siRomics for universal diagnostics of plant viral disease and virus diversity studies

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Silvia Turco
aus Agrigento, Italien.

Basel, 2017.

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Thomas Boller, PD Dr. Mikhail Pooggin and Dr. Olivier Schumpp.

Basel, 14 November 2017.

Prof. Dr. Martin Spiess

General preface

Financial support of this PhD work was provided by the EU Marie-Curie IDP Bridges grant to PD Dr. Mikhail Pooggin from the Department of Environmental Science, Botany at the University of Basel, in collaboration with 13 PIs the Basel-Zurich Plant Science center in the frame of the European Union's Seventh Framework for research, technological development and Demonstration.

This work was done in close collaboration of the team of Dr. Pooggin and the virology research group of Dr. Olivier Schumpp at Swiss Federal Institute Agroscope, Nyon. Dr. Pooggin and his team at the University of Basel provided the expertise in molecular plant virology, plant antiviral defense based on small RNA-directed silencing and bioinformatic analysis. Dr. Olivier Schumpp and Dr. Jean-Sebastien Reynard from Agroscope provided the expertise in molecular plant virology, crop-protection and plant production. The PhD committee was composed of PD Dr. Mikhail Pooggin as Academic Advisor, Prof. Dr. Thomas Boller as Faculty representative and Dr. Olivier Schumpp as Associated Partner.

Abstract

Traditional methods of viral diagnostics using specific antibodies and PCR often fail to identify a viral pathogen. In our EU Marie-Curie IDP bridges project, we used an alternative novel approach called siRomics which allows not only to detect the virus but also to *de novo* reconstruct a complete consensus master genome in the viral quasispecies population.

The main plant antiviral defense system is based on RNA silencing mediated by small RNAs. In plants infected with DNA and RNA viruses, host Dicer enzymes generate 21-24 nucleotide (nt) viral small interfering RNAs (siRNAs) that restrict virus replication and systemic spread. Growing evidence indicates that viral siRNAs are derived from the entire genome sequence of RNA and DNA viruses and accumulate at high levels. Hence it appears feasible to reconstruct a complete viral genome simply from viral siRNA species. Current bioinformatics algorithms enable *de novo* assembly of genomes and transcriptomes from short sequencing reads. In the past years, the siRomics pipeline, developed by Seguin et al. (2014b) in model plants, was further applied in crop plants (Seguin et al. 2014b, 2016, Rajeswaran et al. 2014a, 2014b, Fuentes et al. 2016). Thus, our siRomics approach has the potential for universal diagnostics of plant virus disease and de novo reconstruction of viral genomes in mixed infections.

In this study we applied siRomics for virus detection and virome reconstruction in several case studies of economically-important viral diseases in Switzerland. In naturally-infected *Solanum tuberosum* (potato), one case study revealed a virome comprising *Potato virus Y* (genus *Potyvirus*) and *Potato virus X* (genus *Potexvirus*), which was reconstructed by *de novo* assembling separate genome-size sRNA contigs. Another case study revealed a virome comprising NTN and O strains of *Potato virus Y*, whose sRNAs assembled in chimeric contigs which could be disentangled on the basis of reference genome sequences. Both viromes were stable in vegetative potato progeny. In a cross-protection trial of *Solanum lycopersicum* (tomato), the supposedly protective mild strain CH2 of *Pepino mosaic virus* (*Potexvirus*) was tested for protection against the strain LP of the same virus. Reciprocal mechanical inoculations eventually resulted in co-infection of all individual plants with CH2 and LP strains, reconstructed as separate sRNA contigs. LP invasions into CH2-preinfected plants and vice versa were accompanied by alterations of consensus genome sequences in viral quasispecies, indicating a potential risk of cross-protection measures. Additionally, the study also revealed, by reconstruction from sRNAs, the presence of the mechanically non-transmissible *Southern tomato virus* (*Amalgavirus*) in some plants. Our in-depth analysis of sRNA sizes, 5'-nucleotide frequencies and hotspot maps revealed similarities in sRNA-generating mechanisms in potato and tomato, differential silencing responses to virome components and potential for sRNA-directed cross-targeting between viral strains which could not, however, prevent the formation of stable viromes. Furthermore, by siRomics we characterized the virome present in cultivated and non-cultivated perennial plants including grapevine, cherry, fig, privet and larch. As expected, grapevine samples showed a complex virome, including viroids, in particular *Grapevine Fanleaf virus*, *Grapevine virus A*, *Grapevine leafroll associated virus*, *Yellow speckle viroid 1*, *Yellow speckle viroid* 2, *Hop*

stunt viroid and *Australian grapevine viroid*. In cherry trees affected by little cherry disease, we confirmed that the presence of two *Little cherry virus* (1 and 2, respectively) in one of the samples, induces more severe symptoms compared with the sample where only *Little cherry virus 1* was present. In a fig tree exhibiting virus-like symptoms coming from a private garden, new isolates of Fig mosaic virus and Fig Badnavirus-1 were identified and reconstructed. In the forest bush plant privet (*Ligustrum vulgare*) showing yellow mosaic disease, a novel virus distantly related to Barley yellow strip virus and Lychnis ringspot virus was identified, fully reconstructed and named Ligustrum mosaic virus. Our work combined multi-disciplinary approaches ranging from advanced molecular methods of next generation sequencing to sophisticated bioinformatics algorithms for virus genome reconstruction. The results of our study are informative for further understanding the mechanisms of RNA silencing-based antiviral defense, which would contribute to basic research in the field of plant-pathogen interaction, and for developing novel strategies of virus control, which could potentially be implemented in the future in Swiss agriculture though our recommendations to the policy makers. In modern agriculture, horticulture and (bio-) farming, it becomes critical to assess the risk of emerging plant infections and to control the spread of plant viral diseases.

Table of contents

1. Introduction

1.1 History and classification of plant viruses

Viruses are intracellular parasites that can infect all forms of living organisms. In plants, they can be transmitted by invertebrate animals such as insects and nematodes as well as by protists and fungi or by mechanical inoculation (e.g. sap contact) and they all need the host molecular machinery for replication. The first report of a viral disease goes back to a Japanese poem (Empress Koken, 752 AD) describing the “autumnal” yellow leaves of Eupatorium plants in summer (in 2003, Saunders et al., proved that the disease was due to a geminivirus infection). Later descriptions of viral infections were found in Carolus Clusius's publications (1576) and in Daniel Rabel's illustrations (1662) regarding the “tulip breaking”.
But it's only at the end of the nineteenth century that plant molecular virology sees the light, with the first transmission experiments in tobacco plants by Mayer (1886) and Iwanowski (1892), demonstrating that the sap extracted from a diseased plant can infect an healthy one, even when the sap is passed through a Pasteur-Chamberland filter candle that retains bacteria. In 1898 Beijerinck repeated the filtering experiment and described the cause of Tobacco mosaic disease as “Contagium vivum fluidum”, giving the term *virus* and distinguished it from the corpuscular bacteria.
Nowadays, 4405 species of viruses have been identified (International Committee on Taxonomy of Viruses (ICTV) report 2016), which can be classified in seven major groups according to the nature of the genome, as defined by the Baltimore classification (Fig.1). Within each of the seven groups, viruses are further classified into families, genera and species according to particle morphology, genomic, biological and serological properties. Below the genome organization and replication strategies of representative plant viruses from different groups will be introduced with a focus on the viruses investigated in this PhD work.

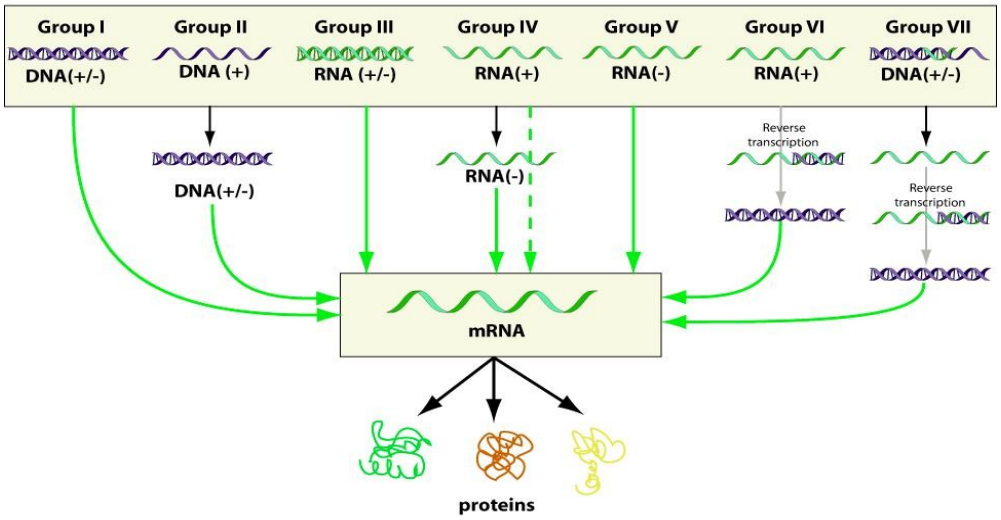


Fig. 1: The Baltimore Classification of viruses.

1.2 DNA viruses

Among the Group I of double-stranded (ds) DNA viruses, the family *Phycodnaviridae* is the only one infecting lower plant species, such as algae. The genome of these viruses is 100-560 Kb long, carrying 700 ORFs and enveloped in an icosahedral capsid of 100-220 nm in diameter and T=169. The dsDNA viruses replicate in the host cell nucleus via DNA-DNA replication. In higher plants, there are dsDNA viruses from family *Caulimoviridae*, which replicate by reverse transcription (RT) of an RNA intermediate and for this reason belong to the Group VII (dsDNA-RT). The monopartite circular dsDNA of 7-8.2 kb in length is released from virions into the nucleus where the discontinuous single-stranded nicks are repaid creating mini-chromosomes that can be then transcribed by the host RNA polymerase II. The RNA intermediate is then transported into the cytoplasm, translated into viral proteins by host ribosomes and converted back to dsDNA by the viral reverse transcriptase and encapsidated in isometric particles (e.g. genus *Caulimovirus*) or bacilliform particles (e.g. genus *Badnavirus*, Bhat et al., 2016). An example of *Badnavirus* investigated in this PhD study is *Fig badnavirus 1* (FBV-1), isolated by Tzanetakis et al. (2010) from a fig tree with leaf mosaic symptoms. The FBV-1 genome is 7.1 kb in length and has four ORF on one strand: ORF1 coding for P1 of 15.3 kDa, with unknown function, ORF2 for P2 of 16.5 kDa with a DNA binding site domain, ORF3 coding for a polyprotein consisting of movement protein (MP), coat protein with zinc-finger domain (CP), pepsin-like aspartate protease (Pro), reverse transcriptase (RT) and RNase H, and ORF4 coding for P4 with unknown function (Fig. 2, Laney et al., 2012).

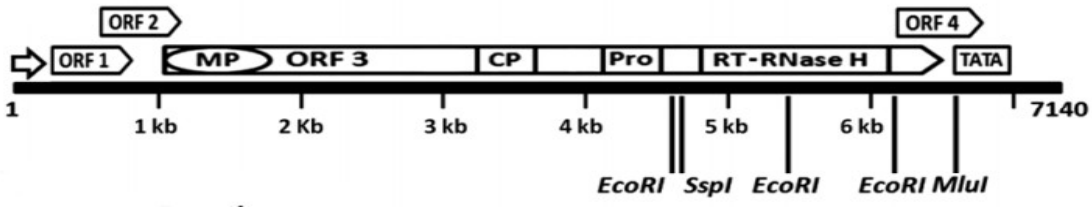


Fig. 2. Linear representation of Fig badnavirus-1 (FBV-1) genome showing MET-tRNA- primer-binding site (denoted by an arrow); TATA-box (denoted by a box); open reading frame (ORF) 1; ORF 2; ORF 3 with movement protein (MP), capsid protein zinc-finger domain (CP), pepsin-like aspartate protease (Pro), reverse transcriptase (RT) and RNase H (RNase H) motifs; and ORF 4 (Copied from Laney et al., 2012)

Single-stranded DNA (ssDNA) viruses belonging to the group II are represented in plants by the families *Geminiviridae* and *Nanoviridae*. The first one is the largest with 7 genera and more than 300 species. The genome of geminiviruses can be either monopartite or bipartite between 2.5-3 Kb, incorporated in an icosahedral capsid of 18-20 mm in diameter. To the *Geminiviridae* family belongs the bipartite begomoviruses such as *Bean golden mosaic virus* (BGMV) and *Cassava mosaic virus* with the DNA segment A carrying one or two genes in the virion sense strand (AV1 or AV1 and AV2) and four genes in the complementary antisense strand (AC1, AC2, AC3 and AC4) coding for proteins involved in replication,

transcription, encapsidation and silencing suppression. DNA B has two genes (BV1 and BC1) involved in movement (Fig.3, Karthikeyan et al., 2016).
Members of the *Nanoviridae* family have a genome composed of 6 or 8 segments of 1 kb each, encapsidated in separate icosahedral particles.
Both gemini- and nanoviruses replicate in the nucleus through a rolling circle replication mediated by the host DNA polymerase.

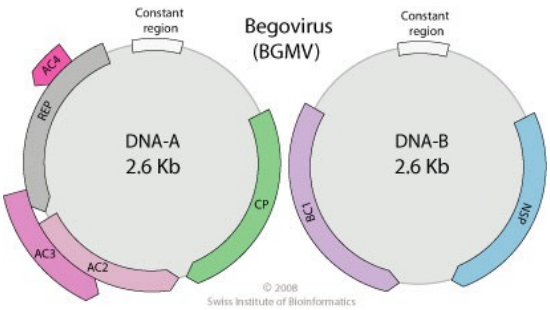


Fig. 3. Begomovirus genome components A and B, circular, ssDNA genome (+) of about 2.6 kb per component (5.2 kb in total). 3' terminus has no poly(A) tract. There are coding regions in both the virion (positive) and complementary (negative) sense strands. (Pictured copied from www.viralzone.expasy.org) .

1.3 RNA viruses

dsRNA viruses

The families *Reoviridae*, *Partitiviridae* and *Amalgaviridae* belong to the group III with double stranded RNA (dsRNA) genome. The genomes of reoviruses infecting plants has 10 (genera *Fijivirus* and *Oryzavirus*) or 12 segments (genus *Phytoreovirus*) of linear dsRNA encapsidated in a double-shelled particle, 65-80 nm in diameter. They differ in the vector (delphacid planthoppers or leafhoppers) and in the host (monocotyledons from *Graminae* and *Liliacea* families or dicotyledons). The members of *Partitiviridae* have a genome of about 4kb divided in two segments, one encoding for the RNA dependent RNA polymerase, the other one for the capsid protein. Two genera of partitiviruses infect plants: *Alphacryptovirus* and *Betacryptovirus*. The family *Amalgaviridae*, to which *Southern tomato virus* belongs, genus Amalgavirus, has a genome composed of ~3.5 kb carrying two overlapping ORFs encoding a putative coat protein (CP) and RNA-dependent RNA polymerase (RdRP): the latter is presumably translated as a CP-PdRP fusion protein by frameshifting (Fig.4)

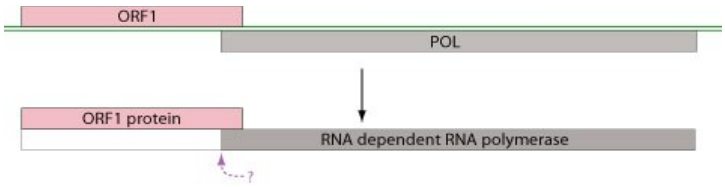


Fig.4. 3.5 kb Linear dsRNA genome of Southern tomato virus. Contains 2 overlapping ORFs with the viral polymerase (POL) translated as an ORF 1-POL fusion protein via ribosome frameshifting (Pictured copied from www.viralzone.expasy.org).

ssRNA viruses

The group IV of positive-sense ssRNA (ssRNA+) viruses is the most represented among plant viruses, with eight families and about 70 different genera.

The members of *Bromoviridae* family, including six genera, have a linear tripartite genome and a subgenomic segment encapsidated in distinct particles with icosahedral or bacilliform geometry. Once in the cytoplasm, genomic ssRNAs from segment 1 and 2 can be directly translated into replication proteins, while RNA3 and subgenomic (sg)RNA4 are translated into movement and capsid proteins, respectively.

The members of *Secoviridae* family have a bipartite genome as well represented by the *Grapevine fanleaf virus* (GFLV), genus Nepovirus, which affects grapevines all over the world, inducing distortion, yellowing and mottling of the leaves, vein-banding, short internodes (Martelli, 1993). The genome is composed of two positive single-stranded RNA segments, with a viral protein linked at the 5' end and a poly(A) tail at their 3'-end. RNA1 is 7.4 kb long, RNA2 is 3.7 kb long and both code for polyproteins P1 and P2 which are then cleaved in the functional proteins: helicase (Hel), viral genome-linked protein (VPg), proteinase (Pro), RdRP, homing protein (HP), MP and CP (Pinck et al. 1988).

Viruses from the *Tombusviridae* are divided among 13 different genera and the name derives from one particular Tombusvirus, the Tomato bushy stunt virus (TBSV). The isometric and icosahedral particles have a monopartite linear genome of 4.6-4.8 kb with one or more subgenomic RNAs.

Members of the family *Luteoviridae* have spherical particles composed of 180 subunits of a single coat protein of ~22 kDa. The monopartite linear genome of 5.3-7 kb has six different ORF and no poly(A) or tRNA-like structure at the 3'-end.

The family *Closteroviridae* is divided in four different genera (*Closterovirus, Ampelovirus, Crinivirus and Velarivirus*) according to the genome type (mono-, bi- or tri-partite) and size (Martelli, 2014). To the Closterovirus genus belong the *Fig mild mottle-associated virus* (FMMaV), *Fig leaf mottle-associated virus 1* (FLMaV-1) and *Fig leaf mottle-associated virus 2* (FLMaV-2).

Little cherry virus-1 (LChV-1, genus *Velarivirus*, 16.9 Kb) and *little cherry virus-2* (LchV-2, genus *Ampelovirus*, 15 Kb, Fig.5) are known to be responsible for little cherry disease (LCD) that affects sweet cherry (*Prunus avium*) and sour cherry (*Prunus cerasus*) inducing reduction of fruit size, color and taste and leaf-reddening.

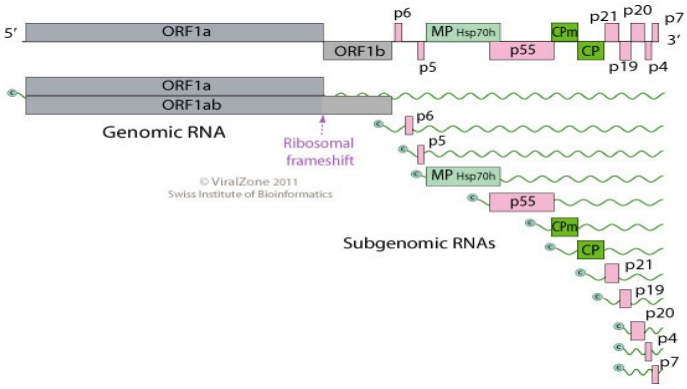


Fig.5 Schematic representation of Ampelovirus genome. Linear, ssRNA(+) genome of 16.9 to 17.9 kb. 3' terminus has no poly(A) tract and the 5' terminus probably has a methylated nucleotide cap. Encodes 13 proteins (Picture copied from www.viralzone.expasy.org)

The grapevine leafroll disease (GLRD) is associated with different viruses belonging to *Closteroviridae* and referred to as grapevine leafroll-associated viruses (Fig. 6, GLRaVs, Reynard et al., 2015). Twelve different GLRaVs have been identified: one member of *Closterovirus* (GLRaV-2), 10 members of *Ampelovirus* (GLRaV-1, -3, -4, -5, -6, -8, -9, -Pr, -De, -Car) and one member of *Velarivirus* (GLRaV-7) genera (Martelli et al.2012, Al Rwahnih et al. 2012).

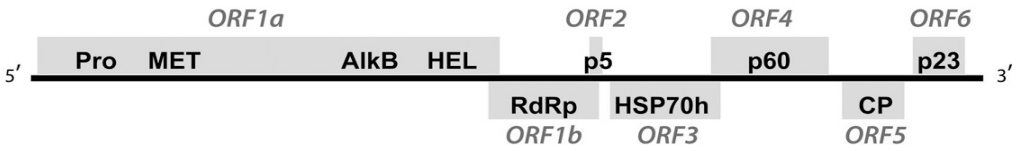


Fig.6 Schematic representation of the GLRaV-4 Ob. Putative open reading frames (ORFs) are shown in boxes: ORF1a with corresponding domains: Pro = protease, MET = methyltransferase, AlkB = 2OG-Fe(II) oxygenase domain, HEL = helicase; ORF1b = RNA-dependent RNA polymerase; ORF2 = small 5 K protein; ORF3 = heat shock 70 protein homolog; ORF4 = 60 K protein; ORF5 = coat protein; and ORF6 = 23 K protein. Copied from Reynard et al., 2015)

The family *Potyviridae* is one of the largest families of plant viruses. Members of this family have a monopartite RNA genome (except for genus *Bymovirus* with bipartite genome) translated in a polyprotein of about 350 kDa. The virions are flexuous and filamentous rods of 680-950 nm in length and 11-15 nm in diameter (Valli, 2015).

The prototype member of this family is *Potato virus Y* (PVY), genus Potyvirus, transmitted by grafting, sap inoculation and at least 25 aphid species with *Myzus persicae* is the most common vector all over the world. The PVY genomic RNA is 9.7 kb in length, with a VPg protein linked at the 5'-end, and a polyadenylated sequence at the 3'-end. A single large ORF codes for the polyprotein which is then auto catalytically cleaved in nine functional proteins: P1/proteinase, HC-Pro (helper component-proteinase) involved in movement and silencing suppression, P3 involved in replication, 6K1, CI/helicase, 6K2, VPg, Nia/proteinase, Nib/RdRP and CP. The ninth protein, P3N-PIPO involved in movement, derives from a ribosomal frameshift at the 5' terminal part of the P3 gene (Fig.7, Chung et al. 2008).

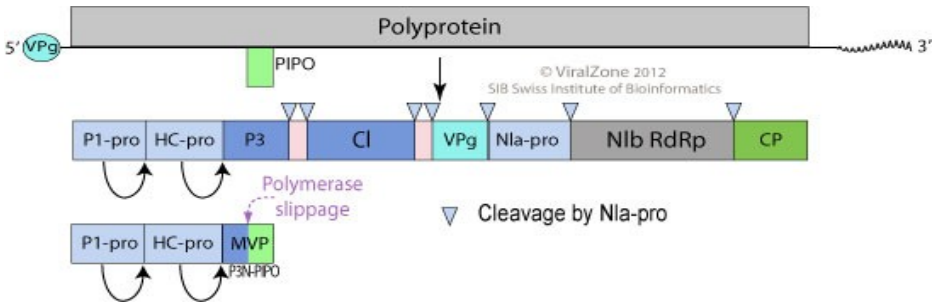


Fig.7: Schematic representation of potyvirus monopartite linear genome of 10 kb in size. 3' terminus has a poly (A) tract. 5' terminus has a genome-linked protein (Vpg). Pictured copied from www.viralzone.expasy.org.

The infection cycle starts with the entrance of the coated potyvirus genomic RNA (virion) inside the host cell from outside (e.g. delivered by the aphid) or from an adjacent infected cell. Inside the cell, the positive-sense genomic RNA is uncoated to be translated into the polyprotein and once the RdRP is produced by self-proteolysis (Fig. 7), synthesis of a negative strand starts viral replication: the negative strand serves as a template for synthesis of multiple positive strands that can be translated or encapsidated by CP into new infectious virions.

Several isolates of PVY have been identified and classified according to the symptoms they induce in the infected potato. The ordinary strain of PVY (PVY^{O}) causes severe symptoms such as crinkling, stunting, mild mosaic and mild tuber damage without leaf necrosis. On the contrary, the necrotic strain ($PVY^{N)}$ induces venial leaf necrosis and no tuber symptoms in susceptible cultivars where the Nc or Ny genes are not present. The common strain PVY^{C} induces mosaic pattern and stipple streak, stimulating a hypersensitive response in potato cultivars carrying the Nc resistance gene. In 1978, a recombinant strain between PVY^{O} and PVY^{N} (PVY^{NTN}), was identified in Hungary as the responsible for the so called tuber necrotic ringspot disease (Ahmadvand et al, 2012).

The family *Flexiviridae* was split in 2009 into three new families *Alphaflexiviridae*, *Betaflexiviridae* and *Gammaflexiviridae*, all classified under the order *Tymovirales.*

Potato virus X (PVX) belongs to genus *Potexvirus* of the family *Alphaflexiviridae* and can infect solanaceous host plants by mechanical transmission (Verchot-Lubicz et al., 2007). It has a monopartite genome with a cap at the 5' end, followed by 84-nucleotide (nt) 5'-UTR arranged in three stem-loop structures (Choi et al, 2016), five ORFs and 72-nt 3'-UTR arranged in three stem loops (Batten et al, 2003). The first ORF encodes the viral RdRP of 160 kDa which is translated from the genomic RNA.

ORF2, ORF3 and ORF4 known as the triple-gene block (TGB), are translated from two subgenomic RNAs into 25kDa (TGB1), 12kDa (TGB2), 8kDa (TGB3) proteins required for cell-to-cell movement. The last ORF is also translated from subgenomic RNA into the viral CP of 25kDa, which encapsidates the whole 6.4 kb long genomic RNA into flexuous filament and is also needed for viral movement (Fig.8).

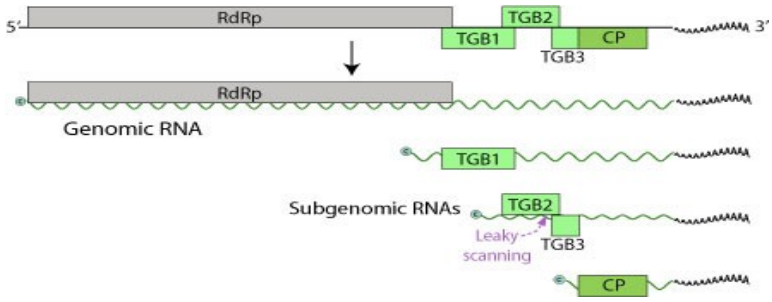


Fig.8 Schematic representation of potexvirus genome of 5.9-7 kb in size. The 5' end is capped and 3' terminus is polyadenylated. Encodes for 5 proteins. (Picture from www.viralzone.expasy.org)

Pepino mosaic virus (PepMV) is a potexvirus isolated from pepino (*Solanum muricatum*) in 1974 in Peru (Jones et al., 1980) that began to spread in tomato crop (*Solanum lycopersicum*) from the 2000s (Hanssen et al., 2009). Like PVX, PepMV possesses a monopartite genome of is 6.4 kb in length with the secondary and tertiary conserved structures at the 3' UTR required for RdRp binding and negative-strand RNA synthesis (Osman et al. 2014). According to a sequence alignment between isolates from all over the world, PepMV strains can been distinguished as the original Peruvian genotype (LP, from *Lycopersicon peruvianum* where it was first isolated), the European genotype (EU), the American genotype US1 and the Chilean genotype CH2. The symptomatology induced by PepMV in the infected crop can be variable, from asymptomatic to severe (fruit marbling, open fruit, fruit discoloration, leaf bubbling, yellow spots) are also influenced by environmental and light conditions, tomato and virus genotype (Hanssen and Thomma, 2010).

The family *Betaflexiviridae* is represented by *Grapevine virus A* (GVA) that belongs to the genus Vitivirus. The genome of GVA is 7.3 kb long, capped and polyadenylated with five slightly overlapping ORFs coding for methyltransferase, helicase, RdRp, movement and coat protein (Minafra et al., 1994). The virus has been associated with a rugose wood (RW) complex, one of the major disease affecting grapevines, together with grapevine leafroll disease (GLD), first reported by Graniti et al., in 1966. It's transmitted by grafting, propagation of infected material but also by mealybugs (Bertin et al., 2016) and induces marked wood alteration (Martelli 2014).

To the same family, but genus *Capillovirus* belongs the *Cherry virus A* (CVA), which genome consists of 7.4 kb polyprotein of 266 kDa (ORF1) encoding for the coat protein and the replication-related proteins, and a 53kDa protein (ORF2) involved in cell-to-cell movement (Jelkmann, 1995).

The famous *Tobacco mosaic virus* (TMV), the first one to be studied and characterized, belongs to genus *Tobamovirus* of the positive-sense ssRNA family *Virgaviridae*. The family name is derived from rod-shaped helical virions (from Latin virga = rod) of 20-25 nm in diameter with a central channel. The genomes of *Virgaviridae* can be monopartite, like in genus *Tobamovirus*, or segmented, like e.g. in genus *Hordeivirus*, and has no poly(A) but a tRNA-like structure at the 3'-terminus. Interestingly, in hordeiviruses such as *Barley stripe mosaic virus* (BSMV, Fig. 9) the 3'-teminal tRNA-like structure is preceded by a long stretch of adenines on each RNA segment (Adams MJ, 2009).

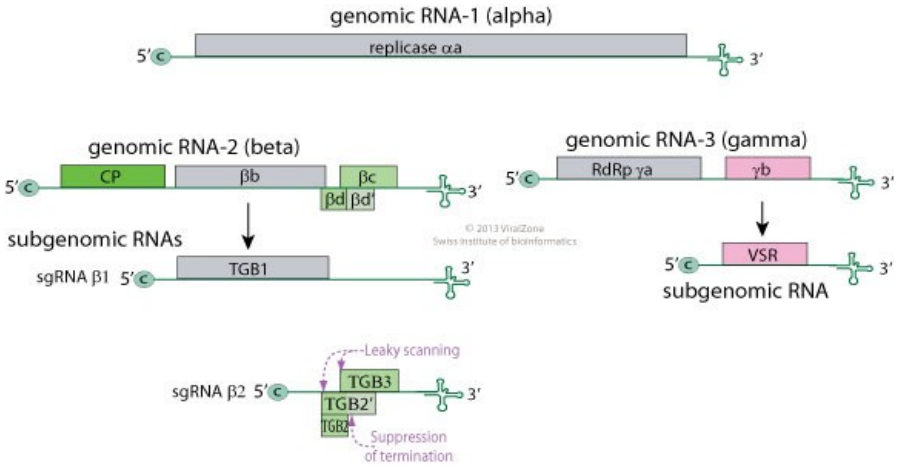


Fig.9 Schematic representation of *Hordeivurus* genome. Segmented, tri- (or possibly quadri-) partite linear ssRNA(+) genome composed of RNA-alpha=3.8 kb, RNA-beta=3.3 kb and RNA-gamma=2.9-3.3. The 3' terminus of each RNA contains a tRNA-like structure. The 5' terminus is capped. There are a total of 7 open reading frames. (Picture taken from www.viralzone.expasy.org)

The group V of negative-sense ssRNA viruses includes three families: *Rhabdoviridae*, *Bunyaviridae* and *Ophiviridae*. They all need to synthesize a positive-sense RNA for translation and replication with an RNA-dependent RNA polymerase they encode.

The family *Rhabdoviridae* is characterized by bacilliform enveloped virions with a monopartite ssRNA(-) genome carrying five ORFs that code for nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and large protein (L/RdRP), flanked by a 5' UTR sequence and a 3' regulatory region (Dietzgen, 2016).

The family *Bunyaviridae* can be represented by *Fig mosaic virus* (FMV), genus Emaravirus, which symptoms were first described in California by Condit and Horne in 1933 but the virus was finally identified only in 2009. The predicted genome was thought to be composed by four negative-sense ssRNA segments (Elbeaino et al., 2009, Walia et al. 2009), but more recently two novel RNA molecules have been identified (Elbeaino et al, 2012). In the light of these latest findings, the genome of FMV has six segments that code for RdRP of 267.8 kDa translated from RNA-1), a glycoprotein precursor translated from RNA-2, which after cleavage gives rise a so called Gn (21.5 kDa) and Gc (51.7 kDa) proteins, a nucleocapsid protein of 35 kDa translated from RNA-3, and finally, p4 of 40,5 kDa from RNA-4, which is involved in cell-to-cell movement (Ishikawa et al, 2014), p5 of 59 kDa (unknown function) from RNA-5 and p6 of 21.5 kDa (unknown function) from RNA-6 (Fig. 10).

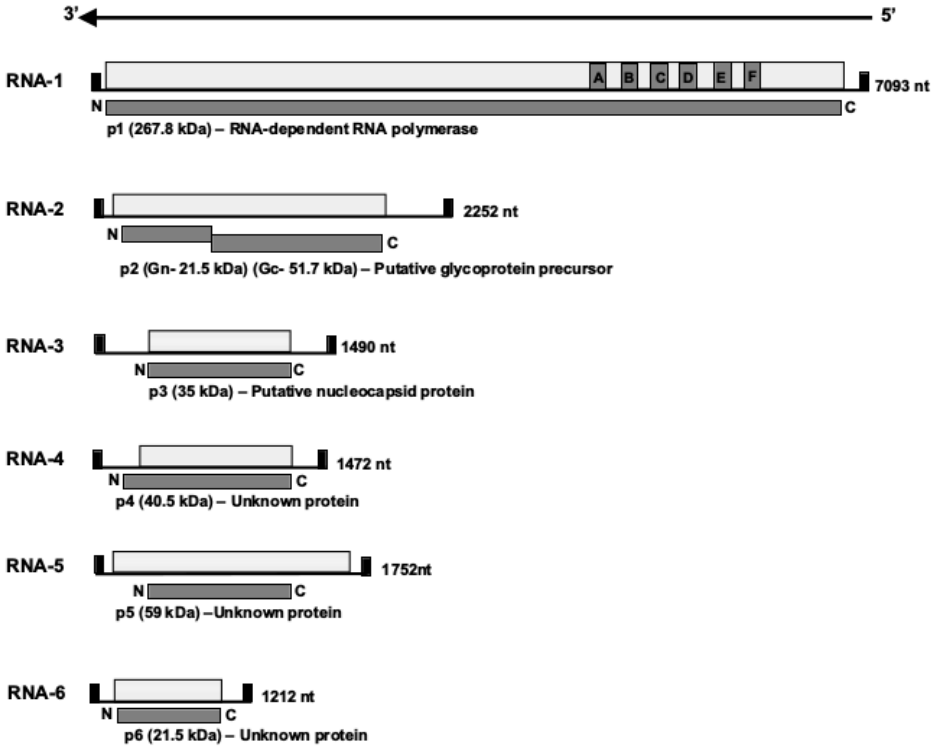


Fig.10. Schematic representation of the organization of six RNA segments constituting the putatively complete genome of FMV. The terminal 13 nucleotides conserved at the 5' and 3' termini are indicated as black boxes on each segment. Letters (A-F) represents the conserved motifs of the RdRp (RNA-1) gene. Expression products of each RNA (p1 to p6) are represented as dark grey boxes. The function and estimated molecular weight of each protein are reported. Figure is not drawn to scale. Copied from Elbeaino et al. 2012.

The family *Ophioviridae*, with the only genus *Ophiovirus*, is characterized by filamentous and flexuous virions covered with nucleocapsid forming open circles. The genome has three to four segments coding for up to seven proteins.

Retroelements that form isomeric virus-like particles are classified into the VI group of ssRNA-RT viruses, composed of the families *Pseudoviridae* and *Metaviridae*. They encode an RNA-dependent DNA polymerase or reverse transcriptase (RT) needed for DNA synthesis and often an integrase and capsid proteins.

1.4 Viroids

Viroids are circular RNA molecules from 246 to 401 nucleotides that not coding for any proteins but are associated with many pathogenic situations. According to the replication mechanism and compartment, viroids are divided in families *Pospiviroidae* and *Avsunviroidae*. Members of the former one, such as *Grapevine yellow speckle viroid* 1 and 2 (GYSVd-1, GYSVd-2), *Hop stant viroid* (HSVd) and *Australian grapevine viroid* (AGV) are copied in the nucleus by DNA-dependent RNA polymerase II via a rolling circle replication, the resulting concatamers are then cleaved by the host RNase III into monomers that assume a rod-like secondary structure thanks to a "central conserved region" CCR (Steger, 2016).

The *Avsunviroidae* don't have a CCR region, replicate in the chloroplast and autocleave themselves via a hammerhead ribozyme.

1.5 RNA silencing

RNA silencing is a conserved mechanism among most eukaryotes, involved in gene regulations at transcriptional and post-transcriptional level and defense against invasive nucleic acids like transposons and viruses (Fukudome and Fukuhara 2016, Borges and Martienssen, 2016). The mechanism is based on the production of small RNA molecules of 21-24 nucleotides in length that can be classified into microRNA (miRNAs) and small interfering RNAs (siRNAs) according to their origin, biogenesis precursor, structure and mode of action (Fig. 11).

miRNAs derive from endogenous *MIR* genes originating from inverted duplication, spontaneous evolution or miniature inverted-repeat transposable elements MITE. They are transcribed by the RNA polymerase II (Pol II) in single-stranded capped and polyadenylated primary miRNAs arranged into hairpin-like structures (Borges and Martienssen, 2015).

siRNAs derive from long double-stranded RNA precursor either with endogenous (secondary structure of mRNA, folded inverted-repeat sequences, sense/antisense transcription of endogenes, transposons and transgenes, and products of RNA-dependent RNA polymerase activity) or exogenous (pathogens like viruses and viroids) origin (Aregger et al. 2012, Boualem, 2016). The most characterized are trans-acting siRNAs (ta-siRNAs), repeat-associated siRNAs (ra-siRNAs), natural-inverted siRNAs (na-siRNAs) and heterochromatic siRNAs (hc-siRNAs) that are categorized into secondary siRNAs, deriving from the plant RNA-dependent RNA polymerase (RDR) activity.

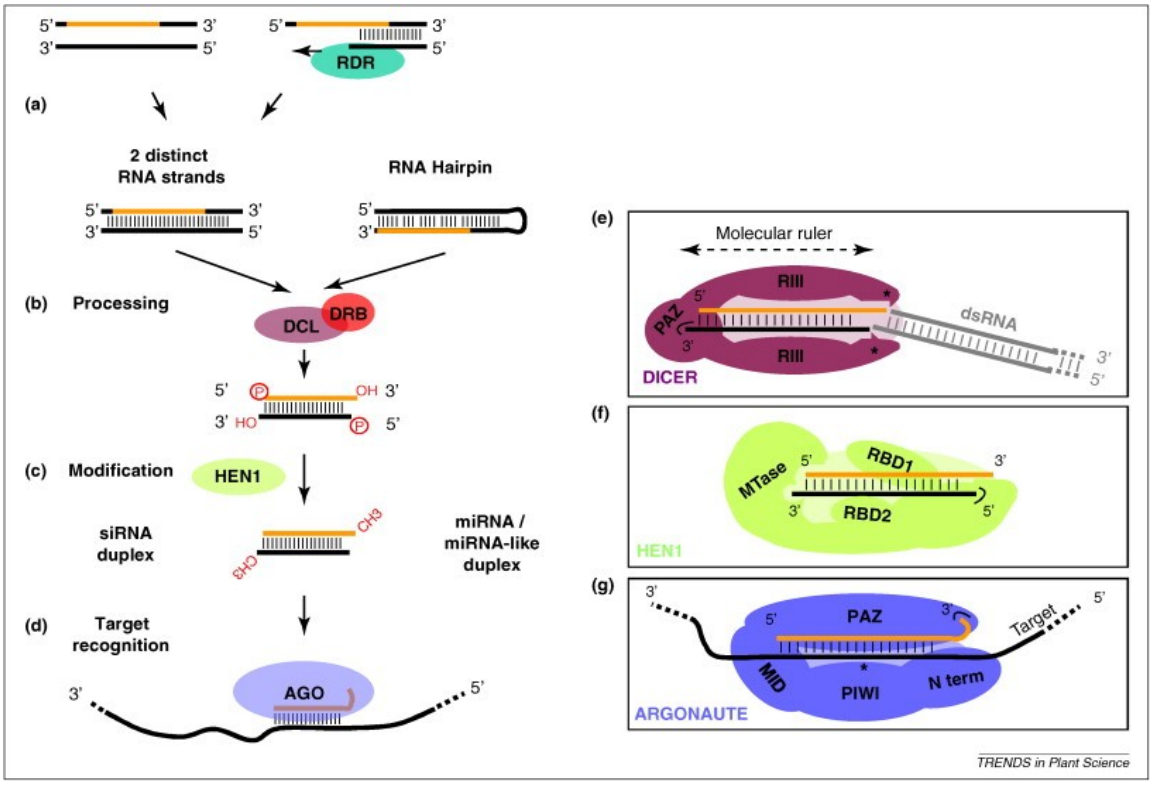


Fig.11. Biosynthetic pathway of small RNA biogenesis. a) dsRNA precursors derived from Pol II, Pol IV, RDR activity or from self-annealed transcript into hairpin structures. b) DCL and DRB produce sRNA duplexes with a 2-nt 3'-overhang, a 3' hydroxyl group and a 5' phosphate group; c) HEN1 methylates the 3' end to protect the dsRNA from degradation; d) The mature sRNA is loaded on RISC complex, the passenger strand discarded and the target recognized by anti-complementarity; e), f), g) DCL, HEN1 and AGO domains interacting with precursors, siRNAs and guide strand, respectively. Picture taken from Vazquez et al., 2010.

In *A. thaliana* there are six different RDR, among which RDR1, RDR2 and RDR6 are involved in the production of secondary siRNAs (Wassenegger and Krczal 2006; Cao et al. 2014).
Both siRNAs and miRNAs are processed from dsRNA or hairpin precursors by the same class of RNase III-like enzymes called Dicers, supported by dsRNA-binding proteins (DRBs). The Dicer family is characterized by the presence of an ATPase/DEAD/H-box domain at the N-terminus necessary to recruit regulatory protein and hydrolyze ATP for the catalytic reaction, a Piwi/Argonaute/Zwill (PAZ) domain which is able to recognize the 2 nucleotides 3' overhang of the dsRNA and it is connected to the two catalytic domains throughout an α-helix structure which flexibility determines the sRNA size product. At the C-terminus there is a RNA-binding domain also used for protein-protein interaction and nuclear localization (Fig. 11, Fukudome and Fukuhara, 2016).
In plants, four different Dicer-like (DCL) proteins are involved in the production of 21 nt (DCL1 and DCL4), 22 nt (DCL2) and 24 (DCL3) sRNAs (Fig. 12, Blevins et al. 2006). DCL1 is the only one that produces miRNAs of 21-22 nts in Arabidopsis thaliana and for this reason its loss of function in mutants is lethal (Schauer et al. 2012). DCL4 is involved in the production of 21-nt viral siRNAs and RDR6-dependent endogenous tasiRNAs and phased siRNAs (Bouche et al. 2006). Like DCL4, DCL2 is cytoplasmatic, involved in secondary siRNA biogenesis and known to be also involved in the synthesis of viral 22-nt siRNAs, specially when DLC4 is down regulated by viral suppressors (Bouche et al. 2006, Pooggin 2016). DCL3 produces 24-nt siRNAs deriving from transposons and repetitive elements and it's involved in the epigenetic modification and chromatin remodeling (Blevins et al. 2006). In rice and other monocots, two DCL3 paralogs mediate biogenesis of 24-nt phased siRNAs and 24 nt miRNAs (Liu et al. 2014).
After DCL processing, the sRNA duplexes are 2'-O-methylated at the 3'-terminal nucleotide of each strand by HUA HENHANCER 1 (HEN1), to increase their stability and prevent degradation by 3'-uridylation via SMALL RNA DEGRADING NUCLEASEs (SDN, Borges and Martienssen, 2016). HEN1 has two RNA-binding domains through which it interacts with the sRNA duplex while the methyltransferase domain (MTase) performs 2'-O-methylation (Fig.11, Vazquez et al., 2010). Once methylated, one of the two strands of the sRNA duplex (the guide strand) is associated with an AGO family protein, to form the RNA-induced silencing complexes (RISC) that contains a single stranded sRNA guiding RISC to complementary target sequences.
In plants multiple AGO genes have been found: ten in *Arabidopsis thaliana*, eleven in *Solanum tuberosum* and *Prunus persica*, fifteen in *Solanum lycopersicum* and *Vitis vinifera* (Mirzaei et al. 2014). They are all characterized by four domains: the N-terminal, PAZ, MID and PIWI (Fig.11). Despite the attempts in vitro and in vivo to better understand the N-terminal domain, its function is still unclear (Wang et al. 2009). The PAZ domain has an OB (oligonucleotide/oligosaccharide binding) fold able to bind and anchor the 3' end of the guide strand that will be incorporated in the RISC complex (Fang and Qi, 2016). The MID domain specifically recognizes the first nucleotide at the 5' end according to the 5'-rule with AGO1 binding 21-22 nt 5'U sRNAs, AGO2 21-22 nt 5'A sRNAs, AGO4, AGO6 and AGO9 24-nt 5'A sRNAs, AGO5 21-22 nt 5'C sRNAs and chooses the strand with lower thermodynamic stability at the 5'-end of the guide strand, while the passenger will be destroyed (Tolia and Tor, 2006, Fang and Qi 2016, Pooggin 2016). The PIWI domain has an RNAse H-like fold which enables

some but not all the AGO proteins to cleave the target RNA. The mechanism of action of the assembled RISC could be at the DNA level via transcriptional gene silencing (TGS), with includes RNA-directed DNA methylation (RdDM) and chromatin modification, or at the RNA level via post-transcriptional gene silencing (PTGS). The ten different AGO proteins from *Arabidopsis thaliana* are organized in three clades, according to their mode of action: AGO1/AGO5/AGO10 and AGO2/AGO3/AGO7 are mainly cytoplasmatic and involved in PTGS through miRNAs and ta-siRNAs, whereas AGO4/AGO6/AGO8/AGO9 are nuclear and involved in RdDM and TGS (Pooggin, 2016). One method of the PTGS is the sequence-specific degradation of the target RNA (also called slicing), performed by the PIWI domain with endonucleolytic activity present in AGO1, AGO2, AGO4, AGO7 and AGO10. The slicing is guided by the degree of complementarity between the guide strand and the target RNA, which in plant is really high compared to the animal miRNAs counterparts (Fang and Qi, 2016, Machado et al., 2017). The cleaved target products are further converted in secondary siRNAs to amplify the silencing effect.
Recently it has been proved that RDR6 is able to discriminate between an aberrant transcript and a regular one by sensing the presence and the length of the poly-A tail. This finding, together with the localization in siRNAs bodies (cytoplasmatic loci where the siRNAs biogenesis occurs), RNA-binding proteins and specific internal structure protecting the RNA from RDR6 activity, assure that PTGS is strictly regulated to avoid silencing of transcript important for the plant fitness (Baeg et al. 2017).
Transcriptional silencing involves the de novo DNA methylation or the maintenance of RdDM at specific loci, like the promoters of retrotransposons. In the first case, the secondary siRNAs produced by Pol II transcripts, RDR6/SGS3, DCL4, DCL2 and DCL3, together with AGO6 or AGO4, interact with the methyltransferase DRM2 to establish the DNA methylation. In the second case, there is a switch of the involved proteins, from Pol II to Pol IV and Pol V, RDR2, DCL3 and 24-nt siRNAs loaded onto the nuclear AGO4/AGO6/AGO9 to reinforce DNA methylation (Borges and Martienssen, 2016).

1.6 RNA silencing against viruses

Two main lines of evidence suggest the involvement of RNA silencing in the plant antiviral defense mechanism: the accumulation of virus-derived siRNAs in the infected plant and the fact that plant viruses encode silencing suppressors. The precursors of viral siRNAs derive either from dsRNA precursors formed during virus replication or transcription, or from single-stranded RNA folded in an hairpin-like structure, both the suitable substrate for DCL activity. In the case of ssRNA viruses, Molnar et al. (2005) showed that the production of viral siRNAs has a strong bias towards the sense RNA strand and that the intermediate would be sequestered (i.e. protected) by the replication machinery and not available for the silencing machinery, suggesting that the precursor of viral siRNAs is mainly coming from secondary-structured RNA (Zhu and Guo, 2012). By contrast, Donaire et al., (2009), Garcia-Ruiz et al., (2010) and Seguin et al., (2014a) proved that there is an equal amount of viral siRNAs coming from both sense and antisense strand, that the hotspots along the genome do not necessarily correspond to secondary-structured region, tipping the balance towards the dsRNA categories as real precursors (Pooggin, 2016).

Plant infected with single-stranded RNA viruses accumulate 21nt siRNAs processed by DCL4, which can be replaced by DCL2 for the production of 22nt siRNAs when DLC4 is knock-out or down regulated by viral suppressors, like P38 of turnip crinkle virus (TCV, Deleris et al., 2006). As well as for the ta-siRNAs, DCL4 substrates come from RDR6 activity, which can be replaced by RDR1 or RDR2 due to their redundancy, proved by the fact that the rdr1/rdr2/rdr6 triple mutants have reduced level of viral siRNAs (Donaire et al., 2009, Gracia-Ruiz et al., 2010). DNA viruses do not replicate via dsRNA, thus they do not code for RDR proteins. The circular single-stranded geminivirus genome is transcribed bidirectionally by Pol II and the transcripts can form a perfect dsRNA substrate for all the DCL proteins (Pooggin, 2016). In pararetrovirus the transcription is monodirectional and therefore the dsRNA precursors may not be involved in the production of viral siRNAs. However, the CaMV pgRNA promoter can accidentally starts the transcription in the antisense direction, leading to a complementary strand suitable for dsRNA formation (Blevins et al., 2006). Furthermore, bioinformatic analysis of viral siRNAs deriving from DNA viruses revealed that they are produced along the whole genome and accumulate at high level in both orientation (Seguin et al., 2014a, Rajeswaran et al., 2014a, 2014b).

The major size classes that accumulate in DNA virus infected plants are 21-nt, 22-nt and 24-nt viral siRNAs (Fig.12). The accumulation level can be different among different host plants, as in the case of cassava mosaic virus infecting *Nicotiana benthamiana*, where 22-nt siRNAs are more abundant compared to the 24-nt and 21-nt siRNAs more abundant in cassava (Akbergenov et al. 2006). Once processed by DCLs, viral siRNAs are also methylated by HEN1, if not inhibited by viral suppressors like in the case of RNA potyvirus, tobamovirus and tombusviruses. In this case the stability of the duplex is probably confered by interaction with the suppressor itself.

Viral siRNAs can associate with multiple AGOs, following the 5' nucleotide rule. In particular, it is known that RNA viruses-derived siRNAs associate with AGO1 (5'U), AGO2 (5'A) or AGO5 (5'C) but also with AGO7. There is no specific evidence for the association of DNA virus-derived siRNAs with specific AGOs, since specific *ago* mutants do not exhibit increased susceptibility to DNA viruses (Pooggin 2016). Immuno-precipitation experiments with AGO1 and AGO4 antibodies revealed that 21-nt viral siRNAs are associated with AGO1, while more abundant 24-nt viral siRNA do not appear to be associated with AGO4 in Arabidopsis plants infected with a DNA pararetrovirus CaMV (Blevins et al. 2011).

