bp Cloned amplicons were Sanger-sequenced, and the resultant chromatograms were analyzed to identify sequence origins. In a subsequent experiment, gradient PCR targeting a 1,737 bp fragment yielded positive reactions across cDNA samples, verifying Wuhan Insect Virus 21 in E. canadensis. Phylogenetic analysis of sequenced data further contextualized these findings. However, deeper investigation into E. canadensis-virus interactions remains necessary.

These findings advance understanding of plant-virus ecology. Functional insights into viral prevalence and genetic relationships underscore the broader value of metagenomics in uncovering hidden viral diversity within agroecosystems, emphasizing its role in proactive pathogen surveillance and ecological risk assessment.

Bibliographic References

Adeleke, B. S., Fadiji, A. E., Ayilara, M. S., Igiehon, O. N., Nwachukwu, B. C., & Babalola, O. O. (2022). Strategies to enhance the use of endophytes as bioinoculants in agriculture. *Horticulturae*, *8*(6), 498.

Al-Snafi, A. E. (2017). Pharmacological and therapeutic importance of E. canadensis (Syn: Conyza canadensis). *Indo American Journal of Pharmaceutical Sciences*, *4*(2), 248-256.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Bachouche, N., Lamri, N., Berghout, K., Oultaf, L., Metna, A. A. F., Bourfiss, N., & Akkouche, S. (2024). Agrochemical protection practices for market gardeners and environmental risks on farms in Bouira (Northern Algeria). *Glasnik Srpskog geografskog drustva*, *104*(2), 185-196.

Bajwa, A. A., Mahajan, G., & Chauhan, B. S. (2015). Nonconventional weed management strategies for modern agriculture. *Weed science*, *63*(4), 723-747.

Balas, P., Pargi, S., Lakhani, A., Mehta, T., & Bambhaniya, V. (2023). Inter and Intra Row Weed and Its Effect: A Review. *Current Journal of Applied Science and Technology*, *42*, 97-105. doi:10.9734/cjast/2023/v42i484336

Bhattacharya, S., Hernández, F., Alves, M. F., Machado, R. M., Sun, Y.-Y., Wang, M.-R., . . . Hao, J.-H. (2022). Genetic diversity and population structure of invasive and native populations of E. canadensis L. *Journal of Plant Ecology*, *15*(4), 864-876.

Biolabs, N. E. (2025). Q5® High-Fidelity DNA Polymerases. Retrieved from https://www.neb.com/en/products/pcr-qpcr-and-amplification-technologies/q5-high-fidelity-dna-polymerases

Bleidorn, C. (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and biodiversity*, *14*(1), 1-8.

Byron, M., Treadwell, D. D., & Dittmar, P. J. (2019). Weeds as Reservoirs of Plant Pathogens Affecting Economically Important Crops: HS1335, 9/2019. *EDIS*, *2019*(5), 7-7.

Chauhan, A., Pandey, N., & Jain, N. (2021). A review of methods for detecting single-nucleotide polymorphisms in the Toll-like receptor gene family. *Biomarkers in Medicine*, *15*(13), 1187-1198.

Davis, M. W. (2011). ApE (A Plasmid Editor). Retrieved from https://jorgensen.biology.utah.edu/wayned/ape/

Dotmatics. (2023). Geneious Prime User Manual. (2023). *Biomatters Ltd.* Retrieved from https://manual.geneious.com/en/latest/

FAO. (2019). What is Conservation Agriculture? Retrieved from https://www.fao.org/conservation-agriculture/overview/what-is-conservation-agriculture/en/

Farooq, M., Flower, K., Jabran, K., Wahid, A., & Siddique, K. H. (2011). Crop yield and weed management in rainfed conservation agriculture. *Soil and tillage research*, *117*, 172-183.

Flora, V. H. o. B. (2018). Canadian Fleabane (Conyza canadensis). Retrieved from http://www.herbiguide.com.au/Descriptions/hg_Canadian_Fleabane.htm

Garibaldi, L. A., Steffan-Dewenter, I., Kremen, C., Morales, J. M., Bommarco, R., Cunningham, S. A., . . . Greenleaf, S. S. (2011). Stability of pollination services decreases with isolation from natural areas despite honey bee visits. *Ecology letters*, *14*(10), 1062-1072.

Gerhards, R., Andujar Sanchez, D., Hamouz, P., Peteinatos, G. G., Christensen, S., & Fernandez-Quintanilla, C. (2022). Advances in site-specific weed management in agriculture—A review. *Weed Research*, *62*(2), 123-133.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature reviews genetics, 17*(6), 333-351.

Grbelja, J., Erić, Ž., & Jeknić, Z. (1988). E. canadensis L.-a potential source of infection by the tomato bushy stunt virus for cultivated plants.

