NATIONAL INSTITUTE OF BIOLOGY

DETECTION OF PLANT VIRUSES BY NGS AT NIB

Maja Ravnikar[1,5], Nataša Mehle[1], Ion Gutierrez Aguirre[1], Anja Pecman[1,4], Ian Adams[2], Adrian Fox[2], Neil Boonham[2,3], Denis Kutnjak[1]

[1]National Institute of Biology, SI
[2]Fera, Sand Hutton, York, UK
[3]IAFRI, Newcastle University, Newcastle upon Tyne, UK
[4]Jožef Stefan International Postgraduate School, SI
[5]University Nova Gorica, SI

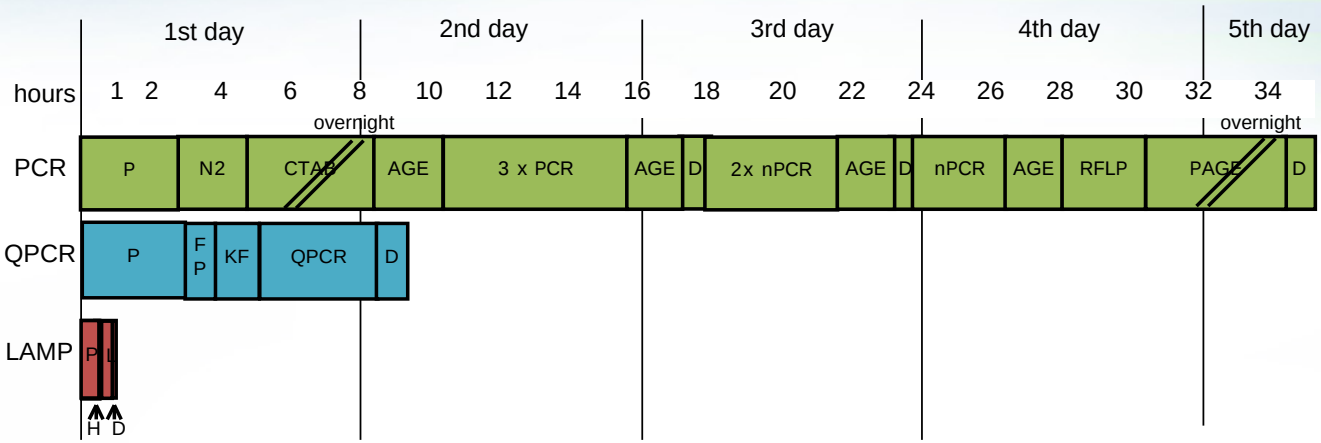
NATIONAL INSTITUTE OF BIOLOGY

Official plant diagnostic laboratory authorised by Plant health administration of Slovenia

Biology, biodiversity, epidemiology of plant pathogens, diagnostics, methods development including automation

Methods in virology: ELISA, test plants, electron microscopy

PCR based (RT-PCR, PCR, qPCR, ddPCR), sanger seq and NGS



NATIONAL INSTITUTE OF BIOLOGY

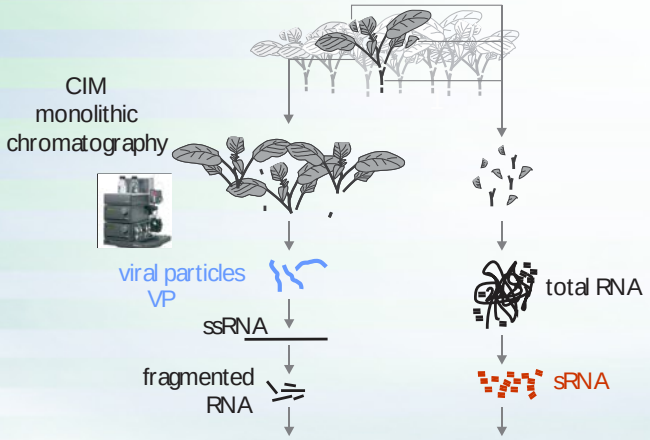
Detection of plant viruses with NGS

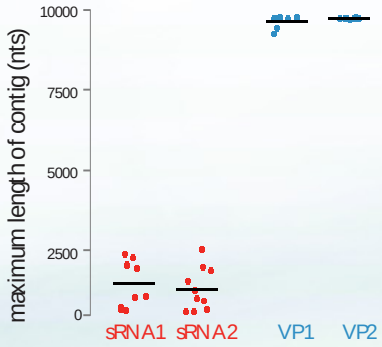
- NGS at NIB is **official diagnostic method from 2015, used on** samples with symptoms but negative results of classical methods

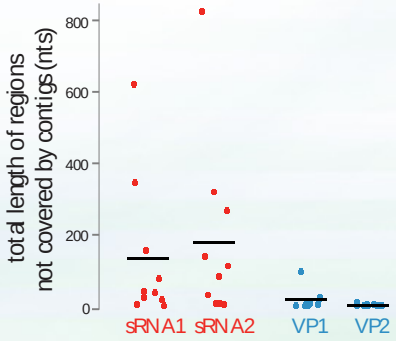
- Different methodological approaches in **viral nucleic acid extraction** step (trizol/RNeazy)
- The comparison of **different viral nucleic acid inputs** (sequencing of purified particles, small RNA (sRNA) and ribosomal RNA (rRNA) depleted total RNA
- **Different data analysis approaches** were compared (host genom removal)
- user-friendly **bioinformatic pipeline in CLC** was developed at NIB
- **Tomato** was shown to be a **new host of Henbane mosaic virus**
- NGS analyses of ornamental and vegetable samples revealed **known and new plant viruses in Slovenia.**

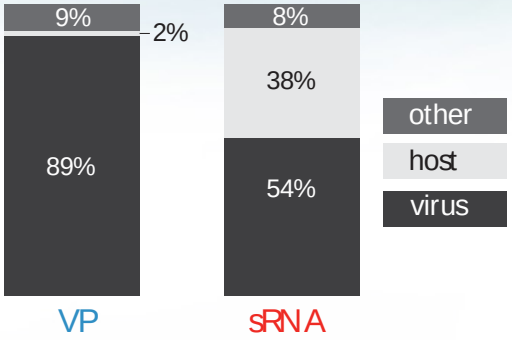
Comparison: RNA from viral particles vs virus derived small RNAs

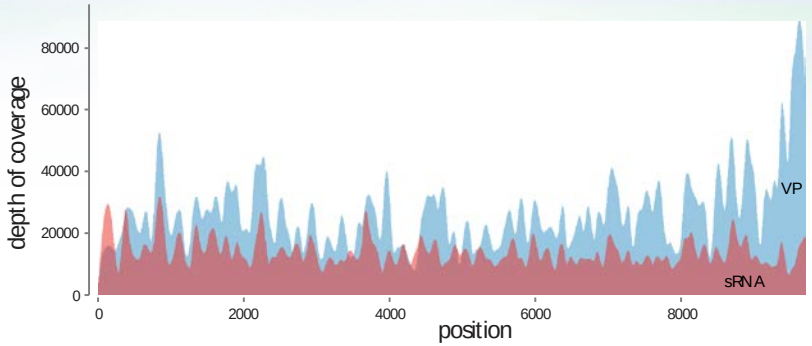
Potato virus Y (*Potyvirus*, *Potyviridae*) infected potato plants











Viral DNA enrichment and length of contigs was higher using VP RNA

Kutnjak et al., 2015, *Journal of Virology*

Characterization of a novel orthoreovirus for human *using monolithic chromatography and NGS*

Case of gastroenteritis with unknown causative agent.