Primary siRNAs target complementary viral transcript for cleavage and the resulting products would enter the RDR6/SGS3/DCL4 pathway for secondary siRNAs production, necessary for restriction of the virus to the infected tissue and suppression of the systemic infection (Fig.12). In fact, in rdr6/sgs3/dcl4 mutants, suppressor deficient viruses can spread systemically (Deleris et al. 2006). They can also regulate host gene expression by a mechanism called Virus-Induced Gene Silencing (VIGS), that can be further classified in PTGS-VIGS when the host transcript is homologous to the virus or TGS-VIGS when the host promoter is homologous to the virus (Pooggin 2016). Thus, VIGS could lead to inhibition of chlorophyll biosynthesis which is associated to yellowing symptoms in CMV-infected plants, downregulation of host-silencing machinery (i.e. DCLs transcripts). Furthermore, viruses developed strategies to escape the plant defenses. By replicating via rolling circle in the nucleus, DNA viruses escape the cytoplasmatic PTGS but also the possible RdDM and TGS, further supported by protein suppressors like AC2

or AC4 in geminiviruses, respectively (Pooggin 2016).
From the other side, virus-encoded suppressors and other viral proteins can be recognized plant immune receptors that mediate ETI and hypersensitive response.
Those defense, counter defense and counter-counter defense, perfectly described by the zig-zag-zig model (Jones and Dangl 2006, Zvereva and Pooggin 2012), resembling an ongoing arm wrestling hosted in a neutral field owned by evolution.

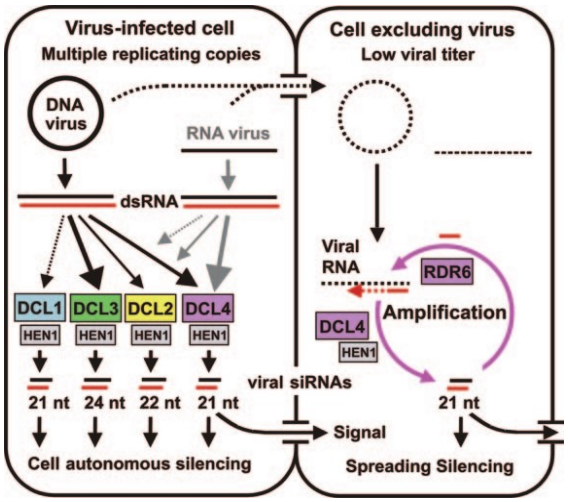


Fig. 12. dsRNAs deriving from both DNA or RNA viruses are processed by DCLs in different size classes. The 21 nt DCL4 product is potentially the signal required for VIGS spread. Both infectious nucleic acids and viral siRNAs move into the right-hand cell. However, the viral titer remains low, because DCL4 and RDR6 amplify incoming siRNA signal and digest viral transcripts. In this manner, VIGS would spread into meristematic tissues from which viruses are normally excluded. Viral siRNAs are stabilized by HEN1-mediated methylation. Picture taken from Blevins et al. 2006.

1.7 Cross-protection

The term cross-protection refers to the phenomenon whereby tolerance or resistance to a virulent pathogen is induced by pre-infection with a second, related but less virulent pathogen, mimicking the mechanism of human vaccination. Even though there is no adaptive immunity in plants, neither circulating cells nor antibodies, cross-protection is well-known to occur in plants, as first described by McKinney in 1929 in tobacco plants infected with a “light green strain” of TMV that recovered from the yellow symptoms after inoculation with a TMV “yellow mosaic strain”. Since then, cross-protection has been applied against *Citrus tristeza virus* (Costa and Muller, 1980), *Tobacco mosaic virus* (Burgyán and Gáborjányi, 1984), *Papaya ringspot virus* (Yeh et al., 1988), *Cucumber mosaic virus (*Kosaka and Fukunishi, 1997, Ziebell et al. 2007) and many more viruses (Ziebell and Carr, 2010). However, it has been proven that also a preinoculation with a severe strain can protect the plants against a following infection with a less virulent strain (Aapola and Rochow, 1971; Bodaghi et al., 2004; Ziebell and Carr, 2009, Ziebell and Carr, 2010).
When choosing the protective strain, some requirements have to be taken into account. First of all, there should be a high percentage of nucleotide sequence similarity between the two strains. The protective strain should induce mild symptoms and should not affect drastically the plant fitness. It should also be easy to propagate, to inoculate and should be able to infect the plant systemically and not only locally (Ziebell and Carr, 2010).
The involvement of viral coat protein in cross-protection has been investigated, with incongruous results. First, it was thought that coat protein of the protective strain would interfere with the replication of the challenging strain, but Zaitlin in 1976 proved that cross-protection was occurring also in CP-defective mutants. On the contrary, in other experiments,

when RNA from an TMV strain was coated with BMV CP or not coated at all, no cross-protection was observed suggesting that CP is required (Sherwood and Fulton, 1982). In 1984 it was suggested by Palukaitis and Zaitlin that RNA homology and accumulation of positive-sense RNA is important for cross-protection against a negative-sense related strain, suppressing transcription and replication and paving the way to RNA silencing (Ziebell and Carr 2010).

The first evidence for an RNA silencing-based mechanism underlying viral coat protein-mediated cross-protection was obtained in transgenic plants expressing untranslatable sense or antisense forms of the tobacco etch virus (TEV) coat protein gene sequence (Lindbo et al 1992; Lindbo et al. 1993; Dougherty et al. 1994). The TEV CP-mediated antiviral resistance of transgenic plants was found to be post-transcriptional and highly sequence-specific (no broad-spectrum virus resistance was observed). Some of the transgenic lines displayed viral symptoms but eventually recovered from viral disease. The highly resistant and recovery phenotypes associated with TEV challenge inoculation and the reduction of steady-state RNA levels in recovered transgenic leaf tissue were proposed to be manifestations of a common mechanism (Dougherty et al. 1994; reviewed in Lindbo and Dougherty 2005). This mechanism was initially called RNA-mediated virus resistance or posttransriptional gene silencing (PTGS). Now it is broadly recognized as RNA silencing or RNA interference (RNAi) which also includes RNA-mediated TGS (reviewed in Baulcombe 2004). As described above the plant RNA silencing machinery generates siRNAs that act in a sequence-specific manner to repress endogenous gene expression and invasive nucleic acids such as transposons and viruses both *in cis* and *in trans*.

In the light of the above, in cross-protection the protective strain could induce the production of siRNAs that systemically move to the adjacent tissues leading to an overall protection of the infected plant from an incoming aggressive strain sharing sufficient identity in genome sequence.

Recently, researchers have been focused on PepMV, the most prevalent virus in tomato production. Even if sources of resistance have been identified in wild tomatoes (*Solanum peruvianum*, Soler-Aleixandre et al. 2007), no resistant cultivars against PepMV are commercially available (Jaroszewska, 2014). For this reason, in December 2011 Belgian and Dutch producers got a temporary authorization (renewed every year since then) to use an attenuated strain of the virus, registered and distributed as PMV-01 (Hansenn et al., 2010).

In Switzerland, growing interest for this attenuated strain coupled with a PepMV disease outbreak in Ticino in 2012, which resulted in an official demand from the association of tomato growers to the office of agriculture to get the right to also use the attenuated strain under a temporary authorization because of exceptional circumstances.

In parallel, the Swiss Federal Institute Agroscope initiated a survey to evaluate the situation (Schumpp et al., 2014). Production sites were monitored all over the year: about half of the monitored sites developed the disease with one or two strains at some point of the year which demonstrated the establishment of the virus in tomato crop production in Switzerland. Phylogenetic analysis grouped some of these Swiss strains within the CH2 clade while others were closely related to the EU strain. Agroscope started to investigate the interactions between PMV-01 that belongs to the CH2 clade and the strain EU-Swi isolated in Switzerland in 2013 and our study is part of this investigation, further described in the Results chapter.

1.8 NGS as viral diagnostic tool

For decades, common methods in viral diagnostics have been based on amino acid or nucleotide sequence-specific molecular and biochemical techniques such as ELISA, Lateral Flow Devices, PCRs, microarrays or more non-specific methods like inspection of visual symptoms (Adams and Fox, 2016). However, the sequence specific methods have a disadvantage, because they all require a previous knowledge of the virus to be investigated, leading to false negative results for those viruses which have differences in the target nucleotide or protein sequences and don't allow identification of unknown, divergent viruses (Adams and Fox, 2016). The methods relying on visual monitoring of the disease symptoms can be useful for a first screening but require an excellent knowledge about viral symptoms. Furthermore, it should be taken into account that different viruses can induce similar symptoms and that there are also asymptomatic viruses.

Since 2009, with the advent of the Next Generation Sequencing (NGS) combined with bioinformatic analysis and its application in diagnostics, virus discovery and evolution studies, a large number of studies have been published, increasing our knowledge in plant virology (Wu et al., 2015). To increase the possibility to detect viral genome, several enrichment strategies have been used, like rRNA depletion, sequencing of polyadenylated RNAs, dsRNA enrichment, viral particles purifications and deep sequencing of small RNAs. However, sequencing of polyadenylated RNA is not useful to detect viroids and those viruses that don't have any poly(A)-tail, while dsRNA enrichment could prevent detection of DNA viruses that don't produce enough dsRNA during replication, and viral particles purification requires specific protocols that could exclude different or new viruses in a mixed infection (Wu et al., 2015). Deep sequencing and de novo assembly of small RNAs is gaining more power among the other strategies. This is due to the universal applicability of the method, since viral siRNAs accumulate at high level in infected cells and cover the whole viral genome of DNA and RNA viruses and viroids, in both sense and antisense direction (Seguin et al. 2014a), as further described in the next section.

One of the first platforms for NGS, the Roche 454 machine based on emulsion PCR and pyrosequencing was released in 2005 (Margulies et al. 2005), and eventually dismissed due to the high cost and low output. The latest version 454 GS-FLX+ Titanium is able to produce 600 M reads up to 1 Kb long (Barba et al. 2014).

In 2005, Solexa released the Genome Analyzer based on sequencing by synthesis, where a fluorescently labeled terminator is detected as each dNTP is added, and then cleaved to allow incorporation of the next base. Solexa was later acquired by Illumina and the first platforms were replaced by the latest HiSeq platform series like HiSeq 2500 (Barba et al. 2014). Nowadays the Illumina Genome Analyser (HiSeq2500 and its follow-up versions such as HiSeq4000) is the most widely platform for sRNA sequencing.

Another NGS approach is used by the Life Technologies SOLiD system, called sequencing by ligation, where a probe bound to a fluorophore hybridizes to a DNA fragment and is ligated to an adjacent oligonucleotide for detection. According to the emission spectrum of the labeled probe and its anti-complementarity with the base, the sequence can be revealed (Goodwin et al., 2016).

Recently, the Pacific Biosciences produced a platform for single molecule real time sequencing

(SMRT) in which the polymerase is attached to the bottom of a well and a single DNA molecule pass through a zero-mode waveguide (ZMW), a structure small enough to observe a single nucleotide being incorporated and the released fluorescence. This technology allows for sequencing long DNA molecules and it is used in combination with Illumina to sequence large genomes containing highly repetitive sequence regions.
One of the latest technologies is the Oxford Nanopore MinION, a portable device with 512 nanopores where the DNA passes through and changes the electronic current according to the nucleotide (Lu et al., 2016). This technology also allows for sequencing very long DNA molecules, but it has a very high error-rate.

1.9 Deep sequencing of small RNAs

Following the pioneering work of Kreuze et al. (2009), small RNA deep sequencing has been applied for viral diagnostic for plant and animal (including human) viruses (review by Wu et al. 2015). The fact that small RNAs accumulate at high level in all the infected cells and no prior knowledge of the pathogen is required are two of the main reasons why this technique has been preferred, even compared to a normal total RNA-seq, which can be limited to some RNA species. Seguin et al. (2014b) has developed the siRNA-omics (siRomics) pipeline for de novo reconstruction of viral genomes which was applied for crop plants (Seguin et al. 2014b, 2016, Rajeswaran et al. 2014a, 2014b, Fuentes et al. 2016) and further used in this PhD study. The pipeline, which is described in Material and Method chapter, relies on the accumulation of small RNA of different size classes deriving from any viral species present in the infected plants (DNA and RNA viruses, subgenomic RNA, satellites, viroids) and makes use of several bioinformatics tools for the *de novo* reconstruction of the consensus virus genome. Furthermore, the information acquired with this analysis can be useful for further understanding the mechanisms of RNA silencing-based antiviral defense, which could contribute to basic research in the field of plant-pathogen interaction and for developing novel strategies of virus control.

2. Material and methods

In this study we applied and further improved the small RNA-omics (sRNA-omics) approach developed earlier in Pooggin lab for virome reconstruction and characterization of sRNA-based antiviral defense (Seguin et al., 2014a, 2014b, 2016).
The sRNA-omics approach consists of total RNA extraction and validation (by blot hybridization), deep sequencing of the sRNA population using Illumina technology, and bioinformatic reconstruction of the consensus viral genome sequence and its quasispecies variants (Fig.13).

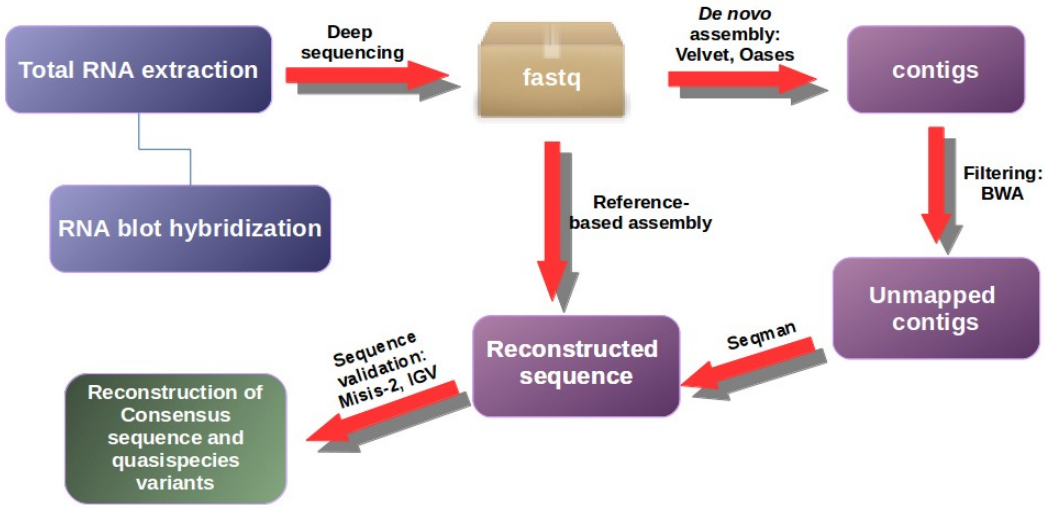


Fig. 13: Schematic representation of sRNA-omics pipeline.

2.1 Plant materials

All the plant materials analyzed in this study were collected in Switzerland and listed in Table 1. In 2014, potato (*Solanum tuberosum*) tubers were purchased in local shops in Basel and planted in the greenhouse of the Institute of Botany (Basel, Switzerland). After three weeks post plantation, five to seven leaves were harvested from the plants exhibiting virus-like disease symptoms and control healthy looking plants and used for total RNA extraction (Fig. 14). Tomato plants (*Solanum lycopersicum*) were analyzed in the frame of a cross-protection trial at Agroscope, Switzerland. In December 2013, plants of the variety Mont Favet were mechanically inoculated with an isolate of PepMV strain LP in an Agroscope greenhouse in Changins (Nyon, Switzerland, isolate originating from a tomato Merlice plant harvested in Champagny, Fribourg, Switzerland), while plants of the variety Merlice were treated with a protecting isolate of PepMV-CH2 in another Agroscope greenhouse (Conthey, Switzerland). In March 2014 the plants infected with the LP strain were transferred from Changins to Conthey and introduced in the same greenhouse where the Merlice plants were growing. The co-cultivation procedure involved deliberate mechanical cross-contamination between all the

greenhouse plants using leaf cutting and fruit harvesting tools, thus allowing repetitive inoculations of CH2-infected plants by sap from LP-infected plants, and vice versa. Upper (systemic) leaves from PMV-CH2 treated plants and from LP inoculated plants were collected from the same sample every month from April till September. The leaf samples were frozen in liquid nitrogen and stored. Pictures of frozen leaves were taken before extraction of total RNA (Figure 15).

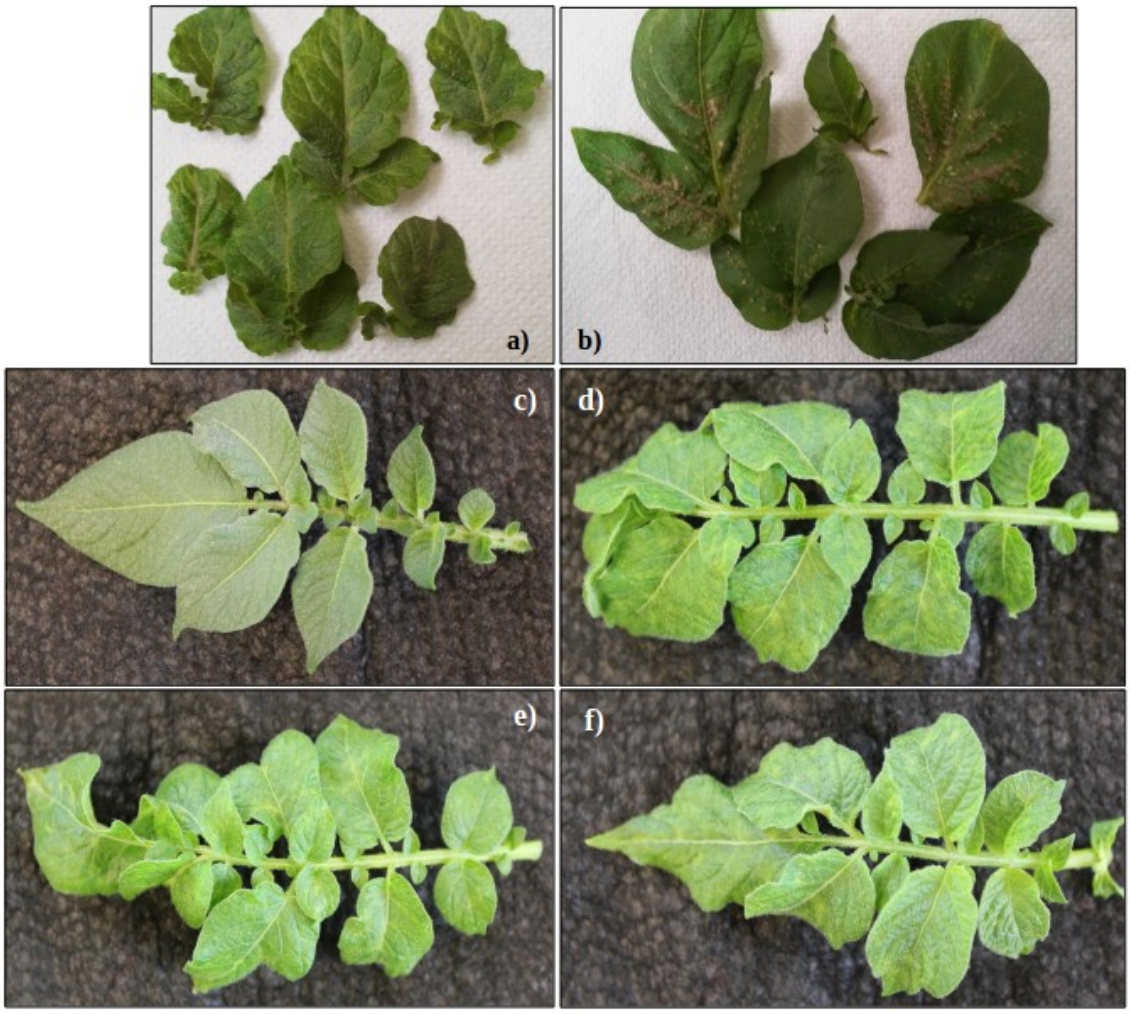


Fig.14. Potato leaves used for RNA extraction. a) Man3, data set BPO-147; b) Den2, data set BPO-148; c) Vic1, data set BPO-149; d) Vic7, data set BPO-150; e) Vic8, data set BPO-151; f) Vic9, data set BPO-152.

Leaf samples from *Vitis Vinifera* (grapevine), *Prunus avium* (cherry), *Larix decidua* (larch), *Ligustrum vulgare* (privet) and *Ficus carica* (fig) were kindly provided by Jean-Sébastien Reynard (Agroscope, Nyon).

Sweet cherry samples (Fig. 16) came from a field experiments in Agroscope (Nyon) where they were graft inoculated with a bud infected with one or two closteroviruses, *Little cherry virus 1 (*LChV-1) alone or in combination with *Little cherry virus 2 (*LChV-2) to evaluate their impact on tree physiology and fruit production and composition. Little cherry disease is a complex viral disease of cherry (Hadidi et al., 2011) but the disease etiology has not been totally clarified. Thus, to further support the phenotypic field trial with molecular evidence

and to better clarify the disease etiology, two different closterovirus-infected samples were collected in Autumn 2015, one asymptomatic (HYT-14) and one showing symptoms (HYT-15), to be analyzed through siRomics approach.

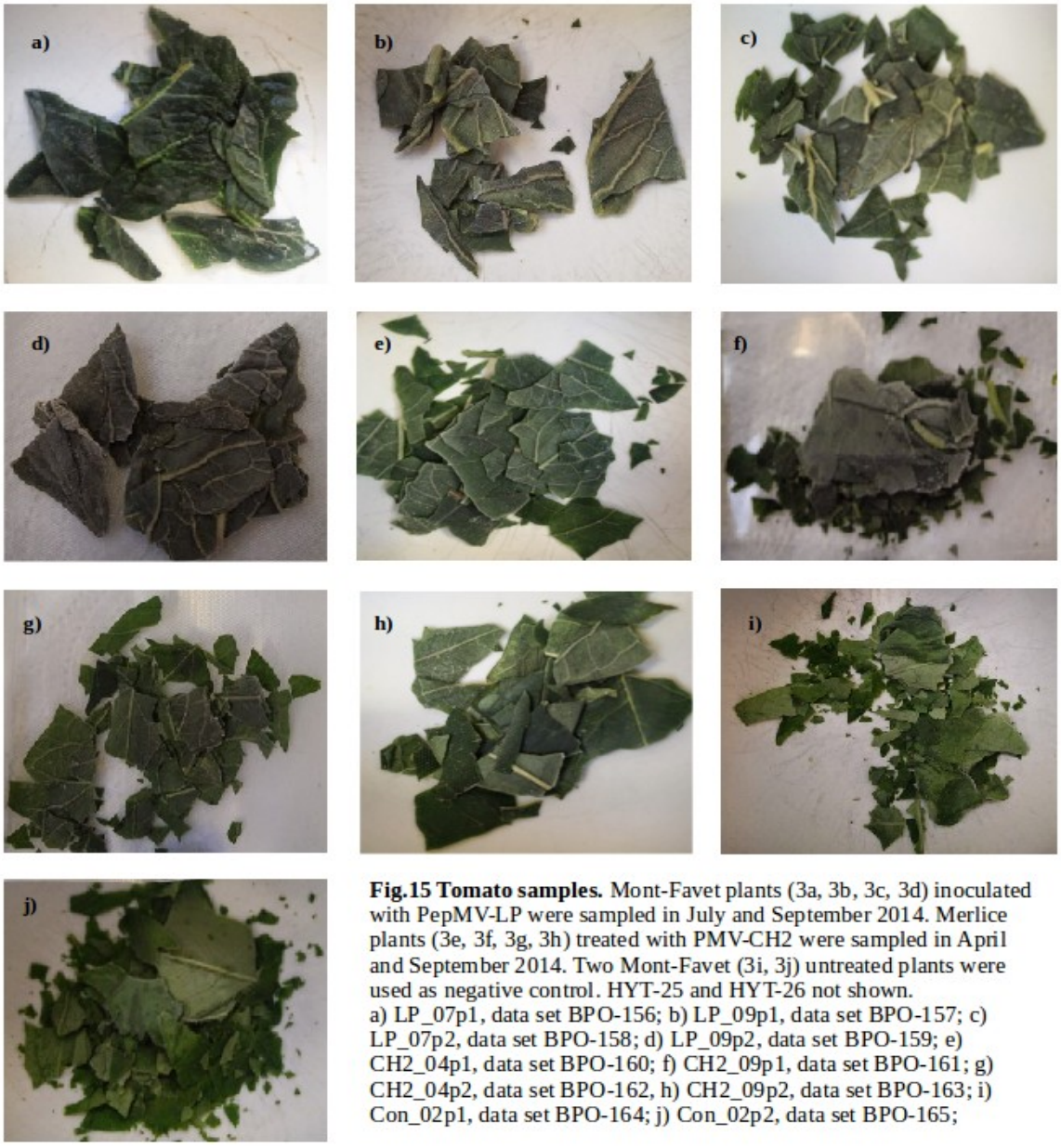


Fig.15 Tomato samples. Mont-Favet plants (3a, 3b, 3c, 3d) inoculated with PepMV-LP were sampled in July and September 2014. Merlice plants (3e, 3f, 3g, 3h) treated with PMV-CH2 were sampled in April and September 2014. Two Mont-Favet (3i, 3j) untreated plants were used as negative control. HYT-25 and HYT-26 not shown.
a) LP_07p1, data set BPO-156; b) LP_09p1, data set BPO-157; c) LP_07p2, data set BPO-158; d) LP_09p2, data set BPO-159; e) CH2_04p1, data set BPO-160; f) CH2_09p1, data set BPO-161; g) CH2_04p2, data set BPO-162, h) CH2_09p2, data set BPO-163; i) Con_02p1, data set BPO-164; j) Con_02p2, data set BPO-165;

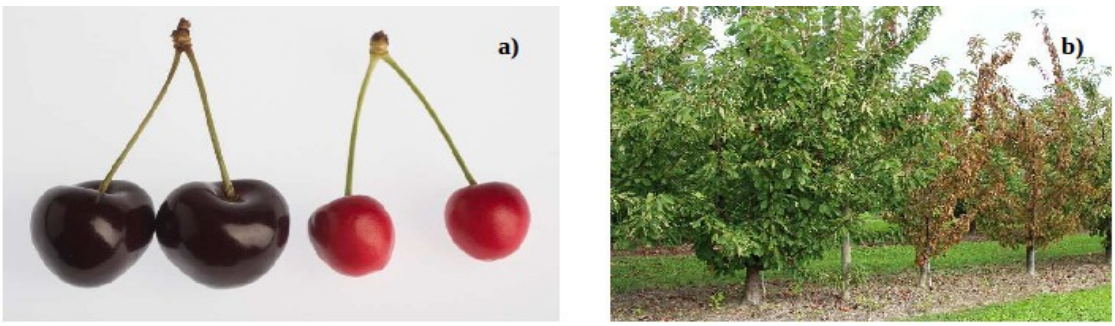


Fig.16. Cherry samples. a) Fruit affected by little cherry disease (right) compared to healthy one (left); b) Two diseased trees on the right, healthy tree on the left.

Leaf samples from grapevine varieties Pinot Noir and Otcha bala (Fig. 17) came from a field experiment in Agroscope (Nyon) and were collected in autumn 2015. The infected Pinot Noir resulted negative to the ELISA and PCR for all the nepoviruses found in Switzerland and no viral structures were identified at the electronic microscope. The infected Otcha bala plant (HYT-18) was previously analyzed by Reynard et al. (2015) by deep sequencing of nucleic acids purified from viral particles. Here we applied deep sequencing of sRNAs to compare these two techniques for virome reconstruction.

Fig.17. Grapevine samples. a) non-infected Pinot Noir; b) virus-infected Pinot Noir plant showing mosaic symptoms; c) virus-infected Otcha bala plant showing interveinal red coloration.

Needles from a larch tree exhibiting virus-like disease symptoms and from a control healthy tree were collected in the forest close to Nyon in Summer 2015. During a previous work performed by Dr. Paul Gugerli (unpublished) in larch needles showing leaf chlorosis (Fig 18B), a virus-like particle was purified and visualized at the EM (Fig 18A). Symptoms and particles resulted graft-transmissible but the virus was not genetically identified. A monoclonal antibody was developed to detect this virus, which gave positive ELISA results but the previous NGS approaches were not successful to identify this virus.

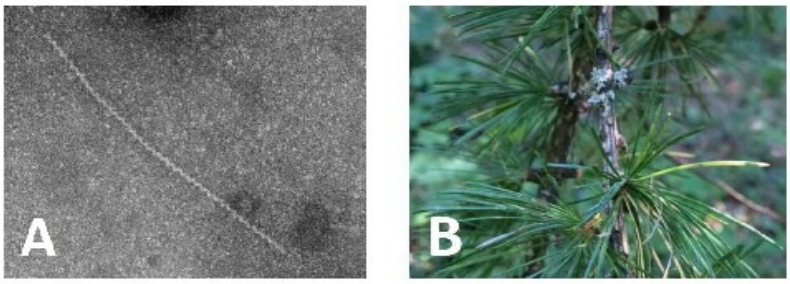


Fig.18. Larch sample. A) Virus-like particles visualized at the EM. B) Leaf chlorosis in infected tree.

Symptomatic leaves and asymptomatic leaves were collected from privet tree in Summer 2015 in the forest close to Nyon (Switzerland), on the edge of Lake Geneva (400 m.a.s.l.) up to 1400 m.a.s.l. in the alps (Valais region). In the privet sample showing yellow mosaic symptoms, virus particles similar to those of *Tobacco rattle virus* (TRV) were purified but, at the time of the first analysis, the virus was still unknown (Fig. 19).

Fig.19. Privet samples. Virus particles similar to those of Tobacco rattle virus (TRV) and typical of the genus tobravirus have been constantly associated with a yellow mosaic disease of privet.

In this study, two samples fig leaves were subjected to siRomics analysis, one symptomless coming from Agroscope collection in Changing (Switzerland, HYT-21) and the other one collected in a private garden in Sion (Switzerland) showing leaf mottling symptoms (HYT-22). Previous electron microscopy analysis at Agroscope of both samples revealed closterovirus-like particles and non-flexuous virus-like particles, but it cannot be excluded that they are endogenous fig structure (Fig.20).

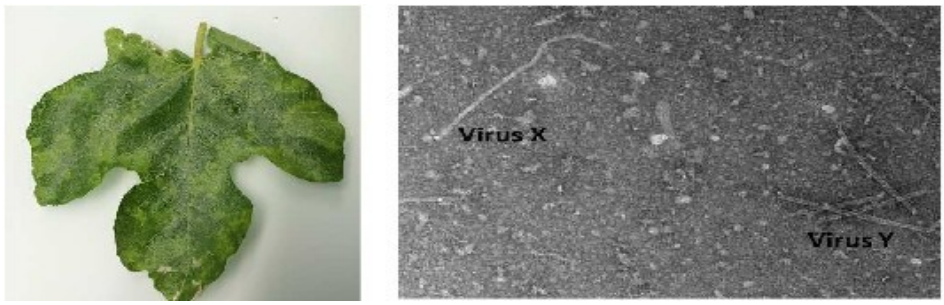


Fig. 20. Fig leaf showing mosaic symptoms on the left. Two types of viral particles have been found in diseased fig trees using EM.

Table 1. List of the samples analyzed in this study

Plant species	Dataset number	Sample name
Solanum tuberosum	BPO-147	Man3
	BPO-148	Den2
	BPO-149	Vic1
	BPO-150	Vic7
	BPO-151	Vic8
	BPO-152	Vic9
Solanum lycopersicum	BPO-156	LP_07p1
	BPO-157	LP_09p1
	BPO-158	LP_07p2
	BPO-159	LP_09p2
	BPO-160	CH2_04p1
	BPO-161	CH2_09p1
	BPO-162	CH2_04p2
	BPO-163	CH2_09p2
	BPO-164	Con_02p1
	BPO-165	Con_02p2
	HYT-25	LP-inf
	HYT-26	Con_08p3
Prunus avium	HYT-13	Cherry H
	HYT-14	Cherry 93
	HYT-15	Cherry 95
Vitis vinifera	HYT-16	Grapevine H
	HYT-17	Grapevine P1
	HYT-18	Grapevine P5
Larix decidua	HYT-19	Larix H
	HYT-20	Larix inf
Ficus carica	HYT-21	Ficus H
	HYT-22	Ficus inf
Ligustrum vulgare	HYT-23	Lig H
	HYT-24	Lig inf

2.2 siRomics pipeline

Total RNA preparation

Two different protocol were followed for total RNA extraction. For potato, tomato, fig and privet samples, TRI Reagent (Sigma) manufacturer's protocol was followed, with some modifications. For the other more recalcitrant samples (grapevine, cherry and larch), a combination of the protocols developed by Morante-Carriel et al. (2014) and Gambino et al. (2008) was followed, slightly modified.

One gram of potato, tomato, fig or privet leaf tissue samples was ground in liquid nitrogen, 5 ml of GHCL solution (6.5 M guanidine hydrochloride, 100 µM Tris HCl pH 8.0, 100 µM sodium acetate pH 5.5, 100 µM beta-mercaptoethanol) was added and the resulting suspension

was transferred into a sterile disposable polypropylene tube and mixed vigorously. After incubation at room temperature for 10 min, the suspension was centrifuged at 12000 rpm for 10 min at 4°C and the supernatant was then transferred into a new tube.
2.5 ml Trizol reagent (Sigma) and 1 ml chloroform were added to the supernatant, the mixture was vortexed thoroughly and the centrifugation was repeated at the same conditions. The aqueous phase was transferred to a new tube and an equal volume of isopropanol was added, followed by vortexing. The mixture was incubated on ice for 30 min and then centrifuged at 12000 rpm, 4°C for 20 min. The pellet was washed in 5 ml of 75% ethanol (pre-chilled on ice) and, for fig and privet, the precipitation was prolonged for 90 min at -20°C. The RNA was then pelleted at 12000 rpm, 4°C for 10 min, vacuum-dried for 10 min and dissolved in DEPC-treated water at 65°C for 10 min.
One gram of grapevine, cherry or larch leaf tissue samples was ground in liquid nitrogen to a fine powder and transferred into a sterile disposable polypropylene tube. Ten milliliter of extraction buffer (2% CTAB, 2.5% PVP-40, 2 M NaCl, 100 mM Tris-HCl pH 8.0, 25 mM EDTA pH 8.0, 2% β-mercaptoethanol) were added to the samples and incubated for 10 minutes at 65°C. The suspension was split into two separated tubes and 5 ml chloroform:isoamyl alcohol (24:1 v/v) were added to each tube, vortexed and centrifuged at 12000 rpm for 10 min at 4°C. The aqueous phase was transferred into a new tube and the chloroform:isoamyl alcohol purification was repeated at the same conditions. The supernatant was transferred into a new tube, 0.1 volume of 3M sodium acetate and 0.6 volume of isopropanol were added and the mixture was incubated at -20°C for one hour. The RNA was precipitated by centrifugation at 15000 rpm for 20 min at 4°C. The pellet was resuspended in 1 ml of DEPC-treated water and transferred into a microcentrifuge tube, 0.3 volume of lithium chloride (3M final concentration) were added and the mixture was incubated at -20°C overnight. The RNA was precipitated again by centrifugation at 20000 rpm for 30 min at 4°C. Following overnight incubation at -20°C, the centrifugation were repeated and the pellet was resuspended in 100 µl DEPC-treated water, 10 µl 3M sodium acetate and 200 µl ethanol (pre-chilled at -20°C). Centrifugation at 20000 rpm was performed for 20 min, at 4°C. The pellet was washed with 75% ethanol (pre-chilled at -20°C) and centrifuged at 4°C, 12000 rpm for 10 minutes. Finally, the pellet was vacuum-dried and resuspended in 30 µl of DEPC-treated water.
Total RNA concentration was measured using Nanodrop Spectrophotometer (Thermo Fisher Scientific) and Qubit RNA HS Assay Kit (Thermo Fisher Scientific).
Nanodrop was mainly used to evaluate the absorbance at 260 nm, the ration 260/230 nm and the ratio 260/280 nm which indicate the purity of the nucleic acids. A ratio for both of ~2.0 indicates a good RNA quality, if lower than 2, it may indicate the presence of contaminants that absorb at the same wavelength, like proteins, guanidine or phenol. However, Nanodrop is not able to distinguish between RNA or DNA and for this reason a further analysis with Qubit was performed. The Qubit kit uses a dye which specifically binds RNA molecules and its fluorescent emission is directly proportional to the RNA concentration in the sample.

Gel electrophoresis and blot hybridization

Formaldehyde and polyacrylamide gel electrophoresis were performed to check the quality and quantity of long and small RNA, respectively, followed by blot hybridization.

Five to ten micrograms of total RNA were vacuum-dried and resuspended in 5 µl DEPC-treated water and mixed with 5 µl RNA Gel Loading Dye (Thermo Fisher Scientific). The samples were incubated at 70°C for 10 minutes and loaded on formaldehyde-containing agarose gel [1.2% agarose, 3% formaldehyde, 1x MOPS buffer (0.02 M MOPS pH 7.0, 1 mM EDTA, 5 mM sodium acetate)], followed by electrophoresis for 2.5 hours at 100V. RNA was transferred by capillarity blotting to a Hybond N+ membrane (Amersham) for 24 hours in transfer buffer (50 mM NaH_2PO_4, 5 mM EDTA, pH 6.5) and cross-linked twice with 1200 µjoules x100 UV light in a Statalinker 1800 (Stratagene).

For the polyacrylamide gel, five to ten micrograms of total RNA were dried in a SpeedVac, resuspended in 10 µl of RNA Gel Loading Dye (Thermo Fisher Scientific) and warmed at 95°C for 3 min. The samples were then loaded on 15% polyacrylamide gel (19:1 acrylamide:bis-acrylamide and 8 M urea) and run at 3 V for 4 hours. RNA was then transferred to a Hybond N+ membrane (Amersham) by electroblotting in 1x TBE buffer at 10 V overnight at 4°C. The RNA was cross-linked to the membrane twice with 1200 µ joules x100 UV light in a Stratalinker 1800 (Stratagene).

Blot hybridization was performed as described by Akbergenov et al. (2006).

To verify the quantity and integrity of sRNAs in all the plant samples, DNA oligonucleotide probes specific for two evolutionarily-conserved miRNA species were used for the sRNA blot hybridization, miR160 and miR169. For the long RNA blot hybridization, strain-specific and species-specific DNA oligonucleotide probes for PVY twins, PepMV strains and LigMV segments were designed based on the reconstructed consensus viral sequences. All the DNA oligonucleotides were synthesized by Eurofins S.A. (Table 2). The DNA oligonucleotides were end-labeled with P^{32} gamma ATP (Hartmann Analytic, Germany) by T4 polynucleotide kinase (Roche, Switzerland) and purified through MicroSpin G-25 columns (Amersham). The hybridization was carried out overnight at 35°C in UltraHyb-oligo buffer (Ambion) followed by three washing with 2X SCC, 0.5% SDS for 30 min at 35°C. The membranes were exposed for 5 days to a phosphor screen and scanned in a GE Typhoon 8600 imager (GE Healthcare Life Sciences). For repeated hybridizations the membrane was stripped with 0.5X SSC, 0.5% SDS for 30 min at 80°C and then with 0.1X SSC, 0.5% SDS for 30 min at 80°C.

Ten micrograms of the total RNA were sent to Fasteris (Fasteris SA, Chemin du Pont-du-Centenaire 109, 1228 Plan-Les-Ouates, https://www.fasteris.com) for Illumina deep sequencing using the platform HiSeq 2500 for all the datasets except for HYT-25 and HYT-26 that were sequenced using the HiSeq 4000 platform.

The small RNA molecules were size-selected on polyacrylamide gel, ligated to adapters at both termini (5' ad 3' end) and reverse transcribed for a cDNA library preparation using TruSeq® small RNA Library Prep Kit (Illumina). The cDNA library was then immobilized on a flowcell, which surface is coated with oligonucleotides complementary to the adapters. Once attached, the DNA strand bends over into a bridge-like shape and the bridge amplification can start to create clusters of identical strands of DNA, in presence of primers, polymerase and

nucleotides (Fig. 21). The clusters were then sequenced using a forward primer generating single reads. Each nucleotides is linked to a reversible terminator and a specific fluorophore, so that after each round of incorporation, the CCD camera can detect which base was incorporated according to the emission. After reading the signal, fluorescent dye and terminator are removed and the sequencing can proceed till the end of the strand (Fig. 22).

Table 2. Oligonucleotides used as probe for blot hybridization sRNA sequencing

Organism	Probe name	Sequence
Plant endogenous miRNA	mir160a_as	TGGCATACAGGGAGCCAGGCA
	mir169a_as	TCGGCaaGTCaTCCTTGGCTG
Potato virus Y and X	Pvy_s	GTTACCAGTGGATATTTTCTGGCGGGCAAG
	Pvy_as	ATGAATCATCCACAACAACGACAACCCGGCAC
	PvyN_s	GGAcAGAGCaGActTGgCtGAaCACAGgcTtG
	PvyN_as	CaAgcCTGTGtTCaGcCAagTCtGCTCTgTCC
	PvyO_s	GGATAGAGCTGATCTGCCAGAGCACAGATTAG
	PvyO_as	CTAATCTGTGCTCTGGCAGATCAGCTCTATCC
	PVX_s	CTACATAACCGACGCCTACCCCAGTTTCATAG
	PVX_as	CTATGAAACTGGGGTAGGCGTCGGTTATGTAG
Pepino Mosaic Virus	Pmv_s	CTTTAAACACTCGAAACTTAATCAGAGTGGGG
	Pmv_as	CCCCACTCTGATTAAGTTTCGAGTGTTTAAAG
	PmvLP_s	TTGGCTAGGCAAAAGATCTCCACAGGTAAT
	PmvLP_as	ATTACCTGTGGAGATCTTTTGCCTAGCCAA
	PmvCH2_s	CTTGCCCGCCAGAAAATATCCACTGGTAAC
	PmvCH2_as	GTTACCAGTGGATATTTTCTGGCGGGCAAG
Ligustrum Mosaic virus	LigMV_3UTR_as	CGTAGGAGTACGCGCTTACACACCAATTCCCAC
	LigMValpha190_as	GCACAGTGTGTCGTGCAAAGGTGATCTCACC
	LigMVbeta225_as	GCATGTCTTTCCAGGACTCTTTGGCGACCCACC
	LigMVgamma191_as	CCCAAACCATCTTTTTAAGACAGAGAGAATAAACG
	LigMVbeta2716_as	GCGATCCCACACAGGAACACTCCCATAGGGGTA
	LigMVgam2341_as	CCATAGAGGCATCACGATACCTCTCCAGCTTCTTC
	LigMV_3UTR_s	GTGGGAATTGGTGTGTAAGCGCGTACTCCTACG
	LigMValpha3419_s	GATCGTTTGTTACTGGACATAGCAAATGAAGATGG
	LigMVbeta2716_s	TACCCCTATGGGAGTGTTCCTGTGTGGGATCGC
	LigMVgamma2341_s	GAAGAAGCTGGAGAGGTATCGTGATGCCTCTATGG

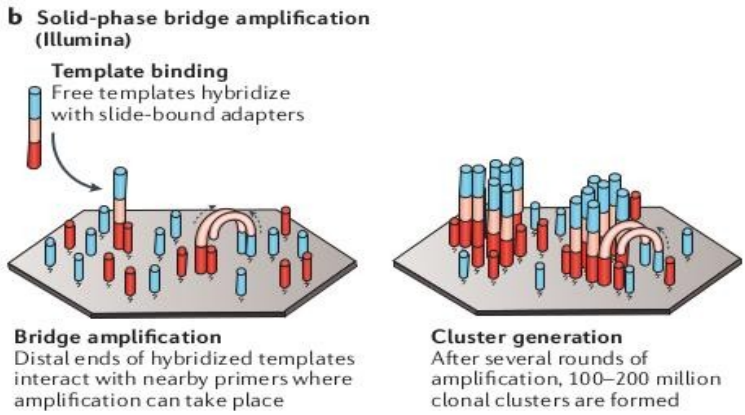


Fig. 21. Illumina solid-phase bridge amplification. Picture taken from Goodwin et al. 2016.

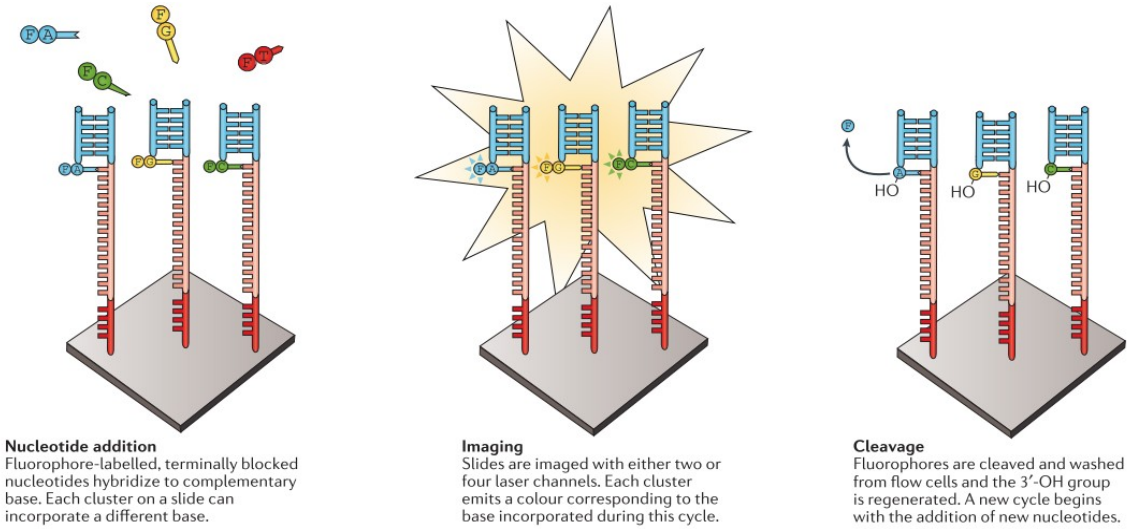


Fig 22. Illumina nucleotide incorporation detection. Picture taken from Goodwin et al. 2016.

The sequencing output results are stored in a FASTQ file, which combines the information of a FASTA file (read sequence) and a QUAL file carrying the quality PHRED score.
The information contained in a fastq file are summarized in four different lines. The first one is a title line, defined by "@", for read identification and optional description. The second one is the sequence line, usually in upper case and without any tabs or spaces. The third line starts with a "+" sign, to make clear the end of the sequence line and the beginning of the quality string. The last line contains the PHRED quality information in ASCII printable character (Table 3). Each character corresponds to a value derived from the formula $Q_{PHRED} = -10 \log_{10}$ (Pe) which estimates the probability that the corresponding base call is incorrect (Cock et al., 2010).
Before further analysis, the fastq file is trimmed to discard the adapter sequences from the reads population (performed by Fasteris using Trimmomatic, Bolger et al. 2014).

Table 3: Fastq file description of the first read in sample BPO-147

1	@HWI-D00393:75:C4EBJACXX:3:1101:5573:1996 1:N:0:ACTGAT
2	AAGGATGCAAAACAAGAGCA
3	+
4	BBBFFFFFFFFFFIIIIIFF

De novo assembly

Genome assembly is a bioinformatic process where fragmented pieces of sequences derived from next generation sequencing are built together in order to reconstruct the complete genome.

Two different categories of genome assemblers can be distinguished: a reference-guided genome assembly, when the reference genome is available, working as an alignment or mapping, and *de novo* genome assembly, when the reference is not available and the genome is built by finding the overlapping regions among the reads.

Furthermore, when talking about short reads assembly, three algorithms can be identified: the Greedy-extension with its string-based model, the overlap-layout consensus (OLC) and the *De Bruijn* graph based on K-mers (Zhang et al., 2011, Fig.23).

The term k-mer stands for all possible subsequences of length=k that are derived from a longer sequence. In computer science, a graph is a set of nodes and edges connecting them, and in a k-mer graph the nodes represent all the fixed-length subsequences while the edges are the fixed-length overlapping between the nodes (Miller et al., 2010).

The siRNA-omics pipeline applied in this work makes use of the *de novo* assembler Velvet, based on the De Bruijn graph, which is able to assemble very short reads (from 20 to 50 bp) in order to produce contigs (Zerbino and Birney, 2008, https://www.ebi.ac.uk/~zerbino/velvet/).

The reads are first hashed according to a predefined k-mer length (for example k=21 for 25 bp reads), then the graph is created with the nodes represented by a series of overlapping k-mers, while the adjacent k-mers overlap by k-1 nucleotides. This means that when two nodes are connected by a direct edge, the last k-mer of a node overlaps by k-1 nucleotides with the first of its next node.

There are three main functions in Velvet. “Simplification” mergers two nodes where one has only one outgoing edge and the other has only one incoming edge. “Tip” removes nodes disconnected to one end. “Bubble” merges redundant paths that start and end at the same nodes and contain similar sequences. With two more functions (Pebble and Rock Band), Velvet distinguishes the unique nodes from the repeat nodes based on node coverage, where the repeat nodes have multiple incoming and outgoing edges (Namiki et al., 2012).

If multiple reads overlap and cover the genome, contigs are created and the coverage for a position is defined as the number of reads that overlap at that position (Namiki et al., 2012).

Velvet 1.2.10 was used for de novo assembly with k-mer set as 13, 15, 17, 19, 21 and 23 on the sciCORE scientific computing core facility of the University of Basel (http://scicore.unibas.ch/).

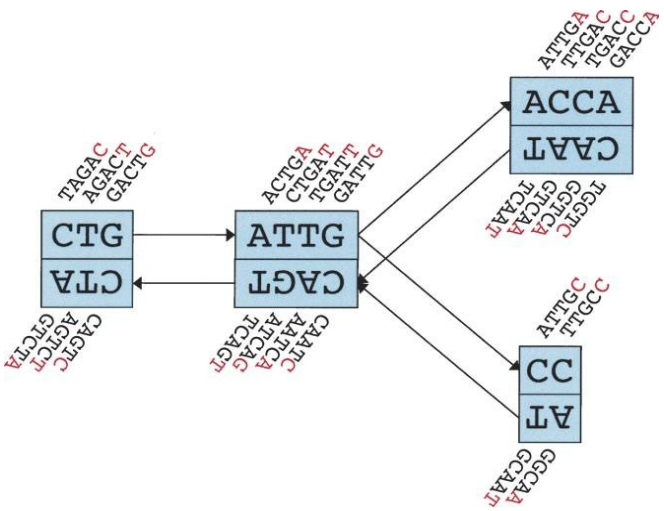


Fig. 23. Schematic representation of our implementation of the de Bruijn graph. Each node, represented by a single rectangle, represents a series of overlapping k-mers (in this case, k = 5), listed directly above or below. (Red) The last nucleotide of each k-mer. The sequence of those final nucleotides, copied in large letters in the rectangle, is the sequence of the node. The twin node, directly attached to the node, either below or above, represents the reverse series of reverse complement k-mers. Arcs are represented as arrows between nodes. The last k-mer of an arc's origin overlaps with the first of its destination. Each arc has a symmetric arc. Note that the two nodes on the left could be merged into one without loss of information, because they form a chain. Copied from Zerbino and Birney, 2008.

