HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

Doctoral School of Biological Sciences

HTS-Based Virome Analysis of Solanaceous Crops and Weeds in Hungary and Kosovo

The Thesis of the PhD dissertation

Burim Ismajli

GÖDÖLLŐ

2026

Name of Doctoral School: **Doctoral School of Natural Sciences**

Head of the Doctoral School: **Prof Dr. Erika Csákiné Michéli**,
full professor, MATE
core member of the Hungarian Academy of Sciences

Name of the Doctoral Program: **Doctoral Program of Biological Sciences**

Disciplinary: **Biological Sciences**

Head of the Doctoral program: **Prof. Dr. Zoltán Nagy, DSc,**
full professor, MATE

Supervisors: **Dr. Éva Várallyay, DSc.**
Scientific Advisor, MATE
corresponding member of the Hungarian Academy of Sciences
Institute of Plant Protection, Department of Plant Pathology, Genomics Research Group, Gödöllő

Dr. habil. András Péter Takács
associate professor, MATE**,** Institute of Plant Protection, Department of Plant Protection, Keszthely

Approval of the head/supervisors: **Dr. Éva Várallyay**

.....

..... **.....**
Dr. Erika Csákiné Michéli **Dr. András Péter Takács**

1. Introduction

The Solanaceae family represents a cohesive group of dicotyledonous plants, encompassing a range of extensively cultivated crops. Members of the Solanaceae family include well-known cultivated crops such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), pepper (*Capsicum annuum*) etc. In addition to the extensively cultivated crops, the Solanaceae family also comprises a notable array of wild and weedy species. For instance, species such as *Solanum nigrum, Datura stramonium*, and *Solanum dulcamara* are frequently encountered in agricultural landscapes, often thriving as pervasive weeds near crop fields. This coexistence underscores the importance of considering both cultivated crops and wild species when studying plant communities, as these weeds can play significant roles in ecological dynamics, pathogen transmission, and the overall genetic diversity of agroecosystems. Solanaceae crops are exposed to various pathogens, with plant viruses causing widespread epidemics that threaten global food security. In the realm of virus research, exploring the role of weeds as potential reservoirs has become pivotal.

2. OBJECTIVES TO ACHIEVE

1. Investigate the viromes of solanaceous weeds in Keszthely, Hungary, using high-throughput sequencing (HTS), and evaluate their role as virus reservoirs
2. To investigate and characterize the virome of cultivated solanaceous crops in Kosovo using HTS.
3. To confirm detected viral infections through RT-PCR and to perform phylogenetic analyses to evaluate their evolutionary relationships.

3. Matherial and Methods

Sample collection in Keszthely

In 2022, the first field study in Keszthely (K2022) focused on asymptomatic weed plants growing naturally around cultivated crops. Sampling was carried out in two fields (Field I and Field II).

In 2023, the Keszthely field study (K2023) was extended to include symptomatic weeds and two additional species, again sampled from two fields (Field III and Field IV), following the same strategy as in K2022.

Symptomatic and asymptomatic potato, tomato, and pepper plants were collected during the summer of 2023 from different locations in Kosovo (Ko2023).

RNA Isolation

Phenol-chloroform extraction method

Total RNA was extracted using a phenol–chloroform method with all steps performed on ice. Plant tissue was homogenized in SDS-containing extraction buffer and purified by sequential phenol, phenol–chloroform, and chloroform extractions. Nucleic acids were precipitated with sodium acetate and ethanol, washed with 70% ethanol, air-dried, and dissolved in sterile water. RNA quality was assessed on a 1.2% agarose gel after heat denaturation, and samples were stored at −70 °C until use.

Total Nucleic Acid Preparation and RNA Sequencing (RNA-Seq)

To ensure sufficient RNA quantity and quality, two independent RNA extractions were performed per plant and pooled. Samples were subsequently combined at the species level and then across all individuals, generating one final pooled sample per dataset (15 µL for K2022, 20 µL for K2023, and 15 µL for Ko2023) for sequencing. This pooling strategy ensured balanced representation and adequate RNA input for high-quality sequencing.