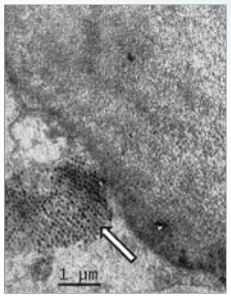
Tests for usual pathogens were negative.

EM: possible reovirus infection.

Viral particles (VP) under EM.

Purification of VP using CIM chromatography.

Ion Torrent sequencing



Steyer et al., JCM, 2013

JCM
Journals.ASM.org

High Similarity of Novel Orthoreovirus Detected in a Child Hospitalized with Acute Gastroenteritis to Mammalian Orthoreoviruses Found in Bats in Europe

Andrej Steyer,[a] Ion Gutiérrez-Aguire,[b,c] Marko Kolenc,[a] Simon Koren,[c] Denis Kutnjak,[b] Marko Pokorn,[d] Mateja Poljšak-Prijatelj,[a] Nejc Rački,[b] Maja Ravnikar,[b,c] Martin Sagadin,[a] Adela Fratnik Steyer,[a] Nataša Toplak[c]

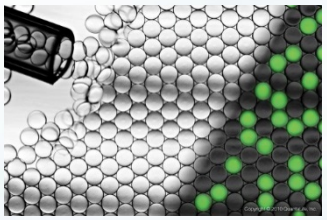
Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia[a]; National Institute of Biology, Ljubljana, Slovenia[b]; Omega d.o.o., Ljubljana, Slovenia[c]; Department of Infectious Diseases, University Medical Centre Ljubljana, Ljubljana, Slovenia[d]; The Centre of Excellence for Biosensors, Instrumentation and Process Control—COBIK, Solkan, Slovenia[e]

Mammalian orthoreoviruses (MRVs) are known to cause mild enteric and respiratory infections in humans. They are widespread and infect a broad spectrum of mammals. We report here the first case of an MRV detected in a child with acute gastroenteritis, which showed the highest similarity to an MRV reported recently in European bats. An examination of a stool sample from the child was negative for most common viral and bacterial pathogens. Reovirus particles were identified by electron microscopic examination of both the stool suspension and cell culture supernatant. The whole-genome sequence was obtained with the Ion Torrent next-generation sequencing platform. Prior to sequencing, the stool sample suspension and cell culture supernatant were pretreated with nucleases and/or the convective interaction medium (CIM) monolithic chromatographic method to purify and concentrate the target viral nucleic acid. Whole-genome sequence analysis revealed that the Slovenian SI-MRV01 isolate was most similar to an MRV found in a bat in Germany. High similarity was shared in all genome segments, with nucleotide and amino acid identities between 93.8 to 99.0% and 98.4 to 99.7%, respectively. It was shown that CIM monolithic chromatography alone is an efficient method for enriching the sample in viral particles before nucleic acid isolation and next-generation sequencing application.

We use CIM for concentration of water samples for metagenom analysis

Samples in work

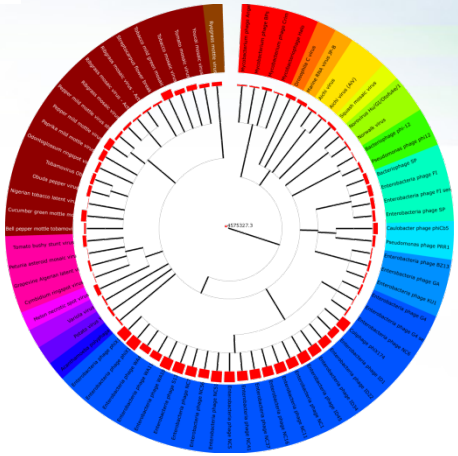
- NGS of viruses from water samples: easy to detect and count (ddPCR)



- after efficient concentration using CIM or filtration,

- Confirmation of infectivity on test plants

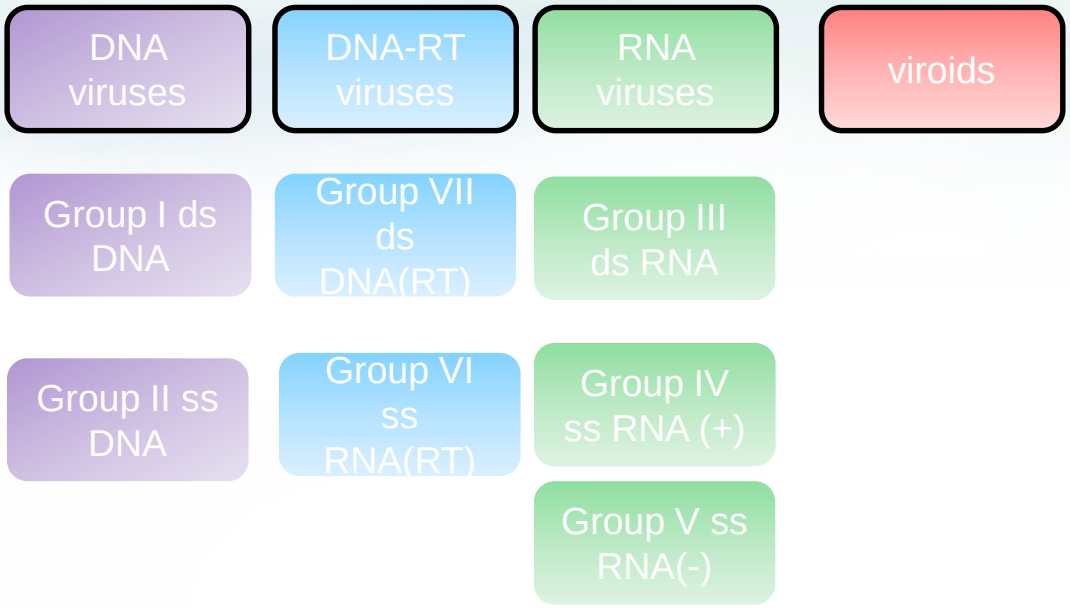
(ToMV, PePMMV)