The graph created by Velvet were then used by Oases 0.2.09, a transcriptome assembler designed to produce scaffolds from the preliminary assembly (Fig.24, Schulz et al., 2012, https://www.ebi.ac.uk/~zerbino/oases/).

After reading the contigs, Oases visits the nodes several times and merges paths that have same starting and end node. For each node, if the coverage is less than 10% of the sum of coverages of outgoing edges from that same node, the node is removed and all contigs with a coverage less than a static cutoff are removed from the assembly. The contigs are then organized in loci based on the idea that all the transcripts from one gene should be assembled into connected components (Schulz et al., 2012).

The Oases contigs derived from different k-mers values (11, 13, 15, 17, 19, 21, 23) were combined in one single FASTA file and used for further analysis described in the following paragraphs.

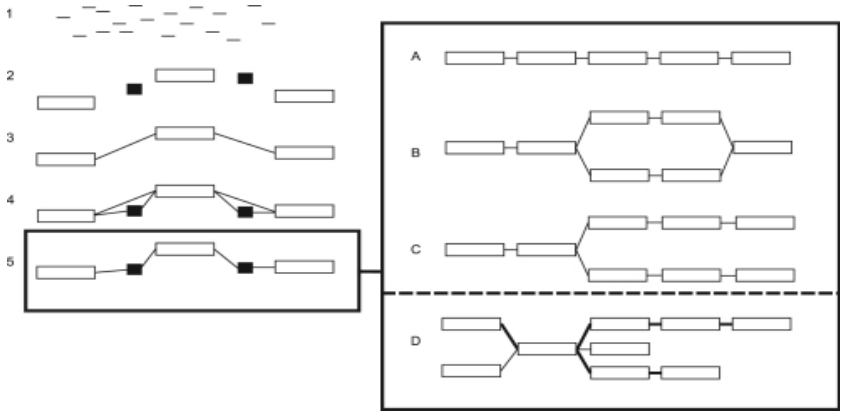


Fig. 24. Schematic overview of the Oases pipeline.

(1) Individual reads are sequenced from an RNA sample, (2) contigs are built from those reads, some of them are labeled as long (clear), others short (dark), (3) long contigs, connected by single reads or read-pairs, are grouped into connected components, called loci, (4) short contigs are attached to the loci, (5) the loci are transitively reduced. Transfrags are then extracted from the loci. The loci are divided into 4 categories: (A) chains, (B) bubbles, (C) forks and (D) complex (i.e. all the loci which did not fit into the previous categories). Copied from (Schulz et al., 2012).

Seqman

Seqman is a sequence assembly tool included in the DNAstar Lasergene package (SeqmanPro 12.0.0, https://www.dnastar.com/t-seqmanpro.aspx).
Giving as input the contigs generated by Velvet and Oases, Seqman extends these sequences by looking at the possible overlapping regions and merging them together. This further scaffold could help to reconstruct the complete genome of the virus.

Mapping tools

Reference host genomes

The short reads population assembled by Velvet and Oases contain also endogenous small RNAs. For this reason, as described by Seguin et al. (2014a) a filtering step to remove all the host small RNA contigs could improve scaffolding of the Oases contigs of viral reads performed by Seqman.
The filtering step was done by mapping of all Oases contigs to the host reference genome. The genomes of potato (*S. tuberosum, PGSC_DM_v3, http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml*), tomato (*S. lycopersicum, GCA_000188115.2.28,ftp://ftp.ensemblgenomes.org/pub/plants/release-28/fasta/solanum_lycopersicum/dna/*), grapevine (*Vitis vinifera,* ftp://ftp.ncbi.nih.gov/genomes/Vitis_vinifera/) and fig (*Ficus carica*, https://www.ncbi.nlm.nih.gov/genome/?term=txid3494[orgn]) were available at the time our analysis. For sweet cherry samples, the sweet cherry and peach genomes were used as host reference (https://www.ncbi.nlm.nih.gov/assembly/GCF_002207925.1/, PAV_r1.0.fasta.gz, and *ftp://ftp.bioinfo.wsu.edu/species/Prunus_persica/Prunus_persica-genome.v2.0.a1/assembly/Prunus_persica_v2.0.a1_scaffolds.fasta.gz* respectively*).*
For larch, the closest related available Norway spruce (*Picea abies*, ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/FASTA/GenomeAssemblies/) was used as reference.

BWA, SAM/BAM format and Samtools

BWA (Burrow-Wheeler Alignment) is a software package based on the Burrow-Wheeler Transform (BWT, Burrow and Wheeler, 1994) for mapping sequences (short, long reads or contigs) to a reference genome. One algorithm of the package is designed for short reads (up to 200 nt) and implemented by the “aln” command which creates the suffix tree, followed by “samse/sampe” command which converts the suffix array in the SAM format output.
The second algorithm is designed for longer reads and uses the command “bwasw” and the output file of the alignment is also in a SAM format.
The Sequence/Alignment/Map (SAM) is a TAB-delimited text format containing a header section and an alignment section. The header contains the reference name and its length, as

defined in the FASTA file used for mapping. The alignment sections is defined by 11 mandatory fields, summarized in Table 4. Among these, the first column is the name of the read (QNAME), followed by a FLAG that can be 4, 0 or 16 if the read unmapped or mapped in the forward or reverse strand, respectively. The third column is the reference sequence name (RNAME) and POS indicates the starting position of the alignment between the read and the reference. MAPQ and CIGAR indicate the quality of the mapping and the possible presence of insert/deletion using the word M (match/mismatch), I (insertion), D (deletion).

Table 4. Mandatory fields in the SAM format

No.	Name	Description
1	QNAME	Query NAME of the read or the read pair
2	FLAG	Bitwise FLAG (pairing, strand, mate strand, etc.)
3	RNAME	Reference sequence NAME
4	POS	1-Based leftmost POSition of clipped alignment
5	MAPQ	MAPping Quality (Phred-scaled)
6	CIGAR	Extended CIGAR string (operations: MIDNSHP)
7	MRNM	Mate Reference NaMe ("=" if same as RNAME)
8	MPOS	1-Based leftmost Mate POSition
9	ISIZE	Inferred Insert SIZE
10	SEQ	Query SEQuence on the same strand as the reference
11	QUAL	Query QUALity (ASCII-33=Phred base quality)

For downstream analysis, the SAM file can be exported by the manipulating SAM-format tool Samtools (version 1.2, http://samtools.sourceforge.net/) in a Binary Alignment/Map (BAM) file, compressed in BGZF format. The file contains the same information but in a binary format readable only by machines (Li et al., 2009). Samtools is also able to extract information like the mapped or unmapped sequences, the consensus sequence and the VCF file containing all the possible variants and SNPs can be exported by combining SAMtools with BEDtools.

BLAST

The Basic Local Alignment Search Tool (BLAST) is an algorithm for comparing a query input sequence with a database of sequences that can be either of proteins or nucleotides. It was developed 1989 at the National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH) (Altschul et al., 1990). There are many programs available that allow alignment between nucleotides (blastn), proteins (blastp), nucleotides towards protein sequence (blastx) or vice versa (tblastx). The quality and reliability of the alignment is depicted by some values in the output information. Among these, the query cover is the

percentage of the query that aligns with the target. The E value describes the number of hits you would expect to find by chance, given the quality of the alignment and the database size. This means that the closer to 0 the e-value is, the more significant is the match. The identity represents the percentage identity between the query and the target. In the siRomics pipeline, blastn is used after the de novo assembly and the scaffolding with Seqman. In this way, we would have already an idea about the viruses present and we can select the most closely related isolate for further analysis.

Tools for visualization and analysis of the mapping data

IGV

The Integrative Genomics Viewer (IGV) is a visualization tool for NGS data developed in 2011 (Robinson et al., 2011). IGV was used to visualize Oases and Seqman contigs mapped by BWA to a reference genome. IGV shows the contigs as gray arrows in forward or reverse orientation with respect of the reference sequence. Single nucleotide polymorphism (SNPs) or indels are indicated using a color code and bars. By zoom in, IGV allows to read the sequence of the single contigs. It is also possible to export the consensus sequence and specific information regarding the contigs, like covered position, length and sequence. Together with MISIS, IGV was used for correction of the viral reconstructed consensus sequence.

MISIS

MISIS is a bioinformatic tool developed by Jonathan Seguin during his PhD studies at the UniBasel for visualization and in-depth analysis of small RNAs mapped on a reference sequence (Seguin et al., 2014b, 2016). By reading the SAM or BAM, MISIS creates a table containing the position along the reference genome and the total counts of the sRNAs on the forward or reverse strand and on both strands (Seguin et al., 2014b). By looking at the CIGAR and MD column in the SAM/BAM file, MISIS creates a table for total counts of sRNA reads mapped with up to two mismatches to the reference genome or with perfect match. These text files are then used by MISIS to display the map of reads as an interactive histogram, where the x-axis indicates the position along the reference and the y-axis the total number of reads mapped at each position, in sense (blue bars) or antisense orientation (red bars). Furthermore, the second version of MISIS (MISIS-2) can determine and visualize the consensus sequence supported by the majority of the mapped reads, the possible SNPs at any position and count the reads based on size and 5' nucleotide identity (Seguin et al., 2016).

RT- Polymerase Chain Reactions

To validate the reconstructed sequence obtained by the siRomics approach for the two strains of the PVY viruses in the potato sample BPO-147 and for the newly discovered *Ligustrum mosaic virus* in sample HYT-24, RT reaction and PCRs were carried out using primers designed on the reconstructed sequences (see Table 5 and 6).
2 µg of total RNA from sample BPO-147 were incubated with 1 µl 10 mM dNTP mix, 2 pmol

of an RT primer complementary to both PVY-A and PVY-B sequences (PVY RT_as, -as2. -as3, -as4 or -as5) and heated to 65°C for 5 minutes. After chilling is ice for 2 minutes, the reverse transcription was performed by adding to the mixture 4 µl of 5x first-strand synthesis buffer (250 mM Tris-HCl [pH 8.3], 375 mM KCl, 15 mM MgCl2, 0.1 M DTT), 1 µl 0.1 M DTT, 1 µl (40 U) of RNase inhibitor RNasin (Promega) and 1 µl (200 U) of Superscript III reverse transcriptase (Invitrogen) and incubated at 55°C for 60 minutes. PCR amplification was performed in 25 µl of 1x PCR buffer, 0.5 µl of 10 mM deoxynucleoside triphosphate mix, 2.5 U of Taq polymerase (Sigma), 20 pmol of PVY-RT primer, 20 pmol of strain-specific primer (B1_s, B2_s, B3_s, B4_s, B5_s, A1_s, A2_s, A3_s, A4_s, A5_s respectively) and 2 µl of reverse transcription reaction mixture as template. The PCR amplification was carried out for 35 cycles (each cycle consisting of 30 s at 95°C, 30 sec at 60°C and 60 sec at 72°C) and the product were analyzed by 1% agarose gel electrophoresis. The PCR bands were extracted from the gel with the GenElute™ Gel Extraction Kit (Sigma) and sequenced at Fasteris (Fasteris SA, CH).

Following this protocol, a genome walking was performed by designing antisense primers (Fig. 25) on the sequenced region in order to amplified the corresponding cDNA together with a strain-specific primer. However, we could not sequence the complete genome of the two PVY strains, but just the last 4 kb starting from the 3' end of the genome.

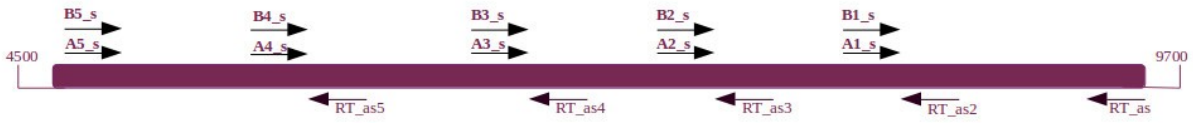


Fig. 25. Schematic representation of RT-PCRs performed on PVY genomes. The RT-primers below the genome were used first to get the cDNA that was then amplified using the same RT-primer in combination with each of the two sense-strain specific primers showed on the top.

Table 5: List of primers used for RT-PCRs and PCRs for PVY genome validation

Viral species	Original name	Primer name	Sequence
PVY	pvyRT_as	RT_as	GTCTCCTGATTGAAGTTTACAGTCAC
	pvyRT_as2	RT_as2	CCAAACCATAAGCCCATTCATCACAG
	147Pvy_as3	RT_as3	AAATCATAGTTCAAACCAAGATCTGAGAA
	147pvy_as4	RT_as4	GCATCCATCAAAGGCCTGAAGAA
	147pvy_as5	RT_as5	TCTCATGAGCGATTTAGCTTCATGCTC
	147pvyO_1s	B1_s	GATACGTGGTATGAAGCGGTACAAC
	147pvyN_1s	A1_s	ATACGTGGTATGAGGCAGTGCGGA
	147PVYO_2s	B2_s	GTTTAGAGGTAATAATAGCGGTCAACCA
	147PVYN_2s	A2_s	GTTCAGAGGAAATAACAGTGGTCAGCCT
	147PVYO_s3	B3_s	GCCAATTAGTAACCAAGCATGTAGTT
	147PVYN_s3	A3_s	GTCAGCTGGTGACAAAGCACGTG
	147pvy_Os4	B4_s	TCGACGTGAAGGACATACCAGCA
	147pvy_Ns4	A4_s	TAGAGGTTGATGTGAAAGACATTC
	147pvy_Os5	not shown	ACAGGAAACAACCAGGAGTCAGCAGA
	147pvy_Ns5	not shown	ATAAAAGACAGCCAGGGGTGAGTAAG
	147Pvy_Ns6	B5_s	GAAATGCTTTGGGAAACGGTTGTCAAG
	147pvy_Os6	A5_s	GAAATGCTTTGGGAAACTGTTGTCAAA

In sample HYT-24, two antisense oligonucleotides were designed on the identified 3'UTR contig and used in a reverse transcriptase reaction to get the cDNA that was further amplified in order to determine if this contig is part of each of the three genome segments and to measure the length of the intervening poly-A stretch.

1 µg of total RNA was incubated with 1 µl 10 mM dNTP mix, 2 pmol of RT primer (3'UTR_79as or 3'UTR_133as), 4 µl of 5x first-strand synthesis buffer (250 mM Tris-HCl [pH 8.3], 375 mM KCl, 15 mM MgCl2, 0.1 M DTT), 1 µl 0.1 M DTT, 1 µl (40 U) of RNase inhibitor RNasin (Promega) and 1 µl (200 U) of Superscript III reverse transcriptase (Invitrogen) and incubated at 55°C for 60 minutes. The reaction was stopped by heating at 70°C for 15min.

PCR amplification was performed in 25 µl of 1x PCR buffer, 1.5 mM of magnesium chloride, 0.5 µl of 10 mM dNTP mix, 2.5 U of Taq polymerase (Sigma), 20 pmol of RT primer, 20 pmol of segment specific primer (listed in Table 6 and schematically represented in Fig. 26) and 1 µl of RT reaction mixture as template. The PCR conditions were as follow: one cycle at 95°C for 10 min, 35 cycles of 95°C for 30 sec, 52°C for 30 sec, 72°C for 1 min and a final cycle at 72°C for 3 min.

The products were analyzed by electrophoresis on a 1.5% agarose gel in TAE buffer (40 mM Tris acetate, 1 mM EDTA, pH 8.0) and stained with 1 µg/ml of ethidium bromide.

The PCR products were purified from the gel with the GenElute™ Gel Extraction Kit (Sigma) and sent for sequencing at Fasteris (Fasteris SA, CH).

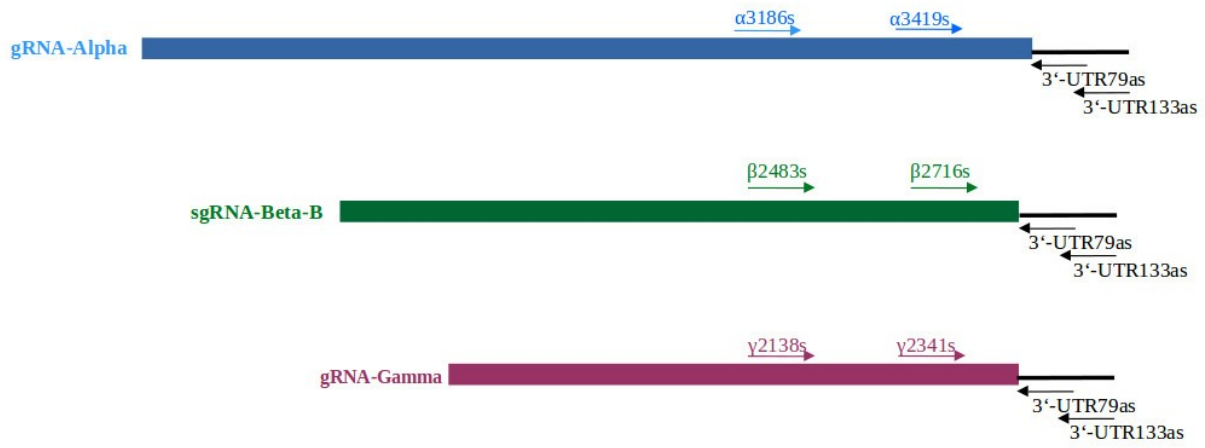


Fig. 26. Schematic representation of RT-PCRs performed on LigMV segments. The cDNA derived from the RT-reaction with the primers designed on the putative 3'UTR was used as template for a PCR with the sense primers designed at the 3'end of the segments to identify the length of the internal polyA-tail.

Table 6: List of primers used for RT-PCRs and PCRs for polyA-tail identification

Viral species	Primer name	Sequence
Ligustrum Mosaic virus	LigMV3UTR_133as	CGGCTCAGTGTGCAACAATC
	LigMV3UTR_79as	CGGCTCAGTGTGCAACAATC
	LigMValpha3186_s	TCTGCAAGCAGAAAAGGAGACA
	LigMValpha3419_s	GATCGTTTGTTACTGGACATAGCAAATGAAGATGG
	LigMVbeta2761_s	ATGGATCATGCGACCCAAT
	LigMVbeta2483_s	CCAGATTGCAAGATCTGTGGG
	LigMVbeta2716_s	TACCCCTATGGGAGTGTTCCTGTGTGGGATCGC
	LigMVgamma2138_s	AGCACGGGCTTACAGGAAAA
	LigMVgamma2341_s	GAAGAAGCTGGAGAGGTATCGTGATGCCTCTATGG

3.1 siRomics applied to *Solanum tuberosum*

During a survey of potato viruses in Switzerland in 2014, potato tubers were purchased in different local shops and planted in the greenhouse of the Institute of Botany (Basel, Switzerland). Upper leaves from the plants which exhibited virus-like disease symptoms (yellow mosaic and deformations), as well as upper leaves of two healthy-looking plants (grown from two separate tubers), were harvested in pools of several leaves per plant and subjected to siRomics, as described in Material and Methods. Once verified the quality and the integrity of the isolated total RNA by separation on formaldehyde-containing 1% agarose gel (Fig.27), a polyacrylamide gel electrophoresis followed by blot hybridization was performed to evaluate the quality of small RNAs. The evolutionarily-conserved plant miRNA miR160 was used as an internal control in all the following blot hybridization (Fig. 28).

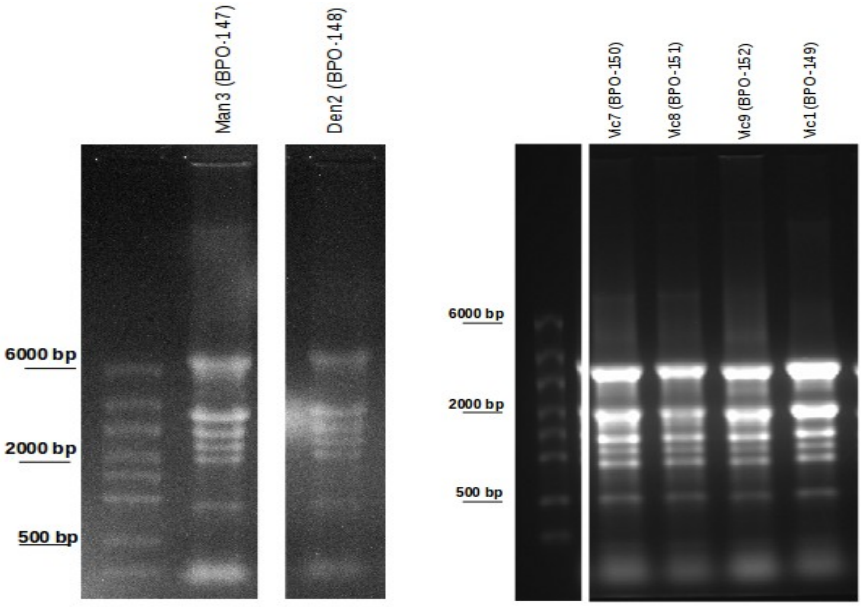


Fig. 27. Potato total RNA gel electrophoresis analysis. 10 µg total RNA extracted from the potato leafsamples BPO-147 and BPO-148 and 8 µg total RNA from the leaf samples BPO-149-150-151-152 were loaded on formaldehyde-containing 1% agarose gel and run for 2.5 hours. The High Range RNA ladder, Riboruler was used to estimate the molecular weight of the RNA. After separation, high molecular weight RNA species (rRNA and tRNA) were stained with EtBr.

The small RNA population was then size selected and converted to cDNA following the Illumina protocol. Sequencing of the resulting cDNA libraries in one lane of Illumina Genome Analyzer HiSeq2500 yielded 23 to 38 M reads per library, with the majority of reads comprised in the range of 20 to 25 nts, which were selected for *de novo* assembly following the siRomics approach (Table 7).

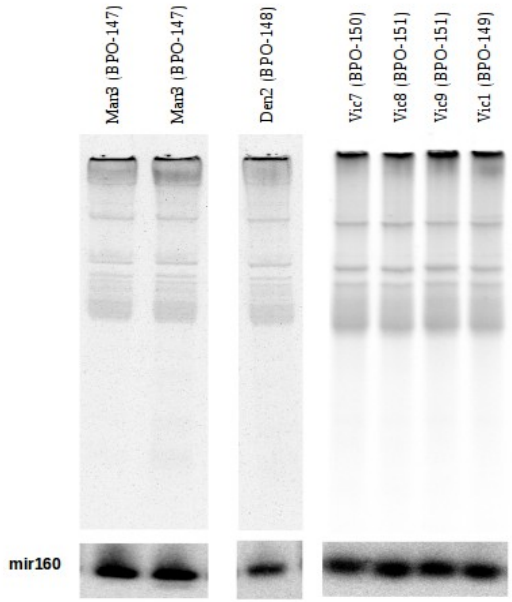


Fig. 28. Small RNA blot hybridization analysis. Ten micrograms of total RNA were separated on 15% polyacrilamid gel for four hours, and after staining with EtBr for loading control (upper scans), transferred to Hybond N+ membrane (Amersham), and then hybridized with the P^{32} end-labeled probe specific for miR160 . The membrane was exposed for four days and scanned (lower scans).

Table 7: Total number of reads per sRNA library prepared from potato leaves

sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
BPO-147	26,731,141	25,316,249
BPO-148	33,304,831	27,869,991
BPO-149	23,527,654	20,006,593
BPO-150	38,808,388	35,765,844
BPO-151	37,115,191	34,131,328
BPO-152	29,276,580	27,040,097

Redundant and non-redundant reads were *de novo* assembled by Velvet and Oases (k-mer length from 11 to 23) using minimum length of the output contigs, 50 or 100 nucleotides respectively. Regardless the redundancy or the cut-off length, for almost all the datasets, the longest Oases contig was obtained with the k-mer set at 17 or 19. The number of contigs per each k-mer obtained with the cut-off length of 50 nts was much higher than the number of contigs obtained with the cut-off length of 100 nt. The contigs derived from all the k-mers were concatenated together in one single FASTA file (oases-all), filtered throughout the host (*Solanum tuberosum*) reference genome and further assembled by Seqman.

3.1.1 siRomics reveals stability of PVY-PVX virome quasispecies and differential silencing responses to PVY and PVX in vegetatively propagated potato plants

In each of the three symptomatic plants stemming from one tuber (BPO-150-152), complete genomes of the potexvirus PVX and the potyvirus PVY were reconstructed, while the healthy-looking plant (BPO-149) didn't show any viral contigs.

Notably, the 6436 nt genome of PVX was reconstructed *de novo* as a single Seqman contig from one library (BPO-152), while in other libraries two (BPO-150) or four (BPO-151) separate Seqman contigs covered a near complete genome, and the gap-filling step was executed to reconstruct the complete genome, by alignment of the Seqman contigs to the reference genome and merging the overlapping regions. Likewise, the 9721 nt PVY genome including a 18 nt poly(A) tail was reconstructed *de novo* as a single Seqman contig (BPO-150 and BPO-152), or as two separate Seqman contigs fused by gap-filling (BPO-151). The single contigs representing a complete PVY genome contained short insertion/deletions, which were corrected by sRNAs re-mapping. The consensus nucleotide sequences of PVX and PVY, supported by the majority of sRNA reads, were found to be identical in all the three plants. Moreover, the profile of single nucleotide polymorphisms (SNPs), deviating from the consensus sequence more than 10% (and less than 50%), was near identical between the plants (see Supplementary Fig.S1 in Annex). This indicates that the PVX-PVY virome quasispecies is stable in vegetative progeny plants grown from different eyes of the same infected potato tuber. The consensus sequence of our PVX isolate (deposited to NCBI Genbank as PVX-CH, accession number MF405302) is 98% identical (122 SNPs) to that of PVX-OG from Japan deposited to Genbank in 2004 (AB196000), while the consensus sequence of our PVY isolate (PVY-CH; MF405303) is 99% identical (34 SNPs) to a IUNG-4 isolate of PVY-NTN strain from Poland deposited in 2011 (JF927752). This illustrates a broad geographic distribution of PVY and PVX strains composing the virome we identified in Switzerland in 2014.

To characterize plant RNA silencing-based response to the PVX-PVY virome we mapped 20-25 nt sRNAs from infected and healthy plants to the reference genome sequences of *S. tuberosum* (superscaffold, unanchored, chloroplast and mitochondrial) as well as PVX and PVY (in combination and individually) with zero mismatches, and counted the endogenous and viral sRNAs sorted by size (20-25 nts), polarity (forward, reverse) and 5'-terminal nucleotide identity (5'A, 5'C, 5'G, 5'U). The virus-derived sRNAs constituted ca. 30% of total 20-25 nt sRNAs in all three libraries (Figure 29a). Notably, in each library, PVY-derived sRNAs were more abundant than PVX-derived sRNAs (Figure 30b), indicating a much stronger silencing response to a PVY component of the virome. As previously described, the majority of the potato-derived sRNAs are characterized by the 21-nt and 24-nt size class (Fig. 29c). The 21-nt sRNAs size-class is the most abundant in both PVY and PVX, with the 22-nt class more abundant for PVX than PVY (Figure 29d, 29e). This suggests that potato DCL4 and DCL2 activities likely generating 21-nt and 22-nt siRNAs, respectively, may target PVY and PVX differentially, with either DCL2 being relatively more active against PVX, or DCL4 being relatively more active against PVY. Analysis of 5'-nt identities of viral sRNAs revealed

a slight difference between PVY and PVX: while both viruses spawn comparable numbers of 5'U, 5'A, and 5'C sRNAs and much less abundant 5'G sRNAs, their relative frequencies were somewhat different in that the third most frequent nucleotide 5'C, being relatively more frequent in PVX than PVY (not shown).

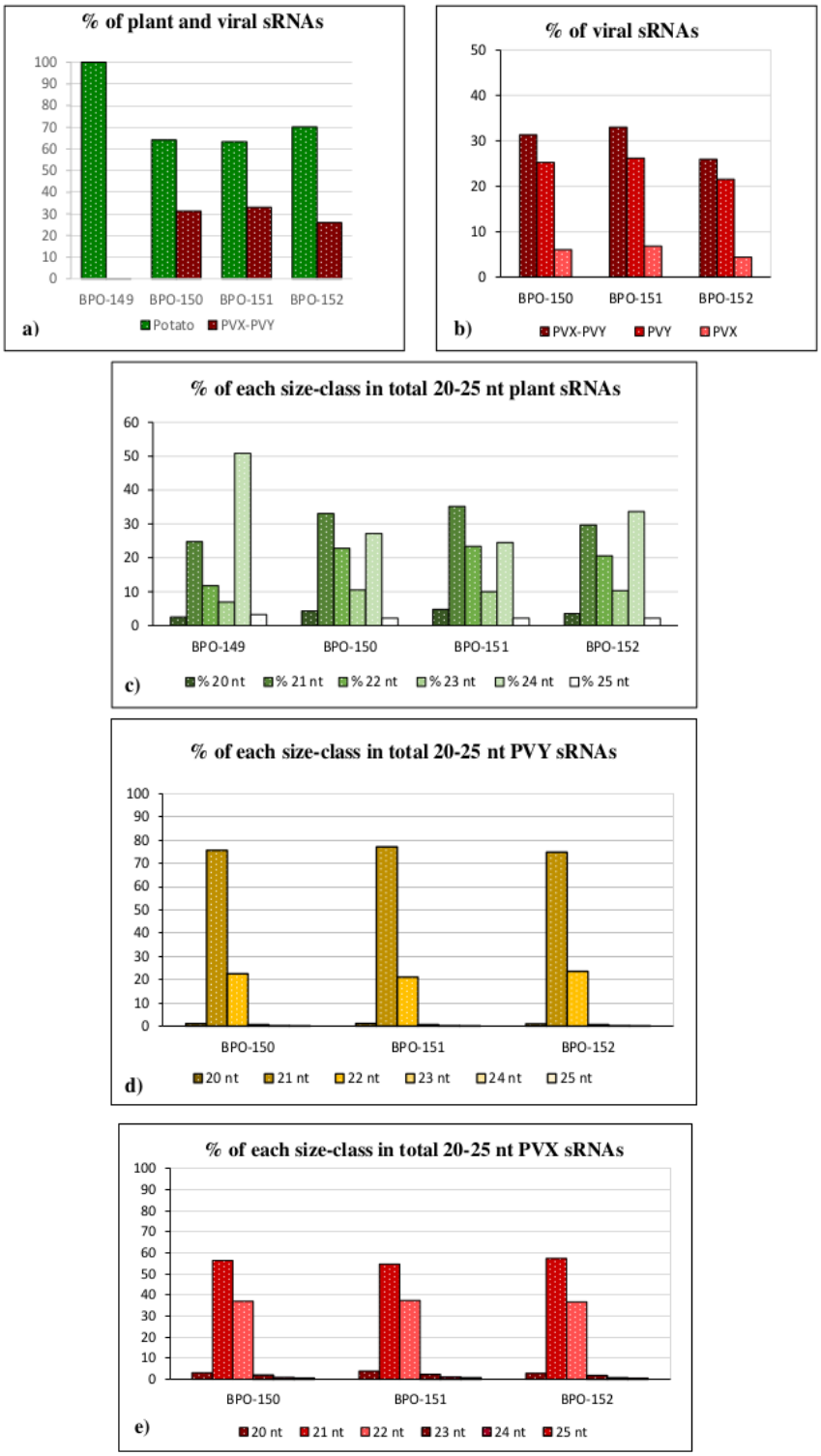


Fig.29. Counting of viral and host 20-25 nt small RNA reads from healthy (BPO-149) and three PVY-PVX virome (BPO_150, BPO-151,BPO-152)-infected potato plants **.** a) Percentage of plant and viral reads in total 20-25 nt sRNAs; b) Percentage of total viral (PVX+PVY), PVY and PVX reads in each datasets; c) Percentage of each size-class of sRNAs derived from the plant genome; d) and e) Percentage of each size-class of viral siRNAs derived from PVX and PVY in each of three different datasets.

This suggests that viral 21-22 nt sRNAs are predominantly associated with potato orthologs of AGO1 (5'U), AGO2 (5'A) and AGO5 (5'C), with the latter having higher affinity to PVX-derived sRNAs. Alternatively, a higher CG content in PVX than PVY could also account for the observed differences in 5'C frequencies, provided that viral siRNAs are randomly processed by DCLs from dsRNA precursors covering the entire virus genome. The 24-nt sRNAs, being under represented in viral, mitochondrial, or chloroplast sRNA populations, are derived predominantly from the potato nuclear genome and have a strong bias to 5'A (71-72%), thus resembling *Arabidopsis* heterochromatic siRNAs (Havecker et al. 2010). Notably, 24-nt potato sRNAs are relatively less abundant in virus-infected plants where 21-nt and 22-nt sRNAs increase compared to the healthy sample (Figure 29c). This shift in size-profile may reflect activation of DCL4 and DCL2 in response to virus infection, which would generate viral siRNAs as well as endogenous secondary siRNAs of respective 21-nt and 22-nt classes. Consistent with this hypothesis, analysis of 5' nt frequencies revealed that the nuclear genome-derived 21-nt sRNAs with 5'U, likely representing miRNAs generated by DCL1 and associated with AGO1, are highly abundant in healthy plants (76-80% of total 21-nt reads) and much less abundant in PVX-PVY virome-infected plants (52-57%)**.**

Analysis of single-nucleotide resolution sRNA maps generated by MISIS-2 (Seguin et al. 2016) revealed that viral sRNA species of major size classes (21-nt and 22-nt) cover the entire genomes of PVY and PVX in both forward and reverse orientations without gaps (Fig. 30). The hotspots of sense and antisense sRNA production are almost equally distributed along the PVY genome, whereas those are more abundant in a 5' portion of the PVX genome containing an RdRP ORF, compared to a 3'-portion of the genome containing triple gene block (TGB1, TGB2, TGB3) and coat protein (CP) ORFs (Fig. 30). Unlike PVY that translates a single large polyprotein from genomic RNA, PVX translates TGB1-3 and CP from two separate subgenomic RNAs (sgRNAs). The PVX genome region transcribed into sgRNAs appears to be more resistant to the sRNA-generating machinery. Almost equal abundance of sense and antisense reads along PVY and PVX genome sequences for each major sRNA size (Fig. 30) suggests that 21-nt and 22-nt viral sRNAs are processed by respective DCLs (likely DCL4 and DCL2) in the form of duplexes from perfect dsRNA precursors covering the entire virus genome. These precursors are likely produced by viral RdRP during gRNA replication, and/or by host RDR activities converting viral single-stranded RNAs into dsRNA. In PVX, the sgRNA region transcribed by viral RdRP appears to generate either less abundant or less optimal dsRNA substrates for DCLs.

Comparison of single-nucleotide resolution maps of viral sRNAs in BPO-150, BPO-151 and BPO-152 (Fig. 30) indicates that the biogenesis mechanisms for PVY- and PVX-derived 21-nt and 22-nt siRNAs are both qualitatively and quantitatively similar in all the three PVY-PVX virome-infected plants. Taken together, the potent antiviral RNA silencing machinery constantly but differentially targeting PVY and PVX does not appear to destabilize the virome complex or modify the consensus sequences in the virome quasispecies during vegetative propagation of potato plants under greenhouse conditions.

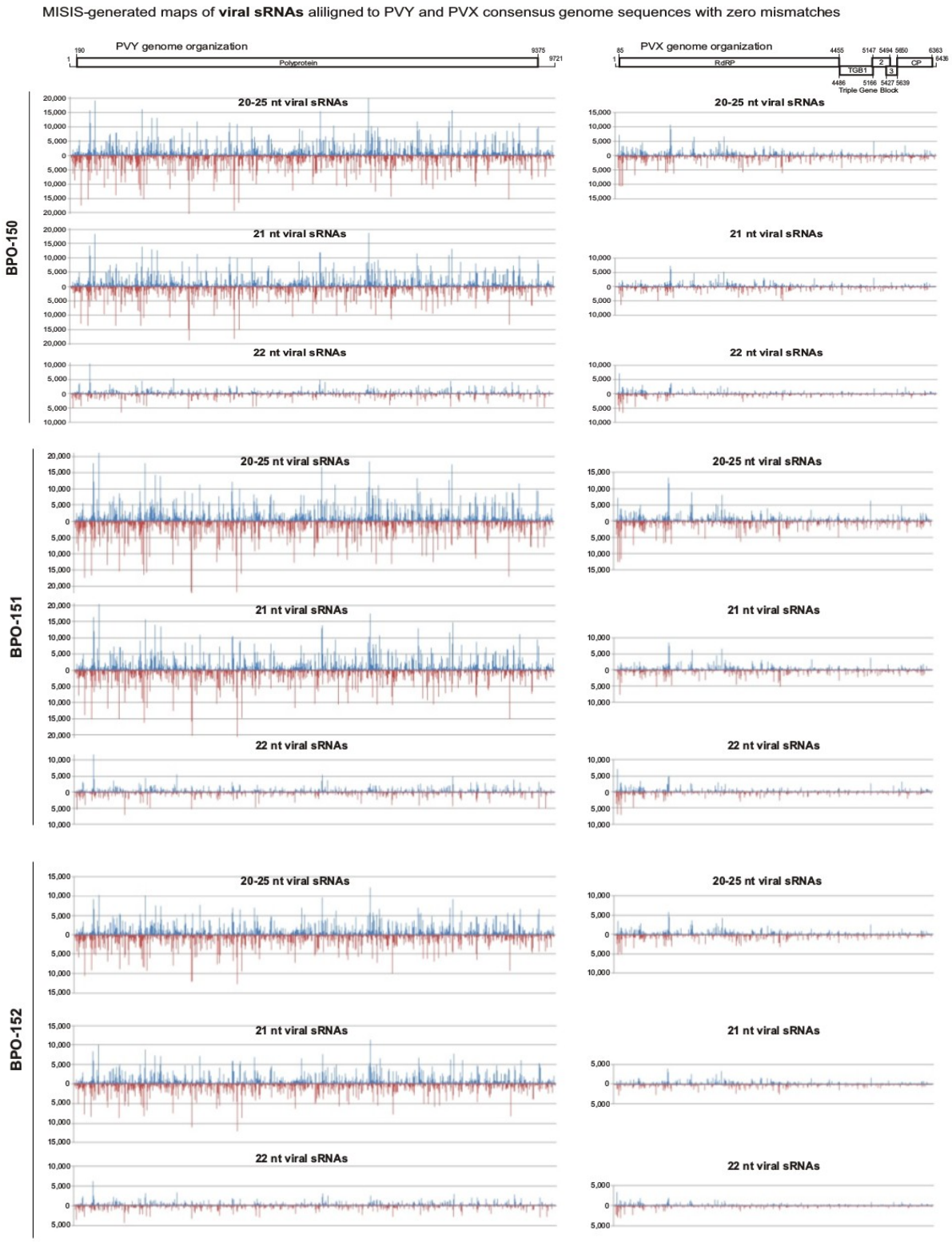


Fig. 30. Single-nucleotide resolution maps of viral sRNAs from leaf tissues of three plants grown from a potato tuber infected with a PVY-PVX virome. For each plant sample (BPO-150, -151 and -152), the histograms plot the numbers of total 20-25-nt, 21-nt or 22-nt viral sRNA reads at each nucleotide position of 9721 nt PVY and 6436 nt PVX genome sequences (mapped with zero mismatches). The bars above the axis represent sense reads starting at each position and those below represent antisense reads ending at the respective position. A scaled PVY and PVY genome diagrams are shown above the histograms, with the ORFs boxed and their encoded proteins and nucleotide positions indicated.

3.1.2 Reconstruction of two distinct PVY strains co-infecting a potato plant

In a plant grown from one of the infected tubers (BPO-147), *de novo* virome reconstruction revealed the presence of two strains of PVY. Redundant reads were assembled in recombinant Seqman contigs, longer than the PVY genome itself (10,654 and 17,315 nts respectively), while the non-redundant sRNAs were assembled in two Seqman contigs: a 9710 nt Contig_6 covering a complete PVY including a 24-nt poly(A) tail and a 1495 nt Contig_9 representing the 5'-terminal portion of PVY. Contig_6 contained numerous degenerate nucleotides throughout its sequence, except for the 5'-terminal portion of 1535 nt, suggesting its chimeric nature.

Indeed, BLAST analysis revealed that the Contig_6 of 9710 nt covered the complete PVY genome including a 24-nt poly(A) tail and it's 94% identical to the Ordinary isolate of PVY. Contig_9 of 1495 nts covered the 5'-terminal portion of PVY and is 99% identical to the corresponding region of many isolates of PVY strain NTN. Mapping of Oases contigs to Contig_6 and Contig_9, followed by map inspection using Interactive Genomic Viewer (IGV) revealed that Contig_6 has numerous degenerate nucleotides in its sequence probably due to two different categories of contigs with high sequence similarity that were merged by Seqman creating a chimeric regions throughout its genome. However, the 5' end of Contig_6 lacks degenerated nucleotides and resulted to be 99% identical to many isolates of PVY-O and only 74% identical to Contig_9, which explain why the two different categories of contigs were merged together to create two different 5' extremities (Fig.31 and Fig. 32).

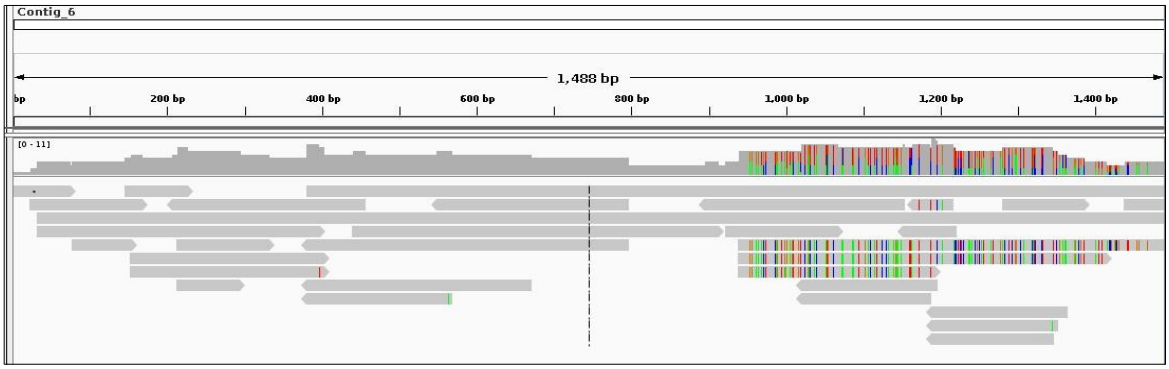


Fig. 31. Zoom in at the 5' extremity of the Oases contig aligned to Contig_6. The gray contigs perfectly match with the reference sequence (Contig_6), whereas the colored contigs represent the second category that align to the other isolate of PVY and Contig_9.

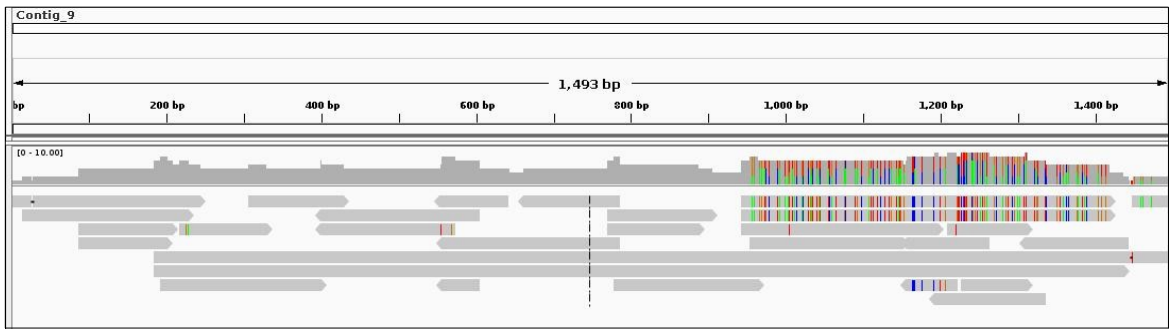


Fig. 32. Oases contigs align to Contig_9. The gray contigs perfectly match with the reference sequence (Contig_9), whereas the colored contigs represent the second category that align to Contig_6.

To confirm the presence of two strains in our virome (operationally named PVY-A^N and PVY-

B^O) we used an RT-PCR sequencing approach in which several overlapping amplicons of ca. 700-900 bps were obtained, in each case using two forward primers corresponding to a region of sRNA sequence dissimilarity in pair with one reverse primer corresponding to a region of sRNA sequence identity (see Materials and Methods). This approach enabled us to disentangle a 4002-nt 3'-terminal parts of PVY-A^N and PVY-B^O, which shared 87% identity and matched best to the corresponding parts of an isolate IUNG-4 of PVY-NTN (JF927752; 99% identical) and an isolate CO2140 of PVY-O (HQ912914; 99% identical).

To disentangled the remaining middle portions of PVY-A^N and PVY-B^O (sharing 95% identity), redundant sRNA reads were mapped to the reference sequences of NTN and CO2140. By looking at the mapping results with Misis and the alignment between NTN and CO2140 sequences, the positions with ~50% of possibilities between one nucleotide or the other were resolved and the overall consensus sequence of PVY-A^N and PVY-B reconstructed (Fig. 33).

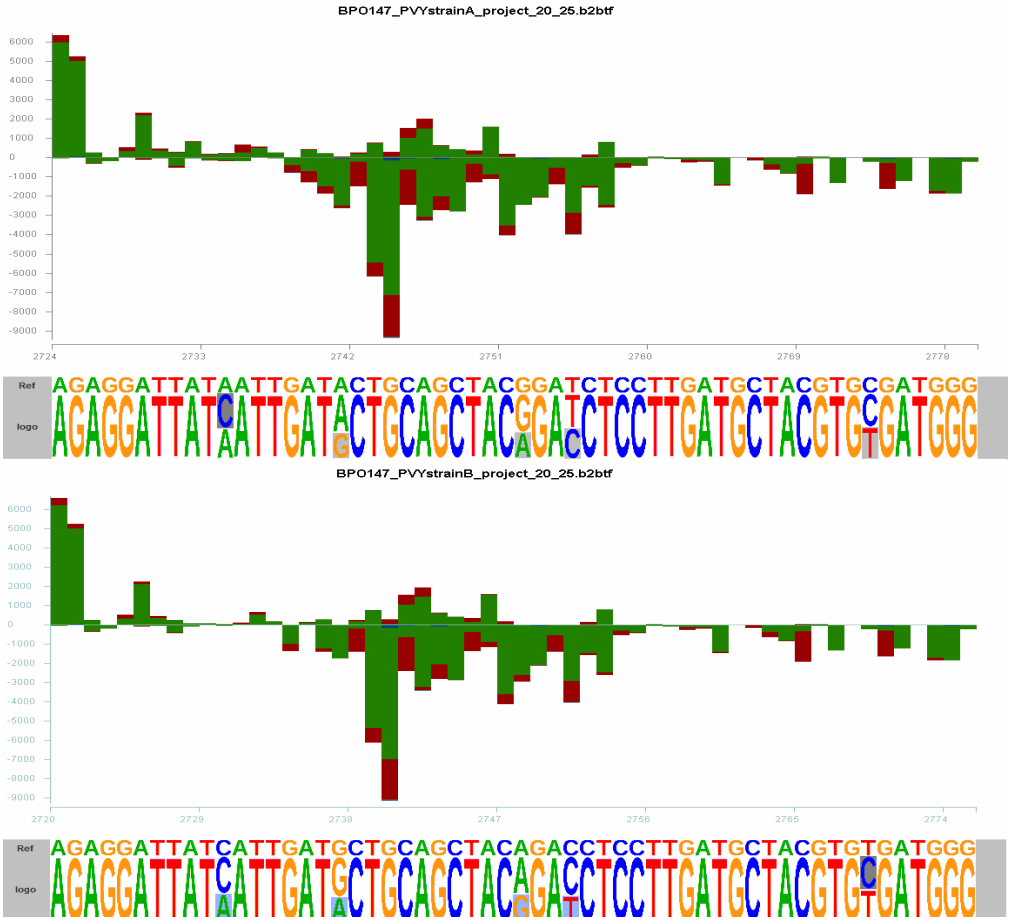


Fig. 33. Representative SNPs with two major nucleotides derived from PVY-A^N and PVY-B^O . The 20-25 nt sRNAs from BPO-147 dataset were mapped to the reconstructed reference sequence of PVY-A^N (upper image) and PVY-B^O (lower image) and visualized by MISIS-2 (Seguin et al. 2016).

It should be noted that *de novo* reconstruction of the complete genome sequences of PVY-A^N and PVY-B^O without the reference sequences of PVY-NTN and PVY-O would have been impossible, given high nucleotide identity throughout the genome sequences, except for 5'-

terminal parts. The latter parts share only 74% identity, allowing the short read assembler Velvet/Oases to generate separate, non-chimeric sRNA contigs. In contrast, the higher nucleotide identities (87 to 95%) in the middle and the 3'-terminal portions of PVY-AN and PVY-BO prevented separation of Oases contigs by Seqman, thus creating a single chimeric contig (Fig. 34). Given the co-existence of PVY-A^N and PVY-B^O in one plant virome, we cannot exclude potential recombination events leading to swaps of short or long sequences between the two strains, further complicating reliable separation of the two consensus sequences. Our attempts to physically separate PVY-A^N and PVY-B^O in vegetative progeny of the infected plant failed, because both strains were detected by RNA blot hybridization with strain-specific probes in seedlings grown from each of the four progeny tubers tested (data not shown). Thus, the virome composed of two PVY strains appears to be as stable as the PVY-PVX virome described above.

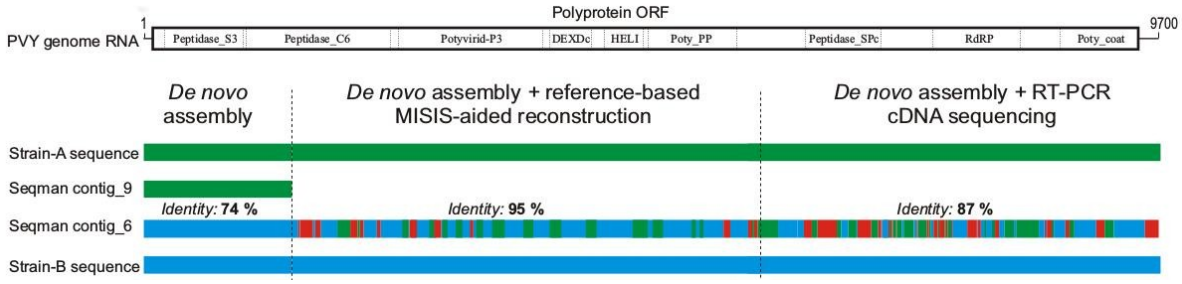


Fig. 34. Reconstruction of PVY twin strains. Seqman contig_6 covers the complete genome of PVY, but except for 5'-end which is similar to the PVY strain O, resulted to be a chimera. Contig_9 represents the 5'-end of a strain similar to PVY strain NTN. Those two regions share 74% of identity. The middle part of the two genomes (sharing 95% identity) was reconstructed with the help of MISIS, whereas the last 4 kb were disentangled by an RT-PCR based method.