DNase treatment prior to RNA-seq

Prior to sequencing, pooled TNA samples were treated with RNase-free DNase I to remove genomic DNA. The resulting DNA-free RNA was subjected to ribodepleted RNA-seq on an Illumina platform (150 bp paired-end reads) by Novogene.

Preparation of Small RNA Libraries

Isolation of the sRNA Fraction from Total RNA

Small RNA libraries were prepared following gel-based size selection using the TruSeq® Small RNA Library Prep Kit (Illumina). The protocol involves adapter ligation, reverse transcription, PCR amplification, and gel purification to generate indexed cDNA libraries suitable for multiplexed Illumina sequencing and small RNA/miRNA profiling.

Small RNA library preparation

Small RNA libraries for Illumina sequencing were prepared using the TruSeq Small RNA Library Prep Kit (Illumina). Although the kit allows library construction from as little as 1 μg of total RNA, higher read quality and yield were achieved by using gel-purified small RNA fractions obtained from 10–30 μg of total RNA. Read quality primarily depends on RNA integrity, whereas sequencing depth is influenced by the sequencing platform and the number of pooled libraries. On average, Illumina HiSeq generates approximately 140 million reads per flow cell, enabling multiplexing of several libraries. During library preparation, 5′ and 3′ adapters were ligated to small RNAs, and sample-specific six-base indexes were introduced via reverse PCR primers to allow demultiplexing of sequencing reads.

Small RNA isolation and library preparation for Illumina sequencing

Small RNAs were isolated by gel-based size selection using an 8% denaturing polyacrylamide gel containing urea. Total RNA was denatured, separated by electrophoresis, and the small RNA fraction (15–30 nt) was excised, eluted overnight, and precipitated using salt–alcohol precipitation. Purified small RNAs were used for library preparation with the TruSeq Small RNA Library Prep Kit (Illumina), which includes sequential 3′ and 5′ adapter ligation, reverse transcription, and PCR amplification with indexed primers. Amplified libraries were size-selected by PAGE, excising fragments corresponding to 140–145 nt, purified, and resuspended in TE buffer. Final libraries were sequenced on an Illumina HiScanSQ platform (50 bp, single-end reads).

Bioinformatic analysis of HTS data

High-throughput sequencing data (sRNA and RNA-seq) in FastQ format were analyzed using CLC Genomics Workbench (Qiagen, Germany). Initial quality control included adapter trimming, removal of low-quality reads, and assessment

of read size distribution, with particular focus on 21–24 nt small RNA populations. Redundant reads were collapsed into non-redundant datasets prior to downstream analysis. High-quality reads were assembled de novo into contigs, which were subjected to BLAST searches against a curated database of plant virus sequences retrieved from NCBI GenBank (accessed 31 July 2023). For RNA-seq data, trimmed paired-end reads were similarly assembled, and viral reads were additionally mapped to reference genomes identified by contig analysis. Consensus sequences were generated from read mappings to determine viral genome coverage.

Validation of virus presence by RT-PCR

Primer design

Viruses identified through bioinformatic analyses were validated by RT-PCR using virus-specific primers. Primers were designed in Geneious Prime (v2024.0.7) based on reference genomes and assembled contigs, primarily targeting the conserved coat protein (CP) region commonly used for virus detection and phylogenetic analysis. Primer specificity was evaluated by BLAST searches against the NCBI database. This approach ensured reliable amplification and accurate validation of HTS-derived virus detections across the analyzed samples.

Gradient PCR and Optimization of Annealing Temperature

Primers were diluted according to the manufacturer's protocol, and a gradient PCR was performed to determine the optimal annealing temperature for each primer pair. This step was necessary to ensure specific amplification and to avoid non-specific products or primer-dimer formation. The gradient ranged from 51.1 to 66 °C, allowing evaluation of primer behaviour across a wide temperature interval. For each primer pair, PCR reactions were set up under different conditions, varying only the annealing temperature, to identify the temperature that produced a strong and specific band of the expected size. This optimization step ensured that only the most suitable temperature was chosen for the subsequent PCR reactions performed on individual samples.