Hasiów-Jaroszewska, B., Boezen, D., & Zwart, M. P. (2021). Metagenomic studies of viruses in weeds and wild plants: a powerful approach to characterise variable virus communities. *Viruses, 13*(10), 1939.

Jang, W. S., Lim, D. H., Choe, Y. L., Nam, J., Moon, K. C., Kim, C., . . . Lim, C. S. (2022). Developing a multiplex loop-mediated isothermal amplification assay (LAMP) to determine severe fever with thrombocytopenia syndrome (SFTS) and scrub typhus. *PLoS One, 17*(2), e0262302.

Jeong, H., Lee, S., Ko, J., Ko, M., & Seo, H. W. (2022). Identification of conserved regions from 230,163 SARS-CoV-2 genomes and their use in diagnostic PCR primer design. *Genes & Genomics, 44*(8), 899-912. doi:10.1007/s13258-022-01264-7

Jiménez-Arias, D., Morales-Sierra, S., Borges, A. A., Herrera, A. J., & Luis, J. C. (2022). New biostimulants screening method for crop seedlings under water deficit stress. *Agronomy, 12*(3), 728.

Kanapiya, A., Amanbayeva, U., Tulegenova, Z., Abash, A., Zhangazin, S., Dyussembayev, K., & Mukiyanova, G. (2024). Recent advances and challenges in plant viral diagnostics. *Frontiers in Plant Science, 15*, 1451790.

Khan, M., Shah, S. H., Salman, M., Abdullah, M., Hayat, F., & Akbar, S. (2023). Enzyme-linked immunosorbent assay versus chemiluminescent immunoassay: A general overview. *Glob. J. Med. Pharm. Biomed. Update, 18*(1).

Kovalskaya, N., & Hammond, R. W. (2014). Molecular biology of viroid–host interactions and disease control strategies. *Plant Science, 228*, 48-60.

Kusumavathi, K., Sarkar, S., Ali, M. A., Bera, S., Krishna, V. J., Maity, S., . . . Bandopadhyay, P. (2025). Effect of diverse crop establishment techniques and weed management approaches in rice-chickpea cropping systems on soil attributes. *BMC Plant Biology, 25*(1), 372.

Kyrychenko, A., Shcherbatenko, I., & Kovalenko, A. (2021). Viruses of wild plants and current metagenomic methods for their investigation. *Cytology and Genetics, 55*(3), 248-255.

Laxman Navi, N. M., Kushal, Manjunath Madhukar Mopagar. (2024). Weed management strategies for conservation agriculture: A review. *International Journal of Research in Agronomy, 7*(4), 313-321. doi:10.33545/2618060X.2024.v7.i4e.569

Manjunatha, L., Rajashekara, H., Uppala, L. S., Ambika, D. S., Patil, B., Shankarappa, K. S., . . . Mishra, A. K. (2022). Mechanisms of microbial plant protection and control of plant viruses. *Plants, 11*(24), 3449.

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., & Candresse, T. (2018). Application of HTS for routine plant virus diagnostics: State of the art and challenges. *Frontiers in Plant Science, 9*, 1082.

Martínez-Ochoa, N., Langston, D. B., Mullis, S. W., & Flanders, J. T. (2003). First report of pepper mild mottle virus in jalapeno pepper in Georgia. *Plant Health Progress, 4*(1), 26.

Mullis, K. B., & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Methods in enzymology* (Vol. 155, pp. 335-350): Elsevier.

Olmos, A., Boonham, N., Candresse, T., Gentit, P., Giovani, B., Kutnjak, D., . . . Moreira, A. (2018). High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities. *EPPO Bulletin, 48*(2), 219-224.

Pabla, S., & Pabla, S. (2008). Real-time polymerase chain reaction. *Resonance, 13*, 369-377. doi:10.1007/s12045-008-0017-x

Ponce-Cusi, R., Bravo, L., Paez, K. J., Pinto, J. A., & Pilco-Ferreto, N. (2024). Host-pathogen interaction: biology and public health. In *Host-Pathogen Interactions: Methods and Protocols* (pp. 3-18): Springer.

Qin, T.-X., You, E.-M., Zhang, M.-X., Zheng, P., Huang, X.-F., Ding, S.-Y., . . . Tian, Z.-Q. (2021). Quantification of electron accumulation at grain boundaries in perovskite polycrystalline films by correlative infrared-spectroscopic nanoimaging and Kelvin probe force microscopy. *Light: Science & Applications*, *10*(1), 84.

Regehr, D., & Bazzaz, F. (1979). The population dynamics of E. canadensis , a successional winter annual. *The Journal of Ecology*, 923-933.