To evaluate RNA silencing in potato plants and the response to PVY twins, we mapped sRNAs with zero mismatches to the host reference genome (complete or separated in superscaffold, mitochondrial, chloroplastic, and unanchored) and to the viral reconstructed reference sequences of PVY-A^N and PVY-B^O either in combination, allowing 50-50% distribution of identical sRNAs between the strains, or separately, allowing identical sRNAs from both strains to be co-mapped on each strain. The healthy control plant (BPO-148) showed only plant-derived small RNAs (Fig. 35a), with the majority of the reads coming from the superscaffold and unanchored genome and only less than 10% coming from chloroplast and mitochondrial DNA (Fig. 35b). The same pattern of siRNAs distribution was observed in the PVY-infected plant (BPO-147), even though the percentage of host-derived siRNAs is reduced (38%) and the sum of viral reads from both strains together represents 62.5% of the total 20-25 nt reads (Fig. 35a, 35b).

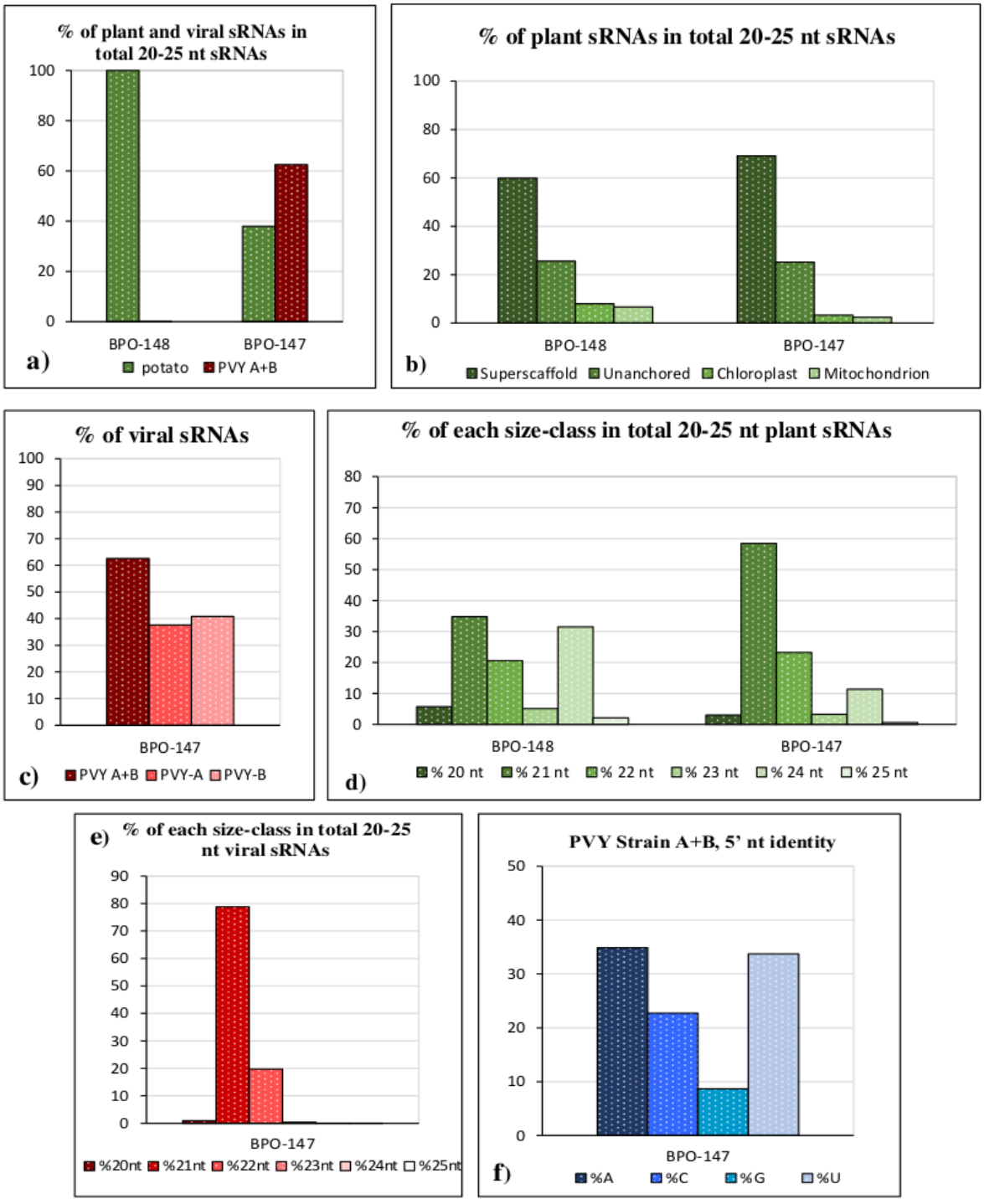


Fig. 35. The 20-25 nt sRNA libraries from healthy (BPO-148) and or PVY-A^N+PVY-B^O twin (BPO-147) virome-infected plants were mapped to the combined virome or each individual viral reference sequences and the plant genome reference sequence with zero mismatches and counted. The graphs show the percentage (%) of virome- and plant-derived sRNAs in the pool of total 20-25 nt reads (a), % each individual host-derived sRNAs (b), % of each virus-derived sRNAs combined and separated (c), % of size class in the 20-25 nt pool of plant-derived (d) or virus-derived (e) sRNA reads, % of nucleotide identity at the sRNA 5' end (f).

The identical sRNAs from both PVY-A^N and PVY-B^O constituted ca. 25% of all viral sRNAs, which elevated the counts of viral sRNAs mapped separately to PVY-A^N (37.6% of total) or PVY-B^O (40.8% of total). This shows a substantial cross-targeting potential of viral siRNAs

derived from PVY-A^N against PVY-B^O, and vice versa (Fig. 35c).
In depth analysis of viral sRNA size profiles, 5'-nt frequencies and hotspot maps revealed that, in the healthy plant, the potato RNA silencing machinery generates 21-nt, 22-nt and 24-nt siRNAs, with the former being the most abundant (35%, 21% and 32%, respectively) and enriched at the 5' nt in uracil and adenine for 21- and 22-nt and 5' adenine for 24-nt sRNAs, which implies possible association of 21-22 nts host sRNAs with potato AGO1 (5'U) and AGO2 (5'A) orthologs, respectively, and of 24-nt host sRNAs with a potato AGO4 (5'A) ortholog (Fig. 35d, 5' identity not shown).
In the PVY-infected sample, our sRNA analysis implies presumptive host DCL4- and DCL2-mediated biogenesis of viral 21-nt and 22-nt siRNAs (58% and 23%, respectively) from perfect dsRNA precursors covering the entire PVY genome (Fig. 36) followed by association with at least three AGOs (AGO1, AGO2 and AGO5). Interestingly, the 24-nt host siRNAs are reduced to 11%, compared to the non-infected sample (Fig. 35d, 35e), suggesting virus-mediated regulation of the plant sRNA biogenesis and/or function.

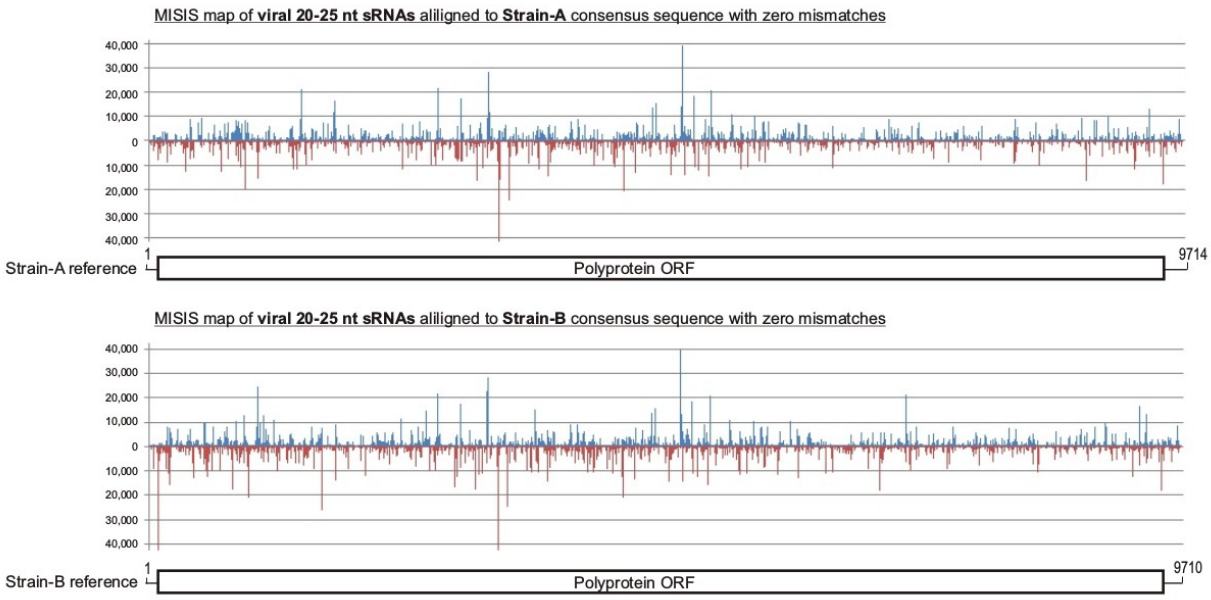


Fig. 36. Single-nucleotide resolution maps of viral sRNAs from two different strain of PVY. The histograms plot the numbers of total 20-25-nt viral sRNA reads at each nucleotide position of 9714 nt PVY-A^N and 9714 nt PVY-B^O (mapped with zero mismatches). The bars above the axis represent sense reads starting at each position and those below represent antisense reads ending at the respective position. PVY genome diagrams are shown above the histograms, with the polyprotein ORF boxed and nucleotide positions indicated.

Our findings that (i) the plant silencing machinery almost equally targets each component of PVY twins generating massive quantities of viral 21-22 nt siRNAs from each strain, and (ii) that the cross-targeting potential of viral siRNAs does not seem to affect the virome integrity or stability in vegetative progeny, suggest that PVY-A^N and PVY-B^O interact with each other neutrally (or synergistically) rather than antagonistically. In a previous study, co-inoculation of potato and tobacco plants with different isolates of three PVY strains have revealed both antagonistic and neutral interactions, depending on a strain/isolate combination and a host

plant (Syller and Grupa 2014). The interactions between PVY-NTN and PVY-O were usually antagonistic, resulting in suppression of PVY-O accumulation by PVX-NTN, compared to single infections (Syller and Grupa, 2014). Nonetheless, it cannot be excluded that dual infections with certain genetic variants of these strains can, under natural conditions, form a stable virome, like the PVY twin virome described here.

3.2 siRomics applied to *Solanum lycopersicum*

3.2.1 Interactions between mild-CH2 and LP strains of *Pepino mosaic virus* (PepMV) in *Solanum lycopersicum*

In the frame of a cross-protection trial at Agroscope in 2014, young plants of tomato cultivar Merlice grown in a production greenhouse of 150 m^2 with hydroponic cultivation (Conthey, Switzerland) were spray-inoculated (on January 23) with a mild isolate of PepMV-CH2 (Hanssen et al. 2010; De Nayer et al. 2011), the protecting strain marketed under tradename PMV-01 (Sciencia Terrae, B). At about the same time in another greenhouse (Nyon, Switzerland), young plants of tomato cultivar Mt Favet were inoculated with an isolate of PepMV-LP, the challenging strain which has originated from a Merlice plant collected in Champagny, Switzerland in October 2013 and is designated here LP based on its genome sequence reconstruction (see below). In about two months (on March 07) four Mt Favet plants infected with LP were transferred from Nyon to Conthey, where they were co-cultivated with CH2-preinfected Merlice plants in order to evaluate a protective effect of CH2 against LP. The co-cultivation procedure involved deliberate mechanical cross-contamination between all the greenhouse plants using leaf cutting and fruit harvesting tools, thus allowing repetitive inoculations of CH2-infected plants by sap from LP-infected plants, and vice versa. Upper leaves of two selected Merlice plants pre-infected with CH2 and challenge-inoculated with LP were sampled at two time points, 1.5 months (April 17) and 6.5 months (September 30) post-introduction of LP. Likewise, upper leaves of the two Mt Favet plants pre-infected with LP and challenge-inoculated with CH2 were sampled at two time points (July 18 and September 30). Total RNA was extracted from a total of eight leaf samples of the infected plants as well as from upper leaf samples of two control non-treated Mt Favet plants (harvested in February), and analysed by sRNA-omics, followed by Northern blot hybridization.
Illumina sequencing of all cDNA libraries in one lane of HiSeq2500 yielded 20.2 to 31.5 M reads per library (Table 8). The 20-25 nt reads were taken for bioinformatic analysis as described above for the potato libraries. *De novo* virome reconstruction revealed the presence of CH2 and absence of LP in both plants pre-infected with CH2 at the 1st time point, indicating CH2-mediated cross-protection during 1.5 months of repetitive challenge-inoculation with LP.

Table 8: Total number of sRNA reads per library from tomato plants

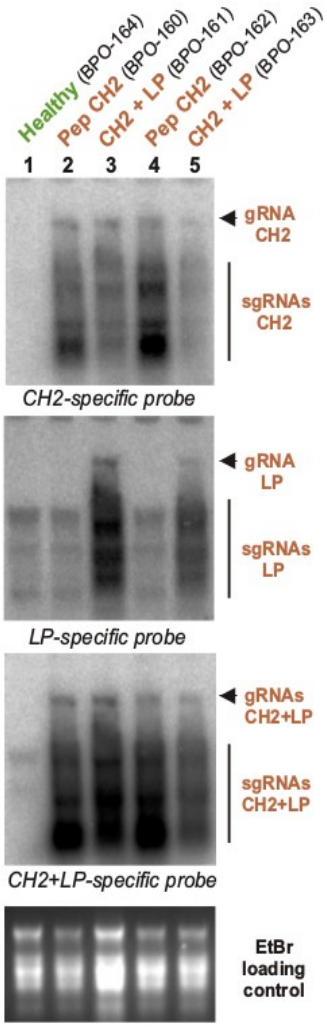
sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
BPO-156	29,211,780	24,661,960
BPO-157	24,492,857	20,429,252
BPO-158	26,995,552	22,249,057
BPO-159	26,544,222	18,497,067
BPO-160	31,559,078	25,894,816
BPO-161	26,699,567	17,930,709
BPO-162	27,676,321	22,291,182
BPO-163	25,012,942	19,753,441
BPO-164	20,274,979	16,468,754
BPO-165	20,744,441	16,916,887

However, at the 2nd time point, both plants were found to be co-infected with CH2 and LP, indicating a breach in cross-protection at some point during the next 5 months of repetitive challenge-inoculation with LP. Likewise, virome reconstruction revealed the presence of both LP and CH2 in both plants pre-infected with LP after 4.5 months of repetitive challenge-inoculation with CH2 as well as at the later time point (6.5 months). Northern blot hybridization analysis confirmed the sRNA-omics findings and further revealed that LP invasion had reduced the accumulation levels of gRNA and, more pronouncedly, sgRNAs of CH2 (Fig. 37; CH2-specific probe, lanes 1 and 3 vs 2 and 4) suggesting an antagonistic interactions between CH2 and LP in mixed infections. In previous studies, competition between PepMV strains in mixed infections did not appear to affect final viral loads, compared to control single infections (Gómez et al. 2009; Hanssen et al. 2010), although sgRNAs were not analysed.

De novo reconstructed consensus sequences of CH2 from the two single infections (BPO160 and BPO-162; deposited to Genbank as MF422611 and MF422613, respectively) differed at three nucleotide positions of the 6412 nt genome. MISIS-aided analysis revealed that, in the plant 2, the nucleotide substitutions A2041G, T3358C and T5224G are supported by 80-85% reads at the 1st time point (BPO-162) and by almost 100% reads at the 2nd time point (BPO-163, see Annex, Supplementary Fig.S2).

These three substitutions were absent in the plant 1 at the 1st (BPO-160) or 2nd (BPO-161) time points as well as in two plants pre-infected with LP at the 1st (BPO-156 and BPO-158) or 2nd (BPO-157 and BPO-159) time points. In one of the latter plants two addition nucleotide substitutions (G2875A and T3265C) were found, both supported by the majority of reads at the two time points (BPO-158 and BPO-159, see Supplementary Fig. S2 in the Annex).

Fig. 37. Blot hybridization analysis of viral RNA accumulation in *S. lycopersicum* plants infected with one (CH2) or two strains (CH2+LP) of PepMV. Five microgram aliquots of the total RNA samples from healthy (BPO-164) and virus-infected plants (BPO-160-163) pre-analysed by sRNA sequencing were separated on formaldehyde-containing agarose gel, stained with EtBr and blotted to a membrane. The membrane was successively hybridized with DNA oligonucleotide probes specific for the reconstructed reference sequences of CH2, LP or both (CH2+LP) strains and scanned. The scanned images after each hybridization are shown with the



probe indicated below and the positions of viral gRNA and sgRNAs indicated on the right. EtBr staining of the gel is shown as loading control.

These findings led us to conclude that in one of the two Merlice plants inoculated with CH2, the replicating virus has acquired the three mutations which started to dominate in viral quasispecies at the early time point and eventually outcompeted an original consensus genome sequence, as evident at the later time point. Only the original genetic variant of CH2 (MF422611) invaded both LP-infected Mt Favet plants, and then either stably persisted in one plant or acquired two mutations in another one. The on-going evolution of CH2 quasispecies is further illustrated by comparison of our BPO-160 variant (MF422611) derived from PMV-01 with the initial sequence of this isolate deposited in Genebank in 2011 (JN835466): CH2 has accumulated 16 SNPs throughout the genome sequence as well as a 1-nt deletion and a 1-nt insertion separated by 9 nts in the RdRP ORF, which together altered 7 amino acids in RdRP and one amino acid in each of the other proteins (TGB1, TGB2, TBG3 and CP).

PMV-01 is known to ensure only partial protection against non-CH2 strains of the virus (De Nayer et al. 2011). We confirm and extend this finding by showing that PMV-01/CH2 stops LP dissemination for at least 1.5 months, but repetitive challenge inoculation does break CH2-mediated cross-protection against LP. Furthermore, we demonstrate that PMV-01 quasispecies evolves on passages to new host plants through appearance and fixation of mutations. This on-going evolution may eventually result in a loss of protectiveness and/or an alteration of symptom severity. Indeed, mild (FJ457096) and aggressive (FJ457097) isolates of PepMV-CH2 differ only by 38 SNPs altering 9 amino acids (Hanssen et al. 2009), and two studies have shown that a single point mutation is sufficient to increase PepMV aggressiveness. In one study, a point mutation altering one amino acid in TGB3 at position 67 (K [AAA] to E [GAA]) could convert a mild PepMV pathotype into a necrotic one (Hasiów-Jaroszewska et al. 2011), while in another study single amino acid substitutions in CP at position 155 (E [GAA] to K [AAA]) or position 166 (D [GAT] to G [GGT]) resulted in severe yellowing symptoms (Hasiów-Jaroszewska et al. 2013). Our analysis the evolving CH2 quasispesies revealed the non-consensus nucleotides A and G at the two respective positions of CP, albeit in a tiny proportion of reads (6 of 30337 and 5 of 29537 total reads from the six CH2-infected plants), while no polymorphism was found at the corresponding nucleotide position of TGB3 in any CH2-infected plant (0 of 2725 total reads). The quasispecies nature of PepMV has previously been established based on analysis of PCR-amplified TGB1 sequences from

several naturally-infected tomato plants (Hasiów-Jaroszewska et al. 2010). Our genome-wide analysis of consensus sequence variation reveals appearance of SNPs and fixation of new consensus nucleotides in PepMV quasispecies within individual plants during 8 months' trial.

De novo reconstruction of the LP consensus sequences was somewhat complicated by the presence of CH2 in all the LP-infected plant samples and required MISIS-aided corrections (Annex, Supplementary Fig.S2) at SNP positions within several stretches of high similarity between LP and CH2 genome sequences (sharing 78% overall identity).

With this reservation, the LP consensus sequence reconstructed from two CH2-preinfected plants, where the CH2 consensus sequence is known from the earlier time point, differ only by one nucleotide: T4191 in BPO-161 (deposited to Genebank as MF422612) vs. C4191 in BPO-163 (deposited to Genebank as MF422614). In both plants pre-infected with LP, the LP consensus sequence appeared, with the above reservations, to be identical to that from BPO-163, suggesting that this LP variant invaded both Merlice plants pre-infected with CH2 and then acquired the point mutation in one of them (BPO-161). BLAST analysis revealed that the 6408 nt LP consensus sequence from BPO-163 (MF422614) is 99.8% identical (14 SNPs) to a sequence of PepMV isolate LP-2001 (AJ606361) obtained in Peru from an asymptomatic Peruvian tomato (*Solanum peruvianum* also known as *Lycopersicon peruvianum*) plant (López et al. 2005; hence the name LP for our strain). Interestingly, a nearly identical isolate SM.74 (AM109896) differing by 18 SNPs from the BPO-163 sequence has been obtained in Peru from a pepino (*Solanum muricatum*) plant (Pagán et al. 2006). Thus, PepMV-LP mutated upon arrival in Europe from Peru and eventually infected the tomato plant that was collected in Switzerland for our cross-protection trial.

To verify if our LP isolate undergoes quasispecies evolution in single infections we repeated (in 2016) inoculation tomato Mt Favet plants with sap from the same LP-infected leaf tissue as was used for the cross-protection trial and sequenced sRNAs from uppermost leaves of LP-infected (HYT-25) and mock-inoculated (HYT-26) plants. *De novo* reconstruction of the consensus genome sequence of LP from HYT25 (deposited to Genebank as MF422616) revealed 13 SNPs compared to that from BPO-163 (A138C, C1548T, A1609G, T1997C, A2272G, C2394A, T3138C, G3225A, T3486C, C4418T, G5861A, C5987T, T6254C), each supported by the majority of redundant and non-redundant sRNA reads. Interestingly, 7 of these 13 substituted nucleotides (C1548, T3138, G3225, T3486, G5861, C5987, T6254) are shared between BPO-163 and the Peruvian isolates LP-2001 and SM.74, indicating that the corresponding substitutions are new genetic variants fixed in viral quasispecies of the HYT-25 plant. On the other hand, the nucleotides C138, G1609, C1997, G2272, A2394, T4418 are shared between HYT-25 and both Peruvian isolates, indicating that those 6 nucleotides were substituted and fixed through independent evolution before pre-inoculation of the tomato plants (followed by stable transmission of this genetic variant into the CH2-infected plant BPO-163). Taken together, similar to CH2, the LP quasispecies undergoes continuous micro-evolution of the consensus genome sequence on passages in tomato plants cultivated under greenhouse conditions.

We then analysed the tomato silencing response to CH2 and LP by mapping 20-25 nt reads to the tomato genome and the reconstructed consensus sequences of CH2 and LP (individually and in combination) with zero mismatches, and counting the endogenous and viral sRNAs based on size, polarity and 5'-nucleotide identity. The proportion of all viral (CH2+LP)

sRNAs in a total (tomato+viral) 20-25 nt sRNA population varied from 1.6 to 4.7% (Fig. 38), thus being comparable to the proportions reported for tomato plants co-infected with EU and US1 strains of PepMV (Li et al. 2012) and also to the proportion of PVX-derived sRNAs in potato plants (Fig. 29). Notably, viral sRNA accumulation was ca. 2 to 2.5 times higher in the plants infected with CH2 alone (3.7-4.7%), compared to the plants co-infected with CH2 and LP (1.6-1.8%), suggesting that LP may suppress siRNA production from CH2. In all co-infected plants, the levels of CH2-derived sRNAs somewhat exceeded those of LP-derived sRNAs, although the difference was less pronounced in the plants pre-infected with LP (Fig. 39).

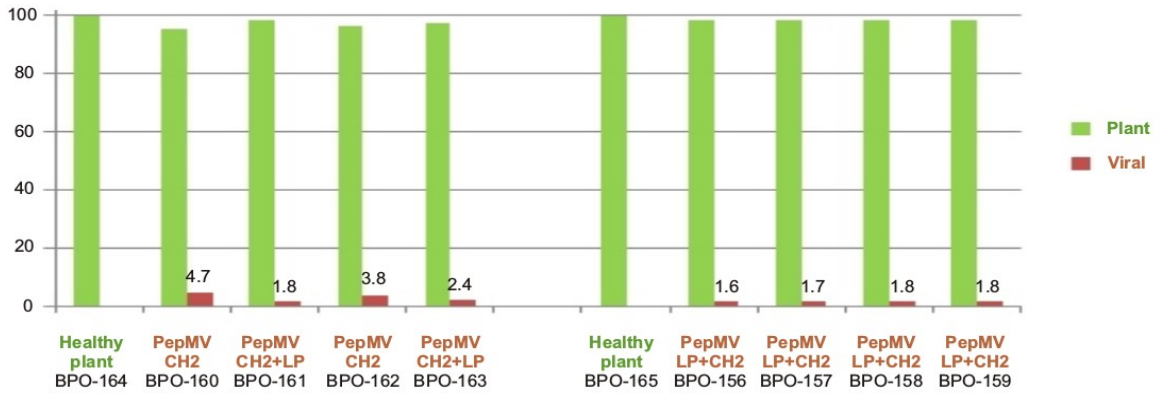


Fig. 38. Percentage of plant and viral sRNAs in total 20-25 nt sRNAs

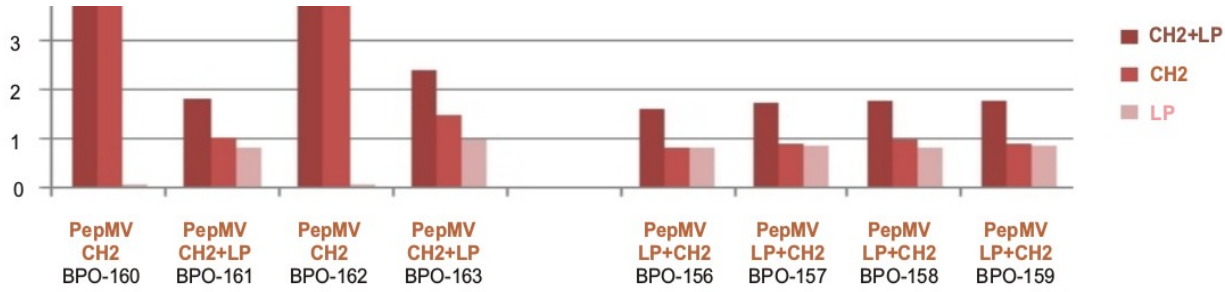


Fig. 39. Percentage of viral sRNAs mapped to both and each of the two viral genomes with zero mismatches

Comparison of size-class profiles revealed a striking difference between single (CH2) and dual (CH2+LP) infections: while 21-nt class is dominant in all cases for both CH2 and LP, 22-nt class of CH2-derived sRNAs is much more prominent in single infections (Fig. 40). This indicates that LP may interferes with production of 22-nt siRNAs from CH2, which contributes to an overall reduction of viral sRNA accumulation in the plants pre-infected with CH2 upon invasion of LP (Fig. 38) and correlates with reduced levels of CH2 genomic and subgenomic RNAs.

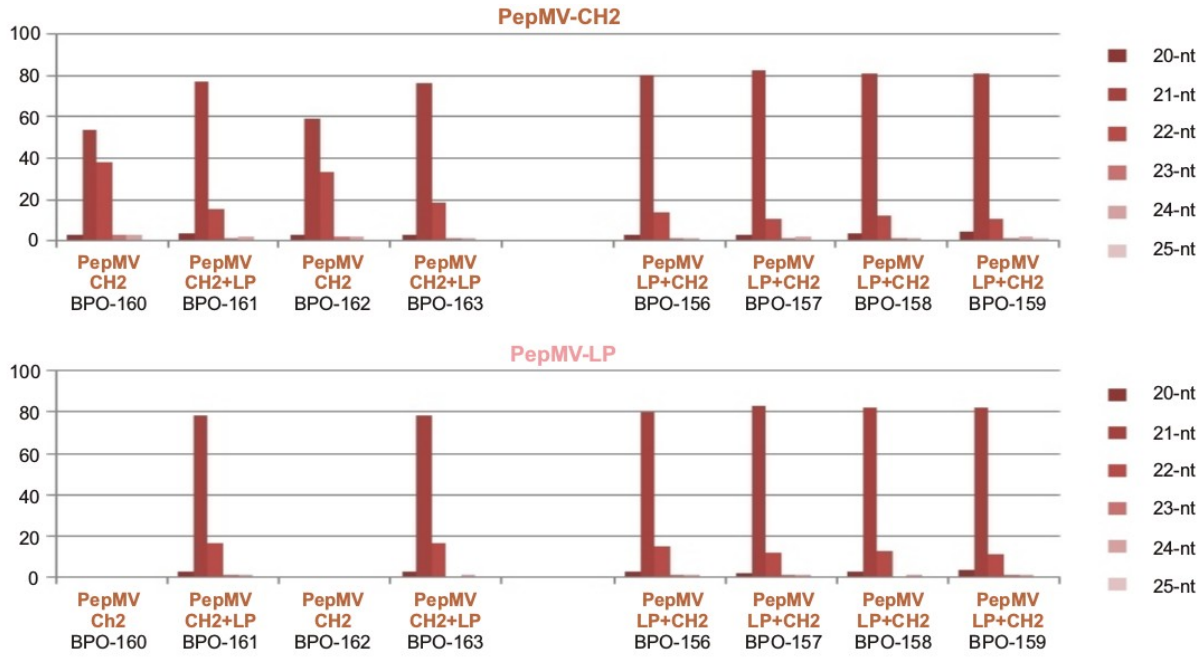


Fig. 40. Percentage of each size-class in total 20-25 nt viral sRNAs. The 20-25 nt sRNA libraries from two healthy (BPO-164-163), two CH2-preinfected (BPO-160-161 and BPO-162-163) and two LP-preinfected (BPO-156-157 and BPO-158-159) plants were mapped to the combined (CH2+LP) virome or individual viral genome sequences and the plant genome reference sequence with zero mismatches and counted. The graphs show the percentage (%) of each size class in the 20-25 nt pool of each virus-derived sRNA reads.

Since 22-nt siRNA biogenesis is likely mediated by DCL2 activity, an LP variant of PepMV-encoded silencing suppressor may target one (or more) of the four tomato DCL2 paralogs (Bai et al. 2012), while its CH2 counterpart may fail to do so. Analysis of sRNAs from the tomato plant infected with LP alone (HYT-25, Fig. 41), confirmed the results for dual infections in that LP-derived sRNAs were dominated by 21-nt class and constituted a smaller proportion in total 20-25 nt sRNAs (0.6%) than CH2-derived sRNAs.

Analysis of 5'-nt frequencies revealed a substantial bias to 5'U and 5'A in both 21-nt and 22-nt viral sRNAs derived from both CH2 and LP in dual infections, while this bias was less pronounced in single infections due to higher frequencies of 5'C. Taken together, PepMV-derived 21-22-nt sRNAs appear to be sorted by tomato AGO1, AGO2 and AGO5, similar to PVX-derived sRNAs in potato plants. Analysis of size profiles and 5'-nt frequencies of endogenous tomato sRNAs further confirmed a differential impact of single (CH2) and dual (CH2+LP) infections of the plant RNA silencing machinery, although the size profile alterations are not always consistent between biological replicates. Notably, 21-nt 5'U-sRNAs which dominate in healthy plants (79%) and likely represent DCL1-dependent AGO1-associated miRNAs (Kravchik et al. 2014b), are less dominant in all infected plants (62-73%). Likewise, 24-nt 5'A-sRNA which dominate in healthy plants (73%) and likely represent DCL3-dependent, AGO4-associated heterochromatic and phased sRNAs (Kravchik et al. 2014a), become less dominant (61-63%) in infected plants. Biological significance of the impact of PepMV infections on endogenous sRNA remains to be investigated.

Analysis of single-nucleotide resolutions of CH2- and LP-derived sRNAs revealed similarities

to those of PVX-derived sRNAs in that both 21-nt and 22-nt viral sRNA species cover the entire viral genome in sense and antisense orientation and that the hotspots generally map to the 5'-portion of RdRP ORF. A notable difference between CH2 and LP is that a sgRNA region (TGB-CP ORFs) of the CH2 genome spawns relatively more abundant sRNAs than that of the LP (or PVX) genome. Furthermore, sRNA hotspot patterns appear to be unique for CH2 and LP and remarkably reproducible in all the plants at both time points (see Supplementary Fig.S3 in Annex), indicating sequence-specificity of the viral siRNA-generating machinery targeting LP and CH2. Similar observations have been reported for two distinct DNA begomoviruses in tomato plants (Fuentes et al. 2016).

To assess a cross-targeting potential of CH2-derived sRNAs against LP, we mapped 20-25 nt sRNA reads from CH2-infected plants at the 1st time point, when no LP was detectable (BPO-160 and BPO-162), to the LP genome. A significant proportion of CH2-derived sRNAs (1.2 and 1.4 %, respectively) mapped to the LP genome with zero mismatches. The majority of these identical sRNAs concentrate in five hotspots of 21-22 nt sRNAs of both polarities across the LP genome (Fig.S3 in Annex, BPO-160 and BPO-162), which contain continuous stretches of more than 20 nts with 100% identity between CH2 and LP sequences. These 21nt and 22-nt siRNAs derived from CH2 can potentially target LP gRNA, gRNA(-) and sgRNAs for cleavage and degradation. Such cross-targeting might have contributed to CH2-mediated cross-protection against LP for at least 1.5 months. However, a relatively small quantity of the cross-targeting sRNAs and/or a relatively short length of the cross-targeted sites in the LP genome may account for the failure of CH2-infected plants to resist repetitive challenge-inoculation with LP for a longer term. Consistent with our findings, mixed infections of distinct PepMV strains have been observed in a previous cross-protection trial (Hanssen et al. 2010) as well as under natural conditions (Gomes et al. 2009), indicating that PepMV strains can form a stable virome. On the other hand, a previous trial has suggested that the protecting effect of a mild variant of CH2 (PVM-01) against an aggressive variant of CH2 could be based on RNA silencing, because of 99.4% sequence identity (Hanssen et al. 2010). However, our above described evidence for the PVY twins indicates that even a very high cross-targeting potential for viral sRNAs does not prevent the co-existence of two viral strains in a stable virome.

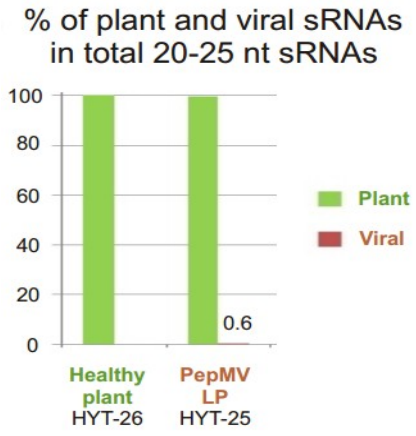


Fig. 41. Counts and single-nucleotide resolution maps of viral sRNAs in *S. lycopersicum* infected with PepMV-LP. The 20-25 nt sRNA libraries from mock-inoculated (HYT-26) and LP-infected (HYT-25) plants were mapped to the plant and viral genome reference sequences with zero mismatches and counted. The graphs show the percentage (%) of virus- and plant-derived sRNAs in the pool of total 20-25 nt reads.

3.2.2 Identification and reconstruction of *Southern tomato virus* by siRomics

In addition to CH2 and LP strains of PepMV, *de novo* virome reconstruction revealed the presence of *Southern tomato virus* (STV) in the two Merlice tomato plants pre-infected with CH2 at both time points. Interestingly, STV was not detected in any of the Mt Favet plants (BPO-160-163 vs BPO-156-159), suggesting that it had persisted in the Merlice plants and was not transmitted to the Mt Favet plants during their co-cultivation. STV belongs to the family *Amalgaviridae*, has a dsRNA genome of ca. 3.4 kb with two overlapping ORFs, and is transmitted vertically via seeds, but not mechanically (Sabanadzovic et al. 2009). The latter observation is consistent with our finding that STV was not co-transmitted with the sap together with PepMV. The consensus sequences of our STV isolates from two plants differ at one nucleotide position of the 3438 nt genome, G1490 in the plant 1 (BPO-160/BPO-161; deposited to Genbank as MF422617) and A1490 in the plant 2 (BPO-162/BPO-163; deposited to Genbank as MF422618), and are 99.3% identical (24 and 23 SNPs, respectively) to the STV isolate CN-12 (KT438549) identified in tomato plants in China in 2012 (Padmanabhan et al. 2015). Interestingly, the Chinese isolate CN-12 has exhibited severe disease symptoms, while our Swiss genetic variants of STV appeared to be symptomless.

Accumulation of STV-derived sRNAs was very low in both plants at the 1st time point in the presence of CH2, while it was strongly elevated at the 2nd time point in the presence of CH2 and LP . This suggests that LP co-infection might have activated replication of STV, which would in turn boost the silencing response to STV. Moreover, the size profile of STV sRNAs was altered dramatically in that 21-nt class being the second most abundant in the presence of CH2 alone became the first most abundant, outnumbering the 22-nt class, in the presence of both CH2 and LP . Taking into consideration the negative effect of LP on production of 22-nt sRNAs from CH2 (as described above), LP (but not CH2) appears to directly interfere with DCL2 activity generating 22-nt viral sRNAs, which would indirectly increase the proportion of 21-nt siRNAs generated by DCL4 from STV and CH2. Comparison of the hotspot distribution for 21-nt and 22-nt sRNAs suggests that DCL4 may preferentially target the middle portion, while DCL2 may prefer targeting the termini of the STV genome (Fig. S3 in Annex).

3.3 siRomics applied to sweet cherry

In a field experiment at Agroscope (Nyon, Switzerland), sweet cherry (*Prunus avium*) trees were graft inoculated with *Little cherry virus 1 (*LChV-1) alone or in combination with *Little cherry virus 2 (*LChV-2) to evaluate the impact of these two closteroviruses on tree physiology and fruit development.

To further support the phenotypic field trial with molecular evidence, we used the siRomics pipeline to reconstruct the consensus viral sequences and evaluate any possible interaction between the two viruses and their impact on the plant antiviral defense based RNA silencing .

Total RNA was extracted from leaves of the cherry trees inoculated with LChV-1 (HYT-15) or a combination of LChV-1 and LChV-2 (HYT-14) and of a healthy control cherry tree (HYT-13). Five µg of total RNA were separated on 1% formaldehyde containing 1% agarose gel,

followed by EtBr staining to validate long RNA quality and on 15% polyacrylamide gel, followed by blot hybridization with miR160-specific probe to validate sRNA integrity. Following the quality control, the total RNA samples were sent to Fasteris for Illumina sequencing of 19-30 nt sRNAs.
The sequencing output was from 11 to 17 M reads per library, with the majority of the reads being between 20 and 25 nts (Table 9).

Table 9: Total number of reads per sRNA library from cherry tree leaves

sRNA library	Sample inoculation	Total number of 1-44 nt reads	Total number of 20-25 nt reads
HYT-13	none	13,380,789	12,248,222
HYT-14	LChV1 + LChV2	17,355,842	14,985,555
HYT-15	LChV1	11,785,847	10,345,927

Following the standard siRomics strategy, redundant and non redundant 20-25 nt reads were *de novo* assembled by Velvet and Oases, with the k-mers set from 11 to 23 and the minimum contig length of 50 and 100 nts. At the time when this analysis was performed for the first time, the sweet cherry (*Prunus avium*) genome was not sequenced yet and the contigs derived from all the k-mer sizes, concatenated in one single file, were filtered through the genome of peach (*Prunus persica)*, which is closely related to sweat cherry. Following the publication of the sweet cherry genome by Shirasawa et al. (2017) the filtering step was repeated. In both cases, the filtered unmapped Oases contigs were further assembled by Seqman and analyzed with NCBI blastn for viral similarity.
As expected, no viral sRNA contigs were found in the asymptomatic sample (HYT-13). In the sample inoculated with two closteroviruses (HYT-14), blastn analysis of the redundant Seqman contigs, confirmed the presence of LChV-1 (58 contigs) and LChV-2 (47 contigs) but also revealed the presence of *Cherry virus A* (CVA, 2 contigs). The number of viral contigs was higher for the cut-off length min50 than min100 (240 compare to 128 contigs). However, mapping of both Oases and Seqman contigs on the closest reference genomes (KR736335.1, AF531505.1 and KU131205.1 respectively) and visualization of the maps using IGV showed that the coverage was incomplete, especially for CVA (Fig.42), as was also evident from Samtools depth-coverage analysis (data not shown). Furthermore, the number of virus-related contigs and the genome coverage were comparable between the cherry and peach filtered contigs (data not shown).

Fig. 42. IGV-visualized maps of Oases min50 contigs along the reference genomes of LChV-1, LchV-2 and CVA**.** The IGV snapshots show the Oases contigs (grey arrow-bars) mapped along the corresponding references (top, LChV-1; middle, LChV-2; bottom, CVA) with the arrow pointing in the forward or reverse orientation.

Because of the incomplete coverage with sRNA contigs, a reference-based approach was attempted to reconstruct the consensus genome sequences of the three different viruses, by mapping redundant 20-25 nt sRNA reads on the reference sequences of the LChV-1, LChV-2 and CVA isolates best matching to the Oases/Seqman contigs.

Six iteration steps were needed for LChV-1 and LChV-2 and five iterations for CVA, to reach the point where the consensus sequence was 100% identical to the one from the previous iteration, with no gain in the number of mapped reads. Consistent with a previous report regarding iterative read mapping and assembly (Dutilh et al. 2009), after each iteration the total number of mapped sRNA reads increased due to a new consensus becoming more similar to the one supported by the majority of the reads. This is shown by the MISIS-2 single base resolution map of the reads mapped on the reference used for the first iteration and the last consensus sequence reached at the end (Fig. 43, Fig. 44 and Fig.45).

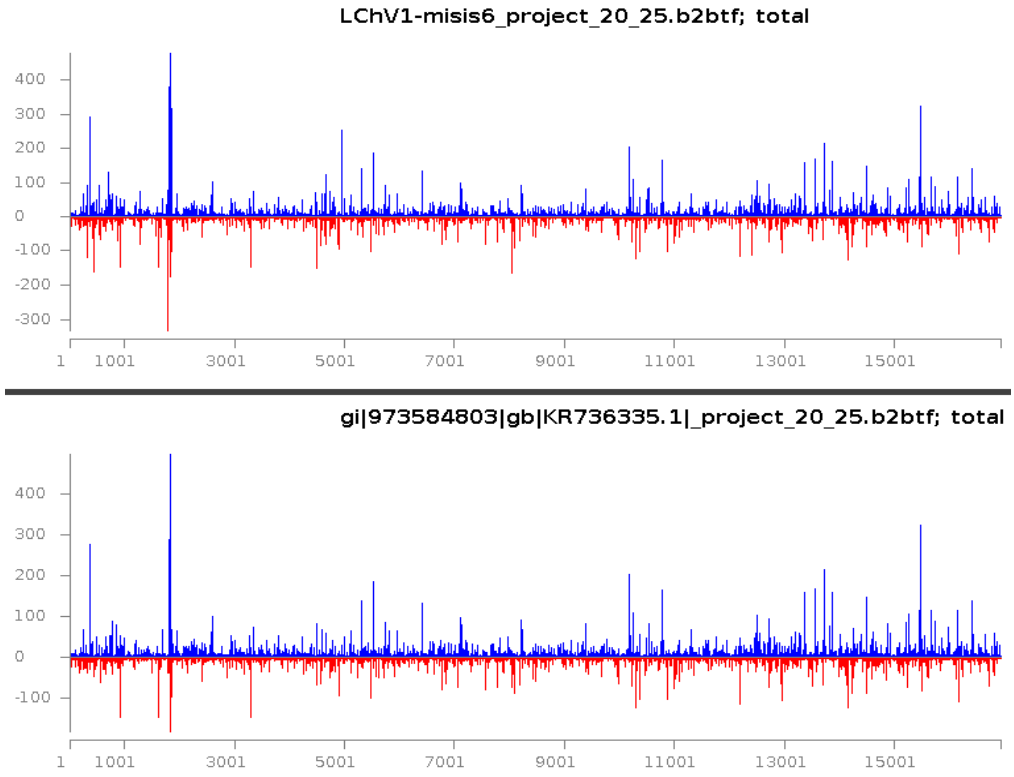


Fig. 43. Single base resolution map of redundant sRNA reads mapped on the reconstructed consensus sequence (top) and on the reference sequence of closely related isolate of LChV-1 (bottom). The x-axis represents the LChV1 genome. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

The 16932 nt consensus genome sequence of LChV-1 was fully supported by non redundant sRNA reads, with few SNPs, where there is an even number of reads supporting either one or the other nucleotide, and resulted to be 97% identical to the Taian isolate (KR736335.1) used as a reference for the first mapping step.

The ORF finder revealed that the thymine at position 4831 (supported by 26 redundant sRNA reads) would terminate the open reading frame 1a encoding 260 kDa protein with helicase, methyltransferase and papain-like protease domains (Wang et al., 2016). To restore the frame. T4831 was changed to G supported by 9 redundant reads and 7 of 16 non redundant reads and also present in the Taian isolate sequence

The 15045 nt reconstructed genome sequence of LChV-2 was 92% identical to the USA6b isolate used as the reference (AF531505.1), by had two gaps in sRNA coverage: the first spanning positions 3814 to 3819 and the second spanning positions 9548 to 9559 and it.

The 7434 nt consensus genome sequence of CVA showed multiple small gaps of 6 to 10 nucleotides along the genome, which were not covered with sRNAs (not shown), and resulted to be 96% identical to the closest isolates present on the NCBI database, with a complete 5' UTR composed of 107 nt starting with G (Koinuma et al., 2016) and two ORFs, ORF1 and ORF2, as expected (Jelkmann, 1995).

Per each reconstructed viral sequence, the percentage of the genome coverage is listed in Table 10.

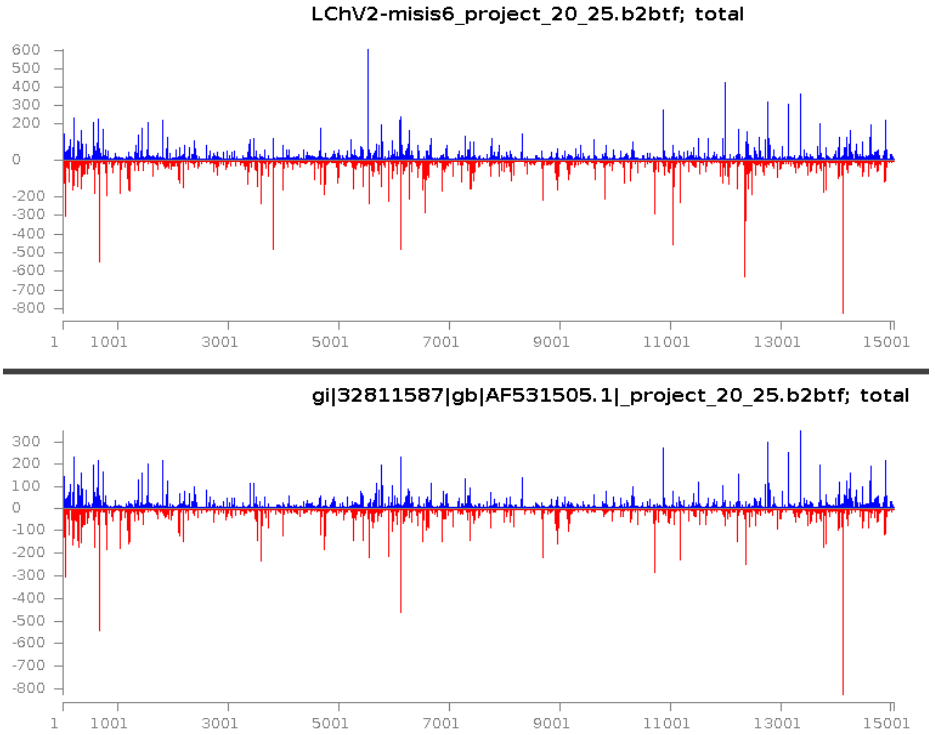


Fig.44. Single base resolution map of redundant 20-25 nt sRNA reads mapped on the reconstructed consensus (top) and the original LChV-2 reference sequences. The x-axis represents the LChV-2 genome. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

Fig.45. Single base resolution map of 20-25 nt sRNA reads mapped on the reconstructed (top) and the closely related reference genome of CVA.

The x-axis represents the CVA genome. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

Table 10: The percentage of reconstructed viral genome sequence covered with sRNA reads or sRNA contigs in HYT-14

Viral Sequence	Reads	Oases-all contigs	Seqman contigs
LchV-1	100	60	43
LchV-2	97.6	84	66
CVA	99.5	14	6

With the aim to unravel any additional components of the virome in sample HYT-14, a second strategy was also applied, consisting of a cascade of filtering steps starting with the mapping of the sRNA reads on the peach genome, in order to remove the host-derived reads, followed by filtering the unmapped reads throughout the reconstructed reference viral genome, starting with LChV-1, followed by LChV-2 and CVA. The unfiltered unmapped reads were de novo assembled with Velvet and Oases (K-mer 17) and the resulting Oases contigs were inspected with blastn. At this step, no other contigs showed sequence similarity to CVA but instead many contigs were still related to LChV-1 and LChV-2, suggesting that more than one genetic variant (strain) of these two viruses could be present in the virome.

After a filtering step for one virus, the reference based iterations were repeated to reach the consensus sequence for the other two viruses. The consensus sequences obtained with the second strategy differred from the first ones. Even though the three viruses share a percentage of similarity below 30%, some short reads could be shared among them and cut off during the BWA filtering steps that allowed up to two mismatches per 20-25 nt sRNA species, thus explaining the differences in the consensus sequences. This suggests that a complete data set of short reads should be used in order to have the consensus sequence supported by the majority of sRNA reads, like it was performed in the first strategy. Furthermore, the reconstructed consensus sequences should be taken as representative of the virome in the samples and supported by the majority of the reads, since evidence of more than one isolate per virus are shown in the MISIS-2 logo, with SNPs through the whole genome. This could be the reason why the assembled contigs are relatively short, many iteration steps were needed and many viral contigs were still present after the filtering cascade in the second strategy.

In the leaf sample HYT-15 from the cherry tree inoculated with only one closterovirus, among the redundant Seqman contigs with both min50 and min100 cut-offs, 13 were closely related to LChV-1. However, as viral sRNA levels were very low, the coverage resulted to be drastically reduced when the reads were assembled in contigs (21% for min50, 11% for min100).

Fourteen iteration steps were needed to reach the point where the consensus was 100% identical to the last reference. Even if the genome coverage increased from 93% to 96%, as also found by Misis logo analysis, many gaps are present along the genome, meaning that a deeper sequencing of sRNAs is needed to reconstruct completely the consensus LChV-1 sequence for this sample. This evidence was further supported by the MISIS-2 single base resolution map of non redundant reads, where a maximal numbers of reads per nucleotide position was below ten (data not shown).