PCR Reaction Setup and Cycling Parameters

PCR reactions were prepared separately for each sample in a final volume of 15 µL, using either Q5 or Phire DNA polymerase. Each reaction contained reaction buffer, dNTPs, forward and reverse primers, polymerase, nuclease-free water, and

0.5 µL of cDNA template. Reaction mixtures were prepared on ice under sterile conditions, briefly centrifuged, and amplified in a thermocycler.

cDNA synthesis and quality control

cDNA was synthesized using the RevertAid™ First Strand cDNA Synthesis Kit or Maxima™ Reverse Transcriptase (Thermo Fisher Scientific) with random primers, following the manufacturers' protocols. RNA extracted from pooled plant samples and individual plants was used as template. Reverse transcription was performed under standard thermal conditions, and the resulting cDNA was diluted prior to downstream applications. cDNA quality was verified by RT-PCR using actin-specific primers. Subsequently, virus-specific RT-PCR assays were conducted using Q5 Hot Start High-Fidelity DNA Polymerase with primer-optimized cycling conditions.

PCR amplification and gel electrophoresis

PCR amplification of RNA-seq–derived cDNA was performed using either Q5 or Phire DNA polymerase in 15 µL reaction volumes containing reaction buffer, dNTPs, gene-specific primers, polymerase, and cDNA template. Amplified products were resolved on TBE agarose gels stained with ethidium bromide, alongside a 100 bp DNA ladder for size estimation. Electrophoresis was performed under standard conditions, and PCR products were visualized and documented using a Bio-Rad ChemiDoc imaging system.

PCR product purification, cloning, and sequence analysis

PCR products were purified using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions and eluted in Buffer NE. Purified amplicons were ligated into the pJET1.2/blunt cloning vector using the CloneJET PCR Cloning Kit (Thermo Scientific) and directly transformed into competent *E. coli* cells. Recombinant plasmids were isolated using the NucleoSpin® Plasmid Kit (Macherey-Nagel). Insert presence was verified by restriction enzyme digestion followed by agarose gel electrophoresis, and confirmed plasmids were submitted for Sanger sequencing.

Analyzing Sanger sequences

During the initial sequence analysis, chromatograms were checked in Chromas, a free viewer suitable for single-read DNA sequencing. Chromas was used to assess peak clarity and identify mixed or ambiguous bases, and sequences were exported in FASTA or text format for further analysis in CLC Genomics

Workbench. Single nucleotide polymorphisms (SNPs) were identified based on consistent base substitutions and confirmed when present in both forward and reverse reads. When mixed peaks or sequence heterogeneity were detected, additional bioinformatics evaluation was performed in CLC.

Phylogenetic analysis

To compare and phylogenetically analyze the virus variants present in the samples multiple sequence alignments were performed using Geneious Prime with the MUSCLE algorithm. Evolutionary relationships were inferred using the Jukes-Cantor model and the Neighbor-Joining method. Phylogenetic trees were constructed based on the optimal model for each alignment, with 1,000 bootstrap replicates to assess reliability. The trees were scaled, with branch lengths representing the number of substitutions per site. The closest relatives of the viruses, as indicated in the tree legend, were used as outgroups.

4. RESULTS

HTS results

Results of K2022 Weed Sampling

For the K2022 sampling, a single sRNA library (255_KSOL) and a single total RNA library (SOL_KES_10) were prepared to represent all sampled plant species, including *Solanum nigrum* and *Datura stramonium* from Field I and *S. nigrum* from Field II. In the case of sRNA for K2022, sequencing produced 21,955,551 reads, of which 21,573,415 high-quality redundant reads remained after trimming. These represented 3,178,957 non-redundant sRNA sequences, from which 3,177 contigs were assembled. For RNA-Seq of the same samples, 16,253,182 reads were generated in total. After trimming and quality control, 15,972,388 reads were retained, resulting in the assembly of 132,523 contigs.

BLAST analysis of contigs derived from sRNA and RNA-Seq revealed the presence of four viruses: Cucumber mosaic virus (CMV; RNA1, RNA2, and RNA3), Broad Bean Wilt Virus 1 (BBWV1; RNA1 and RNA2), Turnip yellow virus (TuYV) and Tobacco vein-clearing virus (TVCV).