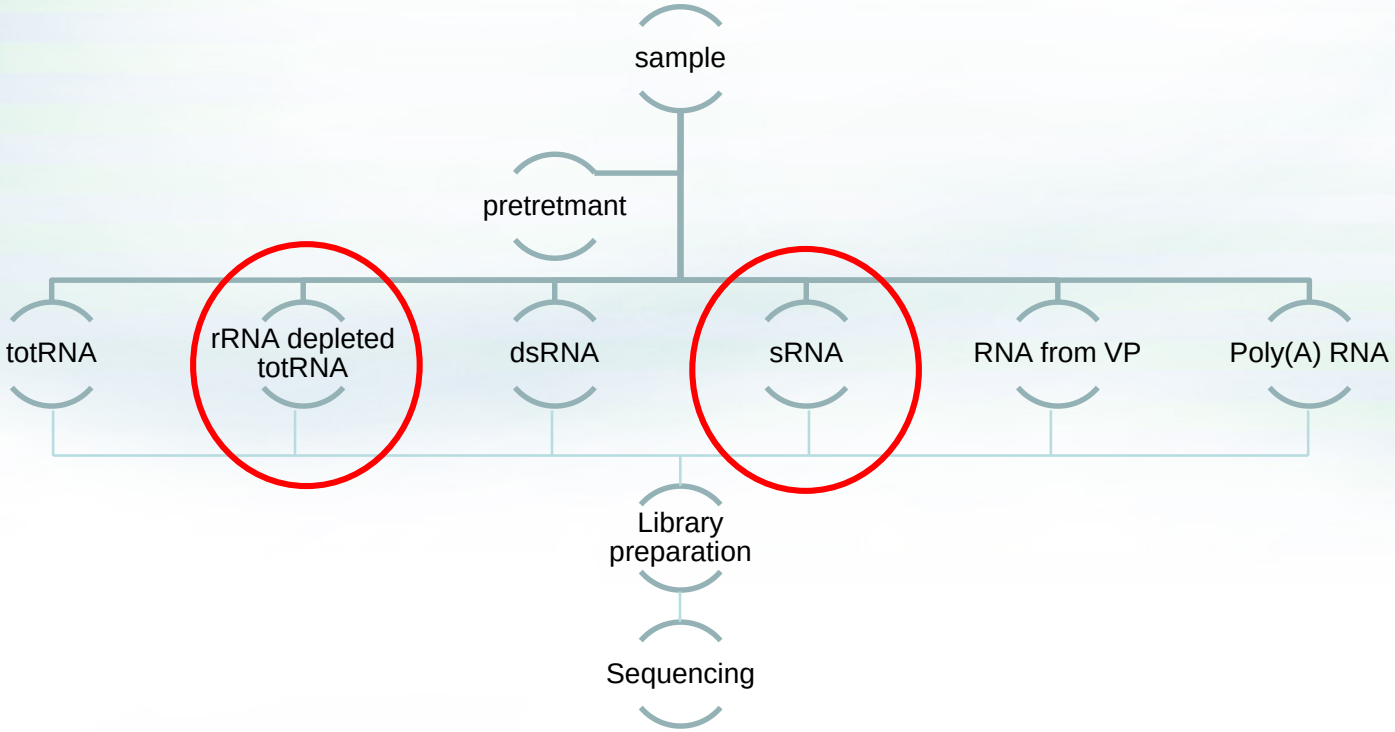


Detection of viruses and viroids with sRNA and rRNA depleted totRNA approaches

Viruses have diverse genome organizations and use different replication strategies. Based on these two characteristics they can be classified into 7 groups (the Baltimore classification).

Viroids are classified into two families: members of *Avsunviroidae* family replicate in chloroplast, whereas members of *Pospiviroidae* family replicate in nucleus



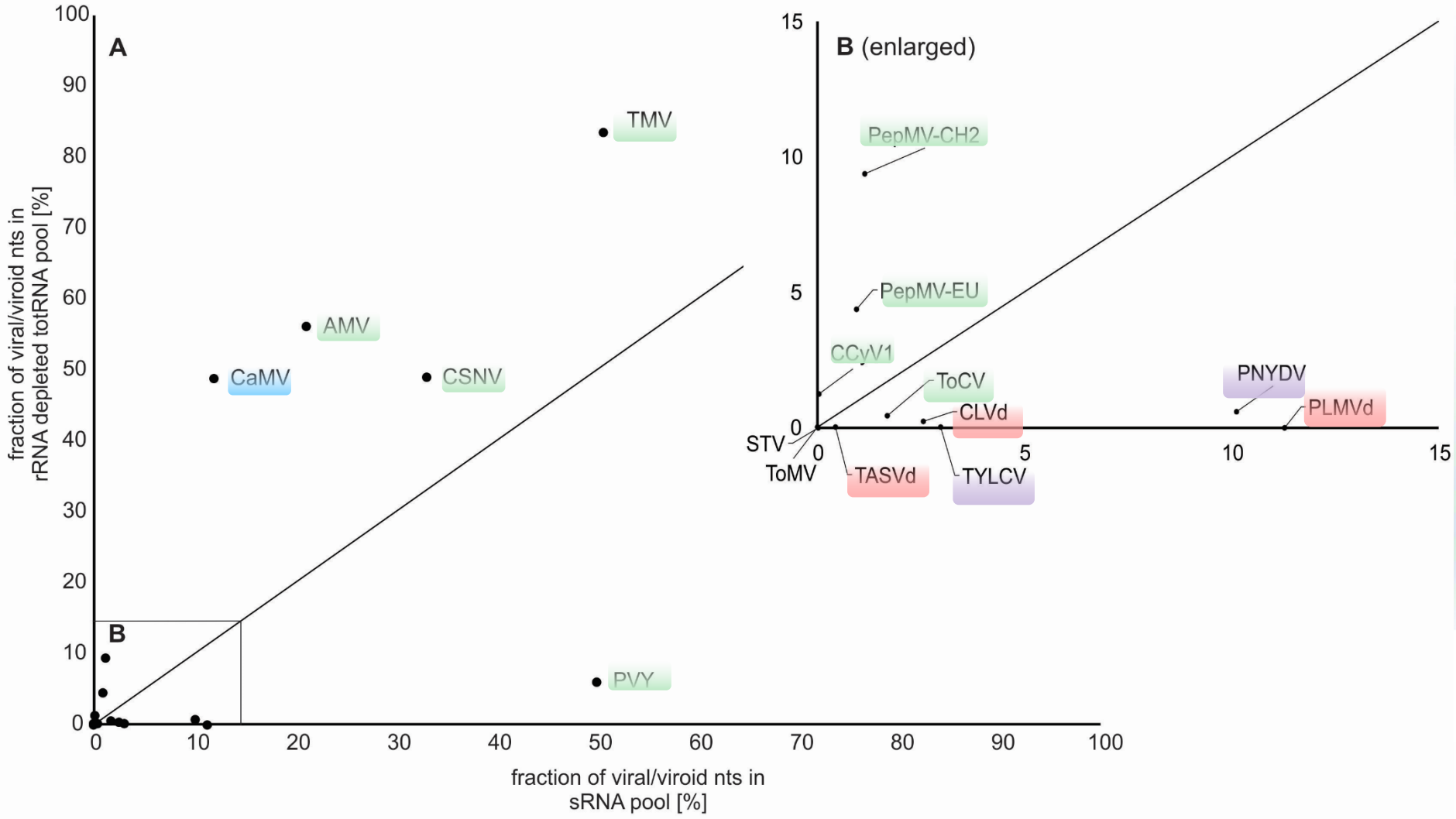

sample
pretretmant
totRNA
rRNA depleted totRNA
dsRNA
sRNA
RNA from VP
Poly(A) RNA
Library preparation
Sequencing