To evaluate the response of the plant RNA silencing machinery to the two closteroviruses LChV-1 and LChV-2 and the capillovirus CVA, the sRNAs were mapped with zero mismatches to the sweet cherry reference genome and to the viral reconstructed reference sequences of LChV-1, LChV-2 and CVA and the numbers of viral and host reads were counted and sorted based or size, polarity and 5'-terminal nucleotide identity. In both datasets, the overall percentage of viral reads in total 20-25 nt reads resulted to be 1% , while the sweet cherry genome-derived sRNAs represented ca. 60% of total 20-25 nt reads in each dataset (data not shown). The remaining ca 39% reads likely derive from unsequenced or unassembled repetitve regions of in the cherry genome. The mapped cherry genome-derived sRNAs predominantly belonged to 21-nt and 24-nt size classes (Fig. 46). This is consistent with the observation that 21-nt and 24-nt size classes are the most abundant in peach (Zhang C. et al., 2016). The 21-nt viral sRNAs are the most abundant in each of the three viruses identified in HYT-14 and one virus in HYT-15, followed by the 22-nt sRNAs in each cases (Fig.47). This suggest that the three different viruses trigger similar antiviral silencing responses mediated by the sweet cherry paralogs of DCL4 and DCL2, respectively, by analogy with positive-sense RNA viruses in *Arabidopsis thaliana* which are targeted by DCL4 and DCL2 generating 21-nt and 22-nt viral siRNA (Bouche, 2006; Blevins et al. 2006). This is further supported by the similar 5' nucleotide frequencies in viral siRNAs, with the most abundant being 5'U, followed by 5'A and 5'C for each of the two major size classes (21-nt and 22-nt), suggesting their association with three orthologs of Arabidopsis AGOs, AGO1 (5'U), AGO2 (5'A) and AGO5 (5'C) (Fig.48).
Finally, the siRomics results supported the ones coming from the field experiments, meaning that the plants infected with the two closteroviruses and the capillovirus CVA, showed more severe disease compared with the single infection.

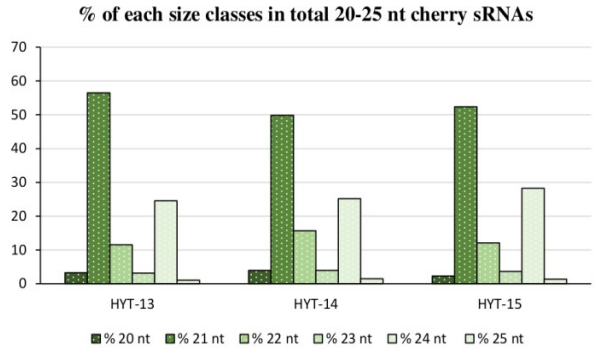
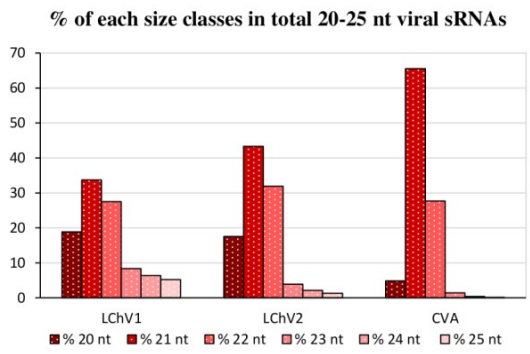


Fig.46. Counts of total endogenous 20-25 nt sRNAs in sweet cherry. The 20-25 nt sRNA libraries from the three datasets were mapped to the sweet cherry genome with zero mismatches and counted. The graph shows the percentage (%) of each size class in cherry-genome derived sRNA reads).



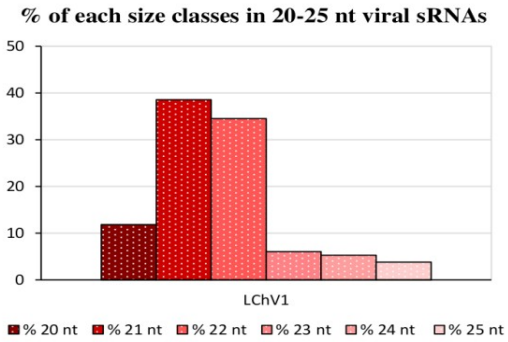


Fig.47. Counts of total 20-25 nt viral sRNAs in cherry samples HYT-14(left) or and HYT-15, (right).
The 20-25 nt sRNA libraries from the three datasets were mapped to each of the three reconstructed consensus sequences with zero mismatches and counted. The graph shows the percentage (%) of each size class in both samples.

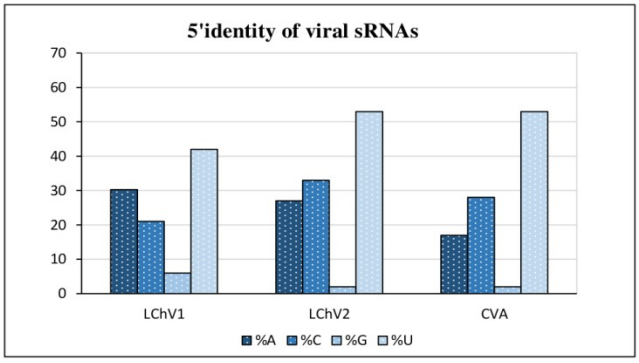


Fig.48. 5'-nucleotide identity (5'A,5'C,5'G,5'U) in sample HYT-14

3.4 siRomics applied to grapevine (*Vitis vinifera*)

Grapevine is prone to be infected by viral diseases, with more than 60 viral species identified (Martelli, 2014), leading to economic loss. The leaf samples from two different grapevine varieties showing viral disease-like symptoms (Pinot noir, HYT-17 and Otcha bala, HYT-18) together with a leaf sample from a healthy looking plant (Pinot noir, HYT-16) were obtained from Agroscope and analyzed by siRomics for virome reconstruction. The diseased plant (HYT-18) was previously analyzed by Reynard et al. (2015) through Virion-Associated Nucleic Acid approach (VANA, Filloux et al., 2015) and 454-pyrosequencing, followed by DNASTAR assembly, and the consensus sequence of the newly identified isolate of *Grapevine leafroll-associated virus 4* (GLRaV-4, KP313764.1) was used in this study as reference, with the aim to compare the two approaches for viral detection and reconstruction.
Total RNA was extracted from leaf samples as described in the Material and Methods. The quality of total RNA samples was evaluated as described in Material and Methods and the small RNA populations were sequenced in one lane of HiSeq2500 (Fasteris, Switzerland) yielding 7.8 to 12.6 M reads per sRNA library (Table 11).

Table 11: Total number of reads per library

sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
HYT-16	12,610,247	11,139,615
HYT-17	10,996,713	8,907,323
HYT-18	7,757,804	6,912,895

Following the siRomics pipeline, 20-25-nt redundant and non redundant sRNA reads from each of the three data sets were *de novo* assembled using Velvet and Oases (k-mers set from 11 to 23 and the minimum contig length of 50 nts). The contigs derived from all k-mers sizes were concatenated in one single file, filtered through the *Vitis vinifera* genome using BWA and Samtools and further assembled by Seqman. The blastn analysis of the Seqman contigs revealed that the asymptomatic Pinot noir sample used here as negative control (HYT-16), is infected with *Hop stunt viroid* (HSVd) and *Grapevine yellow speckle viroid 1* (GYSVd-1). The same viroids were also identified by blastn in the symptomatic sample of the same variety (HYT-17), together with the nepovirus *Grapevine fanleaf virus* (GFLV). Thus, GFLV is associated with the chlorosis of leaves in this case. Blastn analysis of the Seqman contigs from the Otcha bala sample (HYT-18) revealed the presence of three viroids including GYSVd-1, GYSVd-2, *Australian grapevine viroid* (AGVd), and two viruses the vitivirus *Grapevine virus A* (GVA) and the foveavirus *Grapevine rupestris stem pitting-associated virus* (GRSPaV). Surprisingly, no Seqman min50 contigs were related to the expected GLRaV-4 but its presence was confirmed by blastn analysis of the Oases min50 contigs.

3.4.1 Identification of viroids in the asymptomatic grapevine

Grapevine viroids belong to the group of pathogens which often do not induce symptoms in their host, with some notable exceptions reported by Szychowski et al. 1995. Thus, is not surprising that the asymptomatic sample HYT-17 used in this study as negative control, resulted positive to HSVd and GYSVd-1. The closest isolates of these viroids (AB219944.1 and KT000347.1, respectively) matching to the Oases/Seqman contigs were used as references to reconstruct the consensus viroid sequences. One Seqman and four Oases contigs mapped on the HSVd reference without complete coverage. The reference-based reconstruction of HSVd using redundant sRNA reads led to a consensus sequence which was 100% identical to the reference sequence. Seven oases and two Seqman contigs mapped on the GYSVd-1 reference genome with one redundant Oases contig (contig_287) and one non redundant Oases contig (contig_409) covered the complete genome. By running a pairwise alignment and by visualizing with MISIS-2 the redundant sRNA reads mapped on both contigs, two different genetic variants of GASVd-1 were identified with SNPs at positions 126 and 127 (Fig. 49) and an indel of one nucleotide at position 169 (Fig.50). Thus, the reconstructed consensus sequences of the two GYSVd-1 genetic variants were renamed after the contigs in GYSVd-1_287 and GYSVd-1_409 respectively.

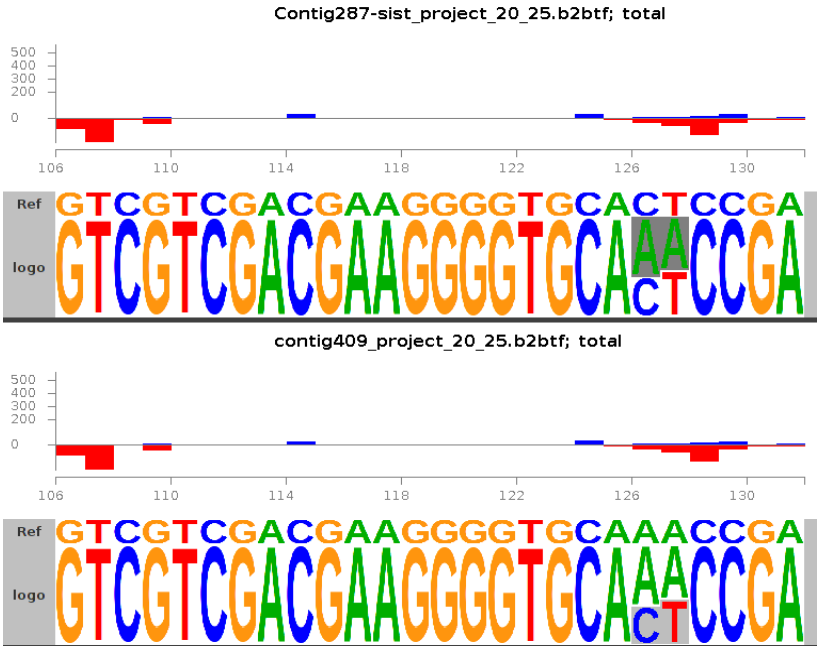


Fig.49. Reconstruction of consensus genome sequence of GYSVd-1. The MISIS-2 mapview zoom-in snapshots shows two SNPs at position 126 and 127 deviating from the consensus sequences of the two genetic variants by more than 10% (light grey) and by more than 50% (dark grey).

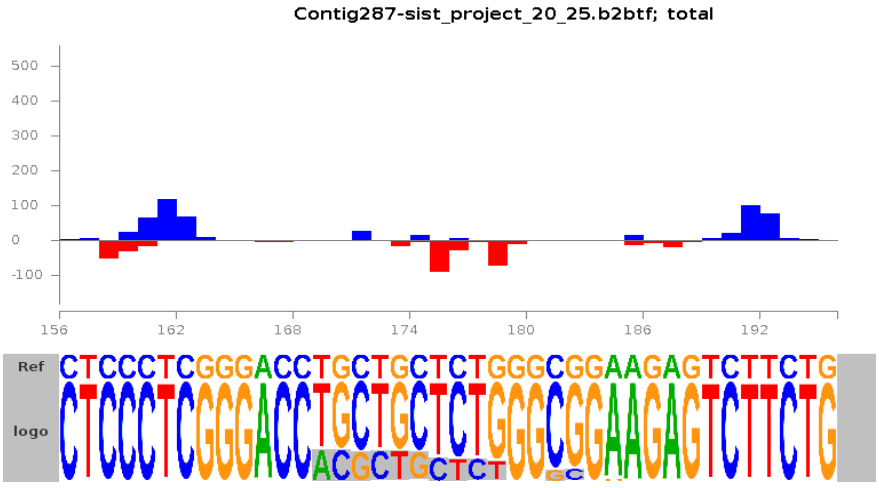


Fig.50. Reconstruction of consensus genome sequence of GYSVd-1. The MISIS-2 mapview zoom-in snapshots shows an indel around position 169.

3.4.2 Virome reconstruction in the symptomatic Pinot noir

As previously reported by Szychowski et al. 1995, three viroids (HSVd, GYSVd-1 and GYSVd-2) in combination with the nepovirus GFLV, could have a synergistic effect leading to a more severe symptomatic phenotype, compared to co-infection of GFLV with a single viroid HSVd which was symptomless. Consistent with this report, 13 Seqman contigs from the symptomatic Pinot Noir (HYT-17) scored positive for HSVd, GYSVd-1 and GFLV with the

best matches to the isolates JN391442.1 (HSVd), JX559643.1 (GYSVd-1), AB219944.1 (GFLV RNA1) and KT000347.1 (GFLV RNA2). In the case of GFLV, the genome coverage for the two segments was only 5% (RNA1) and 23% (RNA2) and a reference-based reconstruction was attempted to reconstruct a complete genome. Mapping redundant sRNA reads on the reference sequences with five iterations steps for RNA1 and six iterations for RNA2 produced near complete consensus sequences, with 94% and 95% sRNA coverage, respectively (Fig.51).

HSVd perfectly matched the reference after one single mapping, whereas three iterations were needed to reach the consensus sequence for GYSVd-1, which resulted 98% identical to the reference isolate (KT000347.1).

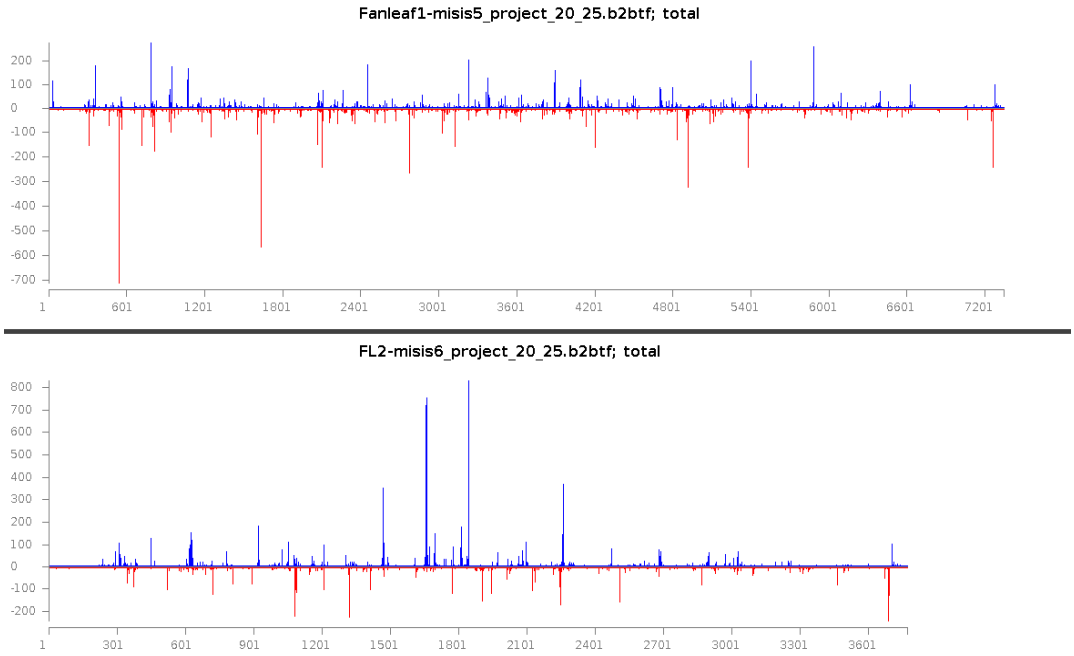


Fig. 51: Misis profile of mapped 20-25nt reads along the two reconstructed segments of GFLV.
The x-axis represents the GFLV genome (segment 1 on top, segment 2 below). The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

3.4.3 Virome reconstruction in the symptomatic Otcha Bala grapevine

As introduced above, the grapevine Otcha Bala plant was previously analyzed using a different RNA sample preparation protocol and bioinformatic tools, which led to the identification of a novel isolate of GLRaV-4, together with the *Grapevine fleck virus* (GFkV), the *Grapevine red globe virus* (GRGV) and the *Grapevine virus A* (GVA, Reynard et al. 2015). The leaf sample from the same plant (HYT-18) was analyzed by the siRomics pipeline. As mentioned above, the blastn analysis of both Oases and Seqman contigs revealed the presence of four different viroids and three viruses and their closely-related sequences from the NCBI Genbank were used here for a reference-based reconstruction.

For the viroid AGVd, five iterations were needed to reach the point when the two following

consensus were 100% identical and supported by majority of the reads. However, by counting the number of mapped reads with Samtools (defined by the -F 4 flag in the bam file), 4,657 reads that mapped after the second iteration decreased to 4,638 reads in the last one, suggesting that the consensus sequence is supported by the majority of the reads but not necessarily by the highest number of mapped reads. The consensus sequence of our AGVd isolate resulted to be 96% identical to the reference (HM211854.1), due to indels and SNPs along the viroid genome.

Four Oases and two Seqman contigs matched a HSVd isolate AB219944.1 without complete coverage and the 241 nt consensus sequence was reconstructed by mapping the redundant sRNA reads on the HSVd reference genome sequence (AB219944.1), with one single iteration, and resulted to have two SNPs between the two sequences.

For GYSVd-1, one single iteration was needed in order to reach a consensus sequence 97.8% identical to the reference (KT000347.1). For GYSVd-2 two iterations were required to reconstruct the consensus sequence, 95.2% identical to the reference (HM211853.1).

In order to reconstruct the consensus sequence of GRSPaV, redundant reads were mapped with six iterations on the closely related isolate of the virus (AY368590.1). The resulting consensus sequence shared 93.6% nucleotide identity with the reference sequence. Being a member of genus *Foveavirus* (*Betaflexiviridae* family), the GRSPaV genome encodes five proteins. In the reconstructed consensus sequence the first large ORF coding for a RdRP was prematurely terminated by stop codons at three positions as identified by the NCBI ORF finder. By looking for possible SNPs in MISIS-2 logo, four nucleotides were corrected in order to restore the complete ORF (A1803 → T, T2707 → C, CT 3455 → TG). Although those changes restored part of the frame, a stop codon was still present at position 3453 and by pairwise alignment with the reference genome, an indel was identified, which was supported by majority of the redundant reads. Same results was obtained with non redundant reads, which showed a gap in the coverage in that specific region. This suggest that a deeper sequencing is needed in order to have a better coverage of the genome, which resulted to be 98% for redundant reads and only 11% for Oases contigs.

Seven iteration were required for GVA which resulted to be only 89% identical to the known isolates present on NCBI (KC962564.1). However, like in the previous case of GRSPaV, the first ORF was interrupted by stop codons in several positions and the non redundant reads could not help to fix this issue, due to low coverage (93% of the genome).

The GLaV4 isolate Ob previously reconstructed by Reynard et al. (2015) was used in this study as reference to reconstruct the consensus sequence using 20-25 nt redundant reads. After four iteration steps, the reconstructed consensus resulted to be 98% identical to the reference, with several SNPs along the genome which shorten the ORFs.

To verify that the complete virome was unraveled, the redundant 20-25nt reads were first filtered through the host genome and then through the seven reconstructed sequences concatenated in one single FASTA file. The filtered (unmapped) reads were de novo assembled by Velvet (k-mer 17) and blastn analysis revealed four contigs related to *Grapevine latent viroid* (GLVd) and two to *Grapevine fleck virus* (GFkV). Going back to check the first blastn analysis, the former viroid appeared to be already present by represented only by one contig, whereas the GFkV was totally missing. Thus, redundant reads were mapped to the closely related reference genome (KR605505.1 and NC_003347.1, respectively) and after four

iteration the consensus sequences were both reconstructed. While the viroid GLVd was totally covered and 94.9% similar to the reference, the GVF was only 86% covered (and 93% similar to the reference), meaning that a deeper sequencing would be required in the latter case to fully reconstruct this viroid.

3.4.4 RNA silencing responses to the grapevine virome components

To evaluate RNA silencing responses to the different viruses and viroids present in the three plants, sRNAs datasets were mapped with zero mismatches to the host reference genome and to the viral reconstructed reference sequences. Thus, the percentage of host-derived sRNAs in the three different datasets resulted to be 30%, 71% and 67%, respectively, whereas the total amount virus-derived sRNAs was equal or lower than 1% in all of them (data not shown). The short reads were also mapped on the reconstructed viral sequences and counted. The percentage of sRNAs deriving from each virus/viroid in total virome-derived sRNAs per each dataset, is shown in Fig. 52.

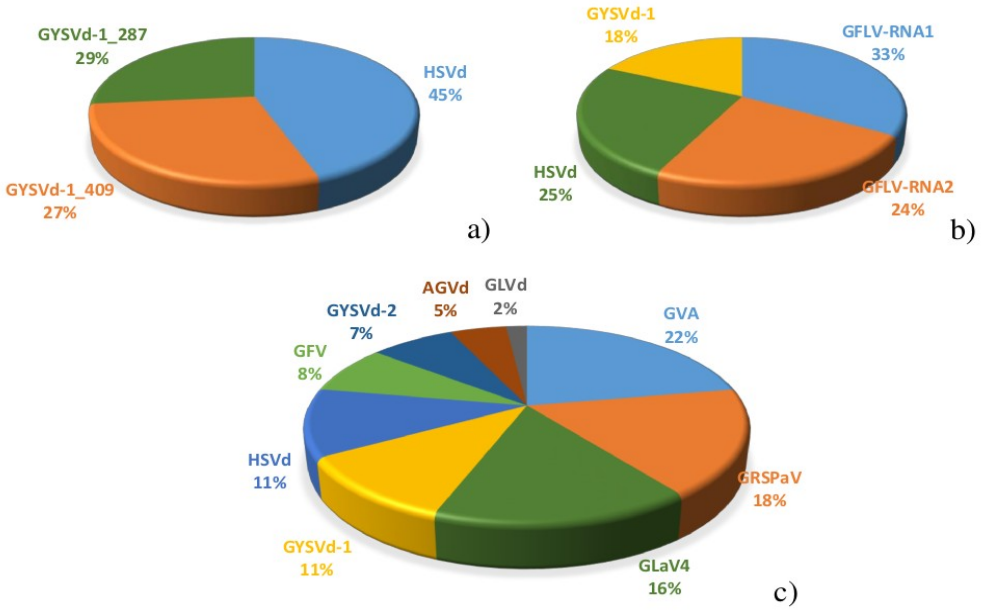


Fig. 52: Percentage of viral sRNAs mapped to each reconstructed consensus sequence.
a) HYT-16 b) HYT-17 c) HYT-18

The host genome-derived sRNAs showed the same size-class profile among the three samples, regardless the infection, with the majority of the reads represented by the 21nt class, followed by the 22nt class (Fig. 53a). The same profiles were obtained from the virus-derived sRNAs, with an increase in the 24nt class, in the total virome (Fig.53b) or when looking at the size-class profile per each virus identified (Fig.54). Interestingly, it seems that HSVd and GYSVd-1 contributed particularly to the 24 nt accumulation (Fig.54, HYT-16), further confirmed in sample HYT-17. Here, however, the main source of 24 nt sRNAs is GLVd, followed by GYSVd-1, HSVd and GYVd-2, suggesting that these viroids are targeted by DCL3 in addition to DCL4 and DCL2, which is consistent with their replication in the plant cell nuclei.

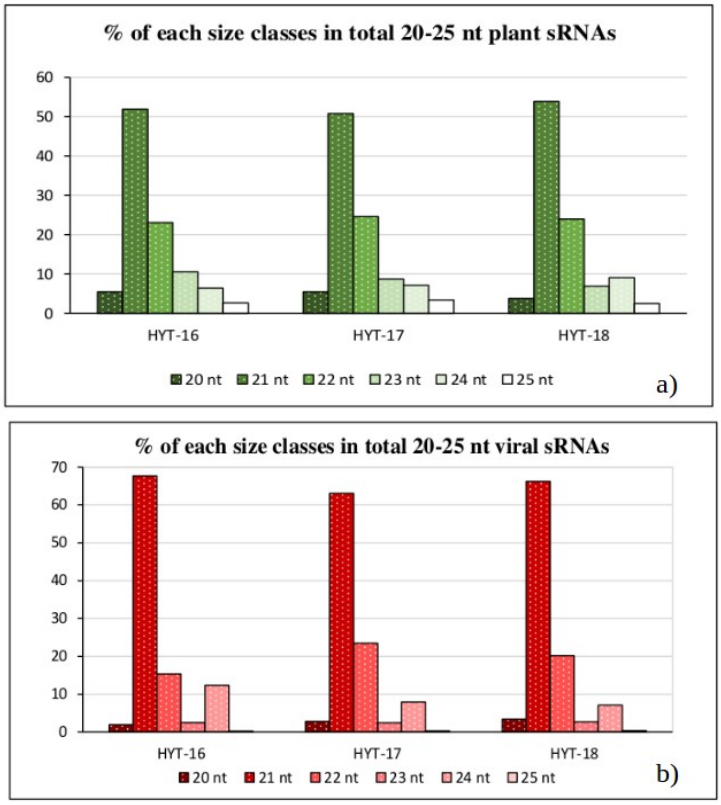


Fig. 53. Percentage of each size-class in total 20-25 nt plant (a) and virus-derived sRNAs (b)

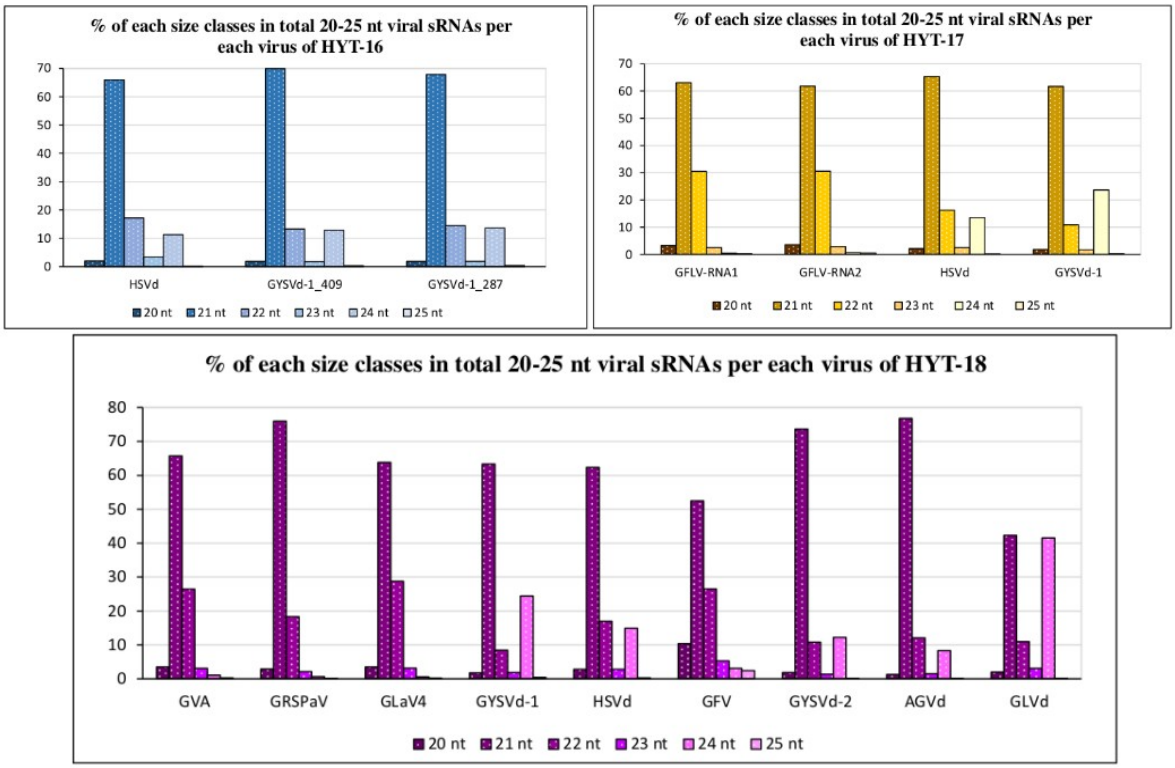


Fig. 54. Percentage of each size-class in total 20-25 nt virus-derived sRNAs in the three datasets

3.5 SiRomics applied to *Larix decidua*

Larch (*Larix decidua*) is a coniferous deciduous tree forming forests in central and eastern Europe, broadly diffused in the Swiss Alps. To date, Larch is known to be susceptible to the larch canker caused by the fungus *Lachnellula willkommii* and to the oomycete *Phytophthora ramorum.* However, the symptoms induced by those pathogens (canker and leaf blight, mostly) differ from the leaf chlorosis observed in the larch tree analyzed in this study. The symptomatic needles from this tree and healthy needles from a neighboring tree were collected in the region close to Nyon (Switzerland) in summer 2015. The same trees were previously studied by Dr. Paul Gugerli, who was able to purify and observe by electron microscopy the virus-like particles of a "zig-zag rod" shape (Fig. 18 in Material and Methods) probably associated with the chlorosis symptoms and to develop a monoclonal antibody to detect this virus-like agent (Gugerli et al., 1986). Since then, Agroscope has been attempting to identify this “zig zag” virus, its nucleotide sequence and molecular characteristics, and to give NGS and our siRomics approach a chance to unravel this mystery.
Following the protocol for recalcitrant samples described in the Material and Methods chapter, total RNA was isolated from the healthy-looking needle sample (HYT-19) and the symptomatic needle sample (HYT-20) for, 10 µg were visualized on 1% of formaldehyde and 15% of polyacrylamide gel, and sent for deep sequencing of the small RNA population at Fasteris, which led to 10 M reads per library, with the majority of the reads from 20 to 25 nt in length (Table 12).

Table 12: Total number of reads per library

sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
HYT-19	10,223,165	7,404,569
HYT-20	10,163,764	9,231,406

As required by the siRomics pipeline, redundant and non redundant 20-25 nt reads were de novo assembled by Velvet and Oases (k-mer 11 → 23, minimum contig length set at 50 or 100) and the resulting contigs filtered through the closest available genome of Norway spruce (*Picea abies*), since the larch genome has not been sequenced yet. The filtered contigs were further assembled by Seqman and analyzed by blastn towards the local viral database and by web blastn. In both samples, the filtered Seqman contigs resulted host-related, with few exceptions of bacterium or fungus-related contigs, which may not be related to the chlorosis because of their presence in the healthy-looking plant and because it may not explain the virus-like particle identified by the EM. Nonetheless, with the aim to make use of all the reads, the siRomics pipeline was slightly modified by expanding the window from 20-25 nt to 1-44nt reads, followed by *de novo* assembly, filtering, Seqman and blastn analysis as described above. Some of the contigs that on the local blast (viral and pathogens database) seemed to be distantly related to phytoplasma or viruses, when blasted towards the web database resulted to be host-related.

Furthermore, a small Larix genome database was created, by concatenating all the 654 records present on NCBI and used as reference to filter the 20-25nt redundant reads, in order to remove more efficiently the host-related reads, instead of the spruce genome. The filtered reads, *de novo* assembled using Velvet (k-mer 17) and checked on web blastn, were still host-related with some hits similar to bacteria or fungi.

Despite the several efforts and despite the virus-related results observed by local blastn analysis, the results are not conclusive, suggesting that the chlorosis may not be caused by a virus. Alternative explanation could be manganese deficiency or phytoplasma, since the latter was found to be responsible of the canker disease of larch. However, these hypotheses would not explain the virus-like particles.

3.6 Identification and characterization of a novel hordeivirus associated with yellow mosaic of privet

Privet (*Ligustrum vulgare*) is a very common plant in natural ecosystem in Switzerland (CH) used mostly for ornamental purpose and hedgerows. Little is known about the possible pathogens infecting privet. Previously, *Tomato black ring virus* (TBRV) and Cucumber mosaic virus (CMV) were isolated from privet (Smith, 1957). Bovey, in his pioneer experiment at the electronic microscope, identified rod-shaped tobravirus or hordeivirus-like particles in privet sample showing yellow mosaicism in the leaves (Bovey, 1976). Later on, Paul Gugerli from Agroscope was able to isolate viral particles from a privet plant with yellow mosaic symptoms and develop an ELISA test that was specific for a virus.

The leaf samples analyzed in this study (the asymptomatic plant, HYT-23 and the symptomatic plant, HYT-24; Fig. 19 in Materials and Methods) were collected in the region close to Nyon, on the edge of Lake Geneva (400 m.a.s.l.) up to 1400 m.a.s.l. in the alps (Valais region). Two different strategies performed in parallel reached the same results that would be further described in the next paragraphs: at Agroscope dsRNA extraction was performed followed by NGS and data analysis (using the metagenomics approach described by Rosslink et al. 2010), while at UniBasel the siRomics approach was applied, with some modifications.

After RNA sample preparation and Illumina sequencing of sRNAs as described in Material and Methods, non- redundant and redundant reads (Table 13) in the range of 20 to 25 were *de novo* assembled by Velvet and Oases (k-mer size from 13 to 23, minimum contig length 50 and 100) in contigs. Since the privet genome has not been sequenced yet, the filtering step required by the siRomics approach was modified in that the contigs deriving from the asymptomatic sample (HYT-23) were used to filtered out the host-related contigs derived from the symptomatic sample (HYT-24) with the aim to keep only the pathogens-related once, assuming that the asymptomatic was pathogen free.

Table 13: Total number of redundant reads per library

sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
HYT-23	11,043,673	9,608,673
HYT-24	11,193,965	10,745,889

Thus, the oases contigs from the healthy sample and the filtered contigs from the infected one were further assembled by Seqman and blasted. The contigs from the symptomatic sample appeared to be distantly related to each of the three RNA segments (alpha, beta and gamma) of *Barley stripe mosaic virus* (BSMV) and *Lychnis ringspot virus* (LRSV), both belonging to the *Hordeivirus* genus of family *Virgaviridae*. These sequences were not detected in the healthy-looking sample, confirming our previous assumption and suggesting that a new, previously-uncharacterized Hordeivirus was present in the sample. Furthermore, besides contigs with no similarity, one contig per strategy-category was related to a DNA geminivirus, but further

analysis of this virus, which might be an integral part of the privet genome, was not pursued. Thus, the contigs derived from the different strategies (redundant, non-redundant, 20-25 nt and R1 datasets, min50 and min100) and belonging to the same viral segment (alpha, beta or gamma) were concatenated in a single FASTA file and further assembled by Seqman. For the alpha segment, starting from 2 redundant contigs min50, 2 non-redundant min50, 2 redundant contigs min100 and 4 non redundant R1 min50, Seqman created one single contig of 3511 nt. For the beta segment, starting from 3 redundant contigs min50, 1 non-redundant min50, 3 redundant contigs min100 and 3 non redundant R1 min50, Seqman created two contigs renamed beta1 of 2884 nt and beta2 of 724 nt, where the beta2 contig was only a more deviating fragment of the beta segment. For the gamma segment, 8 contigs were scaffolded together in a single 3670 long Seqman contig. Furthermore, by comparing and aligning the overall length of the contigs deriving from the different strategies, redundant sRNA contigs were found to be longer than the non-redundant just by few nucleotides. Thus, the two Seqman contigs coming from the redundant min 50 Oases contigs strategy resulted in a longer alpha segment. However, the min50 redundant oases contigs for the gamma segment created a contig, longer than expected from the gamma-hordeivirus segments, which contained an inverted repeat of the same viral region.

In order to verify the reconstructed sequence and to fix those few possible undefined nucleotides positions coming from the Seqman scaffolding, we used these reconstructed segments as references to map redundant reads: two iterations were needed for the alpha segment and only one for beta and gamma. The single base resolution maps deriving from Misis-2 showed that the reconstructed sequences are fully supported by redundant reads in both forward and reverse orientation (Fig.55)

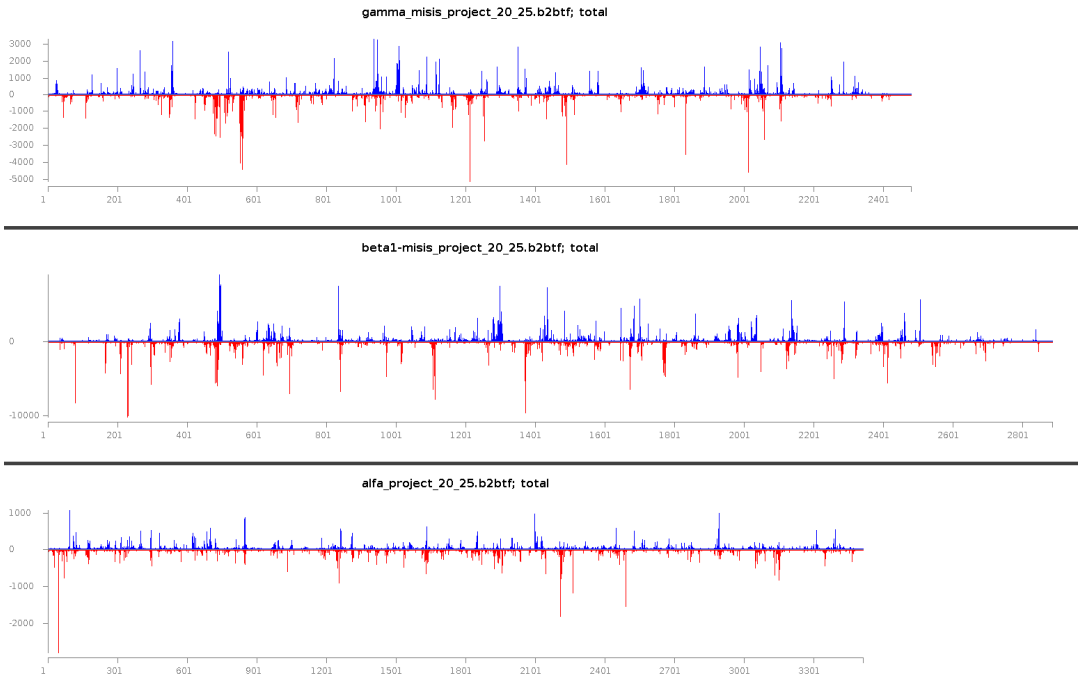


Fig 55. Single base resolution maps of redundant 20-25 nt sRNA reads mapped on the reconstructed consensus sequence of LigMV gamma, beta and alpha genomic RNAs without 3'UTRs .

However, all the three segments appeared to be missing a 3'UTR sequence that is a common sequence for all the segments in other hordeiviruses, which folds into a tRNA-like secondary structure. Screening the Oases contigs for a tRNA-like sequence, led to the identification of a short sequence of 177 nts (Contig_26) with similarities to the 3'UTR of Lychnis ringspot hordeivirus. Thus, this sequence was attached to the reconstructed segment and mapping of redundant reads, followed by Misis-2 inspection was repeated (Fig. 56).

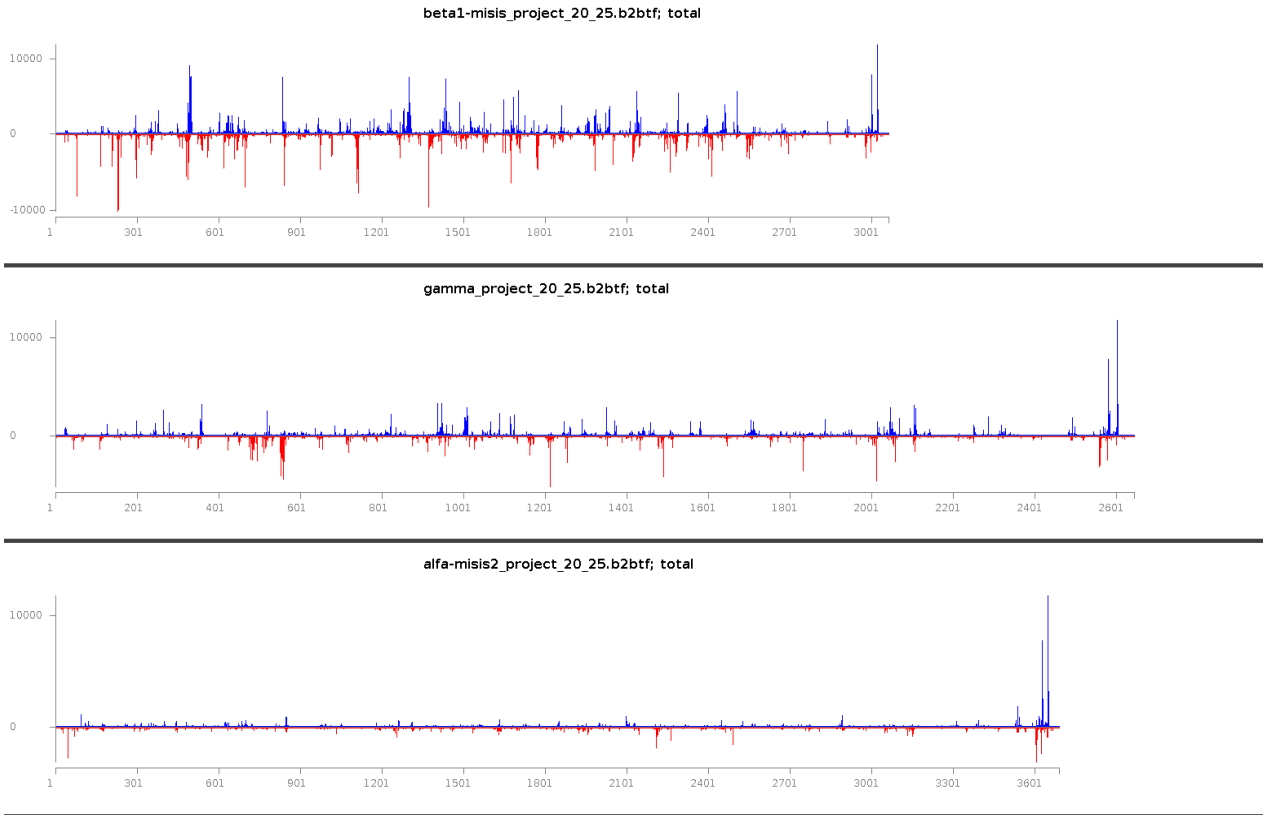


Fig 56. Single base resolution maps of redundant 20-25 nt sRNA reads mapped on the reconstructed consensus sequence of LigMV gamma, beta and alpha genomic RNAs with 3'UTRs attached

To further verify and confirm these bioinformatics findings, specific primers overlapping the junction between the segments and the putative 3'UTR, together with specific probes along the genome were designed for PCRs and blot hybridization analysis as described in Material and Methods. The PCR products resulted to be of the expected sizes for both combination of primers per each segment and were purified for Sanger sequencing (Fig.57). However, during the sequencing of these PCR products from both ends, a poly-A stretch located at the junction between the last ORF and 3'-UTR in each segment caused a slippage of the polymerase, making it difficult to determine exactly the number of adenines in the stretch. Since the 3'UTR contig contained a 20-nt poly(A) stretch at its 5'-end and PCR product sequencing from both ends revealed ca 50-60 nt poly(A)-stretch from one side and ca 50-60 nt poly(T)-stretch from another side, we assume that each segment contains a poly(A) stretch of 50 to 60 nts between the last ORF and the tRNA-like structure.

The blot hybridization analysis with segment- and 3'UTR-specific probes revealed the presence

of all the thee genomic RNA segments of expected size and the subgenomic RNAs for beta and gamma segments, which perfectly matched the size prediction according to the bioinformatics reconstruction and confirmed that the same 3'UTR sequence is part of both the genomic and the subgenomic RNAs (Fig. 58). The three segments reconstructed by siRomics approach, through sRNAs sequencing, were comparable to the three viral contigs assembled from dsRNA sequencing at Agroscope, which however lacked the 3'UTRs and had shorter 5'-UTRs (data not shown).

Thus, this novel hordeivirus was named *Ligustrum mosaic virus* (LigMV). Further analysis was done at Agroscope regarding its geographical distribution in Switzerland and in the rest of Europe, its modes of transmission and host range. Further molecular investigation to confirm the 5' and 3' terminal sequences for each segment using 5'- and 3'-RACE will be performed at Agroscope.

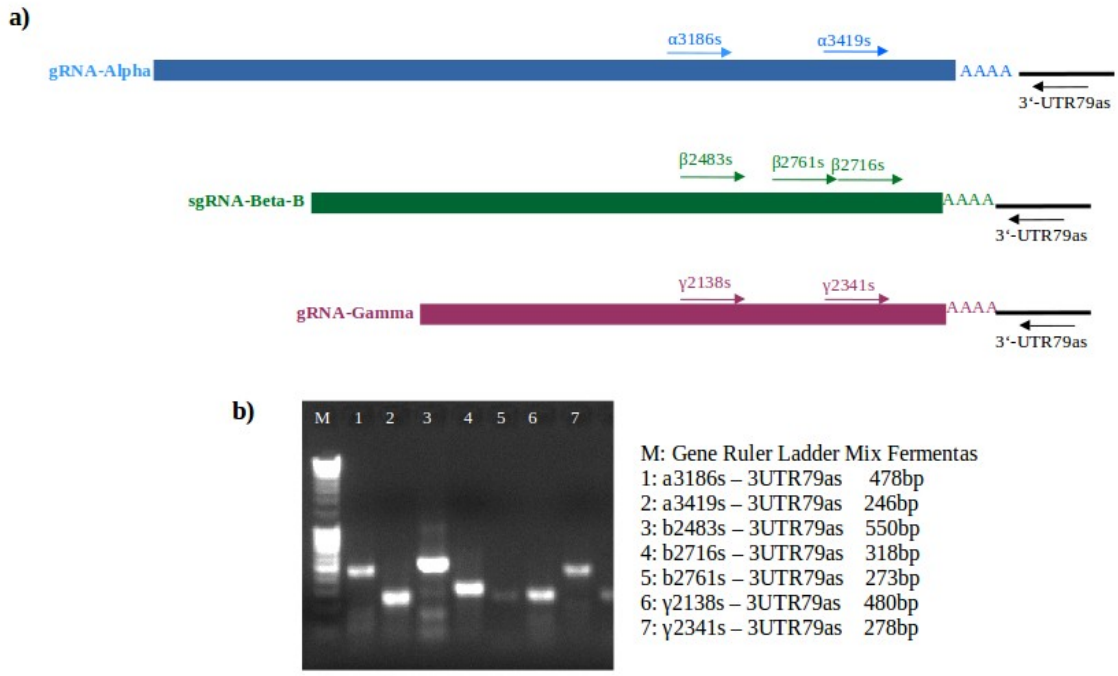
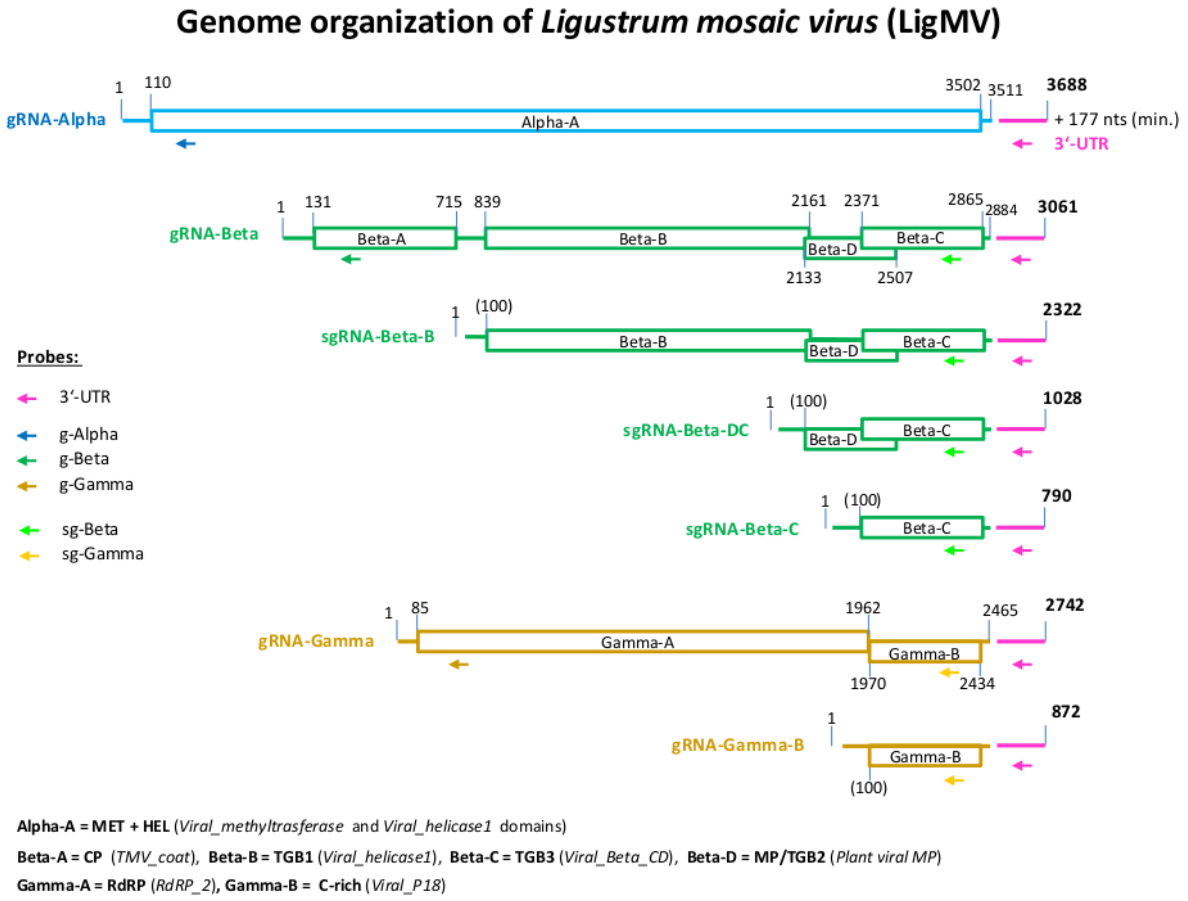


Fig. 57. Schematic representation of RT-PCR analysis of the LigMV segments. a) The arrows indicate the position and the orientation of the primers designed on the reconstructed segments. The positions of the poly-(A) stretch is indicated as AAAA. b) PCR products from different pairs of primers per each segments separated on 1% agarose gel and stained with EtBr. Expected sizes of the PCR products are indicated on the right for each set of primers. 1 kb DNA ladder was used as size marker.

In order to reveal all the possible sequences not yet identified in the virome of HYT-24, the Oases min50 contigs derived from all k-mers assembly and concatenated in one single FASTA file were filtered through the HYT-23 dataset and then through the reconstructed alpha, beta and gamma segments. The unmapped contigs were further scaffolded with Seqman and blastn analysis revealed the contigs related to plant genome, 2 more contigs related to the beta segment of LigMV (probably related with the beta2 segment previously reconstructed), few more for the geminivirus and all the rest with no similarity, which likely represent the unsequenced host genome regions.