Each virus was represented by at least one contig showing significant homology (0 E-value) to reference genomes of known plant-infecting viruses. Mapping of viral reads to the respective genomes indicated high genome coverage in both sequencing approaches. For CMV, sRNA libraries showed 99% coverage for RNA1, RNA2, and RNA3, which was mirrored in the RNA-Seq data. BBWV1 coverage from sRNA reads was 98% for RNA1 and 97% for RNA2, with identical coverage observed in RNA-Seq libraries. TVCV coverage was 80% in sRNA results and 89% in RNA-seq results while TuYV showed lower genome coverage in sRNA sequencing with 43% compared to RNA-seq with 90%.

Results of K2023 Weed Sampling

For the K2023 sampling, a single total RNA library was prepared for all samples (SOL_KES_17). A total of 23,061,440 reads were generated, and after trimming (including redundant sequences), 22,913,754 high-quality reads were retained, from which 71,700 contigs were assembled.

BLAST analysis of the contigs identified seven viruses with significant similarity (E-value = 0): Potato virus M (PVM), Potato virus H (PVH), Obuda pepper virus (ObPV), Tobacco vein clearing virus (TVCV), Lettuce big-vein associated virus

(LBVaV), Oxybasis rubra mitovirus 1 (OxruMV1), and Potato latent virus (PotLV). Genome coverage varied among viruses: 82% for PVM, 90% for PVH, 33% for ObPV, 45% for TVCV, 99% for LBVaV (both RNA1 and RNA2), 72% for OxruMV1, and 31% for PotLV.

Results of Ko2023 Cultivated Crop sampling

For the Ko2023 sampling, a total RNA library (SOL_KOS_18) was prepared to represent all sampled plant species. A total of 21,558,420 reads were obtained. Following trimming and quality control, 21,442,968 reads were retained. De novo assembly of these reads resulted 40,362 contigs.

For all datasets, assembled contigs were analyzed by BLAST against reference genomes of known plant-infecting viruses. In Ko2023, nine viruses were detected, each represented by at least one contig with significant homology (E-value = 0): CMV, BBWV2, PVY, PCV2, BPEV, RWMV, TVCV, TAV, and CMV satellite RNA. Read mapping revealed high genome coverage for most viruses, including complete coverage for all three CMV RNAs and near-complete coverage for BBWV2, PVY, PCV2, BPEV, and RWMV, whereas lower coverage was observed for TVCV (85%) and TAV (65%).

Contigs generally spanned most or all viral genomes, except for TVCV, TAV, and CMV satellite RNA, where contigs were shorter or only partially aligned. BLAST analysis against GenBank indicated high similarity of these contigs to Solanaceae host genomes, suggesting host-integrated viral sequences. This similarity was limited to short regions for TAV and CMV satellite RNA, whereas longer homologous regions for TVCV may indicate remnants of a past infection integrated into the host genome. Hits corresponding to the remaining six viruses were considered reliable and were further validated by virus-specific RT-PCR.

5. DISCUSSIONS

In K2022, four viruses, CMV, BBWV1, TVCV, and TuYV were detected. This site, previously used for horticultural crops such as pepper, tomato, and eggplant, and recently for wheat, with a history of potato cultivation, was surrounded by diverse dicot weed species. These surrounding weeds likely contributed to virus circulation and maintenance within the field. CMV was detected in one *Solanum nigrum* individual from Field I (S1/5), which carried all three viral RNAs. BBWV1 was found in one *Solanum nigrum* plant from Field II (S2/1), and was validated with both RNAs of the virus. TVCV was detected in all *Solanum nigrum* individuals from both fields, whereas none of the *Datura stramonium* plants were infected. TuYV was detected in one *Datura stramonium* plant from Field I (D1/5), and in one individual of *Solanum nigrum* from Field II (S2/5). Although all sampled plants appeared asymptomatic at the time of sampling, all viruses were successfully amplified, indicating that *S. nigrum* and *Datura stramonium* can act as an efficient reservoir host capable of sustaining latent infections