Sabri, K., Mokabli, A., Mokrini, F., Khayi, S., Laasli, S.-E., Smaha, D., . . . Hadj-Sadok, D. N. (2022). Diversity of nematophagous fungi associated with vegetable crops in Northern Algeria. *Archives of Phytopathology and Plant Protection, 55*(4), 405-419.

Saqib, M., Ali, S., Ijaz, M., Latif, M., Ahmad, M., Akbar, N., & Ghaffar, A. (2015). *The influence of weed management on the growth and yield of direct seeded rice (Oryza sativa L.)* (2167-9843). Retrieved from

Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture, 198*, 107119.

Sharma, R. K., Verma, N., Jha, K., Singh, N. K., & Kumar, B. (2014). Phytochemistry, pharmacological activity, traditional and medicinal uses of Erigeron species: A review. *IJARI, 2*, 379-383.

Sinha, A., John, J., Singh, S., & Johri, P. (2022). Microarray-Based Detection and Identification of Bacterial and Viral Plant Pathogens. In U. B. Singh, P. K. Sahu, H. V. Singh, P. K. Sharma, & S. K. Sharma (Eds.), *Rhizosphere Microbes: Biotic Stress Management* (pp. 47-69). Singapore: Springer Nature Singapore.

Suchman, E. (2011). Polymerase chain reaction protocol. *American Society for Microbiology, 1*, 14.

Sun, J., Xiao, S., & Xue, C. (2023). The tug-of-war on iron between plant and pathogen. *Phytopathology Research, 5*(1), 61.

Tang, R., Tan, H., Dai, Y., Li, L. a., Huang, Y., Yao, H., . . . Yu, G. (2023). Application of antimicrobial peptides in plant protection: making use of the overlooked merits. *Frontiers in Plant Science, 14*, 1139539.

Team, B. (2025). DNA Virus Replication and Immune Evasion Mechanisms. Retrieved from https://biologyinsights.com/dna-virus-replication-and-immune-evasion-mechanisms/

Tennant, P., Fermin, G., & Foster, J. E. (2018). *Viruses: molecular biology, host interactions, and applications to biotechnology*: Academic press.

ThermoFisherScientific. (2019). Phire Hot Start II DNA Polymerase. Retrieved from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0012385_Phire_HotStartII_DNAPolymerase_200rxns_UG.pdf

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., . . . Jones, J. B. (2020). Xanthomonas diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology, 18*(8), 415-427.

Zhang, Y., Kuang, X., Yi, J., Sun, T., Guo, Q., Gu, H., & Xu, H. (2024). Revolutionizing the capture efficiency of ultrasensitive digital ELISA via an antibody oriented-immobilization strategy. *Journal of Materials Chemistry B, 12*(39), 10041-10053.

Zheng, J., & Xu, Y. (2023). A review: Development of plant protection methods and advances in pesticide application technology in agro-forestry production. *Agriculture*, *13*(11), 2165.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

WHCCII13077/ WIV21: Wuhan Insect Virus 21

WHAV2: Wuhan aphid Virus 2 (or Wuhan Hydroperus Aphid Virus 2)

SnTV-1: Solanum nigrum Torradovirus 1 (associated with Physalis sp.)

CMV: Cucumber Mosaic Virus (a well-characterized plant pathogen)

AMVT1: Alternaria arborescens Mitovirus 1 (fungal-associated virus)

AgLV: Ageratum Leaf Curl Virus (a begomovirus infecting Ageratum species)

EndU-V82: Endornavirus U-V82 (isolated from tomato, Solanum lycopersicum)

APPENDIX

Virus abbreviation	GenBank ID	RNA X	HOST	Geographical origin	Orientation on genome
WHCCII13077	NC_033491	LINEAR	insect	China	full genome
WHCCII13077	NC_033481	LINEAR	insect	China	full genome
WHAV2	NC_028382	LINEAR	Hyalopterus pruni	China	full genome
WHAV2	NC_028386	LINEAR	Hyalopterus pruni	China	full genome
WHAV2	NC_028383	LINEAR	Hyalopterus pruni	China	full genome
WHAV2	NC_028387	LINEAR	Hyalopterus pruni	China	full genome
SnIV-1	OL472060	RNA 1	Physalis sp.	Slovenia	assembelled complete sequnce
SnIV-1	OL472062	RNA 3	Physalis sp.	Slovenia	assembelled complete sequnce
CMV	NC_001440	RNA 3	NA	NA	full genome
CMV	OL472039	LINEAR	BER19SW1 tomato	Slovenia	assembelled complete sequnce
AaMV1	NC_030747	LINEAR	Alternaria arborescens	USA	full genome
AgLV	NC_022127	LINEAR	Ageratum houstonianum	Australia	full genome
BVF(BBV-3X)	NC_029303	DNA circular	BBV-3X leaf	USA	full genome
EnOLV82	OL472300	LINEAR	KOP20AT tomato	Slovenia	full genome

Appendix01: Origin and source of the RNA-seq results obtained from NCBI databases

Acknowledgments

I extend my deepest gratitude, first and foremost, to Almighty Allah. O Lord, all praise and thanks be to You, in a manner befitting the majesty of Your countenance, the vastness of Your dominion, and the supremacy of Your station. I am profoundly grateful for the health and patience You have bestowed upon me throughout these years of study.