To investigate the plant response to LigMV, 20-25 nt sRNA reads were mapped on each of the reconstructed segments and counted based on size class, polarity and 5'-nt identity. As shown in Fig. 59, the viral sRNAs represented the 18% of the total 20-25 nt sRNAs, with the majority (12%) coming from the beta segment, which encodes coat protein, viral replicase and movement protein. The 21-nt and 22-nt viral siRNAs are the most abundant, suggesting a specific involvement of the privet orthologs of DCL2 and DCL4 (Fig. 59b). Analysis of 5'-nt frequencies revealed a substantial bias to 5'U and 5'A in each of the viral segments (Fig. 59c).



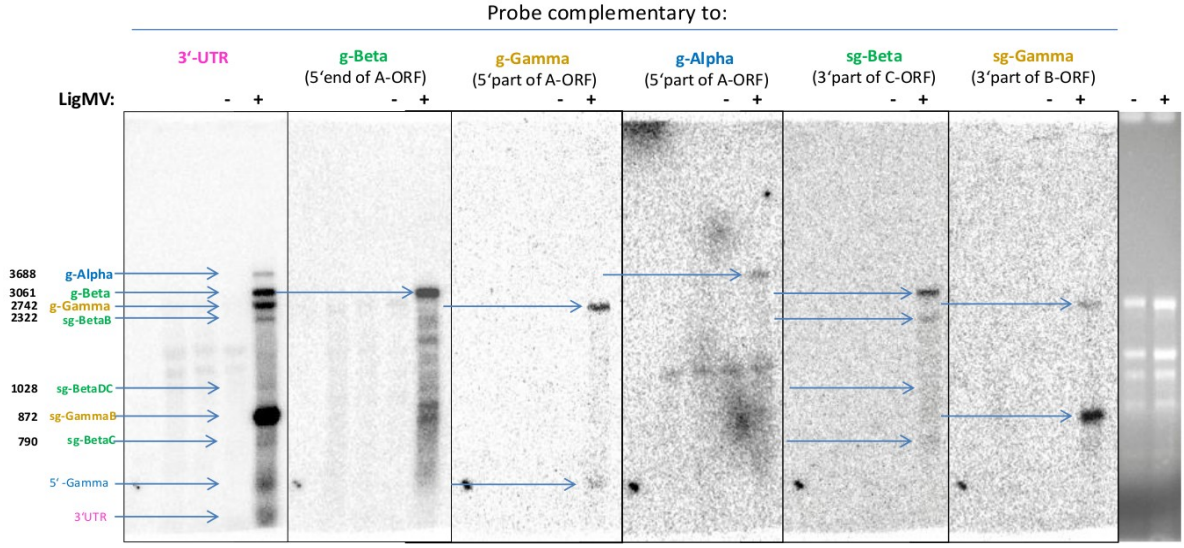


Fig 58. **RNA blot hybridization analysis of LigMV genomic and subgenomic RNAs.** Specific probes were designed for each segments and on the putative 3'UTR sequence, represented by colored arrows in the upper picture which signal is indicated by blue arrows in the second picture, with the expected size written aside.

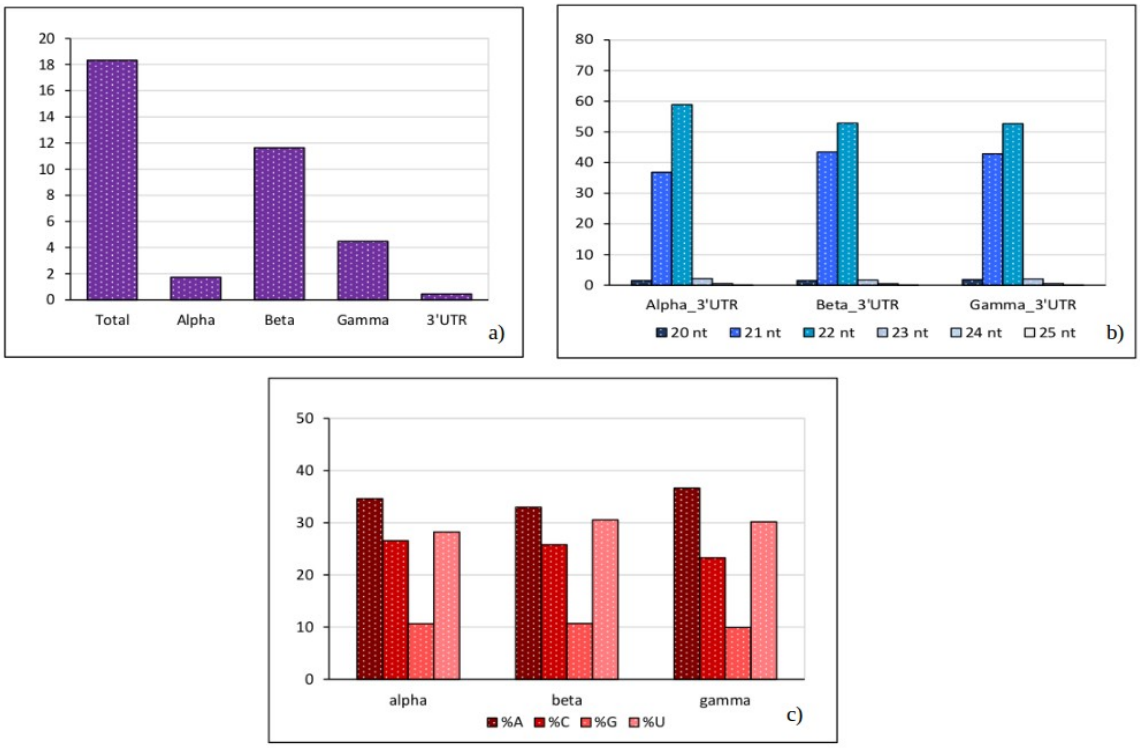


Fig. 59. Counts of total 20-25 nt viral sRNAs. a) Percentage of 20 – 25nt sRNAs deriving from LiMV alpha, beta and gamma segments and 3'UTR, separately. b) Percentage of each size-class siRNAs for each segment extended with 3'UTR. c) Percentage of 5' nucleotide identity.

3.7 siRomics applied to *Ficus carica*

Common fig (*Ficus carica*) is a fruit tree susceptible to the fig mosaic disease (FMD), first described by Condit and Horne in 1933 and subsequently spotted wherever figs are grown, especially in the Mediterranean basin, where it appeared to be associated to *Fig mosaic virus* (FMV) but also to several species of closteroviruses (Elbaino et al. 2006, 2007, 2010, Walia et al.2009). In this study, two samples fig leaves were subjected to siRomics analysis, one symptomless (HYT-21), coming from the Agroscope collection in Changing (Switzerland, and the other one showing leaf mottling symptoms (HYT-22), collected in a private garden in Sion (Switzerland). Previous electron microscopy analysis at Agroscope of both samples revealed non-flexuous virus-like particles and closterovirus-like particles which may also represent endogenous fig structure (Fig.20). Furthermore, by standard diagnostic PCR, FMV was identified in HYT-22 sample, suggesting its involvement in the mosaic symptoms. Thus, with the aim to reconstruct the virome, following the standard siRomics strategy, small RNA population from both samples was deep-sequenced and the resulting redundant 20-25 nt output reads were *de novo* assembled by Velvet and Oases, (Table 14, k-mers set from 11 to 23 and the minimum contig length set to 50 or 100). Per each dataset and length cut-off (min50 and min100), the resulting Oases contigs were merged together in a single FASTA file (oases-all.fasta), filtered through the Ficus carica genome sequence (Mori et al., 2017, https://www.ncbi.nlm.nih.gov/genome/?term=txid3494[orgn]) and further assembled with Seqman. As expected the number of min50 contigs was much bigger than that of min100 contigs (Table 13). Seqman contigs were then aligned to the local viral database by blastn command lines and on the web blastn. Surprisingly, blastn analysis of the contigs from the symptomless sample HYT-21 revealed matches to Fig badnavirus 1 (FBV-1, JF411989.1) and several closteroviruses including Fig closterovirus (FJ211077.1), Fig mild mottle-associated virus (FMMaV, FJ611959.1), Fig leaf mottle-associated virus 1 (FLMaV-1, AM113547.3) and Fig leaf mottle-associated virus 2 (FLMaV2, FJ473383.1) for which only partial genome sequences (covering up to about 1/3 of a typical closterovirus genome) are available in the database. HYT-22 Seqman contigs matched to FMV, as expected from PCR diagnostics results as well as to FBV-1, while no contigs matched to any fig closteroviruses.
The two datasets were then further analyzed separately as described in the following sections.

Table 14: Total number of reads per library

sRNA library	Total number of 1-44 nt reads	Total number of 20-25 nt reads
HYT-21	11,842,553	9,678,397
HYT-22	11,639,727	9,657,937

Table 15. Numbers and lengths of Oases and Seqman contigs

	Redundant reads, min50			Redundant reads, min100		
HYT-21	**Oases-all**	**Oases filtered**	**Seqman**	**Oases-all**	**Oases filtered**	**Seqman**
Number of contigs	3,398	1,576	330	1,220	576	127
N50 of contigs	111	103	168	177	151	235
Average length of contigs	116	106	155	183	155	216
Contig maximum length	1,542	604	800	1,542	604	800
HYT-22	**Oases-all**	**Oases filtered**	**Seqman**	**Oases-all**	**Oases filtered**	**Seqman**
Number of contigs	1,860	879	168	811	352	63
N50 of contigs	168	143	213	354	318	423
Average length of contigs	156	143	186	257	240	322
Contig maximum length	2,415	2,415	2,487	2,415	2,415	2,487

3.7.1 siRomics approach applied to the asymptomatic fig sample HYT-21

In order to reconstruct the consensus sequences, Oases contigs were mapped to the above-mentioned reference sequences (using BWA with default parameters), visualized with IGV and the relative genome coverage calculated by Samtools depth. As shown in Table 16, when mapping the Oases and Seqman contigs on each identified virus, the minimum contig length parameter made a difference in the number of mapped contigs, giving an higher number of mapped contigs for min50, compared with min100, which is translated in an higher coverage of the genome itself (Table 17).

Table 16: Number of Oases and Seqman contigs mapped on reference sequences.

Reference genome	Oases contigs		Seqman contigs	
	min50	min100	min50	min100
FBV-1	70	43	12	8
FLMaV1	9	3	2	1
FLMaV2	27	17	9	5
FMMaV	99	56	28	15
Fig closterovirus	14	9	2	2

Table 17: Percentage of the reference sequence covered by Oases and Seqman contigs.

Genome/partial sequence length	Genome length	Oases contigs		Seqman contigs	
		min50	min100	min50	min100
FBV-1	7,140	58	52	34	26
FLMaV1	616	75	39	46	20
FLMaV2	5,868	29	20	17	11
FMMaV	6,290	79	67	63	37
Fig closterovirus	586	92	80	76	76

Since the coverage of the reference viral sequences by contigs was incomplete and there were gaps between contigs, we used a reference-based reconstruction approach, in which redundant 20-25 nt reads were mapped to the reference sequences using BWA with default parameters, followed by exporting the reconstructed consensus sequence supported by the majority of the reads with MISIS-2 v. 2.7 (Seguin et al. 2016). The redundant reads were then re-mapped to the consensus sequence to generate a new consensus sequence. Such iterated re-mapping steps were repeated until the two sequences (the last reference used and a new consensus) became 100% identical and there was no more increase in the number of mapped reads.

By this approach, a large continuous 5'-portion (68.5%) of the FBV-1 genome sequence was reconstructed after seven iterations. Notably a large portion of the FBV-1 was not represented in the small RNA population as illustrated by the small RNA single nucleotide resolution map generated by MISIS-2 in Fig.60, which is consistent with Oases min50 contigs coverage of FBV-1 genome (Fig. 61).

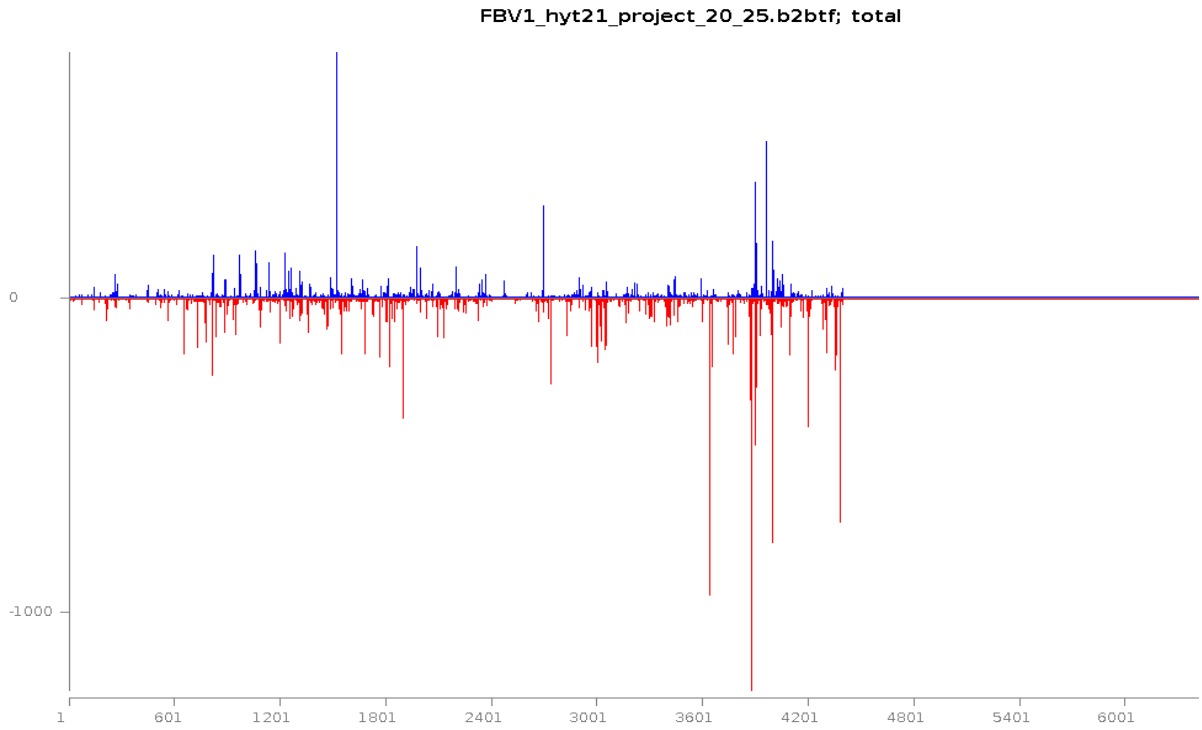


Fig. 60: Misis profile of mapped reads along the FBV-1 genome

The x-axis represents the FBV-1 genome reference sequence partially reconstructed by siRomics. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

This map profile suggested that viral sRNAs are probably derived from a host genome-integrated form of FBV-1, since episomal badnaviruses are known to spawn sRNAs from the entire genome sequence as e.g. was shown for six distinct badnaviruses in *Musa acuminata* banana plants (Rajeswaran et al. 2014a). In fact, FBV-1 was found to be integrated in the

genome of most fig cultivars, while in some of them it was also present as a circular episomal virus (Laney et al. 2012). Blast analysis of the fully-reconstructed 4418 nt portion of the FBV-1 genome showed 99.9% identity with the FBV-1 sequence (JF411989.1; 24 SNPs) isolated from a fig tree in Arkanzas, USA in 2008 (Laney et al. 2012), which was used for the reference-based reconstruction, and even higher identity with three of the five FBV-1 sequences isolated from fig trees in Iran in 2015 (Alimoradian et al. 2016), including KT809307.1 (14 SNPs), KT809305.1 (17 SNPs), KT809305.1 (17 SNPs), KT809306.1 (25 SNPs) and KT809303.1 (25 SNPs). Such a high identity shared among our FBV-1 sequence with episomal FBV-1 isolates from USA and Iran suggests that the same fig genome-integrated FBV-1 locus (yet to be fully sequenced) may give rise to episomal infections worldwide.

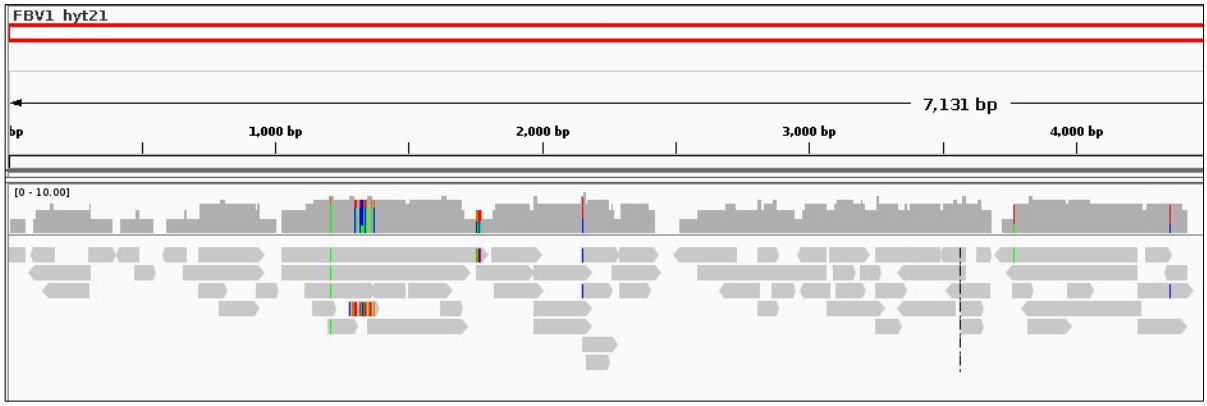


Fig 61: IGV profile of mapped Oases min50 contigs along the reconstructed FBV-1 genome.
The gray bars represent the mapped contigs along the genome, with the arrow pointing in the forward or reverse orientation.

Alignment of our partially reconstructed FBV-1 genome to the draft genome sequence of the fig cultivar Horaishi from Japan (Mori et al. 2017) revealed multiple but very short stretches of similarity, indicating that FBV-1 sequences were not included in the draft genome assembly which covered approximately 75% of the genome size (Mori et al. 2017). Alternatively FBV-1 may not be integrated in Horaishi cultivar's genome. To confirm the integration of FBV-1 possible presence of episomal virus in our genotype of *Ficus carica* (HYT-21), Southern blot hybridization combined with RCA would be required as described by Laney et al. 2012.

Closterovirus infections have been reported in fig trees (Elbaino et al. 2006, 2007, 2010, Walia et al.2009) and partial sequences of at least four distinct viral species (FLMaV-1, FLMaV-2, FMMaV and Fig closterovirus) have been deposited on the NCBI Genbank. For FLMaV-1, two available short reference sequences (AM113547.3 related to the Hsp70 gene and AM279676.1 related to a polyprotein) were used to map redundant reads. The former was much more covered compared to the latter, and four iterations were needed to reach a consensus sequence 92% similar to the reference. Interestingly, 3 ORFs were identified by ORF finder, one more compared to the reference: ORF1 (from position 472 to 615) codes for Hsp70-homolog protein, ORF2 (from position 284 to 382) codes for an unknown protein as

well as ORF3 on the reverse strand (from position 285 to 190). Short reads cover the whole reference sequence, but the coverage is reduced to 50% when assembled in contigs (Fig 62).

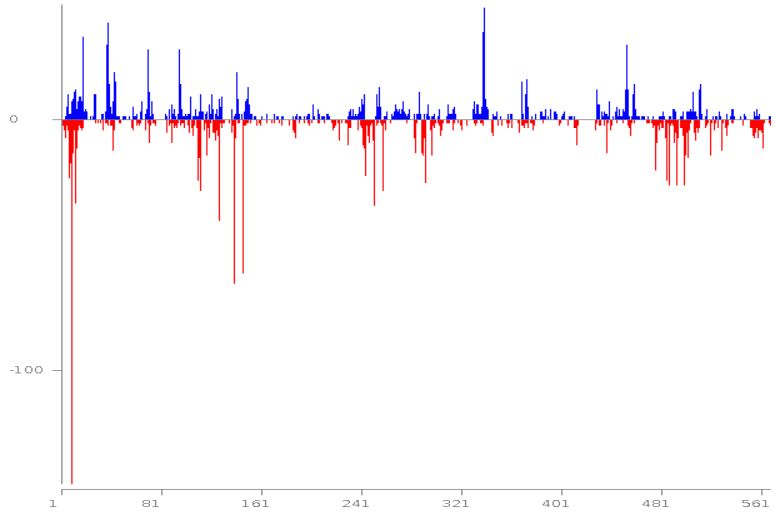


Fig. 62: Misis profile of mapped reads along the FLMaV-1 partial sequence

The x-axis represents the partial hsp70 gene for heat shock protein 70-like. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

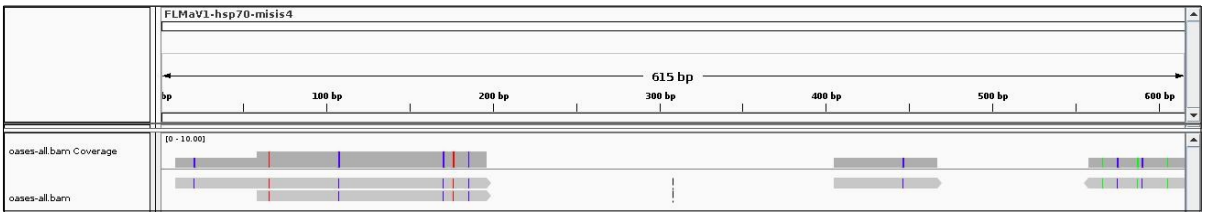


Fig 63: IGV profile of Oasis min50 contigs along the FLMaV-1 partial sequence.

The gray bars represent the mapped contigs along the genome, with the arrow pointing in the forward or reverse orientation. The red and blue vertical bars represent the mismatches between the genome and the contigs.

One of the oases contigs (Contig_548) mapped at the 3' end of the reference sequence and could extend the sequence by 32 nts. The Contig_152 was re-aligned to the reconstructed sequence to extend the 5'-terminus. The extended sequence was verified by re-mapping the redundant reads, non redundant reads (with 4 SNPs) and redundant Oases contigs, with an increased coverage of the reference sequence (Fig. 64).

Even though one contig hit the second partial reference FLMaV-1 (AM279676.1), the coverage in terms of mapped short reads was low and not increasing after iterations, suggesting that this sequence of FLMaV-1 may not generate sufficient amount of sRNAs.

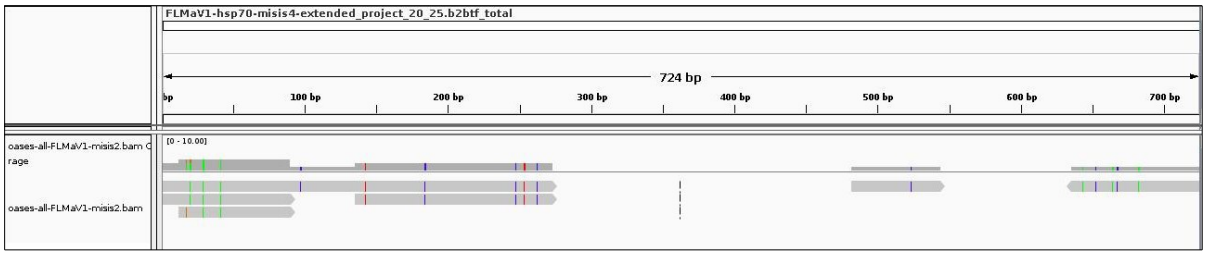


Fig 64: IGV profile of Oasis min50 contigs along the reconstructed extended FLMaV-1.
The gray bars represent the mapped contigs along the genome, with the arrow pointing in the forward or reverse orientation. The red and blue vertical bars represent the mismatches between the genome and the contigs.

The partial sequence of Fig closterovirus (FJ211077.1) was used as reference for the iteration steps which lead to a consensus sequence fully covered by short reads, 80% covered by oases contigs and 94% identical to the reference. IGV inspection of the contigs revealed that the consensus sequence could have been extended by 132 nts coming from the oases Contig_280 at the 5' end and by 158 nts from Contig_310 at 3' end, extending the consensus sequence by 290 nts, further verified and confirmed with one more iteration mapping step (Fig. 65). ORF finder revealed, among others, an ORF of 615nt/204 aa covering near-complete reference sequence, matching the short part of a closterovirus polyprotein ORF codifying for replicase, like the one of *Beet yellow virus* (BYV). However, using the complete genomes of BYV isolate Ukrainian (X73476.1) or of another closterovirus *Citrus tristeza virus* (CTV, U16304.1) for reference-based assembly was not successful, due to the high genetic distance between the viral species, even though they belong to the same genus of the family. Thus, this suggests that the reference-based small RNA assembly approach could be applied only inside the same species but not genus or family. Furthermore, it cannot be excluded that this fig closterovirus consensus (in this work named as fig closterovirus extended FCE), together with the Hsp70-like coding sequence from FLMaV-1, could be parts of one closterovirus, since they have been only partially characterized.

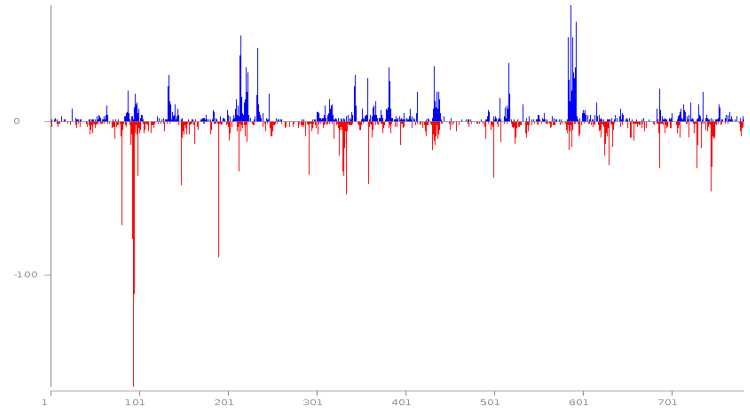


Fig. 65: Misis profile of mapped sRNA reads along the extended Fig closterovirus genome.
The x-axis represents the fig closterovirus genome, extended at the extremities. The y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

Twelve iteration steps were needed for reference-based reconstruction of FLMaV-2 partial sequence, which resulted to be only 86% identical to the 5868 nt reference (FJ473383.1), 92.7% covered by reads and 86% covered by Oases contigs (Fig. 66). ORF finder showed the

ORF organization similar to the one for the reference sequence, with 5 main large ORFs and many shorter ones. The sequence could not be extended by contigs, because no contigs covered the 5'- and 3' termini of the reference sequence (IGV profile not shown).

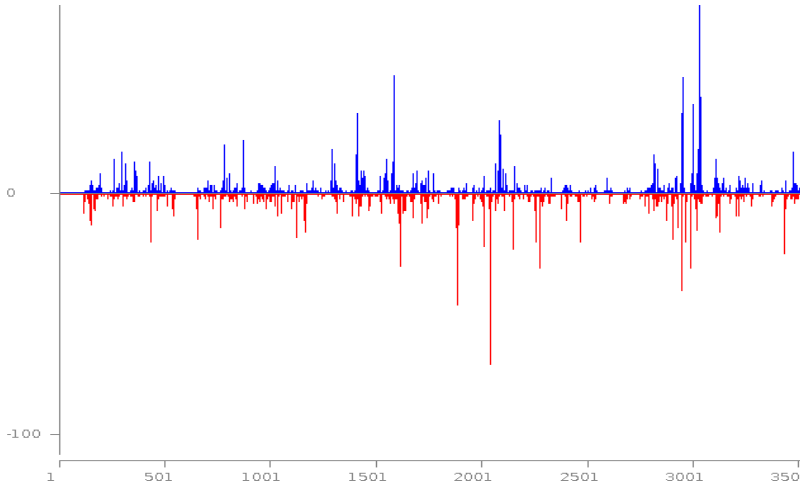


Fig. 66: Misis profile of mapped reads along the reconstructed part of the FLMaV-2 genome.
The x-axis represents the FLMaV-2 genome, the y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

Finally, the consensus sequence of FMMaV's partial genome was reconstructed in four iterations, starting from mapping on the 6290 nt reference (FJ611959.1) with 99.5% coverage in sRNA reads (Fig.67) and 69% in min50 contigs (Fig.10) and 90% sequence identity to the original reference. The ORF finder analysis showed a difference between the consensus sequence and the reference: a stop codon around position 2734 terminated one of the longest Hsp70-like protein ORF in the reference.

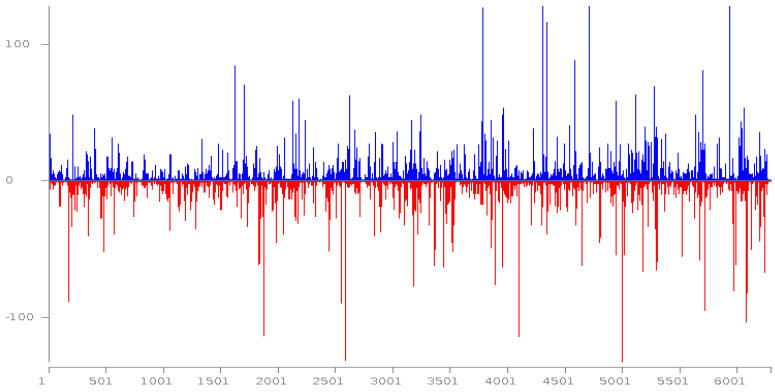


Fig. 67: Misis profile of mapped 20-25 nt sRNA reads along the reconstructed FMMaV partial sequence.
The x-axis represents the FMMaV sequence, the y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

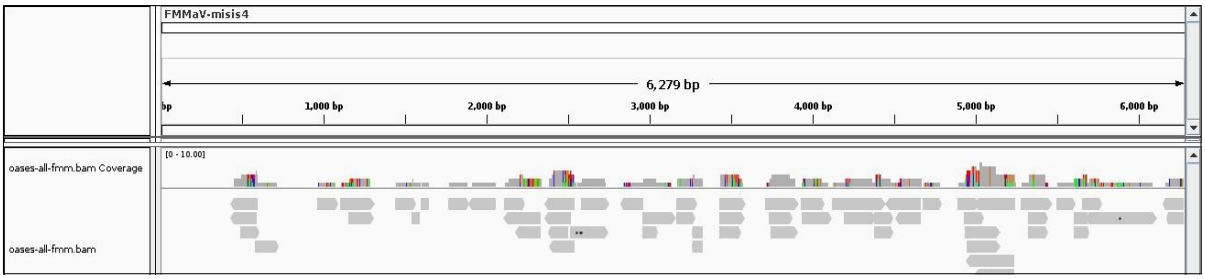


Fig.68: IGV profile of mapped Oases min50 contigs along the FMMaV partial sequence.
The gray bars represent the mapped min50 contigs along the reference sequence, with the arrow pointing in the forward or reverse orientation. The colored vertical bars represent the mismatches between the genome and the contigs.

To verify if any additional virus was present in HYT-21, the 20-25 nt sRNA reads were filtered through the fig host genome and all the five reconstructed sequences of FBV-1, FLMaV-1, FLMaV-2, Fig closterovirus and FMMaV, respectively. The filtered reads were then assembled by Velvet (k-mer 17) and Oases and the resulting Oases min50 contigs were checked for viral similarity using the blastn command line. The results revealed twenty contigs still related to viruses but mostly due to short similarity region, suggesting that the virome may contain additional components or that those contigs may correspond to the unsequenced regions of the four closteroviruses.
In conclusion, the siRomics diagnostic analysis applied to the asymptomatic fig sample HYT-21, used here as *healthy control* revealed the presence of at least two closteroviruses whose available partial genome sequences of 5.8 kb (FLMaV-2) and 6.3 kb (FMMaV) are well represented by sRNAs. In addition, the badnavirus FBV-1 was identified for which a large portion of the 7.1 kb genome (starting from position 1 of the RT primer binding site till position 4418) is well represented by sRNAs, while the remaining 3.7 kb portion is not supported by sRNAs, thus suggesting that this endogenous pararetrovirus is not released from the fig genome as episomal replicating virus (see below).
Further molecular analysis and deeper sequencing to increase the number of reads, are needed to have a better reconstruction of the virome and to verify whether more than two species of the closteroviruses are present.

3.7.2 siRomics approach applied to the symptomatic fig sample HYT-22

The redundant 20-25 nt reads from HYT-22 were reference-based assembled by mapping on the complete genome sequences of FMV composed of six segments (AM941711.6, FM864225.2, FM991954.1, FM992851.1, HE803826.2, HE803827.1) and on the Arkansas isolate of FBV-1 (JF411989.1), which had the best matches to the *de novo* assembled Seqman contigs.
Twelve iteration steps were needed for reconstruction of the FBV-1 consensus sequence. However, since a reference sequence region between position 260 and 727 was poorly represented in the viral siRNA population, each iteration resulted in alteration of the consensus sequence and eventually led to a scrambled non-viral sequence where highly abundant non-FBV-1 sRNAs mapped with zero mismatches, creating a hot spot (not shown).

This region of FBV1 genome was therefore replaced with a corresponding region of the FBV1 consensus sequence reconstructed from HYT-21. The resulting reconstructed FBV1 sequence resulted to be 97% covered by sRNA reads and 70.3% covered by Oases min50 contigs scattered along the genome (Fig. 69, Fig. 70), suggesting its episomal nature as discussed above. Pairwise comparison of the Arkansas isolate reference and the consensus sequence generated from redundant sRNA reads showed 99.6% identity. The ORF finder revealed a profile similar to the reference genome, with a long ORF3 from position 1049 to 6562, codifying for movement protein (MP), capsid protein with zinc-finger domain (CP), pepsin-like aspartate protease (Pro), reverse transcriptase (RT) and RNase H (RNase H) motifs, and the two shorter ORFs, ORF1 and ORF2 in front of ORF3 (Laney et al., 2012).

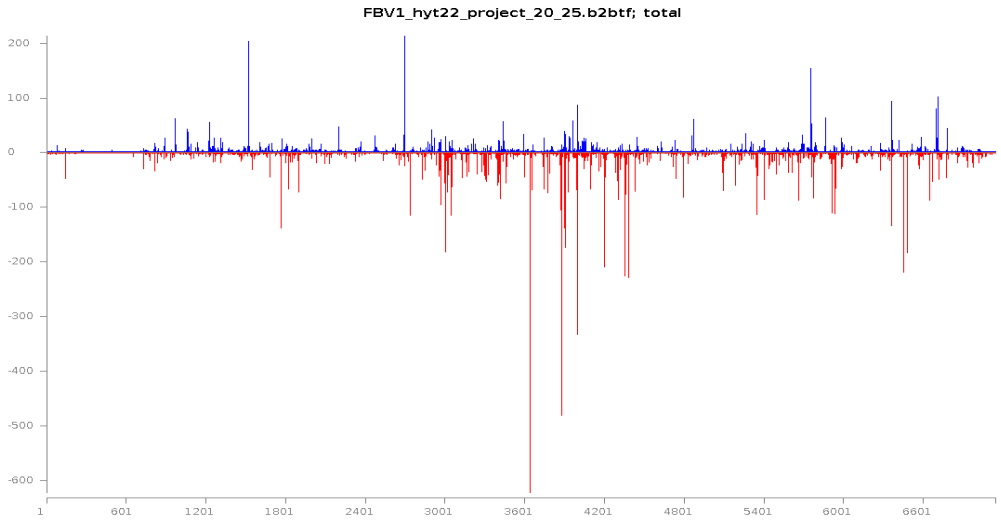


Fig. 69: Misis profile of mapped reads along the reconstructed FBV-1 genome.
The x-axis represents the FBV-1 genome, the y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

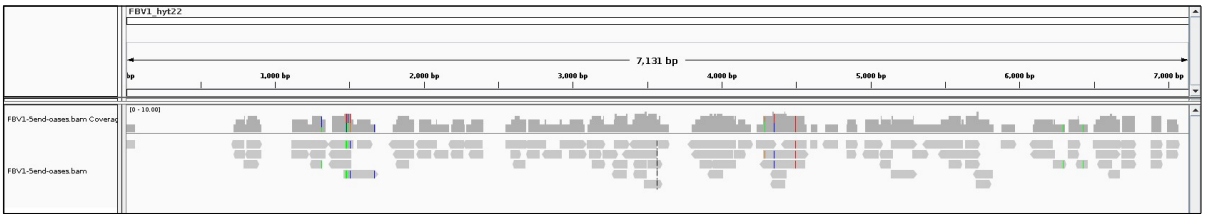


Fig 70: IGV profile of Oases min50 contigs along the FBV-1 genome.
The gray bars represent the mapped contigs along the genome, with the arrow pointing in the forward or reverse orientation. The colored vertical bars represent the mismatches between the genome and the contigs.

Short redundant reads were mapped to each of six segments of the FMV reference genome, and several iteration steps were performed. After more than 14 iterations, the consensus sequence for the first segment carrying the RdRp gene (RNA 1) was still not reached, due to a problematic region that was identified by multi-alignment of all the consensus created by MISIS-2: at each iteration, the one single nucleotide was modified, creating the bottleneck profile shown in Fig.71. Moreover, pairwise alignment of Misis14-consensus with the NCBI

reference showed a region completely distorted (Fig. 72), suggesting that the iterations steps may create a wrong sequence.

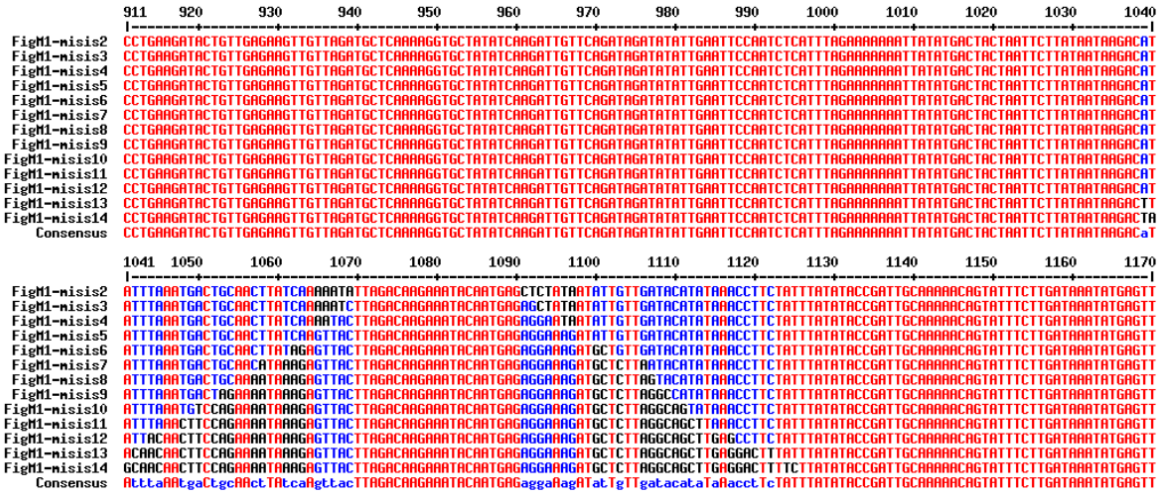


Fig. 71: Multi-alignment of the first 14 consensus sequences derived from iterations steps.
In red the sequence supported by at least 90% of the nucleotides at that position, in blue the low consensus value for 50% reads, neutral black color when the sequence is more diverse (http://multalin.toulouse.inra.fr/multalin/).

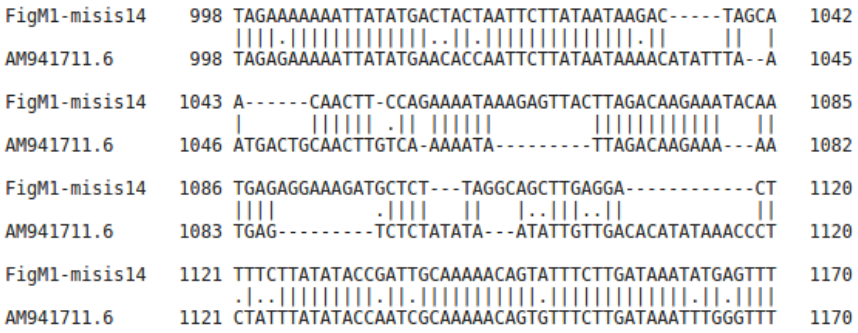


Fig. 72: Pairwise alignment of RNA1-misis14 consensus sequence and reference from NCBI.

Nonetheless, more iterations were performed and the sequence was deviating more and more from the first iteration steps. Thus, the non-redundant reads were mapped on the consensus sequence at iteration 14 and visualized with Misis and IGV. Two position were modified according to the non-redundant consensus (1037C → A, 1038T → A) and verified back with redundant reads and appeared to need more iterations, suggesting a limitation of the reference-based assembly method when there is high variability in a specific region. However, the Oases contigs supported the reconstructed sequence after 14 iterations in the problematic region, as shown by Fig.73. The mapped contigs (position 1040 → 1120) were blasted and appeared to be related to FMV-RNA 4 segment. This was further confirmed by a pairwise alignment between the reconstructed sequences of RNA1 and RNA4 shown in Fig. 74, confirming the previous assumption that during the iteration steps there was an incorporation of RNA4-derived sRNA reads into RNA1 genome, which led to a deviating consensus sequence.

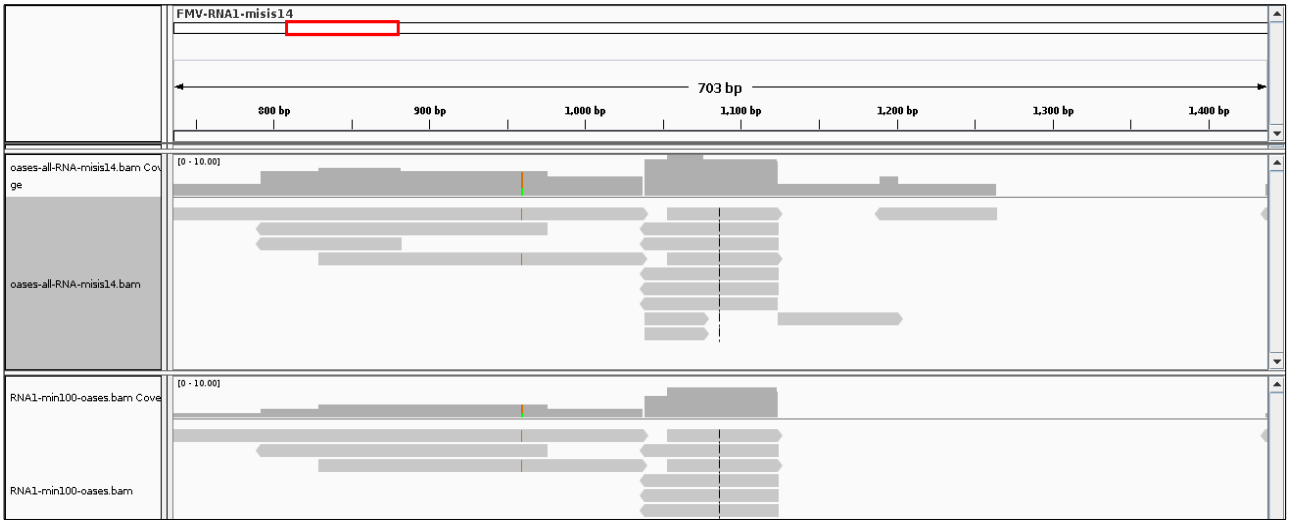


Fig. 73. Oases contigs (min50 on top, min100 bottom) mapped on reconstructed FMV RNA1 sequence.

```
FMV-RNA1-misi    982 GAATTCCAATCTCATTTAGAAAAA-AATTATATGACT--ACTAATTCTTA   1028
                     |..|||.|||||| |||||..|| ||  |.||..|||.|
FMV-RNA4-misi    726 ----------TACTTTTGAAAAACAATTACCTG-CTGAAGTAGCTCTAA    763

FMV-RNA1-misi   1029 TAATAAGACT---AGCAACAACTTCCAGAAAATAAAGAGTTACTTAGACA   1075
                     |     |.||   ||||||||||||||||||||||||||||||||||||
FMV-RNA4-misi    764 T-----GTCTGGAAGCAACAACTTCCAGAAAATAAAGAGTTACTTAGACA    808

FMV-RNA1-misi   1076 AGAAATACAATGAGAGGAAAGATGCTCTTAGGCAGCTTGAGGACTTTTC-   1124
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
FMV-RNA4-misi    809 AGAAATACAATGAGAGGAAAGATGCTCTTAGGCAGCTTGAGGACTTTTCC    858

FMV-RNA1-misi   1125 ------TTATAT-------ACCGATTGCAAA------AACAGTATTTCTT   1155
                           ||| ||        ||.||   |.||      |||
FMV-RNA4-misi    859 AATGCTTTA-ATCTGGGGCACTGA---CCAAGGGCTTAAC----------    894
```

Fig. 74. Pairwise alignment between RNA1 and RNA4 reconstructed consensus sequences.

Thus, the reconstructed region between positions 1000 and 1140 was substituted with the corresponding region from the reference sequence and the modified consensus genome further verified by redundant reads and contigs. However, the region was again not supported by contigs and created gaps on the Misis logo sequence (not shown). In light of the above, we concluded that this problematic region remains to be unraveled by RT-PCR, cloning and sequencing using specific primers designed on the supported reconstructed sequence flanking the problematic region shown in Fig. 71. Furthermore, ORF finder showed a long ORF1, interrupted at position 5421, showing that the latter region should also be further verified.

For the other segments of FMV the consensus sequence reconstruction was more straightforward: six iterations for RNA2 (94% identity to the reference), four iterations for RNA3 (93.6% identity), three iterations for RNA4 (95.6% identity) and five iterations for RNA5 and RNA6 (94.8% and 91.8% identity, respectively). ORF finder analysis of the segments showed ORFs perfectly matching those of the reference genome and corresponding to the expected proteins. Furthermore, the consensus sequences resulted to be fully covered by both sRNA reads and contigs, proving that the deepness of the sequencing in this case was sufficient to fully reconstruct these five shorter segments of FMV. The Misis profile of the

mapped reads along each of the reconstructed segment is shown in Fig. 75, while the IGV mapping profile of the oases contigs is shown in Fig. 76.

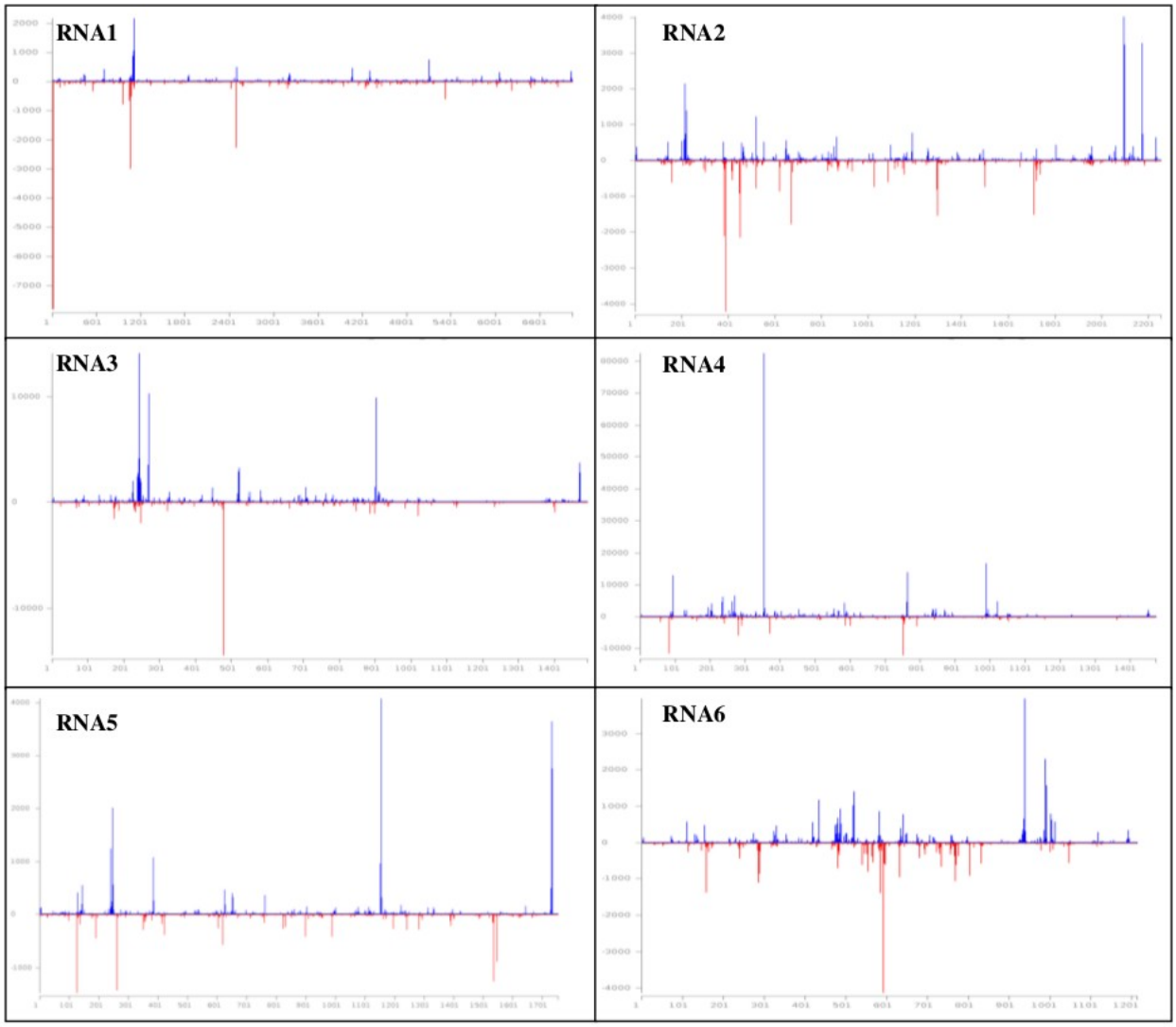


Fig. 75: Misis profile of mapped reads along the reconstructed segments of FMV.
The x-axis represents the FBV-1 genome, the y-axis represents the count of mapped redundant reads with blue bars in the sense orientation and the red bars on the antisense orientation.

Fig. 76: IGV profile of Oases min50 contigs along the reconstructed segments of FMV.

3.7.3 Plant RNA silencing responses to the viromes in asymptomatic and symptomatic fig leaves

To evaluate the virome interactions with the plant RNA silencing machinery, the 20-25 nt sRNAs were mapped with zero mismatches to the *Ficus carica* reference genome and to the reconstructed viral sequences.

The percentage of total viral reads in total 20-25 nt sRNA reads resulted to be 1% for HYT-21 and 10% for HYT-22, whereas the host genome-derived sRNAs represented the 39% and 29% of the total 20-25 nt reads, respectively (Fig. 77a). This indicates much stronger silencing response to the virome in the symptomatic fig leaf than in symptomless. The remaining unmapped sRNAs are likely derived from the host genome sequences that were not assembled in the draft genome (Mori et al. 2017). Interestingly, the majority of the fig genome-derived sRNAs belong to the 21-nt and 24-nt size classes (Fig. 77b), while the 21-nt sRNAs are the most abundant for all the viruses identified in both infected samples, followed by the 22-nt and/or 20-nt sRNAs (Fig. 77c and 77d). The finding is indicative of host DCL4 and DCL2 activities generating 21-nt and 22-nt viral siRNAs as established in model plant *Arabidopsis thaliana*. Interestingly, 24-nt class being negligible in the closterovirus- and FMV-derived sRNAs is quite abundant in FBV-1-derived sRNAs. In fact, the sRNA profiles of FBV-1 and fig genomes are quite similar, except that 22-nt sRNAs are much less abundant in the latter. This is consistent with FBV-1 being a member of DNA pararetroviruses which are transcribed in the nucleus and therefore targeted by the nuclear DCL3 generating 24-nt siRNAs (reviewed in Pooggin 2013). Interestingly, analysis of the 5' nucleotide identity of 21-nt, 22-nt and 24-nt viral sRNAs revealed a bias towards 5'U, 5'A, and 5'C suggesting an involvement in the fig antiviral defense of putative AGO1, AGO2 and AGO5 clade proteins binding 21-22 nt siRNAs with respective 5'-identities and AGO4 clade proteins binding 24-nt 5'A-siRNAs. (Fig. 78).