In K2023, seven viruses, PVM, PVH, ObPV, TVCV, LBVaV, OxruMV1, and PotLV were detected, demonstrating broader viral diversity and greater reservoir potential in the field. This increase reflects the capacity of weed and wild plant species to harbor multiple viruses simultaneously, acting as persistent sources of infection that facilitate virus spread to adjacent crops. *Solanum nigrum*, which had tested positive for CMV, BBWV1, TVCV and TuYV in 2022, was positive for PVM in 2023, with all individuals from Field III infected (S3/1–S3/10), illustrating temporal shifts in virus prevalence that may result from vector dynamics, crop rotations, or competitive interactions among viruses. TVCV was detected in all *S. nigrum* individuals from Field III except S3/5, highlighting widespread infection within the population. *Datura stramonium*, was positive for Obuda pepper virus in 2023, with two individuals from Field III infected (D2/2 and D2/9) and two individuals also positive for TVCV (D2/1 and D2/7), indicating yearly changes in virus circulation influencing infection patterns. *Brassica napus*, grown alongside infected weeds, tested negative for all viruses, suggesting that viruses present in surrounding wild hosts either do not efficiently infect Brassica or were below detectable levels. *Solanum dulcamara* showed the highest virus diversity, with all individuals from Field IV infected with PVM (Sd1–Sd8). PVH was detected in three *S. dulcamara* individuals (Sd1, Sd6, Sd7), ObPV in one individual (Sd4), LBVaV in Sd6 with both RNAs confirmed, and

OxruMV1 in three individuals (Sd2, Sd3, Sd6). For PotLV, no further analyses were performed, BLAST of its contigs showed higher similarity to PVM and PVH, likely due to conserved genomic regions among these related viruses. Sampling in 2023 was more extensive, including leaves, flowery shoots, seeds, and fruits, likely contributing to higher detection rates and observed diversity.

Ko2023 showed the highest virus diversity, with CMV, BBWV2, PVY, PCV2, BPEV, RWMV, TVCV, TAV, and the CMV satellite RNA. This site was used for potato cultivation and bordered by uncultivated or mixed-use areas such as bean, pepper, and tomato fields, creating conditions favorable for virus spillover between crops and weeds. All sampled pepper plants were positive for several viruses, including CMV, BBWV2, PVY, PCV2, BPEV, and RWMV reflecting strong inoculum pressure, high host susceptibility, and efficient vector-mediated transmission. In contrast, only one individual of tomatoes was infected with CMV, likely due to differences in host susceptibility or vector exposure. For TVCV, TAV, and CMV satellite, assembled contigs were shorter than full genomes or partially aligned with known viruses. BLAST comparisons indicated high similarity to Solanaceae host sequences, suggesting many hits originated from the plant genome rather than active infections. TAV and CMV satellite regions were particularly short, likely representing artifacts or remnants, while TVCV contigs showed longer identity regions, possibly reflecting host-integrated viral sequences from past infections.

NEW SCIENTIFC RESULTS

1. The virome of solanaceous weeds in Keszthely was evaluated using high-throughput sequencing (HTS), providing the first sequences of CMV, BBWV1, TVCV, and TuYV obtained from *Solanum nigrum* in Hungary.
2. The sequence of the ObPV variant detected from *Datura stramonium* represents the first sequence of this virus from this host.
3. We described *Solanum dulcamara* as a host of PVH, LBVaV, and OxruMV1 for the first time.
4. Mixed infections occurred in pepper plants in Kosovo, where CMV, BBWV2, PVY, PCV2, BPEV, and RWMV were confirmed by HTS and RT-PCR, providing the first partial molecular sequences of these viruses in the country.
5. BBWV2 was reported for the first time in the Balkan region.
6. The partial sequence of CMV infecting tomato in Kosovo was described for the first time.