Sample number	Virus name (Abbreviations)	Host	NGS results	
			sRNA	rRNA depleted totRNA
I	*Potato virus Y (PVY)	Solanum tuberosum	+	+
II	*Cauliflower mosaic virus (CaMV)	Brassica oleracea	+	+
	Novel cabbage cytorhabdovirus 1 (CCyV1)	Brassica oleracea	-	+
III	*Tomato Yellow Leaf Curl Virus (TYLCV)	Solanum lycopersicum	+	+
	Tomato chlorosis virus (ToCV)	Solanum lycopersicum	+	+
	Pepino mosaic virus (PepMV)	Solanum lycopersicum	+	+
	Tomato mosaic virus (ToMV)	Solanum lycopersicum	+	+
	Southern tomato virus (STV)	Solanum lycopersicum	+	+
	Columnea latent viroid (CLVd)	Solanum lycopersicum	+	+
IV	*Alfalfa mosaic virus (AMV)	Nicotiana tabacum	+	+
V	*Pea necrotic yellow dwarf virus (PNYDV)	Pisum sativum	+	+
VI	*Tobacco mosaic virus (TMV)	Nicotiana sp.	+	+
VII	*Peach latent mosaic viroid (PLMVd)	Prunus sp.	+	+
VIII	*Tomato apical stunt viroid (TASVd)	Solanum lycopersicum	+	+
IX	*Chrysanthemum stem necrosis virus (CSNV)	Nicotiana benthamiana	+	+

Compering the yield of viral/viroid sequences obtained by sRNA and rRNA depleted totRNA approaches

All reads

MAPPING READS to consensus viral/viroid genome


A
B
B (enlarged)
TMV
AMV
CaMV
CSNV
PVY
PepMV-CH2
PepMV-EU
CCyV1
ToCV
CLVd
PNYDV
PLMVd
STV
ToMV
TASVd
TYLCV
fraction of viral/viroid nts in rRNA depleted totRNA pool [%]
fraction of viral/viroid nts in sRNA pool [%]

NATIONAL INSTITUTE OF BIOLOGY

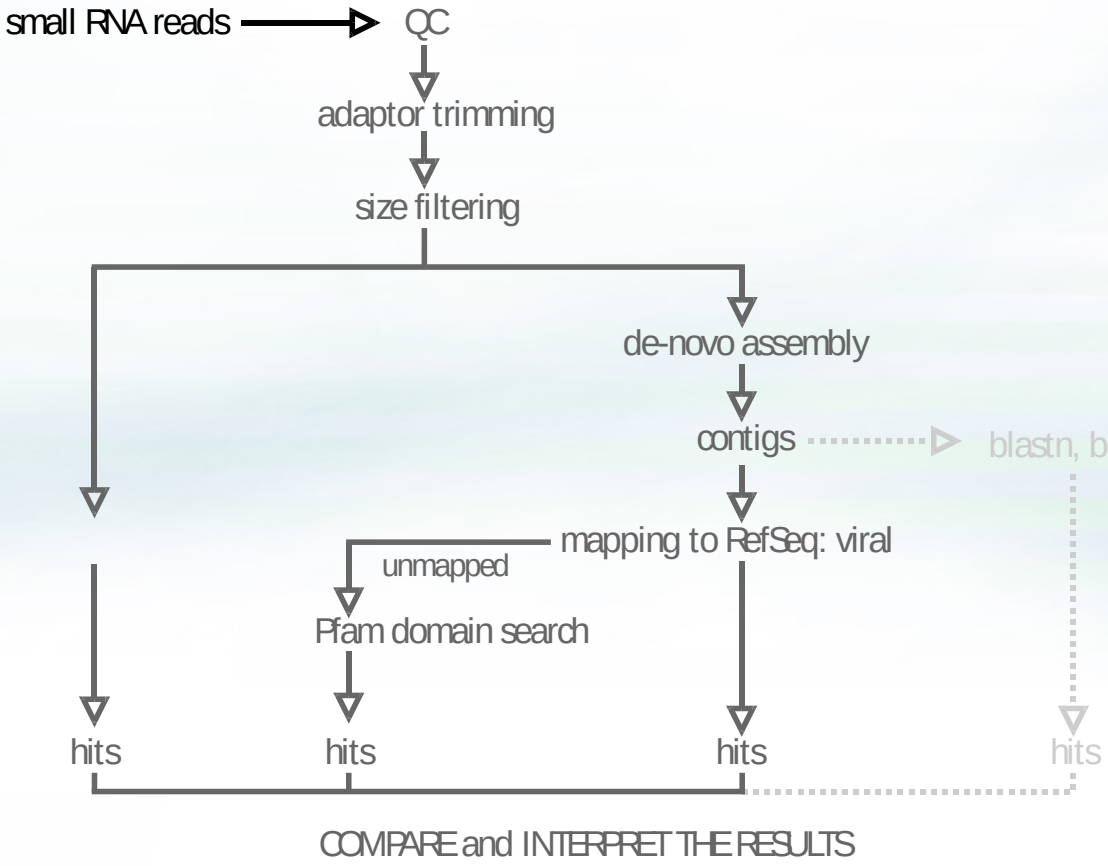
FINAL RESULTS AND CONCLUSIONS

- the outcomes presented in this study showed that all included known viruses/viroids could be identified by both NGS approaches
- for the viruses/viroids under study, the results showed higher yields of viral sequences in small RNA pool for viroids and viruses with no RNA replicative intermediates (ssDNA viruses)
- putative novel *Cytorhabdovirus,* discovered in this study, was only detected by analysing the data generated from ribosomal RNA depleted total RNA and not from the small RNA dataset, due to the low number of short reads in the latter
- finally, the results revealed the strength of NGS technology for the simultaneous detection and identification of several different known/unknown plant viruses from a different sample material, with a different amount of viral/viroid nucleotides and in a different host plants.
- Paper with this study was accepted in the journal Frontiers of Microbiology: Pecman et al.