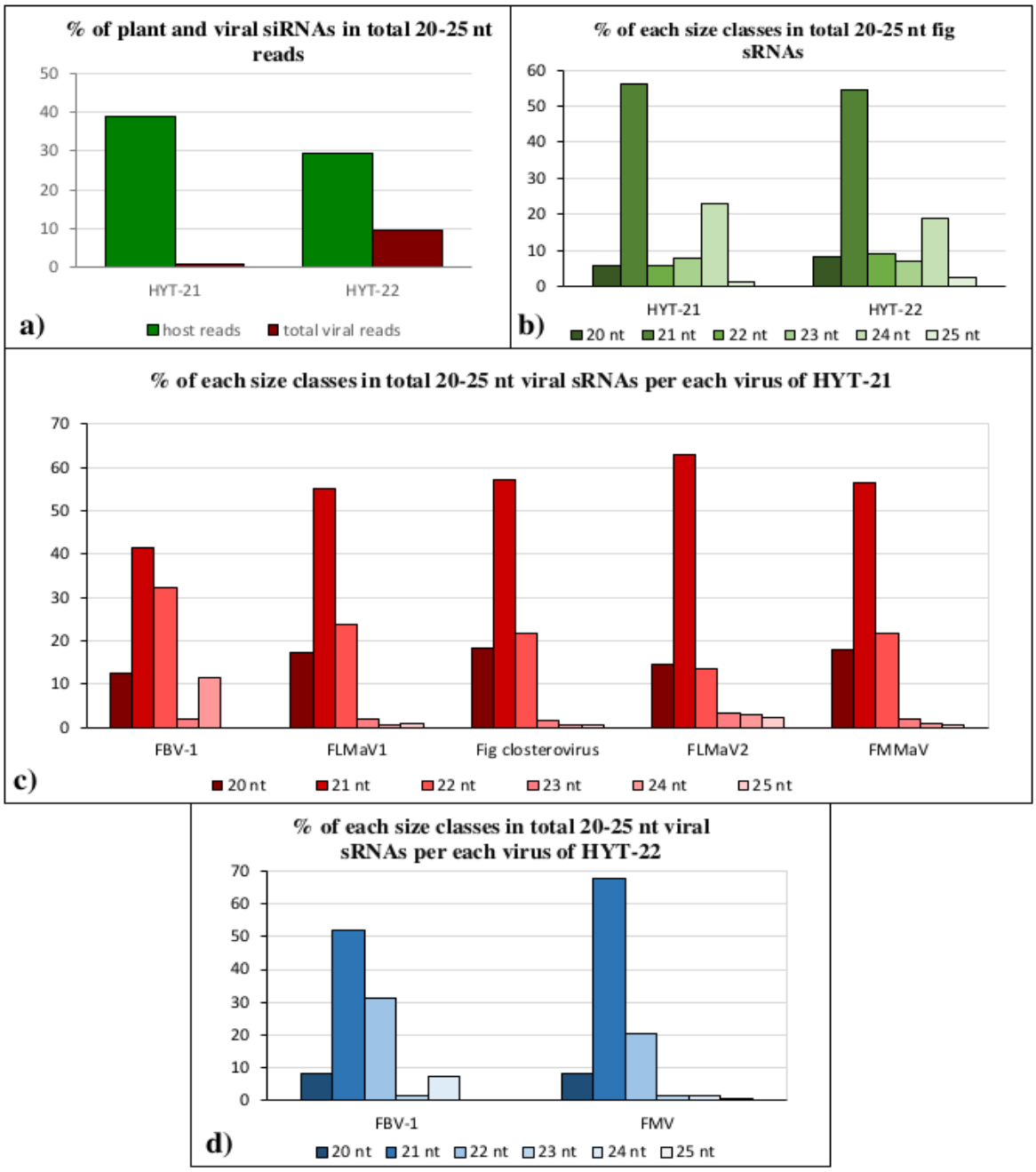


Fig.77 Counts of total 20-25 nt host and viral sRNAs. a) Percentage of host and viral sRNAs in total 20-25 nt population. b) Percentage of each size class of host sRNAs. c) Percentage of each size class of each virus identified in sample HYT-21. d) Percentage of each size class of each virus identified in HYT-22.

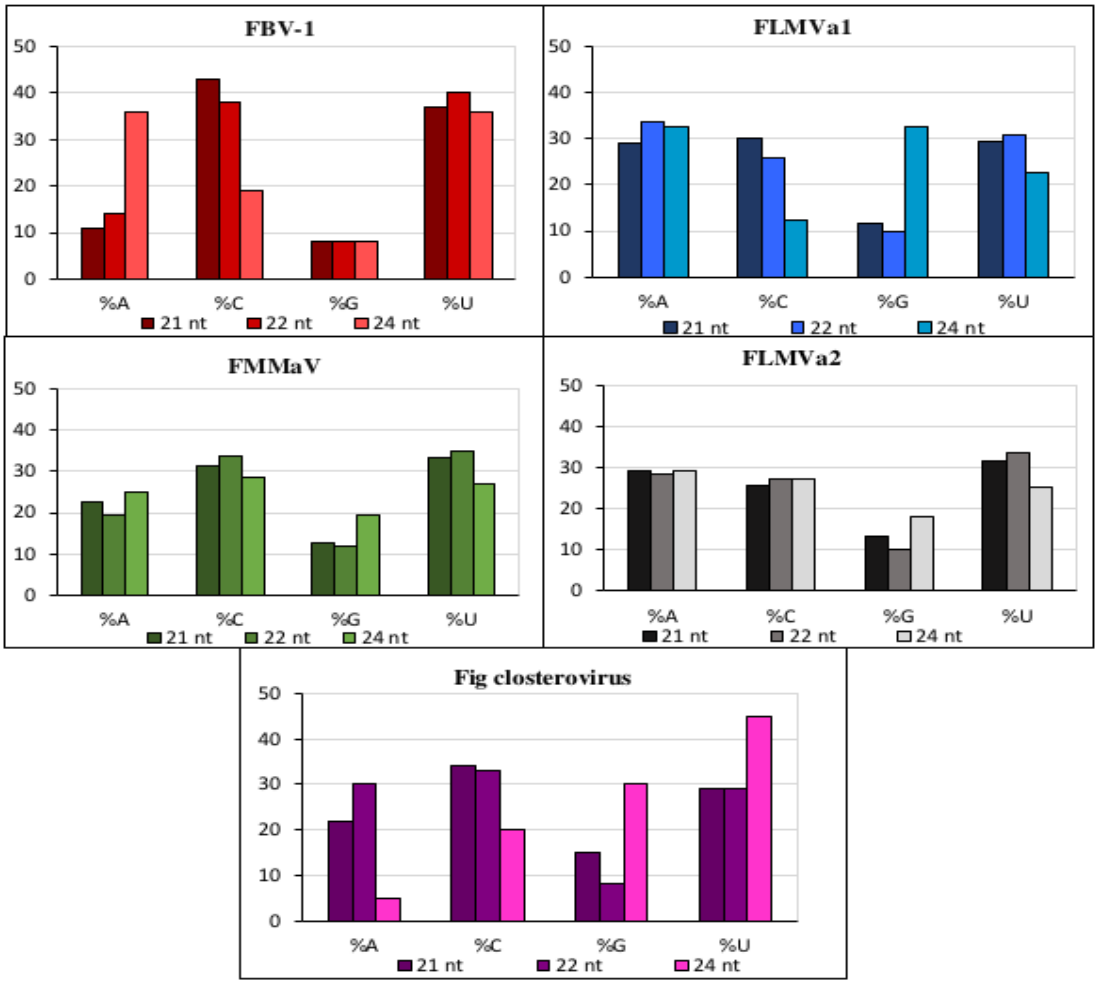


Fig. 78. Percentage of 5' nucleotide identity in 21-nt, 22-nt and 24-nt size class.

4. Discussion and conclusions

4.1 siRomics for virus detection and virome reconstruction

The siRNA-omics (siRomics) pipeline for *de novo* reconstruction of viral genomes developed by Seguin et al. (2014b, 2016) and already applied for crop plants (Seguin et al. 2014a, 2016, Rajeswaran et al. 2014a, 2014b, Fuentes et al. 2016), was further used in this PhD study to identify and reconstruct the viromes present in cultivated and non-cultivated plant species. The pipeline requires a combination of molecular biology techniques (including total RNA isolation and blot hybridization validation), deep sequencing of sRNAs and bioinformatic analysis. Thus, the choice of a good RNA isolation protocol suitable for each type of plant samples represents the first important step that would influence all the downstream analyses. In fact, for the leaf samples from potato, tomato, fig and privet, the standard Trizol protocol (TRI Reagent, Sigma) was applied successfully, with some modifications as described in Material and Methods. The same protocol was not efficient when applied for leaf samples from grapevine, cherry and larch, due to their recalcitrant nature, enrichment in polysaccharides or other secondary metabolites. Thus, a CTAB-based method was applied to get better yield and quality of total RNA (Morante-Carriel et al. 2014, Gambino et al. 2008).

A good quality RNA is also a prerequisite for Illumina deep sequencing, the technology allowing to sequence several multiplexed samples in one lane of the HiSeq2500 or HiSeq4000 platforms, providing up to 280 or 350 million reads per full lane, respectively. Hence, besides the RNA quality, we found that the sequencing depth (i.e. number of million reads per library) was a key point in the whole pipeline, since for some of the samples and the viral species it imposed limitations on the sRNA contigs number and length and thus, on the possibility for *de novo* reconstruction of a complete viral genome. In fact, the 450 M reads obtained from the datasets BPO-147-165 (ca. 25-30 M reads per library) were sufficient for *de novo* assembly of contigs covering the complete genome of the viruses identified in the infected potato and tomato leaf samples (i.e. PVY and PVX) or for reference-based assembly of the consensus sequence without gaps (PepMV isolates). Likewise, the 11 M reads were sufficient to completely *de novo* reconstruct the three segments of a previously unknown virus (LigMV) from the privet leaf sample, where the viral reads constituted 11% of total 20-25 nt reads. On the contrary, the sequencing depth was not sufficient to fully reconstruct, neither by *de novo* nor reference-based assembly, the complete genomes for some of the virome components from cherry, grapevine and fig leaf samples, even if the datasets span from 7 to 14 M reads, likely because viral reads constituted ca. 1% of total 20-25 nt reads. This suggests that indeed the sequencing depth is an important parameter. However, there was no direct correlation between the number of million reads per library and the viral genome coverage, meaning that many more factors are involved such as the total RNA quality and the host-virus interactions resulting in differences in viral titers and most importantly the strength of RNA silencing-based antiviral defenses generating variable

amounts of virus-derived sRNAs. In fact, the number of sRNA reads representing each virus or viroid and thus, required for detection or reconstruction, was influenced by the host plant RNA silencing machinery and its tendency to target differentially the virome components as for example was found for the PVY-PVX virome in potato plants. This would also explain the different percentage of viral sRNAs coming from each different virus or even from different segments of the same virus, as observed for LigMV in privet and FMV.

The *de novo* assembly was performed using the tools previously tested, namely Velvet alone (Kreuze et al. 2009) or Velvet in combination with Oases and Seqman, Seguin et al. 2014b, 2016, Rajeswaran et al. 2014a, 2014b, Fuentes et al. 2016). Two different parameters were used related to the minimum length of contigs (Velvet) or transcripts (Oases), set at 50 or 100 (meaning that the contigs shorter than 50 nt or 100 nt would be discarded). As example of comparison, Table 18 shows four different parameters obtained by the two different cut-off parameters in cherry sample HYT-14 and in fig sample HYT-21. The number of contigs per each k-mer (11 → 23) is indeed higher for min50 value and, as already shown in Table 14 and 15, this increased the viral genome coverage. However, the overall length of the contigs is comparable, meaning that this parameter did not substantially influence the assembly by Seqman and thus, would have not been able to assemble the complete viral genome, leading us to an alternative, reference-based reconstruction approach.

Table 18: Influence of cut-off parameter in contig length

	HYT-14 redundant reads, min50								HYT-14 redundant reads, min100							
K-mer	**11**	**13**	**15**	**17**	**19**	**21**	**23**	**all**	**11**	**13**	**15**	**17**	**19**	**21**	**23**	**all**
Number of contigs	1	247	1,614	1,185	612	113	11	3,782	1	125	919	523	205	26	4	1,802
N50 of contigs	0	103	159	132	97	87	92	132	0	117	199	183	147	126	107	178
Average length of contigs	0	103	146	128	104	91	90	129	0	122	198	188	153	127	109	184
Contig maximum length	0	271	1,205	**1,766**	1,036	195	128	1,766	0	271	1,205	**1,765**	1,036	195	128	1,765
	HYT-21 redundant reads, min50								**HYT-21 redundant reads, min100**							
K-mer	**11**	**13**	**15**	**17**	**19**	**21**	**23**	**all**	**11**	**13**	**15**	**17**	**19**	**21**	**23**	**all**
Number of contigs	1	122	1,881	1,055	281	52	7	3,398	1	47	782	291	77	22	1	1,220
N50 of contigs	0	99	123	100	93	114	85	111	0	116	179	215	162	142	100	177
Average length of contigs	0	98	121	113	102	107	85	116	0	123	182	206	165	143	100	183
Contig maximum length	0	207	898	**1,542**	585	225	100	1,542	0	207	898	**1,542**	585	225	100	1,542

Furthermore, the extension of 5' and 3' extremities by mapping Oases contigs on the reference or reconstructed sequence (as done in sample HYT-21, FLMaV-1 genome) would have been possible using both cut-off parameters, as shown in Fig. 79. However, the coverage resulted to be higher for min50 Oases as well as for Seqman contigs.

Regarding Seqman, we noticed that this tool is not always able to build a longer contig starting from an heterogeneous pool of shorter Oases contigs. Furthermore, it could create chimeric or even rearranged contigs, as for example observed in the potato sample BPO-147 infected with the two PVY strains, where the longest min50 and min100 filtered Seqman contig resulted to be 10,654 and 17315 nt long, clearly an assembly artefact, since the PVY genome is long only 9.7 kb.

Taking together all of this observations and considering that the min100 cut-off did not bring any advantage, compared to the min50 where at least the genome coverage is increased, it sounds reasonable to suggest to use the minimum contigs and transcripts length set at 50 for

future analysis. This parameter was found to play an essential role in a proficiency test organized in the frame of COST Action on NGS and plant virus diagnosis, in which the cut-off min100 in our pipeline prevented the detection of a low-titre virus in an apple leaf sample (Massart et al. 2017, submitted)

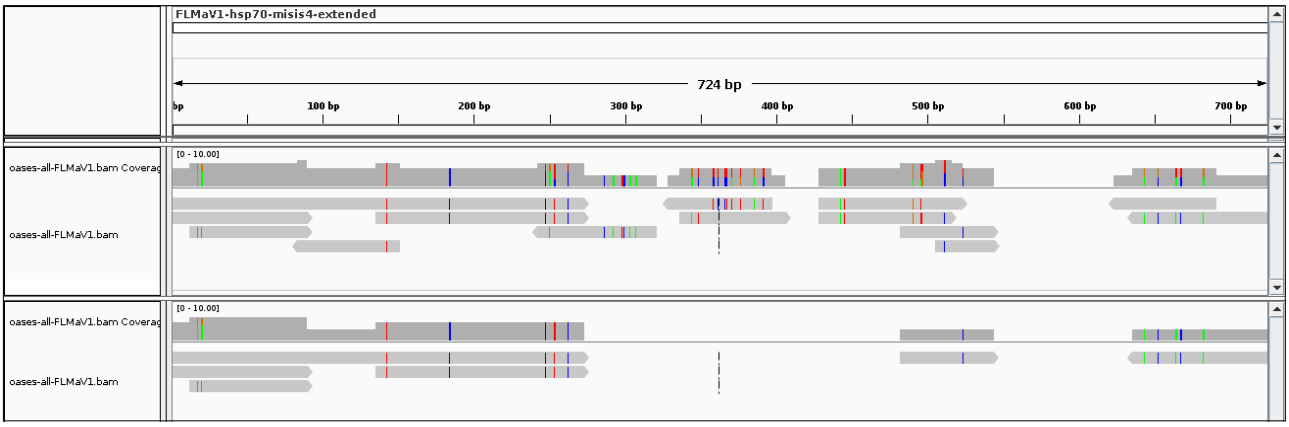


Fig 79. IGV profile of mapped redundant contigs on reconstructed FLMaV-1 virus. On the top, Oases contigs derived from min50 cut-off that cover more the reconstructed genome. On the bottom,Oases contigs derived from min100 cut-off, missing coverage at the central part of the genome. The contigs at the 5' and 3'l extremities share 100% of similarity and they could have used to extend the reconstructed sequence. The gray bars represent the mapped contigs along the genome, with the arrow pointing in the forward or reverse orientation.

Moreover, with the aim to overcome troubleshooting of gaps, contigs length and related coverage, other *de novo* assembly tools like SOAPdenovo, Trinity, Geneious, CAP3, SPAdes were also used for some of the samples in a parallel analysis, reaching overall the same results in term of contig length and genome coverage, suggesting that if the team is not qualified enough (i.e. RNA quality, sequencing depth, reads quality), changing the coach (i.e. bioinformatic tools) would have not helped to win the league (i.e. virome reconstruction). Indeed, the above mentioned troubleshooting, especially for the HYT libraries sequenced in half lane of the HiSeq platform, could also be due to the difficulty in assembling heterogeneous reads, suggesting that a depletion of ribosomal RNA or enrichment of the viral reads could increase the sensibility for viral reconstruction, as previously tested by Visser et al., 2016. However, heterogeneity should not be confused with reads redundancy, since the first term is referred to an heterogeneous RNA population within the sample (the host RNA including ribosomal RNA, transfer RNA, chloroplastic RNA and the non-host RNA including different viruses, viroids and others) while the redundancy is referred to the multiplication of a genomic region in several identical reads. This redundancy can arise during preparation of the cDNA library (adapter ligation) and clusters (PCR amplification) and thus, being a technical error or can represent a biological relevant alternative (i.e. variant). Hence, both categories of redundant and non redundant reads were *de novo* assembled in contigs and compared. In general, the number of redundant read contigs resulted to be higher and the contigs longer than the non-redundant ones. However, the above-mentioned assembly artefact observed for the PVY-related Seqman contig, was observed with redundant reads, but not with non-redundant reads, indicating that both

strategies can be useful. Furthermore, by aligning the consensus sequences obtained from redundant or non redundant reads, several SNPs along the genome were observed in any sample analysed, due to the presence of viral quasispecies more than amplification artefact. Thus, the non-redundant sRNA contigs were used in this study to further verify the reconstructed redundant sRNA consensus sequence by mapping and IGV visualization, which would reveal real SNPs supported also by non-redundant contigs. Furthermore, if supported by the contig coverage, the non-redundant reads contigs could also help to unravel a possible problematic region deriving from an assembly artefact. One of this assembly artefact was observed during the reconstruction of FMV RNA1 by reference-based assembly and iteration mapping steps. As previously reported by Dutilh et al. (2009), after each iteration the total number of mapped reads increase due to a consensus sequence becoming more similar to the one supported by the majority of the reads. However, a random incorporation of wrong reads, due to similarity with the reference, would led to a wrong sequence that would be further modified in the next iteration steps, creating an artificial non-viral sequence like the one in Fig. 71.

The consensus sequence obtained by reference-based reconstruction with sRNA mapping iterations should be taken as representative of the virus quasispecies supported by the majority of the reads. Evidence of more than one genetic variant per a given virus in the virome could then be revealed by searching for SNPs in the MISIS-2 logo at certain positions of the viral genome where two or more alternative nucleotides are supported by substantial fractions of the redundant and non-redundant reads.

Such a reconstruction of two (or more) genetic variants of the same virus was extensively discussed above for the PVY twin virome. In that case, a region with a similarity above 87% was bioinformatically difficult to disentangle into consensus sequences representing the two PVY strains and a PCR-based approach was needed to validate the reconstruction. In the case of PepMV-CH2 and PepMV-LP isolates (78% of similarity) the SNPs were identified by alignment of the consensus sequences and comparison with corresponding reference NCBI sequences. Another way of reference-based reconstruction of twin strain viromes could have been to map the reads on one of the two references allowing zero mismatches and then export the unmapped reads to be mapped on the second strain. In this way we could have lost all the reads shared between the two isolates and thus, creating a consensus sequence with gaps except for those unique parts. However, this analysis was done with the previous version of MISIS, which did not support the exporting of consensus with gaps. The MISIS-2 newest version (v. 2.7, Seguin et al. 2016; used for HYT datasets) is taught to keep the reference nucleotides in those positions not supported by reads (gaps), so to have a complete sequence for downstream analysis and molecular validation.

For the same above-mentioned reasons, the “filtering cascade” should be used just for viral diagnostic purposes and not for virus genome reconstruction, since the short reads could be shared among host and viral genome and also between viral species, as seen for example in the case of cherry viruses, where the three viruses share a percentage of similarity below 30% but nonetheless some short reads derived from a given virus were removed by filtering through the reference genomes of the host and two other viruses, leading to a different consensus sequence of the viral genome, compared to that obtained using just one filtering step through the host genome. Thus, if one is interested in reconstruction of a viral genome,

the complete data sets of short reads should be used in order to obtain the correct consensus sequence, supported by majority of the viral reads. Furthermore, in the *de novo* assembly strategy, the filtering throughout the host genome should be performed after the sRNA assembly itself, to ensure that those shared host/viral reads are kept to eventually contribute to the viral contig assembly, as was demonstrated by Seguin et al. (2014b).

Furthermore, we found that a reference-based approach, using BWA with default parameters (maximum edit distance 0.04), could be applied only using a closely related reference genome (i.e. same species and not genus), since default BWA mapping of the fig closteroviruses-derived sRNA on the related closterovirus CTV or LigMV-derived sRNAs on the related hordeivirus BSMV did not bring any mapped reads, besides the few ones with the highest similarity. Note that also the default parameters of alignment tools could influence either the reconstruction or the diagnostic analysis, as seen when the contigs were blasted online or against local viral databases. In fact, we noticed that the number of virus-related contigs could be different between BWA/Samtools and NCBI blastn due to different default settings.

In conclusion, NGS is indeed a powerful tool for sRNA-based virus detection and virome reconstruction but precautions have to be taken when planning the experiment, by choosing carefully the RNA isolation protocol according to the sample nature, the software and assembly parameters for *de novo* reconstruction or the reference genome and mapping parameters for the reference-based reconstruction.

4.2 siRomics for characterization of the host RNA silencing-based antiviral defences

In this study, we further verified that RNA silencing is a conserved mechanism among different cultivated and non-cultivated (wild) plant species against viruses belonging to different genera and families. Even though the relative abundance, in term of percentage of each virus-specific siRNAs in the total 20-25 nt reads, could be influenced by host-machinery, sensitivity, degree and timing of infection, the overall sRNA size-class profile pattern appeared to be conserved. In fact, the viral 21-nt sRNAs size-class resulted to be always the most abundant, followed by the 22-nt class, with a bias towards 5'U, 5'A and 5'C, suggesting a general involvement of DCL4 and DCL2 orthologs in processing of viral sRNA precursors and AGO1, AGO2 and AGO5 clade proteins in sorting the viral sRNAs to form antiviral RISCs. In addition for the DNA viruses such as FBV-1 detected in fig and the nuclear viroids such as HSVd, GYSVd-1 and other viroids detected in grapevine, the presence of 24-nt viral sRNAs implicates DCL3 orthologs in targeting the viral agents visiting the plant nucleus, in addition to DCL2 and DCL4 that target those agents in the cytoplasm.

Our cross-protection trial of cultivated tomato plants, infected with LP strain of PepMV and treated with the commercial protective mild isolated strain CH2, PMV-01, revealed that CH2 did not protect the plants from the invasion with LP, and vice versa LP did not protect plants from the invasion with CH2. This is despite the potential reciprocal cross-targeting of one strain by viral siRNAs derived from another strain. The co-existence in a natural potato

virome of two closely-related strains of PVY with even larger cross-targeting potentials of viral siRNAs further illustrates that RNA silencing-based cross-protection may fail under field conditions. Moreover our reconstruction of the consensus genome sequences from tomato plants infected either with one and two strains of PepMV revealed a fast evolution the two strain quasispecies within the same plant analysed at different time points, thus raising further concerns regarding the cross-protection approach. Since the genetic stability is one of the key requirements in cross-protection, the appearance and fixation of point mutations observed in our study, highlight the possibility that the protective strain may eventually evolve to a more aggressive genetic variant, causing even more severe disease symptoms. Mutations in attenuated isolates of PepMV used for cross-protection have already been recorded (Vermunt et al. 2017). However, our findings of the ongoing quasispecies evolution within a plant are related to only four out of 100 plants in the greenhouse and no information is available for the rest of the plants. Future research should therefore address the stability of PepMV quasispecies at the plant population level in the field.

In conclusion, using the siRomics approach we were able to fully reconstruct the consensus sequence of PVY and PVX in single sRNA contigs and to discover the new hordeivirus LigMV and reconstruct de novo its three genomic segments. From a diagnostic point of view, the pipeline was helpful to identify components of the viromes present in cherry, grapevine and fig, in which the consensus sequences were only partially reconstructed due to gaps in the sRNA coverage. Two strains of the same potyvirus PVY sharing nearly identical regions virus could not be *de novo* reconstructed, but the viral contig analysis gave the necessary information for reference-based reconstruction and for designing an RT-PCR experiment, which allowed to reliably disentangle the 3' region of the viral strains. Our findings for FBV-1 in FMV-infected and healthy fig trees suggest that a host genome-integrated DNA badnavirus such as FBV-1 can be released as replicating episomal virus upon infection with an unrelated RNA virus such as FMV.
Thus, the siRomics pipeline is a powerful tool for viral diagnostics, virome reconstruction, virus diversity and quasispecies evolution studies as well as for characterization of RNA silencing-based antiviral defences in both cultivated and non-cultivated wild plants.

5. List of abbreviations

AGO Argonaute protein
AGV Australian grapevine viroid
BAM Binary Alignment/Map
BMV Brome mosaic virus
BSMV Barley stripe mosaic virus
BWA Burrows-Wheeler Alignment
BWT Burrows-Wheeler Transform
CaMV Cauliflower Mosaic Virus
CH2 Chilean genotype
CTV Citrus tristeza virus
CVA Cherry virus A
CP Coat Protein
DCL DiCer-Like protein
dNTP deoxy nucleoside triphosphate
dsDNA double-stranded DNA
dsRNA double-stranded RNA
ELISA Enzyme-Linked ImmunoSorbent Assay
ETI Effector triggered immunity
EU European genotype
FMV Fig mosaic virus
GFLV Grapevine fanleaf virus
GLRaV Grapevine leafroll-associated virus
GYSVd Grapevine Yellow Speckle Viroid
GVA Grapevine virus A
HEN1 Hua enhancher 1
HSVd Hop Stunt Viroid
ICTV International Committee on Taxonomy of Viruses
IGV Integrative Genomics Viewer
LChV-1 Little cherry virus 1
LChV-2 Little cherry virus 2
LP Lycopersicon peruvianum
miRNA microRNA
MP Movement Protein
NGS Next-Generation Sequencing
ORF Open Reading Frame
PAMP Pathogen associated molecular patterns
PAZ Piwi Argonaute Zwille
PCR Polymerase Chain Reaction
PepMV Pepino Mosaic virus

pgRNA pregenomic RNA
PMV-01 Pepino mosaic virus vaccine
PTI PAMP/pattern-triggered immunity
PIWI P-element induced Wimpy testis
pri-miRNA primary miRNA
pre-miRNA precursor miRNA
PTGS Post-Transcriptional Gene Silencing
PVX Potato virus X
PVY Potato virus Y
RCA Rolling Circle Amplification
RdRP RNA-Dependant RNA polymerase
RISC RNA-Induced Silencing Complex
rRNA ribosomal RNA
RT reverse transcription
RT-PCR Real-Time Polymerase Chain Reaction
SAM Sequence Alignment/Map
SGS3 Suppressor of gene silencing 3
siRNA short interfering RNA
SNP Single Nucleotide Polymorphism
ssDNA single-stranded DNA
ssRNA single-stranded RNA
TGS Transcriptional Gene Silencing
TMV Tobacco Mosaic Virus
VIGS Virus-induced gene silencing
vsiRNA viral siRNA

6. Acknowledgements

Aware of the fact that a few lines are not enough, I would like nevertheless to express my gratitude towards all the persons that have helped me in achieving this important goal in my academic career.
My first mention goes to my supervisors, PD Dr. Mikhail Pooggin and Thomas Boller, for giving me the opportunity to conduct a stimulating and challenging research at the University of Basel. I want to thank Dr. Pooggin for all the exciting discussion about molecular biology, the friendly environment, the encouragement and support. His guidance and knowledge, together with the experience and the wisdom of Thomas Boller, were essential and precious.
I would like to thank Dr. Olivier Schumpp and Jean-Sebastien Reynard, for their interest towards my project and for believing in me. Plus, by sharing their expertise regarding plant viruses and their ongoing projects at Agroscope, they gave me the opportunity to further apply the siRomics pipeline in different infected crops.
I thank all my lab members, especially Jonathan Seguin for teaching me all the bioinformatics tools and helping me with troubleshooting.
There are not enough words to thank all the other members of the Institute for the amazing time we spent together. Michi, Lukas, Markus, Maura, Jürg, Olivier, with you guys I felt part of a big family.

All my love goes to my family and my friends. To my parents, *meravigliose creature*, who have always supported me, reminding me I am never alone, even if a thousand kilometres away. To my brother and my sister, because *together we stand, divided we fall*.
To my girls, Elena, Simonina, Alessia and Simona, because *you see all my light, you love my dark* and *we are so happy together.* To Laura, because *a friend like you is better.* To Willeke, because when I am with you *all my troubles seem so far away.*
To all the other people in Basel, specially Roberta and Rosario, with whom I have spent unforgettable time in the last years.
Thank you all for standing by my side every day of my PhD life.

7. References

Aapola, A. I. E. & Rochow, W. F. Relationships among three isolates of barley yellow dwarf virus. *Virology* 46, 127–141 (1971).

Adams I., Fox A. Diagnosis of Plant Viruses Using Next-Generation Sequencing and Metagenomic Analysis. In: Wang A., Zhou X. (eds) Current Research Topics in Plant Virology. Springer. (2016).

Adams, M. J., Antoniw, J. F. & Kreuze, J. Virgaviridae: A new family of rod-shaped plant viruses. *Arch. Virol.* 154, 1967–1972 (2009).

Ahmadvand, R., Takács, a., Taller, J., Wolf, I. & Polgár, Z. Potato viruses and resistance genes in potato. *Acta Agron. Hungarica* 60, 283–298 (2012).

Akbergenov, R. *et al.* Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Res.* 34, 462–471 (2006).

Al Rwahnih, M., Dolja, V. V., Daubert, S., Koonin, E. V. & Rowhani, A. Genomic and biological analysis of Grapevine leafroll-associated virus 7 reveals a possible new genus within the family Closteroviridae. *Virus Res.* 163, 302–309 (2012).

Alimoradian, M., Rakhshandehroo, F. & Shams-Bakhsh, M. Prevalence and phylogenetic analysis of Fig mosaic virus and Fig badnavirus-1 in Iran. *J. Plant Prot. Res.* 56, 122–128 (2016).

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Altschul et al.. 1990. Basic Local Alignment Search Tool.pdf. *Journal of Molecular Biology* 215, 403–410 (1990).

Aregger, M. *et al.* Primary and Secondary siRNAs in Geminivirus-induced Gene Silencing. *PLoS Pathog.* 8, (2012).

Baeg, K., Iwakawa, H. & Tomari, Y. The poly(A) tail blocks RDR6 from converting self mRNAs into substrates for gene silencing. *Nat. Plants* 3, 17036 (2017).

Bai, M.,Yang, G., Chen, W., Mao, Z., Kang, H., Chen, G., Yang, Y., Xie, B.,. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in Solanum lycopersicum. *Gene* 501, 52–62 (2012).

Barba, M., Czosnek, H. & Hadidi, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses* 6, 106–136 (2014).

Batten, J. S., Yoshinari, S. & Hemenway, C. Potato virus X: A model system for virus replication, movement and gene expression. *Molecular Plant Pathology* 4, 125–131 (2003).

Baulcombe, D. RNA silencing in plants. *Nature* 431, 356–363 (2004).

Bertin, S. *et al.* Transmission of *Grapevine virus A* and *Grapevine leafroll-associated virus 1* and *3* by *Heliococcus bohemicus* (Hemiptera: Pseudococcidae) Nymphs From Plants With Mixed Infections. *J. Econ. Entomol.* 109, 1504–1511 (2016).

Bhat, A. I., Hohn, T. & Selvarajan, R. Badnaviruses: The current global scenario. *Viruses* (2016).

Blevins, T. *et al.* Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* 34, 6233–6246 (2006).

Blevins, T. *et al.* Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Res.* 39, 5003–5014 (2011).

Bodaghi, S., Mathews, D. M. & Dodds, J. A. Natural Incidence of Mixed Infections and Experimental Cross Protection Between Two Genotypes of Tobacco mild green mosaic virus. *Phytopathology* 94, 1337–41 (2004).

Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).

Borges, F. & Martienssen, R. A. The expanding world of small RNAs in plants. *Nat. Publ. Gr.* 16, 1–15 (2015).

Boualem, A., Dogimont, C. & Bendahmane, A. The battle for survival between viruses and their host plants. *Curr. Opin. Virol.* 17, 32–38 (2016).

Bouché, N., Lauressergues, D., Gasciolli, V. & Vaucheret, H. An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* 25, 3347–56 (2006).

Bovey, R. (1976). Particules de type ‘Tobravirus’ associees a une mosaique jaune du troene (Ligustrum vultare L.). Poljoprivredna znanstvena smotra, 39, 569-73. 1976 (1976).

Burgyán, J. & Gáborjányi, R. Cross-Protection and Multiplication of Mild and Severe Strains of TMV in Tomato Plants. *J. Phytopathol.* 110, 156–167 (1984).

Cao, M. *et al.* Virus infection triggers widespread silencing of host genes by a distinct class of endogenous siRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14613–8 (2014).

Choi, H., Cho, W. K. & Kim, K. Two homologous host proteins interact with potato virus X RNAs and CPs and affect viral replication and movement. *Nat. Publ. Gr.* 1–12 (2016).

Chung, B. Y.-W., Miller, W. A., Atkins, J. F. & Firth, A. E. An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci.* 105, 5897–5902 (2008).

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771 (2009).

Condit, I. J., and Horne, W. T. A mosaic of fig in California. Phytopathology, 23:887-896 (1933).

Costa, A. S. & Muller, G. W. Tristeza control by cross protection: a U.S.-Brazil cooperation sucess. *Plant Dis.* 64, 538–541 (1980).

De Nayer, F. *et al.* Cross-protection as a control strategy for Pepino mosaic virus (PepMV) in greenhouse tomato. in *Acta Horticulturae* 914, 163–169 (2011).

Deleris, A. Hierarchical Action and Inhibition of Plant Dicer-Like Proteins in Antiviral Defense. *Science (80-.).* 313, 68–71 (2006).

Dietzgen, R. G., Mann, K. S. & Johnson, K. N. Plant virus-insect vector interactions: Current and potential future research directions. *Viruses* 8, (2016).

Donaire, L. *et al.* Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* 392, 203–214 (2009).

Dougherty, W. G. *et al.* RNA-mediated virus resistance in transgenic plants: exploitation of a cellular pathway possibly involved in RNA degradation. *Mol Plant Microbe Interact* 7, 544–552 (1994).

Dutilh, B. E., Huynen, M. A. & Strous, M. Increasing the coverage of a metapopulation consensus genome by iterative read mapping and assembly. *Bioinformatics* 25, 2878–2881 (2009).

Elbeaino, T., Digiaro, M. & Martelli, G. P. Complete nucleotide sequence of four RNA segments of fig mosaic virus. *Arch. Virol.* 154, 1719–1727 (2009).

Elbeaino, T., Digiaro, M. & Martelli, G. P. RNA-5 and -6, two additional negative-sense RNA segments associated with fig mosaic virus. *J. Plant Pathol.* 94, 421–425 (2012).

Elbeaino, T., Digiaro, M., Heinoun, K., Stradis, A. De & Martelli, G. P. Fig mild mottle-associated virus , a novel closterovirus infecting fig. 92, 165–172 (2010).

Elbeaino, T., Digiaro, M., Stradis, A. De & Martelli, G. P. Identification of a second member of the family closteroviridae in mosaic-diseased figs. 89, 119–124 (2007).

Fang, X. & Qi, Y. RNAi in Plants: An Argonaute-Centered View. *Plant Cell* 28, 272–85 (2016).

Filloux, D. ,Dallot, S., Delaunay, A., Galzi, S., Jacquot, E., Roumagnac, P., Metagenomics Approaches Based on Virion-Associated Nucleic Acids (VANA): An Innovative Tool for Assessing Without A Priori Viral Diversity of Plants. Plant Pathology: Techniques and Protocols, Pages 249-257. (2015).

Fuentes, A. *et al.* Field Trial and Molecular Characterization of RNAi-Transgenic Tomato Plants That Exhibit Resistance to Tomato Yellow Leaf Curl Geminivirus. *Mol. Plant-Microbe Interact.* 29, 197–209 (2016).

Fukudome, A. & Fukuhara, T. Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *J. Plant Res.* 130, 1–12 (2016).

Gambino, G., Perrone, I. & Gribaudo, I. A rapid and effective method for RNA extraction from

different tissues of grapevine and other woody plants. *Phytochem. Anal.* 19, 520–525 (2008).

Garcia-Ruiz, H. *et al.* Arabidopsis RNA-Dependent RNA Polymerases and Dicer-Like Proteins in Antiviral Defense and Small Interfering RNA Biogenesis during Turnip Mosaic Virus Infection. *Plant Cell* 22, 481–496 (2010).

Gonsalves, D., Station, G. & Wang, H. Control of papaya ringspot virus by cross protection. *Plan disase* 72, 375–380 (1988).

Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17, 333–351 (2016).

Graniti, A. & Martelli, G.P. 1966. Further observations on legno riccio rugose wood, a graft transmissible stem pitting of grapevine. Proc. Int. Conf: Virus Vector Perennial Hosts and Vitis, 1965, p. 168-179. Div. Agric. Sci., Univ. Calif., Davis. 1966 (1966).

Hanssen, I. M. & Thomma, B. P. H. J. Pepino mosaic virus: A successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Mol. Plant Pathol.* 11, 179–189 (2010).

Hanssen, I. M. *et al.* Cross-protection or enhanced symptom display in greenhouse tomato co-infected with different Pepino mosaic virus isolates. *Plant Pathol.* 59, 13–21 (2010).

Hanssen, I. M. *et al.* Pepino mosaic virus isolates and differential symptomatology in tomato. *Plant Pathol.* 58, 450–460 (2009).

Hasiów-Jaroszewska, B. *et al.* Ratio of mutated versus wild-type coat protein sequences in Pepino mosaic virus determines the nature and severity of yellowing symptoms on tomato plants. *Mol. Plant Pathol.* 14, 923–933 (2013).

Hasiów-Jaroszewska, B., Minicka, J. & Pospieszny, H. Cross-protection between different pathotypes of Pepino mosaic virus representing Chilean 2 genotype. *ACTA Sci. Pol.* 13, 177–185 (2014).

Havecker, E. R. *et al.* The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *Plant Cell* 22, 321–334 (2010).

Heitefuss, R. Virus and Virus-like Diseases of Pome and Stone Fruits. *J. Phytopathol.* 160, 508 (2012).

Huang, J., Yang, M., Lu, L. & Zhang, X. Diverse functions of small RNAs in different plant-pathogen communications. *Front. Microbiol.* 7, 1–17 (2016).

Ishikawa, K. *et al.* Nucleocapsid protein from fig mosaic virus forms cytoplasmic agglomerates that are hauled by endoplasmic reticulum streaming. *J. Virol.* 89, 480–91 (2015).

Jelkmann, W. herry virus A: cDNA cloning of dsRNA, nucleotids sequence analysis and serology reveal a new plant capillovirus in sweet cherry. *J. Gen. Virol.* 76, 2015–2024 (1995).

Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* 444, 323–329 (2006).

Jones, R. A. C., Koenig, R. and Jesemann, D. E. Pepino mosaic virus, a new potexvirus from pepino (Solanum muricatum). *Ann. Appl. Biol.* 94, 61–68 (1980).

Karthikeyan, C. *et al.* Emergence of a latent indian cassava mosaic virus from cassava which recovered from infection by a non-persistent Sri Lankan cassava mosaic virus. *Viruses* 8, (2016).

Kosaka, Y. & Fukunishi, T. Multiple inoculation with three attenuated viruses for the control of cucumber virus disease. *Plant Dis.* 81, 733–738 (1997).

Kravchik, M., Damodharan, S., Stav, R. & Arazi, T. Generation and characterization of a tomato DCL3-silencing mutant. *Plant Sci.* 221–222, 81–89 (2014).

Laney, A. G., Hassan, M. & Tzanetakis, I. E. An integrated badnavirus is prevalent in fig germplasm. *Phytopathology* 102, 1182–9 (2012).

Li, R. *et al.* Deep Sequencing of Small RNAs in Tomato for Virus and Viroid Identification and Strain Differentiation. 7, 1–10 (2012).

Lindbo, J. A. & Dougherty, W. G. Pathogen-derived resistance to a potyvirus: immune and resistant phenotypes in transgenic tobacco expressing altered forms of a potyvirus coat protein nucleotide sequence. *Molecular plant-microbe interactions : MPMI* 5, 144–53 (1992).

Lindbo, J. A. & Dougherty, W. G. Plant Pathology and RNAi: A Brief History. *Annu. Rev. Phytopathol.* 43, 191–204 (2005).

Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. & Dougherty, W. G. lnduction of a Highly Specific Antiviral State in Transgenic Plants: lmplications for Regulation of Gene Expression and Virus Resistance. *Plant Cell Am. Soc. Plant Physiol.* 5, 1749–1759 (1993).

Liu, X., Lu, T., Dou, Y., Yu, B. & Zhang, C. Identification of RNA silencing components in soybean and sorghum. *BMC Bioinformatics* 15, 4 (2014).

Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics* 14, 265–279 (2016).

Machado, J. P. B., Calil, I. P., Santos, A. A. & Fontes, E. P. B. Translational control in plant antiviral immunity. *Genet. Mol. Biol.* 1, 0–0 (2017).

Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–80 (2005).

Martelli, G. P. Directory of virus and virus-like diseases of the grapevine and their agents. *J. Plant Pathol.* 96, 1–136 (2014).

Martelli, G. P. *et al.* Taxonomic revision of the family Closteroviridae with special reference to the Grapevine leafroll-associated members of the Genus Ampelovirus and the putative species unassigned to the family. *J. Plant Pathol.* 94, 7–19 (2012).

Miller, J. R., Koren, S. & Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315–327 (2010).

Milton, Z. Viral Cross Protection: More Understanding is Needed. *Phytopathology* 66:382, (1976).

Minafra, A. & Hadidi, A. Sensitive detection of grapevine virus A, B, or leafroll-associated III from viruliferous mealybugs and infected tissue by cDNA amplification. *J. Virol. Methods* 47, 175–187 (1994).

Mirzaei, K., Bahramnejad, B., Shamsifard, M. H. & Zamani, W. In silico identification, phylogenetic and bioinformatic analysis of argonaute genes in plants. *Int. J. Genomics* 2014, (2014).

Molnar, A. *et al.* Plant Virus-Derived Small Interfering RNAs Originate Predominantly from Highly Structured Single-Stranded Viral RNAs. *J. Virol.* 79, 7812–7818 (2005).

Morante-Carriel, J. *et al.* RNA isolation from loquat and other recalcitrant woody plants with high quality and yield. *Anal. Biochem.* 452, 46–53 (2014).

Mori, K., Shirasawa, K., Nogata, H., Hirata, C. & Tashiro, K. Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (Ficus carica). 1–12 (2017).

Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40, (2012).

Osman, T. A. M., Olsthoorn, R. C. L. & Livieratos, I. C. Role of the Pepino mosaic virus 3'-untranslated region elements in negative-strand RNA synthesis in vitro. *Virus Res.* 190, 110–117 (2014).

Padmanabhan, C. *et al.* Complete Genome Sequence of Southern tomato virus Identified in. 3, 5–6 (2015).

Pagán, I. *et al.* Genetic Structure of the Population of Pepino mosaic virus Infecting Tomato Crops in Spain. *Phytopathology* 96, 274–279 (2006).

Palukaitis, P. and Zaitlin, M (1984) A model to explain the cross protection phenomenon shown by plant viruses and viroids. Plant Microbe interactions, Vol. 1, pp.420-429 Macmillan, New York. 1, 1984 (1984).

Pinck, L., Fuchs, M., Pinck, M., Ravelonandro, M. & Walter, B. A Satellite RNA in Grapevine Fanleaf Virus Strain F13. *J. Gen. Virol.* 69, 233–239 (1988).

Pooggin, M. M. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *Int. J. Mol. Sci.* 14, 15233–15259 (2013).

Pooggin, M. M. in *Plant-Virus Interactions: Molecular Biology, Intra- and Intercellular Transport* (ed. Kleinow, T.) 161–189 (Springer International Publishing, 2016).

Rajendran Rajeswaran, Victor Golyaev, Jonathan Seguin, Anna Zvereva, L. & Farinelli, and M. M. P. Interactions of rice tungro bacilliform pararetrovirus and its protein P4 with plant RNA silvencing machinery. *Mol. plant -microbe Interact.* 2, 1–45 (2014b).

Reynard, J.-S., Schneeberger, P. H. H., Frey, J. E. & Schaerer, S. Biological, Serological, and Molecular Characterization of a Highly Divergent Strain of *Grapevine leafroll-associated virus 4* Causing Grapevine Leafroll Disease. *Phytopathology* 105, 1262–1269 (2015).

Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 (2011).

Roossinck, M. J., Martin, D. P. & Roumagnac, P. Plant virus metagenomics: Advances in virus discovery. *Phytopathology* PHYTO12140356RVW (2015).

Sabanadzovic, S., Valverde, R. A., Brown, J. K., Martin, R. R. & Tzanetakis, I. E. Southern tomato virus: The link between the families Totiviridae and Partitiviridae. *Virus Res.* 140, 130–137 (2009).

Saunders, K., Bedford, I. D., Yahara, T. & Stanley, J. Aetiology: The earliest recorded plant virus disease. *Nature* 422, 831–831 (2003).

Schauer, S. E., Jacobsen, S. E., Meinke, D. W. & Ray, A. DICER-LIKE1: Blind men and elephants in Arabidopsis development. *Trends Plant Sci.* 7, 487–491 (2002).

Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092 (2012).

Schumpp, T. O. *et al.* Tomato spotted wilt virus, TSWV. Ed. Agroscope, Amtra, Changins. 1-2 pp. (2014).

Seguin, J. *et al.* De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS One* 9, 1–8 (2014a).

Seguin, J., Otten, P., Baerlocher, L., Farinelli, L. & Pooggin, M. M. MISIS-2: A bioinformatics tool for in-depth analysis of small RNAs and representation of consensus master genome in viral quasispecies. *J. Virol. Methods* 233, 37–40 (2016).

Seguin, J., Otten, P., Baerlocher, L., Farinelli, L. & Pooggin, M. M. MISIS: A bioinformatics tool to view and analyze maps of small RNAs derived from viruses and genomic loci generating multiple small RNAs. *J. Virol. Methods* 195, 120–122 (2014b).

Sempere, R. N., Elena, S. F. & Aranda, M. A. Mixed Infections of Pepino Mosaic Virus Strains Modulate the Evolutionary Dynamics of this Emergent Virus. 83, 12378–12387 (2009).

Sherwood, J. L. & Fulton, R. W. The specific involvement of coat protein in tobacco mosaic virus cross protection. *Virology* 119, 150–158 (1982).

Shirasawa, K. *et al.* The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding. 0, 1–10 (2017).

Smith, B. Y. K. M. The present status of plant virus research. Biological Reviews, 8: 136-179. (1933).

Smith, K.M. Textbook of Plant Virus Disease, 2nd edn. Churchill, London (1957)

Soler-Aleixandre, S., López, C., Cebolla-Cornejo, J. & Nuez, F. Sources of resistance to Pepino mosaic virus (PepMV) in Tomato. *HortScience* 42, 40–45 (2007).

Steger, G. & Perreault, J. P. in *Advances in Virus Research* 94, 141–172 (2016).

Syller, J. & Grupa, A. The effects of co-infection by different *Potato virus Y* (PVY) isolates on virus concentration in solanaceous hosts and efficiency of transmission. *Plant Pathol.* 63, 466–475 (2014).

Szychowsky, J. a. *et al.* The vein-banding disease syndrome: A synergistic reaction between grapevine viroids and fanleaf virus. *Vitis* 34, 229–232 (1995).

Tolia, N. H. & Joshua-Tor, L. Slicer and the argonautes. *Nat Chem Biol* 3, 36–43 (2007).

Tzanetakis, I. E., Laney, A. G., Keller, K. E. & Martin, R. R. New viruses found in fig exhibiting mosaic symptoms. *21st Int. Conf. Virus other Graft Transm. Dis. Fruit Crop.* 79–82 (2010).

Valli, A., García, J. A. & López-Moya, J. J. Potyviridae. *eLS* 1–10 (2015). doi:10.1002/9780470015902.a0000755.pub3

Vazquez, F., Legrand, S. & Windels, D. The biosynthetic pathways and biological scopes of plant small RNAs. *Trends in Plant Science* 15, 337–345 (2010).

Verchot-Lubicz, J., Ye, C. M. & Bamunusinghe, D. Molecular biology of potexviruses: Recent advances. *Journal of General Virology* 88, 1643–1655 (2007).

Vermunt, A. M. W. & Kaarsemaker, R. C. Multi-genotype cross-protection against Pepino mosaic virus in tomato. *Crop Prot.* 96, 116–122 (2017).

Visser, M., Bester, R., Burger, J. T. & Maree, H. J. Next-generation sequencing for virus detection: covering all the bases. *Virol. J.* 13, 85 (2016).

Walia, J. J., Salem, N. M. & Falk, B. W. Partial Sequence and Survey Analysis Identify a Multipartite, Negative-Sense RNA Virus Associated with Fig Mosaic. *Plant Dis.* 93, 4–10 (2009).