Author's Scientific Publications

Publications in Peer-Reviewed Journals with Impact Factor

- **Ismajli, B.**, Galbács, Z. N., Takács, A. P., & Várallyay, É. (2025). The First High-Throughput Sequencing-Based Study of Viruses Infecting Solanaceous Crops in Kosovo Reveals Multiple Infections in Peppers by Six Plant Viruses. Plants, 14(9), 1273, https://doi.org/10.3390/plants14091273. IF: 4.1 (Q1).
- **Ismajli, B**., Pásztor, G., Takács, A., & Várallyay, É. (2024). Investigation of Viromes of Solanaceous Weeds. GEORGIKON FOR AGRICULTURE, 28(Suppl. 1), 13-17, https://journal.uni-mate.hu/index.php/gfa/article/view/6098.
- **Ismajli, Burim**, Zsuzsanna N. Galbács, Lilla Dorottya Péri, György Pasztor, András Takács, Várallyay, Éva. Reinvestigating the viromes of solanaceous weeds in Hungary after several decades confirmed the original results and added new insights. (Under review)

Scientific Talks Presented at Conferences

- **Burim Ismajli**, György Pásztor, András Takács, Éva Várallyay
  Investigation of viromes of Solanaceae weeds
  „MATE Növényvédelmi Intézet Növényvédelmi Tanszék PhD hallgatóinak bemutatkozása" Keszthely, 2023.november 14.
- **Burim Ismajli**, György Pásztor, András Takács, Éva Várallyay
  Investigation of viromes of Solanaceae weeds
  Sloven-Hungarian PhD Student Forum 2023-11-22 Ljubljana, Szlovénia (MTMT 34414572)
- **Burim Ismajli**, György Pásztor, András Takács and Éva Várallyay
  Investigation of viromes of Solanaceae weeds. XXXIII. Keszthelyi Növényvédelmi Fórum, 2024. január 17-19. (MTMT 34414568).
- **Ismajli, Burim**, András Takács, Várallyay, Éva
  Revealing Plant Virus Presence in POTATO, TOMATO and Pepper Cultivars: Kosovo's Debut Virus Report on First-Ever Findings. 2024 LXV. Georgikon Napok Tudományos Konferencia, Keszthely, május 17-18. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;35059876.
- **Burim Ismajli**, György Pásztor, András Takács, Éva Várallyay
  Investigation of viromes of Solanaceae weeds, GEORGIKON FOR AGRICULTURE: A MULTIDISCIPLINARY JOURNAL IN

AGRICULTURAL SCIENCES 28 : Suppl. 1 pp. 13-17. , 6 p. (2024), (MTMT 34532123).

- **Burim Ismajli**; Zsuzsanna, N. Galbács, András Takács, Éva Várallyay
  REVEALING PLANT VIRUS PRESENCE IN PEPPER CULTIVARS: KOSOVO'S DEBUT VIRUS REPORT ON FIRST-EVER FINDINGS. 11th International Plant Protection Symposium at University of Debrecen, 15-17 October 2024. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;35484954.
- **Burim Ismajli**; Nagyné Galbács Zsuzsanna, András Takács, Éva Várallyay
  Spicy Trouble in the Fields: Characterization and First-Ever Report of Plant Viruses in Kosovo's Peppers. 34th Plant Protection Forum, Keszthely, 15-17 Jan 2025.
- **Ismajli, Burim**[1]; Pásztor, György[2], András Takács Péter[2],Várallyay, Éva[1]
  30. Tiszántúli Növényvédelmi Fórum: The 12th Plant Protection Symposium
  Conference: Târgu Mureș, Romania 2025.10.14. - 2025.10.16. (Hungarian Chamber of Plant Protection Engineers and Phytopathologists, MTA-DAB Agricultural Committee, Plant Protection Working Committee, Sapientia EMTE Marosvásárhely Faculty, Plant Protection Institute of the University of Debrecen, University of Debrecen MÉK Plant Protection Institute), pp 128-129 (2025) (Transtizan Plant Protection Forum ; 30th TNF). Awarded the 3rd Prize for Best Conference Presentation.
- **Ismajli, Burim**, Zsuzsanna N. Galbács , Lilla Dorottya Péri, György Pasztor, András Takács, Várallyay, Éva
  Exploring Viromes of Solanaceous Weeds at Keszthely using HTS. XXXV. Keszthelyi Növényvédelmi Fórum 2026. január 14-16, MATE Georgikon Campus, Keszthely - A épület 8360 Keszthely, Deák Ferenc utca 16. novenyvedelmi.forum@uni-mate.hu.