frontiers in Microbiology

Diagnostic pipeline

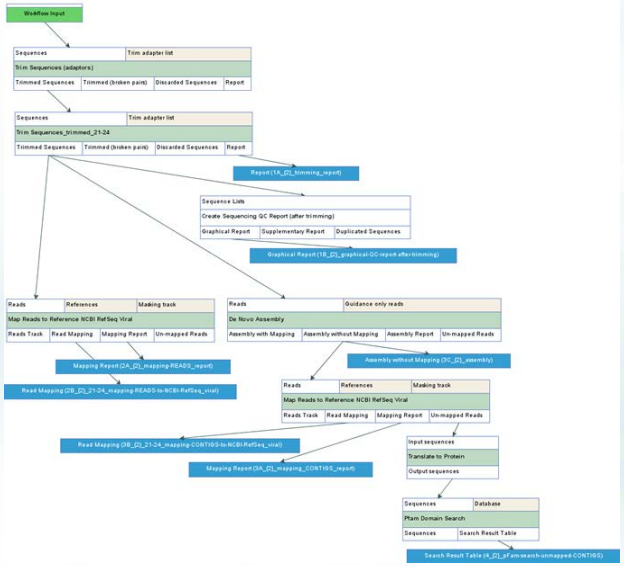
How to make the analysis simple and suitable for routine use in diagnostic set



Diagnostic pipeline

How to make the analysis simple and suitable for routine use in diagnostic set

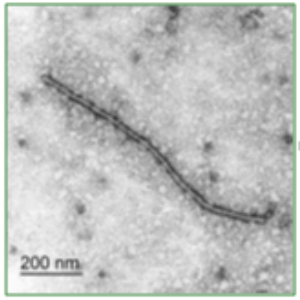
CLC Genomics Workbench allows to connect analyses into pipelines:



New viruses for Slovenia

- See poster Mehle et al

D566/16 (parsley)



ELISA (PVY, WMV, ZYMV): neg

other potyvirus or carlavirus

NEW VIRUSES IN SLO

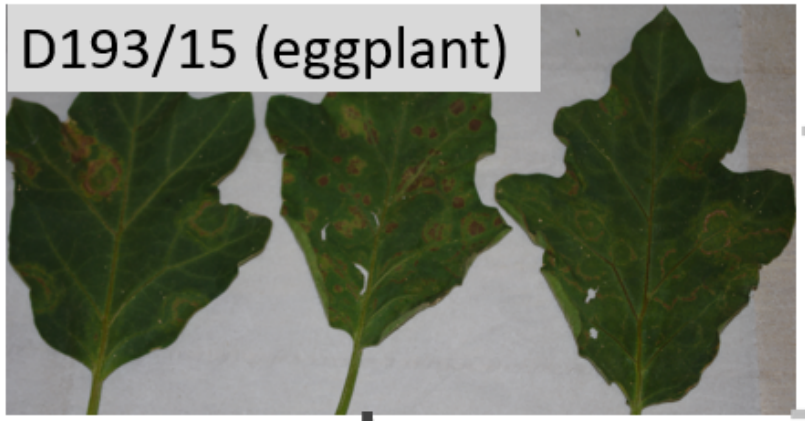
NGS:
Apium virus Y (ApVY) &
Carrot thin leaf virus (CTLV)
(both: *Potyviridae*, *Potyvirus*)

confirmed by ELISA, PCR and mechanical inoculation of test plants

ApVY was found also in asymptomatic parsley —> Is responsible for symptomatology CTLV alone or in mixed infection with ApVY?

???

D193/15 (eggplant)



mechanical inoculation of test plants: severe symptoms on several test plants

EM: neg

ELISA and PCR-based methods for several viruses: neg

unidentified virus or other pathogen

IDENTIFICATION OF VIRUS IN A SINGLE ASSAY

NGS:
Cucumber mosaic virus (CMV)
(*Bromoviridae*, *Cucumovirus*)

confirmed by ELISA on sample and test plants

D64/16 (petunia)

ELISA and PCR-based methods (CMV, CSNV, INSV, PVY, TMV, ToMV, ToRSV, TRSV, TSWV): neg

EM & mechanical inoculation of test plants: neg

NEGATIVE

NGS:
Petunia vein clearing virus (PVCV)
(*Caulimoviridae*, *Petuvirus*)

NEW VIRUS IN SLO

Additional research is needed to evaluate whether the detected viral sequences exist in an episomal form and not only integrated in the plant genome

???

Survey of viruses in tomato in Serbia, overlooked by targeted detection methods

In collaboration with University of Belgrade, Serbia, see poster by Ana Vučurović et al.

Preliminary Results

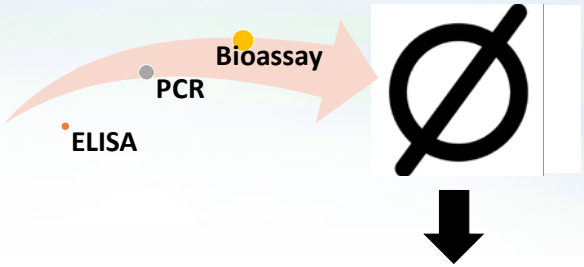
- 6/11 pooled samples with 1 or more viruses detected
- viruses from 5 different genera, 4 viruses are NEW findings for Serbia

4 new

- *Tomato torrado virus*
- *Physostegia chlorotic mottle virus ?*
- *Southern tomato virus*
- *Spinach latent virus*

1 known

- *Tomato spotted wilt virus*



Detection and identification of (new) viruse strain in plant with NGS

1. IDENTIFICATION of new strain of Henbane mosaic virus (HMV) in mixed infection with Potato virus M (PVM) and Southern tomato virus (STV)
2. Confirmation: PCR and test plants; the others: EM, ELISA, qPCR

Sampled plants – *S.lycopersicum* (photo: Patricija Pirnat)

Henbane mosaic virus:

- ✓first finding in *S.lycopersicum* (new host)
- ✓first finding in Slovenia

2. HMV SEPARATION from mixed infection on the selected test plants

3. HOST RANGE analysis:

severe symptoms of different plants from Solanaceae family:

S. lycopersicum cv. MoneyMaker, Roma, Riogrande;

S. melongena;

P. floridana;

D. stramonium;

H. niger;

N. tabacum cv. Samsung; *N. glutinosa; N. benthamiana*)

4. first COMPLETE GENOME SEQUENCE ASSEMBLED with sRNA and rRNA depleted totRNA dataset ~ 10112 nt long consensus

…in process: determination of genomic sequences of four additional known strains of HMV and phylogenetic analysis

HMV strain	average % of identity to HMV - Slovenia
HMV-146 (IPSP-Italy)	87,7
HMV-R (IPSP-Italy)	92,3
HMV-PV-76 (ATCC)	91,9
HMV-PV-79 (ATCC)	92,3

Reporting and discussion with authorities

- First report was: 9 days after sample acceptance (PVM detection with ELISA)
- Second information: 1 month (PVM confirmation with successful inoculation of test plants, EM)
- Third information: few months (results of NGS)
- Reaction of authorities: well accepted since specialist from administration is involved in research project on NGS
- HMV and STV would not be found without use of NGS

- Information of inspectors and other specialists on new tool and expected results on the national symposium

Key factors/problems/bottlenecks for the application of NGS in diagnostic

Pros:

1. better diagnosis (PVM was not shown to cause the symptoms on tomato, STV - no confirmation to cause symptom since it is not mechanically transmissible)
2. efficient first detection of new Potyvirus HMV for Slovenia, first founding on tomato
3. Indication of virus identity led to correct measures in the production site
4. HMV and STV would not be found without use of NGS

Cons:

- Distinction between integrated viral nucleic acid and viral particles is still a challenge

- Pooling of samples in order to reduce price can results in crosstalk
- One positive control of known infected plant material could be at the same time negative control for the whole procedure
- QA in progress

NATIONAL INSTITUTE OF BIOLOGY

Acknowledgement

NATIONAL INSTITUTE OF BIOLOGY

Denis Kutnjak, Nataša Mehle, Ion Gurtierrez Aguirre, Anja Pecman, Katarina Bačnik

COST-STSM

DIVAS COST Action FA 1407

Euphresco NGS-Detect

Slovenian Research Agency (Ph.D. research grant)

Heiko Ziebell

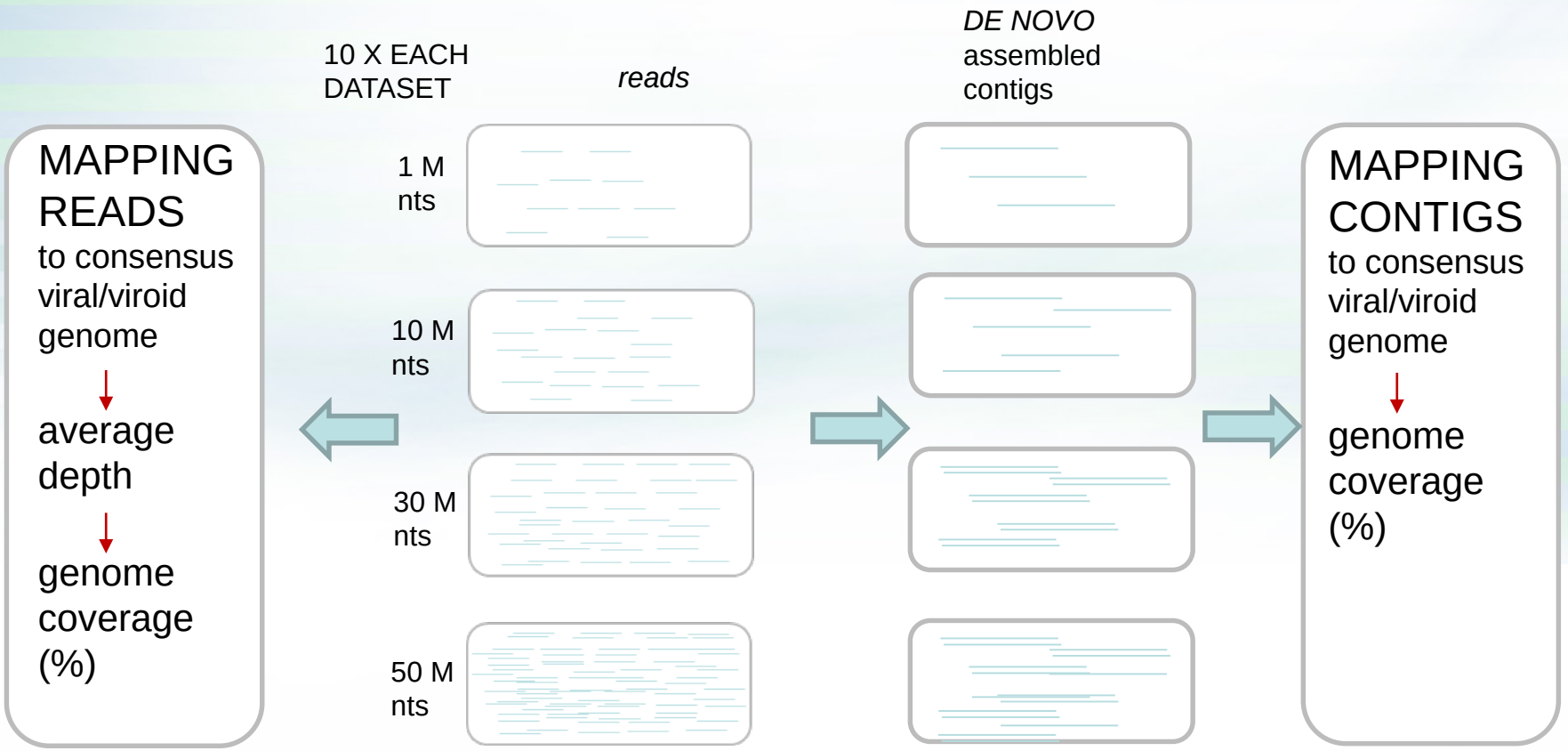
Neil Boonham, Ian Adams, Adrian Fox, Ummey Hany

fera

NATIONAL INSTITUTE OF BIOLOGY

THANK YOU FOR YOUR ATTENTION!!!

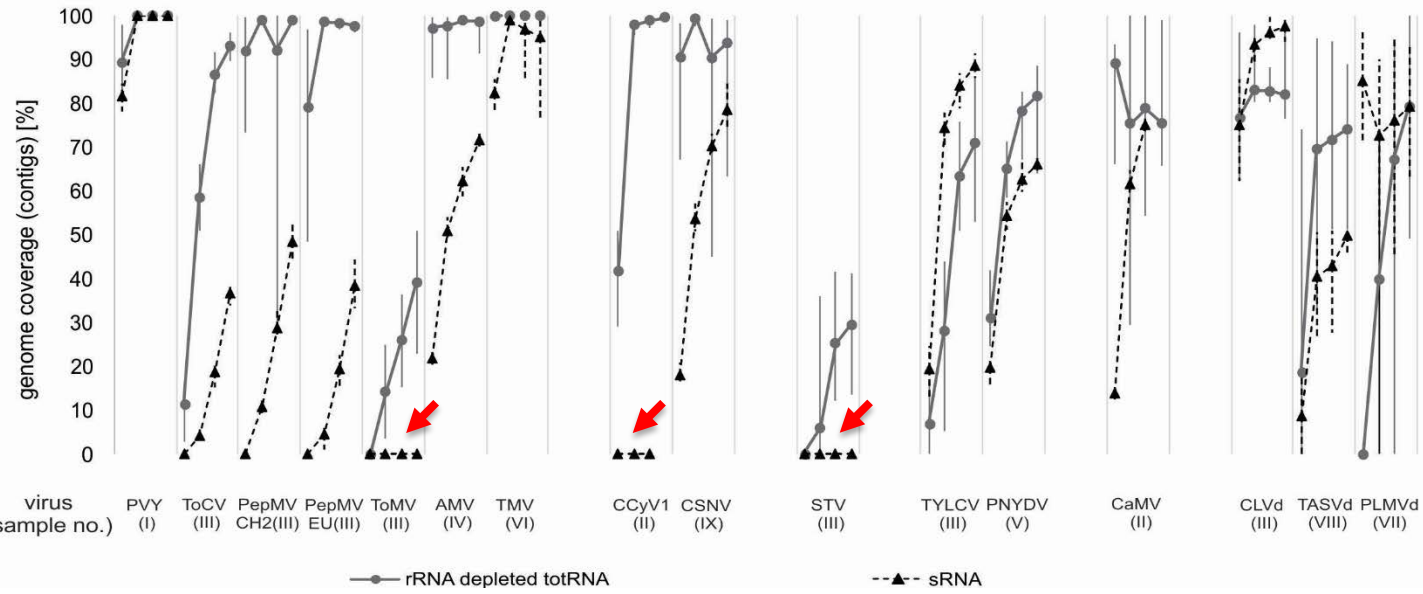
Comparison of sRNA and rRNA depleted totRNA approaches using data size-normalized subsamples