I would like to express my heartfelt appreciation to my thesis supervisor, **Dr. Vàrllyày Éva**, Scientific Advisor at the Institute for Plant Protection, MATE, for the invaluable guidance she has provided over the years. With her humility, kindness, scientific rigor, insightful advice, unwavering patience, and constant encouragement, she has significantly enriched my research. It has been both an honor and a privilege to have my work overseen by her.

I wish to convey my sincerest gratitude to **Dr. Zsuzsanna Galbács** Research fellow, Institute for Plant Protection, MATE as I appreciate her meticulous help and guidance in the laboratory.

A special and warm tribute goes to my beloved mother, **Remadna Nadhira**, who has stood by me unwaveringly throughout these years of study. Thank you for your boundless love, support, and reassurance, for always being present, and for doing everything in your power to assist, uplift, and make my journey as smooth and beautiful as possible.

My deepest and most affectionate gratitude also goes to my dear husband, **Khaled Rachad Mimi**, my pillar of success. Thank you for your invaluable encouragement, unwavering support, and steadfast presence throughout this journey.

I am immensely grateful to my beloved siblings—**Yasmine, Asma, Imen, Mahdi, and Oussama**—for their constant encouragement and for brightening my days with their unwavering affection.

Lastly, I extend my heartfelt appreciation to my friends for their steadfast support throughout these two years.

Finally, I wish to express my sincerest gratitude to all those who, in one way or another, have contributed to the completion of this work.

DECLARATION

the public access and authenticity of the thesis/dissertation/portfolio [1]

Student's name: TIAR SARA

Student's Neptun code: GF92ZK

Title of thesis: INVESTIGATION OF VIROMES OF CROPS AND WEEDS USING HIGH THROUGHPUT SEQUENCING BASED METAGENOMICS APPROACHES

Year of publication: 2025

Name of the consultant's institute: Dr. Éva Várallyay; Dr. Zsuzsanna Galbács

Name of consultant's deparment: Institute for Plant Protection, MATE

I declare that the final thesis/thesis/dissertation/portfolio submitted by me is an individual, original work of my own intellectual creation. I have clearly indicated the parts of my thesis or dissertation which I have taken from other authors' work and have included them in the bibliography.

If the above statement is untrue, I understand that I will be disqualified from the final examination by the final examination board and that I will have to take the final examination after writing a new thesis.

I do not allow editing of the submitted thesis, but I allow the viewing and printing, which is a PDF document.

I acknowledge that the use and exploitation of my thesis as an intellectual work is governed by the intellectual property management regulations of the Hungarian University of Agricultural and Life Sciences.

I acknowledge that the electronic version of my thesis will be uploaded to the library repository of the Hungarian University of Agricultural and Life Sciences. I acknowledge that the defended and ‑ not confidential thesis after the defence ‑ confidential thesis 5 years after the submission will be available publicly and can be searched in the repository system of the University.

Date: 2025 year 04 month 15 day

Student's signature

DECLARATION

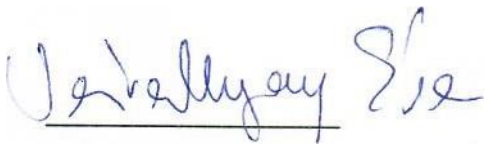
TIAR SARA student Neptun code: GF92ZK

as a consultant, I declare that I have reviewed the final thesis/thesis/dissertation/portfolio [1] and that I have informed the student of the requirements, legal and ethical rules for the correct handling of literary sources.

- I recommend / do not recommend [2] the final thesis / dissertation / portfolio to be defended in the final examination.

The thesis contains a state or official secret: yes no

Date: 2025 year 04 month 15 day



Dr. Éva Várallyay

Insider consultant

DECLARATION

TIAR SARA student Neptun code: GF92ZK

as a consultant, I declare that I have reviewed the final thesis/thesis/dissertation/portfolio [1] and that I have informed the student of the requirements, legal and ethical rules for the correct handling of literary sources.

I recommend / do not recommend [2] the final thesis / dissertation / portfolio to be defended in the final examination.

The thesis contains a state or official secret: yes no

Date: 2025 year 04 month 15 day

Dr. Zsuzsanna Galbács

Insider consultant