A

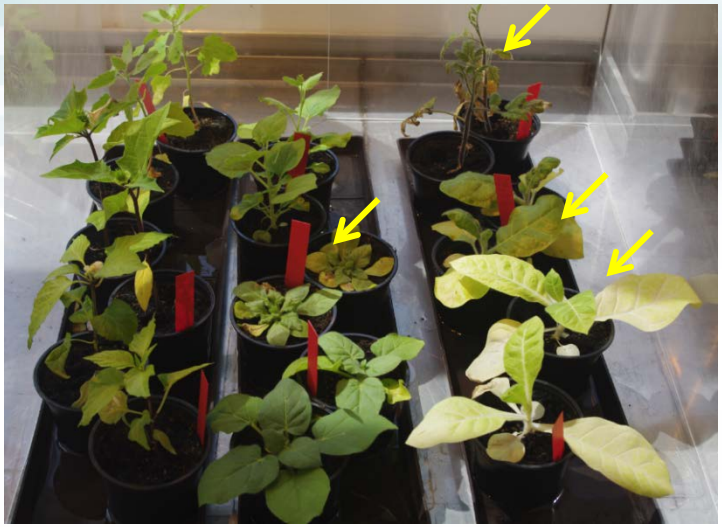
Baltimore group	ssRNA +							ssRNA -		dsRNA +/-	ssDNA +		dsDNA -RT	viroids		
sRNA [%]	49.9	1.7	1.1	0.9	0.008	21.1	50.6	0.05	33.1	0.013	3.0	10.1	12.0	2.6	0.4	11.3
rRNA dep. totRNA [%]	6.0	0.4	9.4	4.4	0.015	56.0	83.5	1.23	48.8	0.006	0.029	0.6	48.7	0.2	0.015	0.003

D

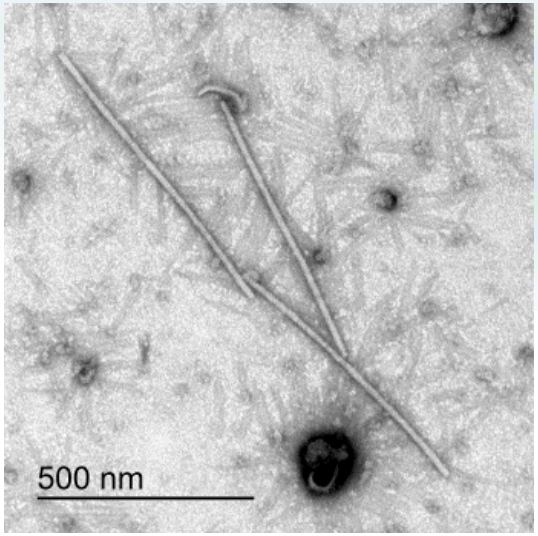


Sample number	Virus, Genus, Family	Baltimore classification	Genome organization	Abbreviations	Host	NGS results		results of confirmatory testing
						sRNA	rRNA depleted totRNA	
I	**Potato virus Y*	Group IV (ssRNA +)	Linear	PVY	*Solanum tuberosum*	+	+	+[a]
II	**Cauliflower mosaic virus*	Group VII (dsDNA-RT +/-)	Circular	CaMV	*Brassica oleracea*	+	+	+[a]
	Novel cabbage cytorhabdovirus 1	Group V (ssRNA -)	Linear	Novel CCyV1	*Brassica oleracea*	-	+	+[b]
III	**Tomato Yellow Leaf Curl Virus*	Group II (DNA +)	Circular	TYLCV	*Solanum lycopersicum*	+	+	+[a]
	Tomato chlorosis virus	Group IV (ssRNA +)	Linear	ToCV	*Solanum lycopersicum*	+	+	+[b]
	Pepino mosaic virus	Group IV (ssRNA +)	Linear	PepMV	*Solanum lycopersicum*	+	+	+[c]
	Tomato mosaic virus	Group IV (ssRNA +)	Linear	ToMV	*Solanum lycopersicum*	+	+	+[c]
	Southern tomato virus	Group III (dsRNA +/-)	Linear	STV	*Solanum lycopersicum*	+	+	+[b]
	Columnela latent viroid	viroid	Circular	CLVd	*Solanum lycopersicum*	+	+	+[b]
IV	**Alfalfa mosaic virus*	Group IV (ssRNA +)	Linear, segmented	AMV	*Nicotiana tabacum*	+	+	+[a]
V	**Pea necrotic yellow dwarf virus*	Group II (ssDNA +)	Circular segmented	PNYDV	*Pisum sativum*	+	+	+[b]
VI	**Tobacco mosaic virus*	Group IV (ssRNA +)	Linear	TMV	*Nicotiana sp.*	+	+	+[b]
VII	**Peach latent mosaic viroid*	viroid	Circular	PLMVd	*Prunus sp.*	+	+	+[b]
VIII	**Tomato apical stunt viroid*	viroid	Circular	TASVd	*Solanum lycopersicum*	+	+	+[b]
IX	**Chrysanthemum stem necrosis virus*	Group V (ssRNA -)	Linear	CSNV	*Nicotiana benthamiana*	+	+	+[c]

- *Potato virus M* (PVM) detected with ELISA on sampled plants and subsequently confirmed by electron microscopy and ELISA on 4/9 species of test plants used (*Solanum lycopersicum* cv. Moneymaker (similar symptoms as on the sample), *Nicotiana rustica*, *Nicotiana tabacum* cv. White Burley, *Nicotiana clevelandii*). We did not find any data on PVM symptoms, just report on asymptomatic systemic infection of tomato.



Mechanically inoculated test plants



TEM of mechanically inoculated test plants

Wet lab, sequencing and IT pipeline

- RNA isolation from tomato field sample and inoculated test plant with Trizol and RNeasy → sent for Illumina sequencing; sRNA library preparation
- HiSeq2500, 1/12 of a lane, sequencing mode:1x50 bp, sRNAs: 20 – 24 bp; ~20 million reads;
- reads were subsampled to enhance de-novo assembly (the coverage was too high for optimal algorithm performance): 250.000, 750.000, 500.000, **1.000.000,** 1.500.000
- Raw data processing in CLC Genomics (Qiagen); results were compared between Trizol and RNeasy isolation

RESULTS

- RNA isolation with Trizol gave better outcomes, when comparing de novo assembly results. Nevertheless, RNeasy isolation also enabled confident virus detection.

- In silico host removal enhanced de novo assembly of viral genomes, yet it is not necessary for confident virus detection, - diagnostics in plants without sequenced genome.

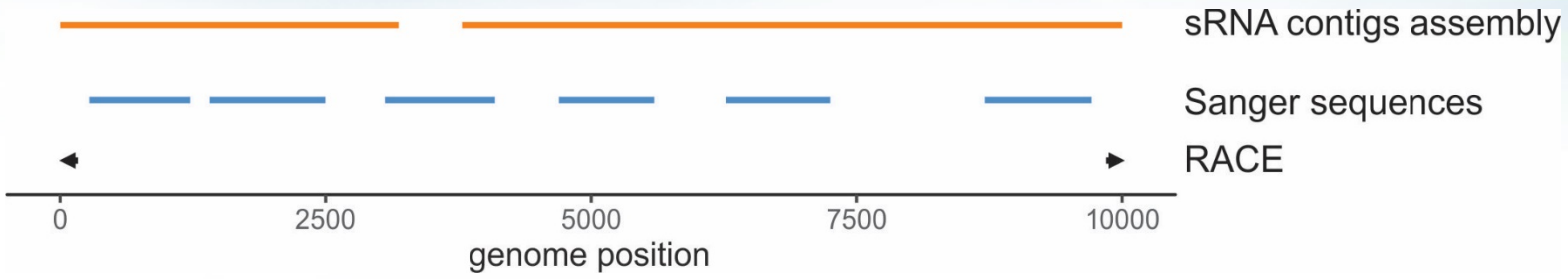
- Blast of de novo assembled contigs confirmed presence of PVM and STV (high number of hits with high % of identity), however many other contigs matched viruses from Potyvirus genus with low % of identity

Genome assembly and confirmation the presence of STV and Potyvirus

- PVM consensus genome assembled
- STV consensus genome assembled: confirmation with Sanger sequencing
- Unknown potyvirus : one of the contigs matched Henbane mosaic virus (blastn 86% identity, blastx 94% identity, query coverage 13%(1273 nt)) → Sanger sequencing for confirmation was done

Assembly of complete Henbane mosaic virus genome

- Defining position (order) and scaffolding of the contigs using blastx to the most similiar potyvirus – Chilli veinal mottle virus
- Designing PCR primers, PCR, Sanger sequencing
- Assembly: contigs + Sanger sequences
- Re-maping reads to the new, near complete genome
- RACE (Rapid amplification of cDNA ends): presented by Anja Pecman, done at FERA)



- still needs to be done: sequence of the „original“ virus, only part of the genome (~1000 nts, partial CP sequence) is published in database

NATIONAL INSTITUTE OF BIOLOGY

BIOINFORMATIC WORKFLOW

Raw reads

↓

Trimming: adapters

1. ASSAY

↓

Trimming: size selection
< 20 nt & >24 nt

↓

Mapping trimmed
reads to host genome

↓

Subsampling unmapped
reads: 200 000 nt (4x)

2. ASSAY

↓

Trimming: size selection
< 20 nt & >24 nt

↓

Subsampling trimmed reads:
200 000 nt (5x)

1. Mapping to viral ref seq (NCBI)
2. Mapping to PVM
3. Mapping to STV
4. De novo assembly → Contig multi blast (NCBI)

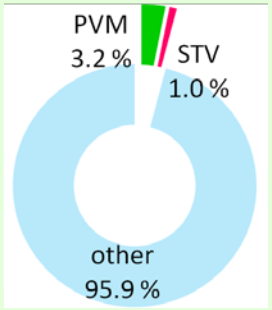
NATIONAL INSTITUTE OF BIOLOGY

Trizol isolation

1. Mapping (trimmed reads) to NCBI ref seq: viral

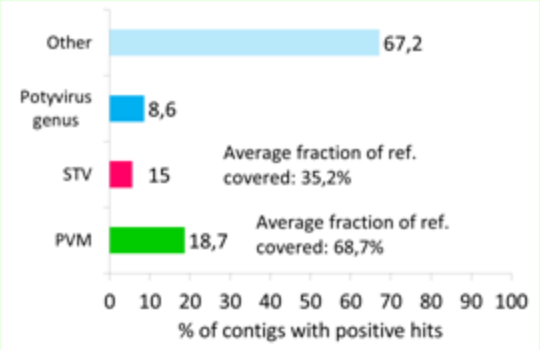
Two longest consensus sequences with good coverage:
Potato virus M (PVM) and *Southern tomato virus* (STV)

2. Mapping to PVM
Average depth of coverage: 15,57
Average fraction of ref. covered: 90,36 %



3. Mapping to STV
Average depth of coverage: 11,72
Average fraction of ref. covered: 94,15 %

4. De novo assembly and multi blast of 268 contigs (N50=136 nt)



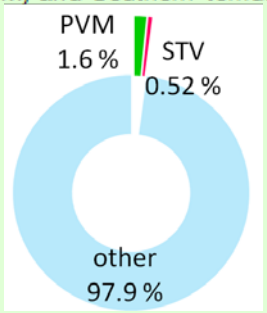
MULTI BLAST RESULTS:
The results with most hits and with high % of identity: PVM. Other hits map to plant genomes, STV and to viruses from Potyvirus genus.

RNeasy isolation

1. Mapping (trimmed reads) to NCBI ref seq: viral

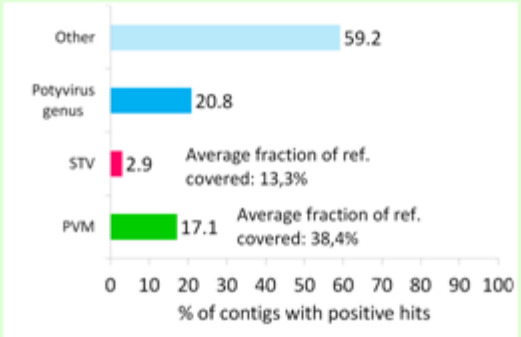
Two longest consensus sequences with good coverage:
Potato virus M (PVM) and *Southern tomato virus* (STV)

2. Mapping to PVM
Average depth of coverage: 7,84
Average fraction of ref. covered: 85,85 %



3. Mapping to STV
Average depth of coverage: 6,33
Average fraction of ref. covered: 81,47%

4. De novo assembly and multi blast of 240 contigs (N50=109 nt)



MULTI BLAST RESULTS:
The results with most hits and with high % of identity: PVM. Other hits map to plant genomes, STV and to viruses from Potyvirus genus.