Wang, J. *et al.* Complete nucleotide sequence of little cherry virus 1 (LChV-1) infecting sweet cherry in China. *Arch. Virol.* 161, 749–753 (2016).

Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461, 754–761 (2009).

Wassenegger, M. & Krczal, G. Nomenclature and functions of RNA-directed RNA polymerases. *Trends in Plant Science* 11, 142–151 (2006).

Wu, Q., Ding, S.-W., Zhang, Y. & Zhu, S. Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms. *Annu. Rev. Phytopathol.* 53, 150605182533006 (2014).

Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829 (2008).

Zhang, W. *et al.* A practical comparison of De Novo genome assembly software tools for next-generation sequencing technologies. *PLoS One* 6, (2011).

Zhu, H. & Guo, H. The role of virus-derived small interfering RNAs in RNA silencing in plants. *Sci. China Life Sci.* 55, 119–125 (2012).

Ziebell, H. & Carr, J. P. *Cross-Protection*. *Natural and engineered resistance to plant viruses, Part B* 76, (Elsevier Inc., 2010).

Ziebell, H. & Carr, J. P. Effects of dicer-like endoribonucleases 2 and 4 on infection of Arabidopsis thaliana by cucumber mosaic virus and a mutant virus lacking the 2b counter-defence protein gene. *J. Gen. Virol.* 90, 2288–2292 (2009).

Ziebell, H., Payne, T., Berry, J. O., Walsh, J. A. & Carr, J. P. A cucumber mosaic virus mutant lacking the 2b counter-defence protein gene provides protection against wild-type strains. *J. Gen. Virol.* 88, 2862–2871 (2007).

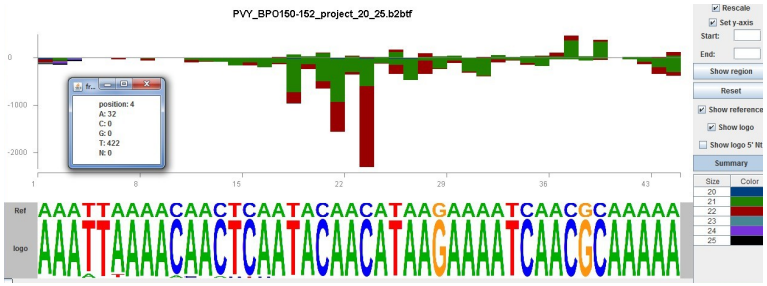
Zvereva, A. S. & Pooggin, M. M. Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* 4, 2578–2597 (2012).

Annex: Supplementary Figure S1:
SNPs in PVY quasispecies of upper leaf samples from three shoots of a potato tuber infected with the PVY-PVX complex

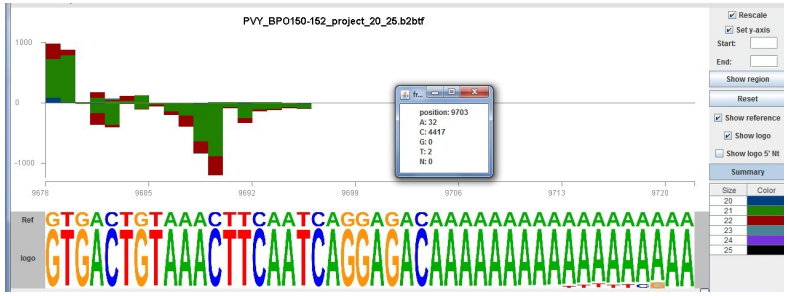
5‘-terminal region of the PVY genome

3‘-terminal region of the PVY genome

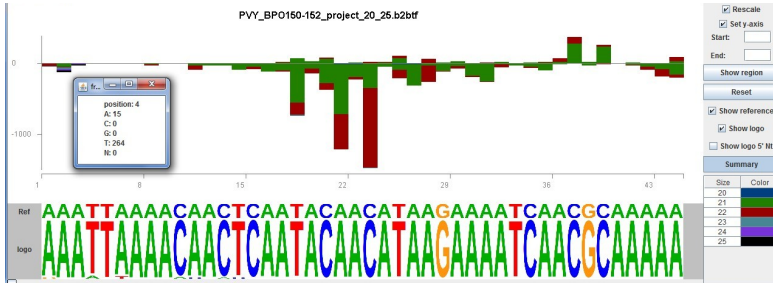
BPO-150



BPO-151

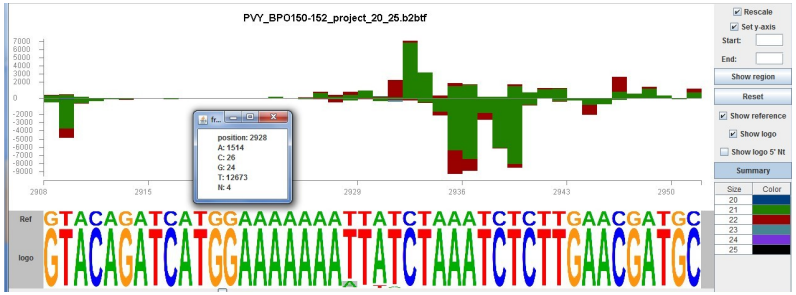


BPO-152

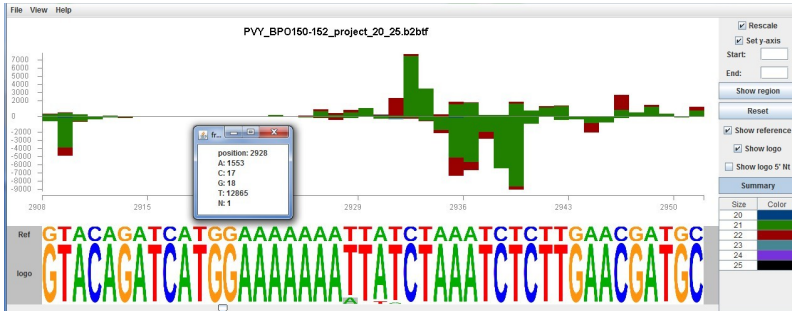


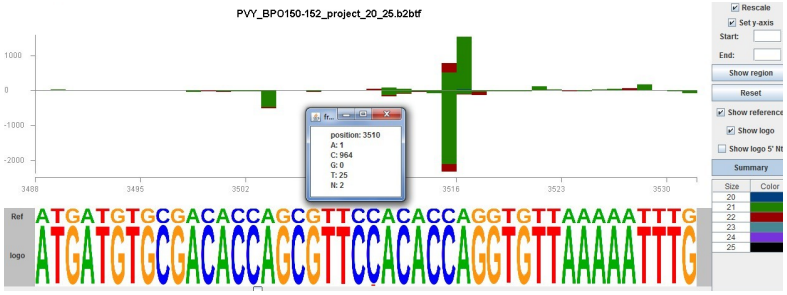
Middle regions of the PVY genome

BPO-150

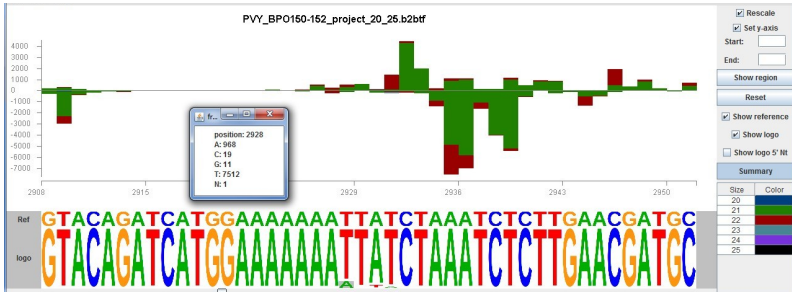


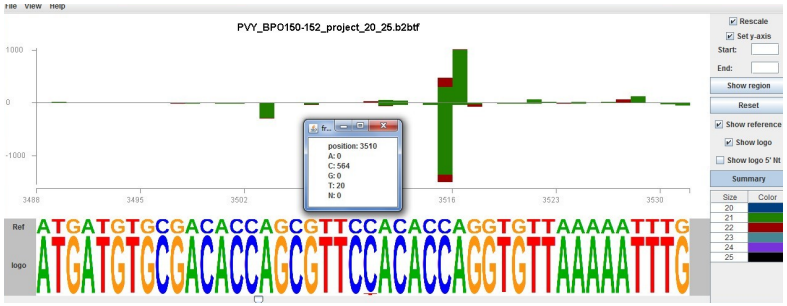
BPO-151





BPO-152

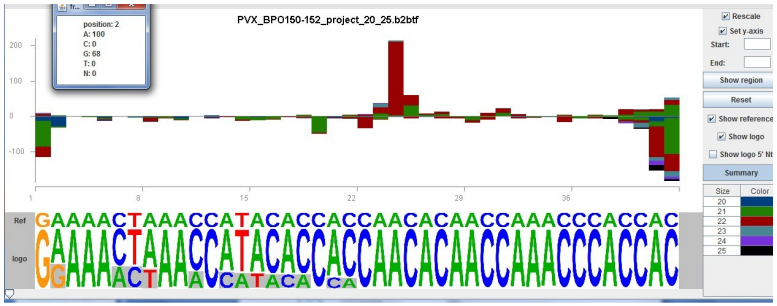


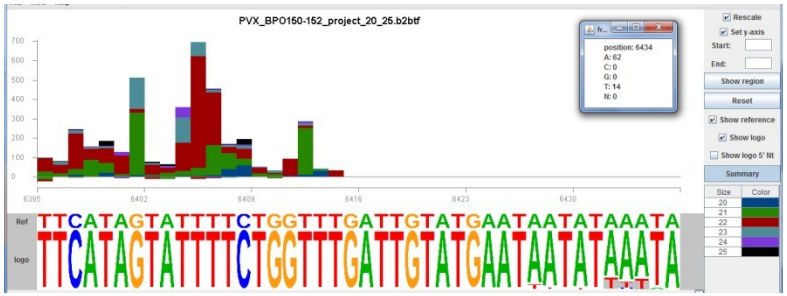


5‘-terminal region of the PVX genome

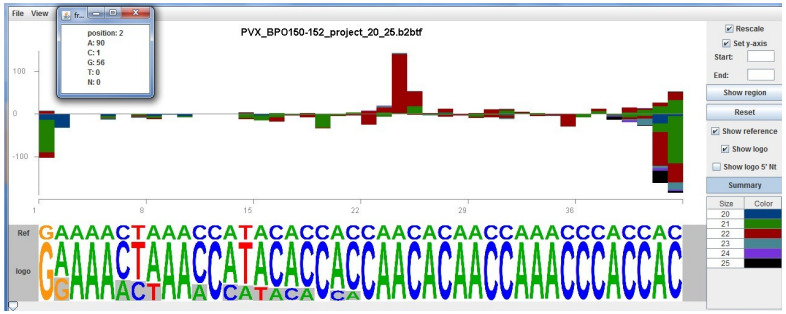
3‘-terminal region of the PVX genome

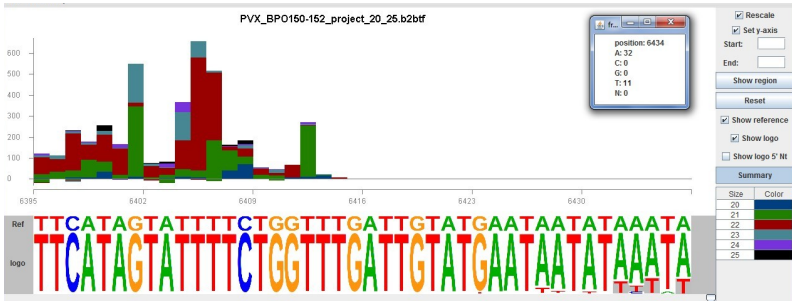
BPO-150



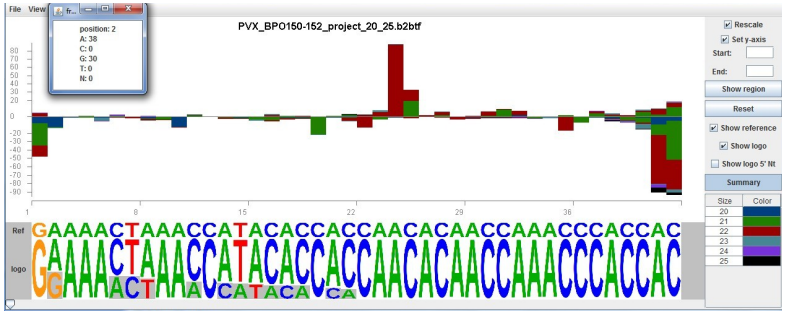


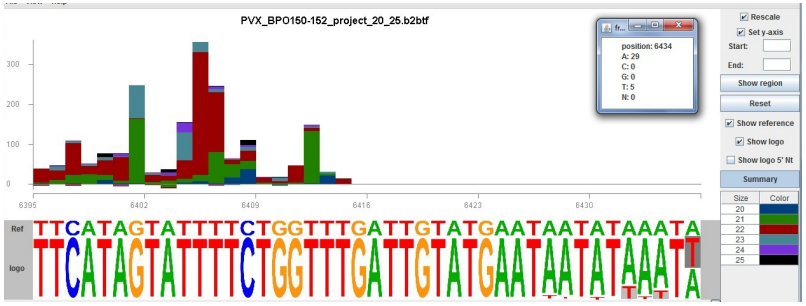
BPO-151





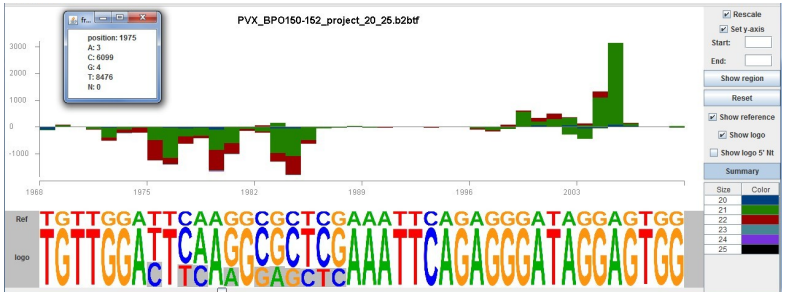
BPO-152

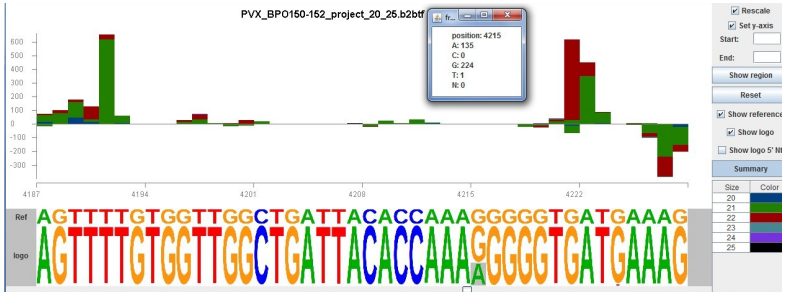




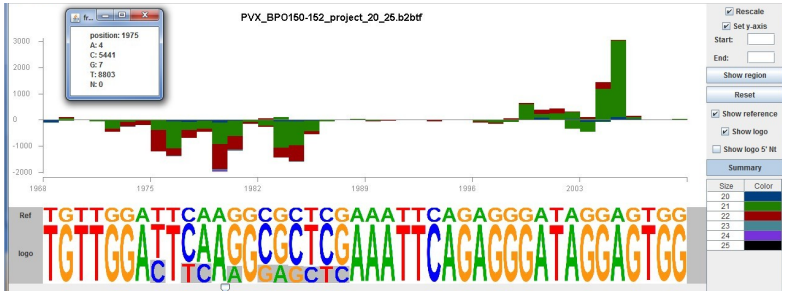
Middle regions of the PVX genome

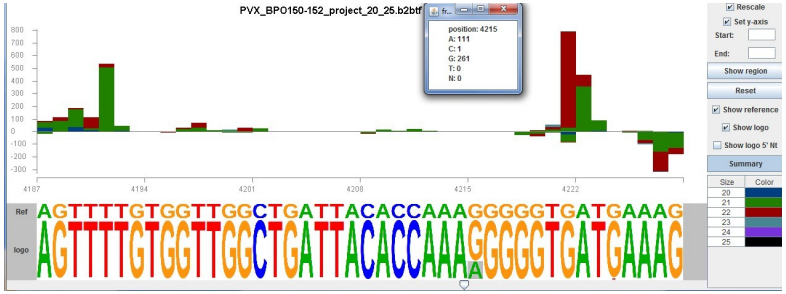
BPO-150



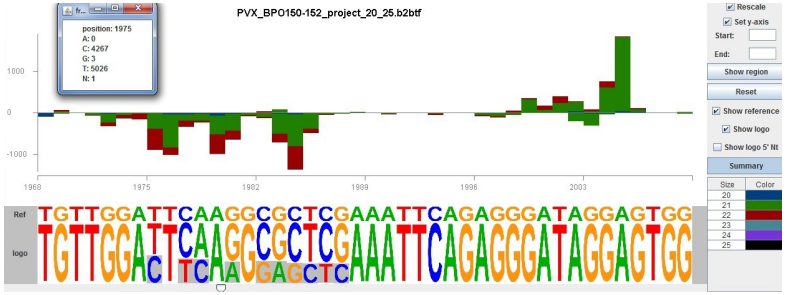


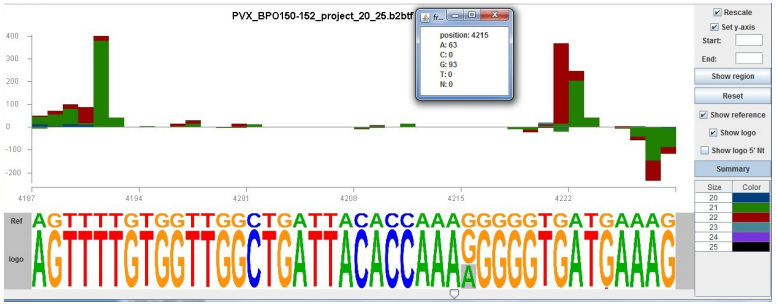
BPO-151





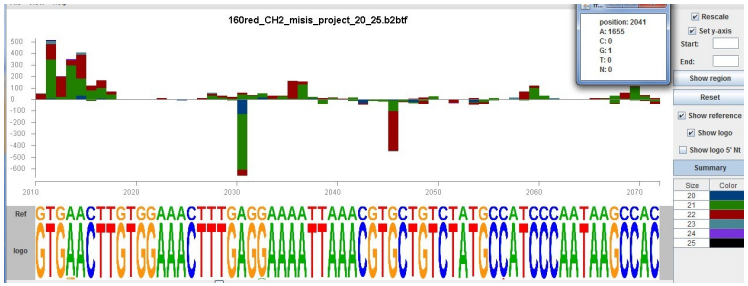
BPO-152



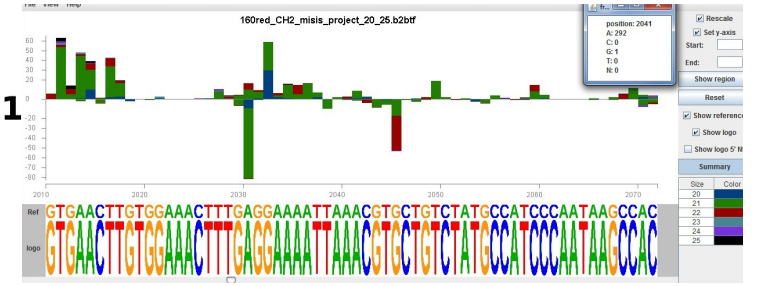


Annex: Supplementary Figure S2: Reconstruction of consensus master genome variants of two distinct PepMV strains
Strain CH2 - Position 2041 (A-to-G variant appeared and fixed in plant 162/163)

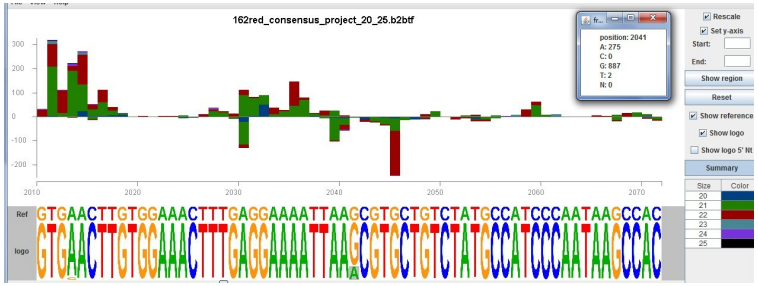
BPO-160



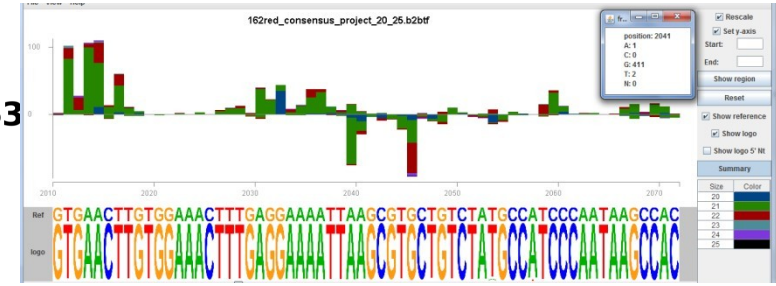
BPO-161



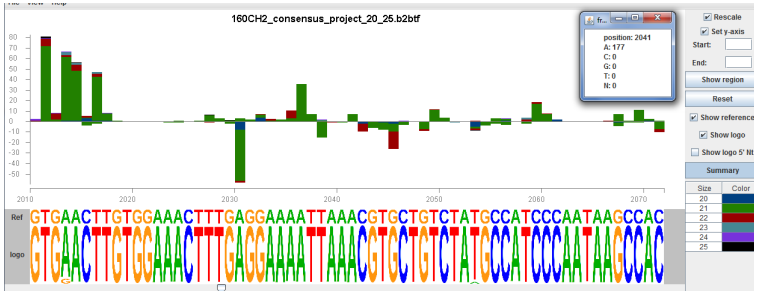
BPO-162



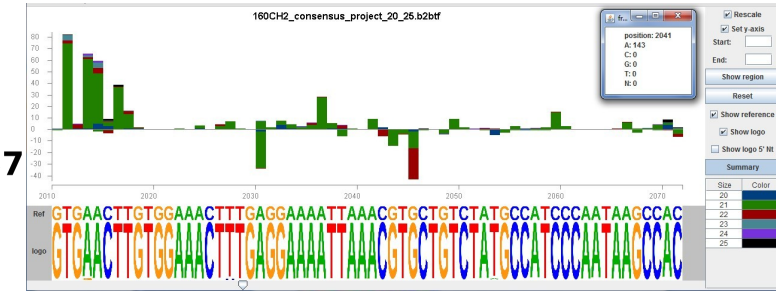
BPO-163



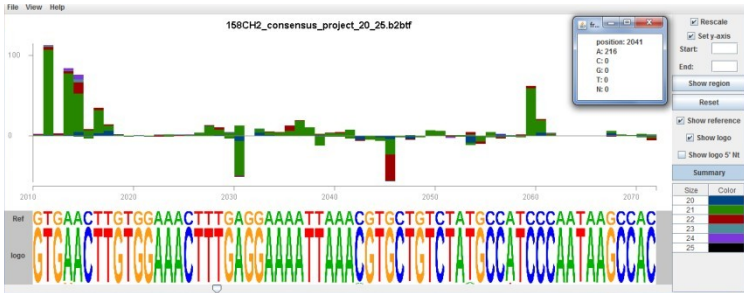
BPO-156



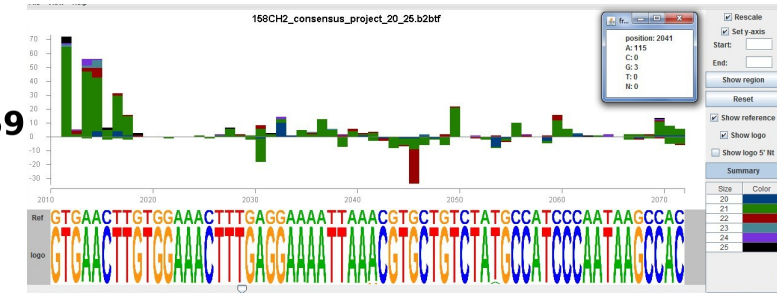
BPO-157



BPO-158

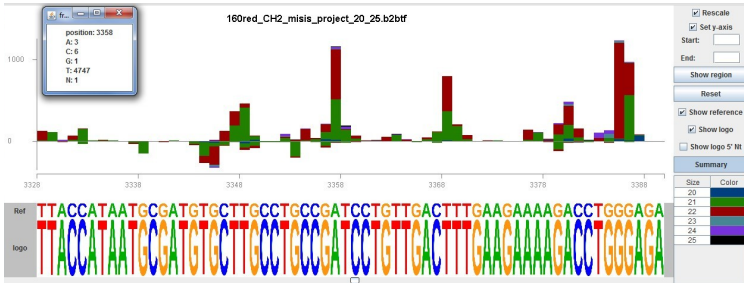


BPO-159

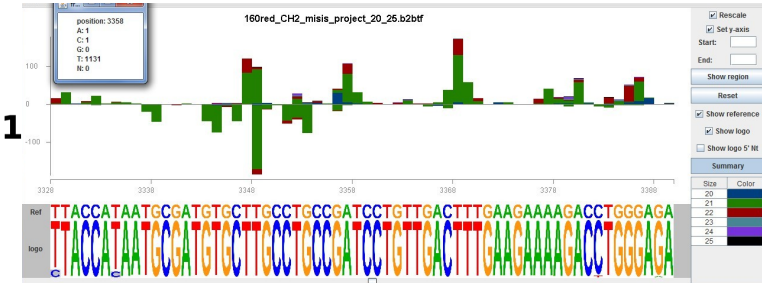


Strain CH2 - Position 3358 (T-to-C variant appeared and fixed in plant 162/163)

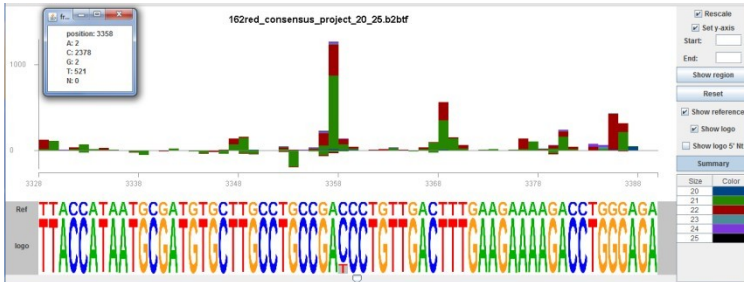
BPO-160



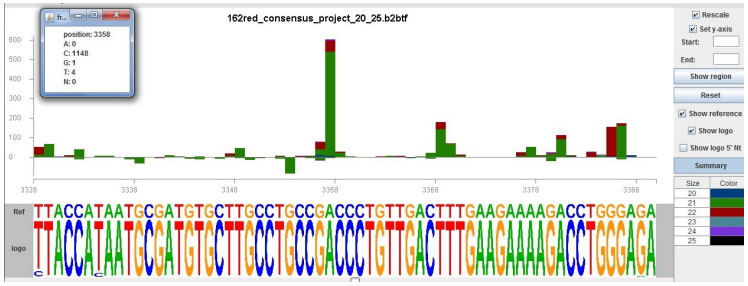
BPO-161



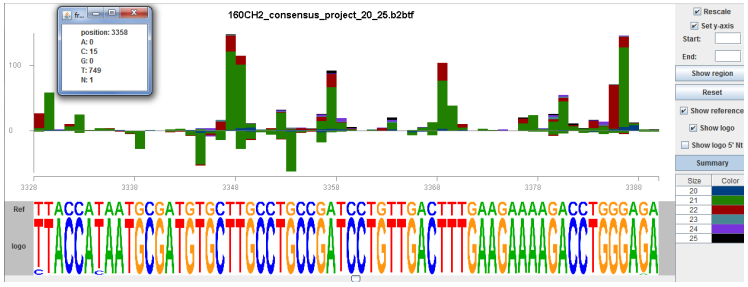
BPO-162



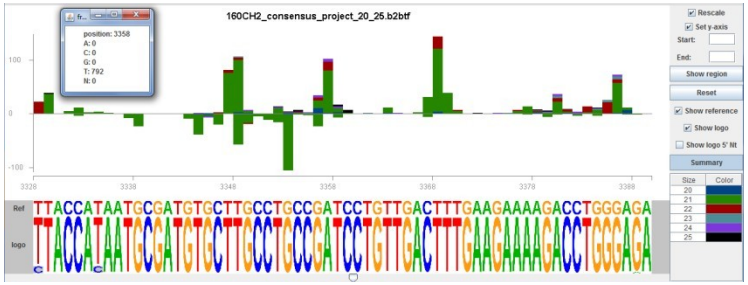
BPO-163



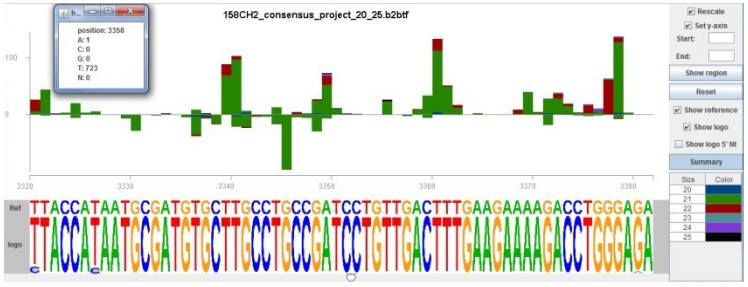
BPO-156



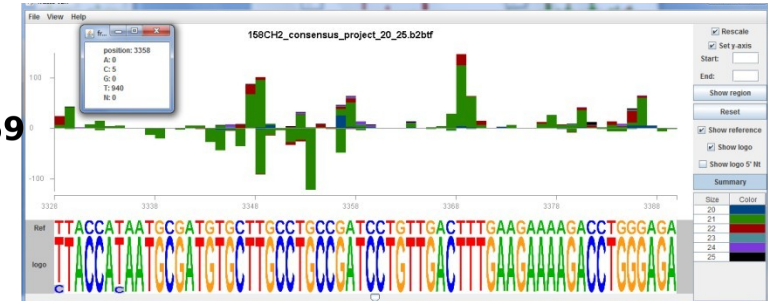
BPO-157



BPO-158

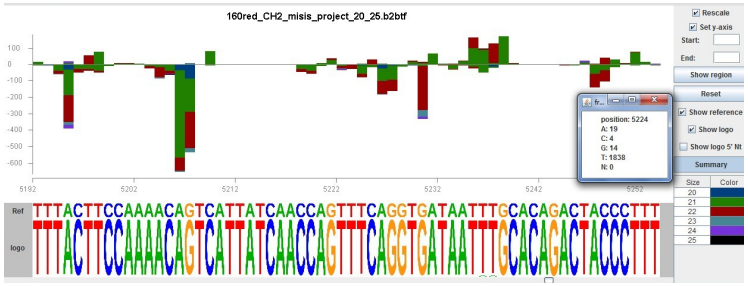


BPO-159

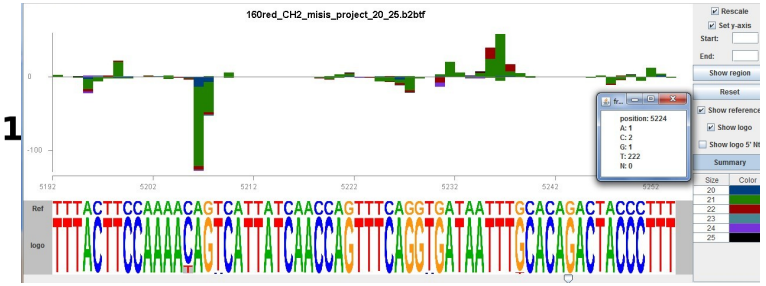


Strain CH2 - Position 5224 (T-to-G variant appeared and fixed in plant 162/163)

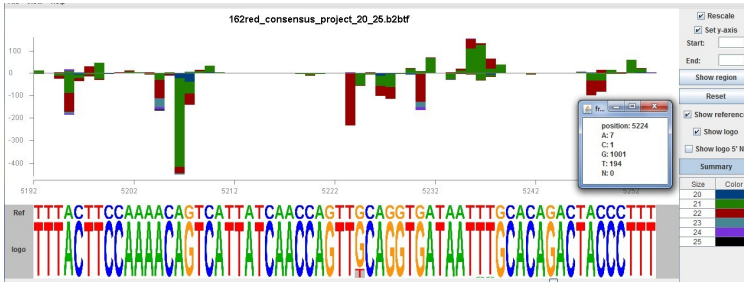
BPO-160



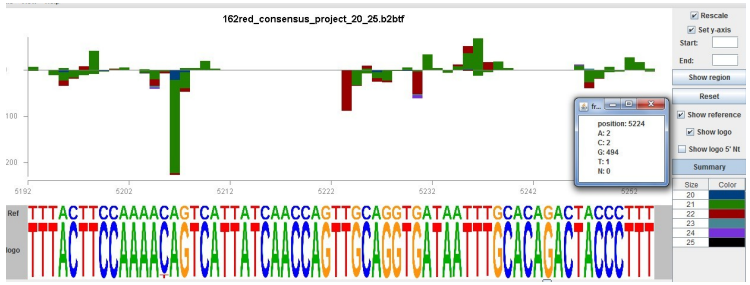
BPO-161



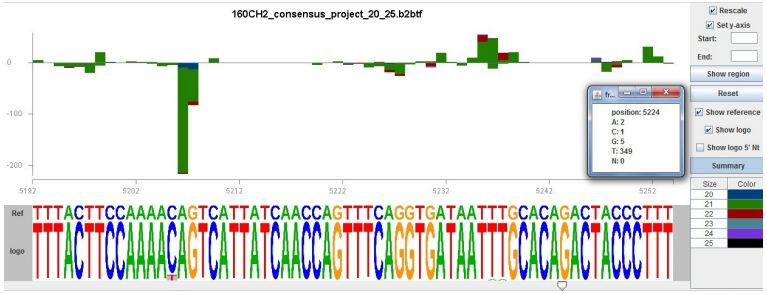
BPO-162



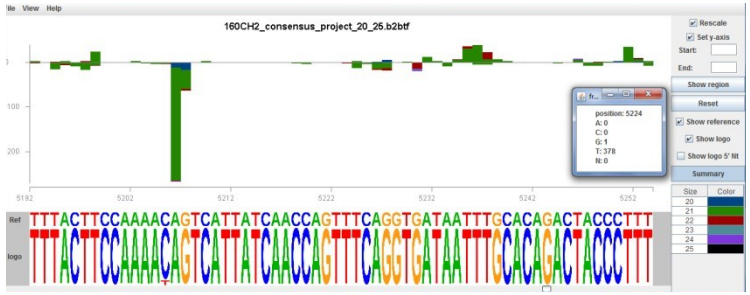
BPO-163



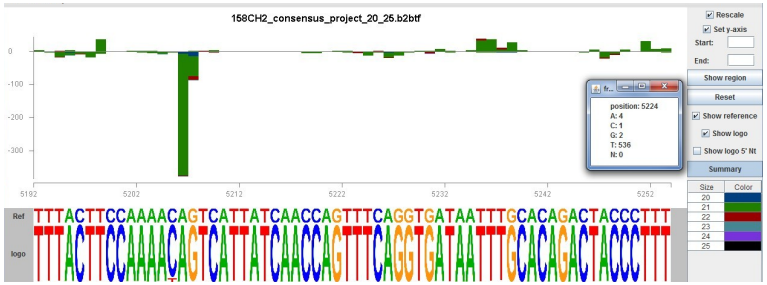
BPO-156



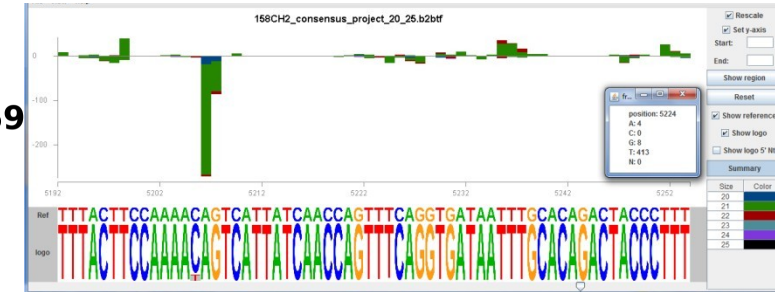
BPO-157



BPO-158

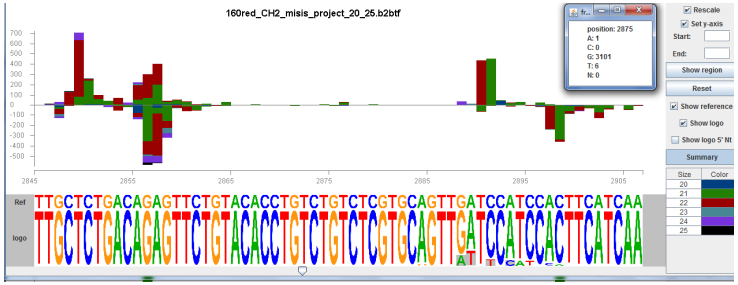


BPO-159

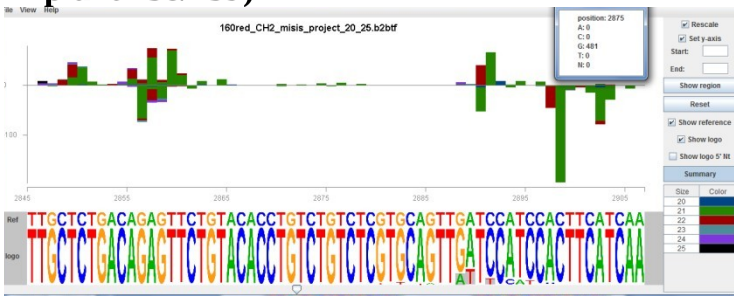


Strain CH2 - Position 2875 (G-to-A variant appeared and fixed in plant 158/159)

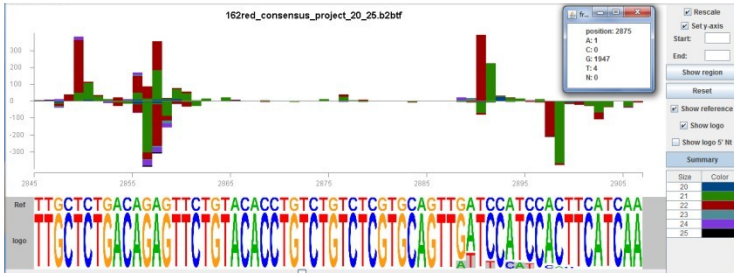
BPO-160



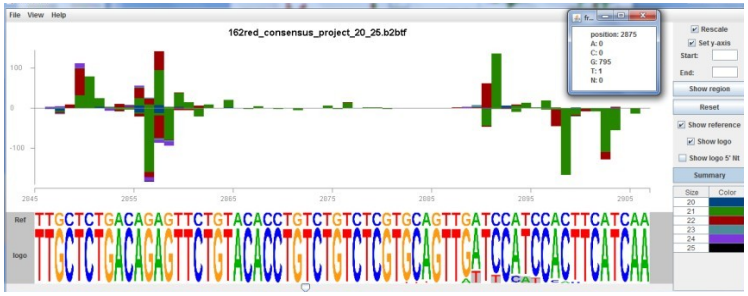
BPO-161



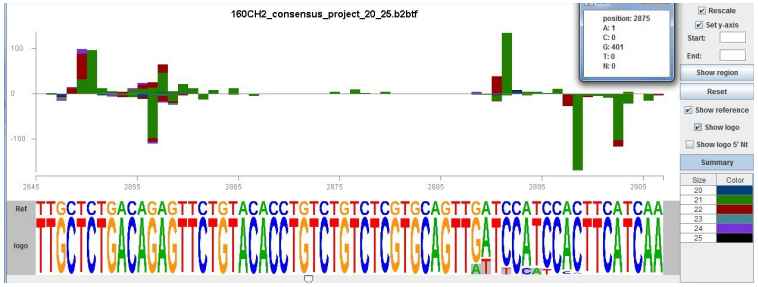
BPO-162



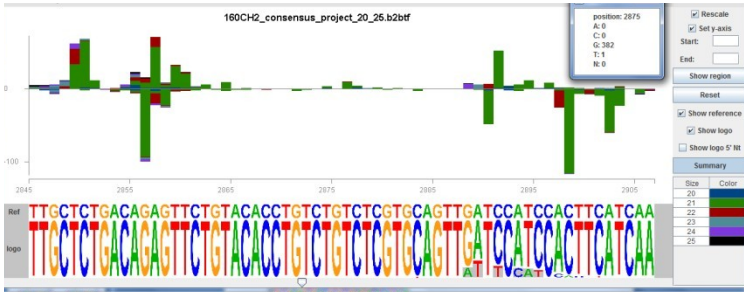
BPO-163



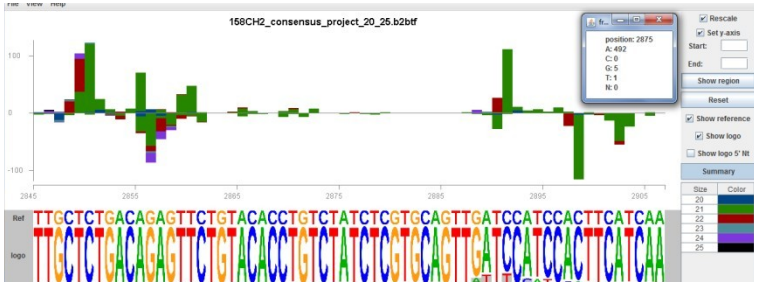
BPO-156



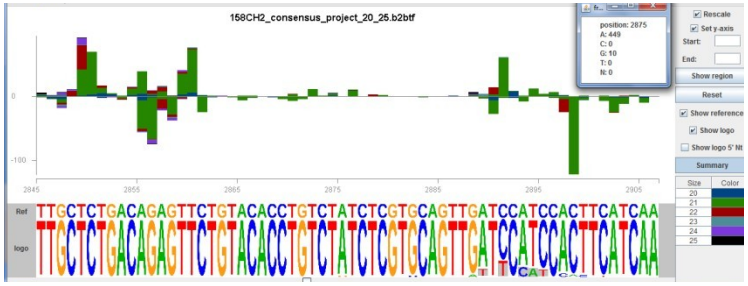
BPO-157



BPO-158

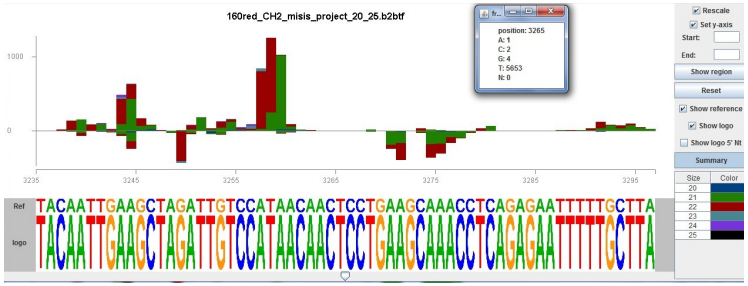


BPO-159

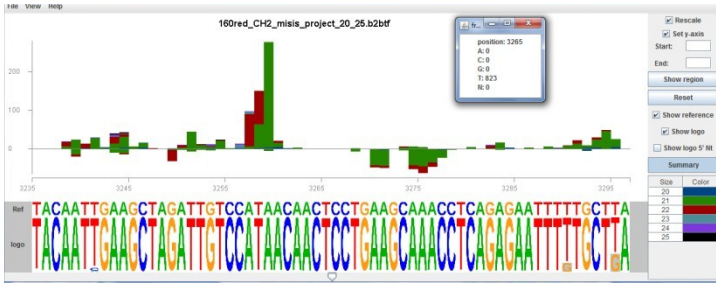


Strain CH2 - Position 3265 (T-to-C variant appeared and fixed in plant 158/159)

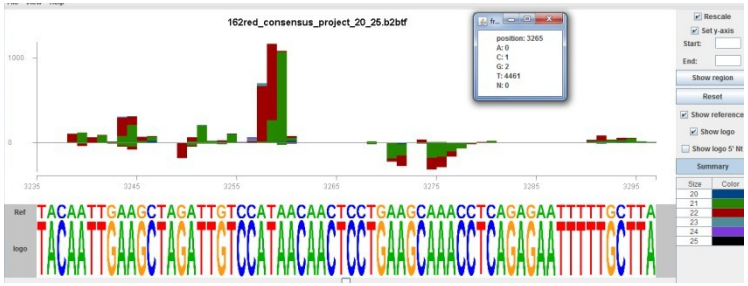
BPO-160



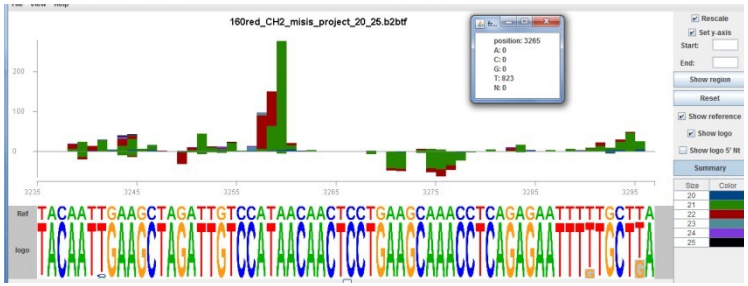
BPO-161



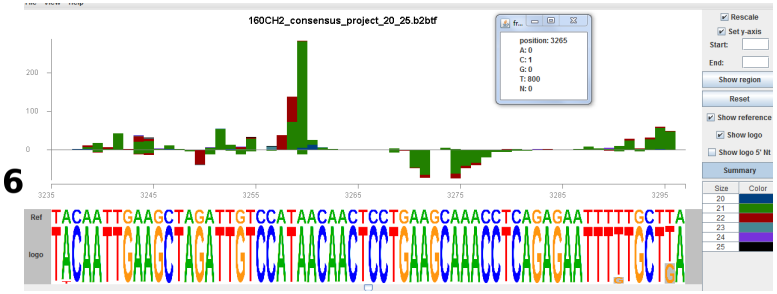
BPO-162



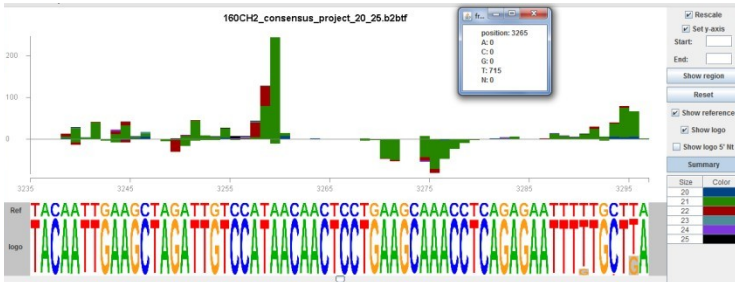
BPO-163



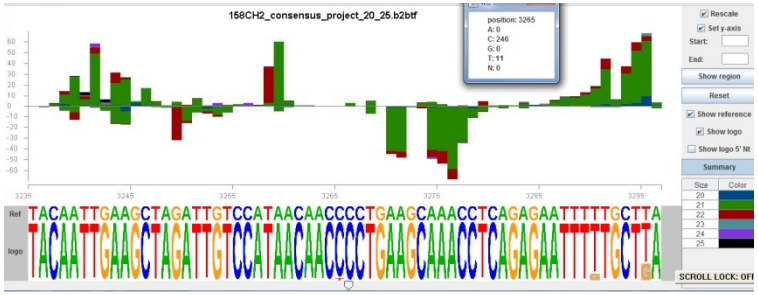
BPO-156



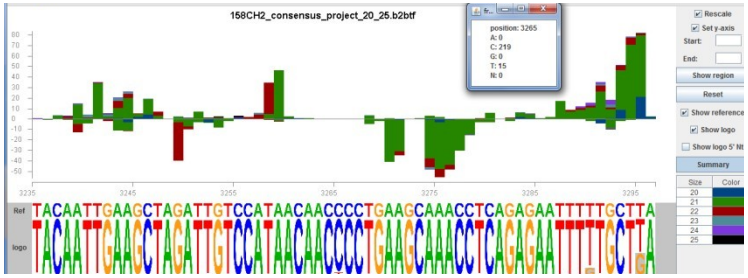
BPO-157



BPO-158

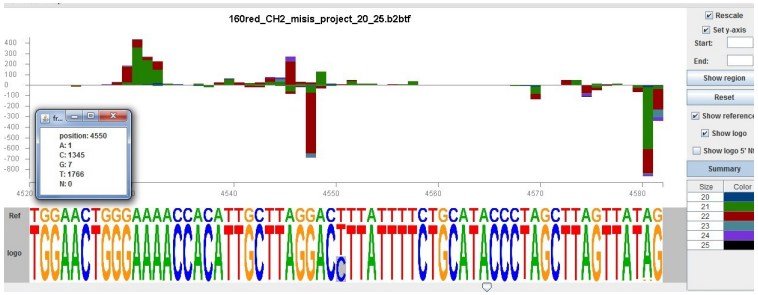


BPO-159

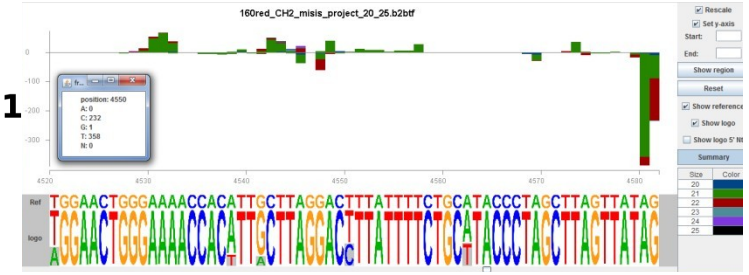


Strain CH2 - Position 4550 (T-to-C variant persisting in plant 160/161)

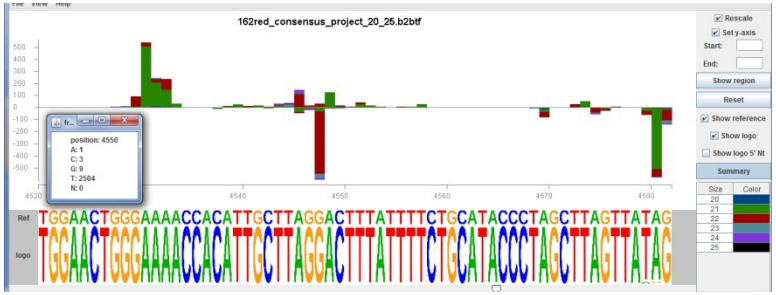
BPO-160



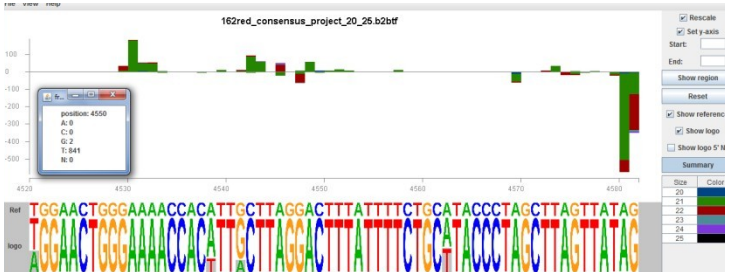
BPO-161



BPO-162



BPO-163



BPO-156

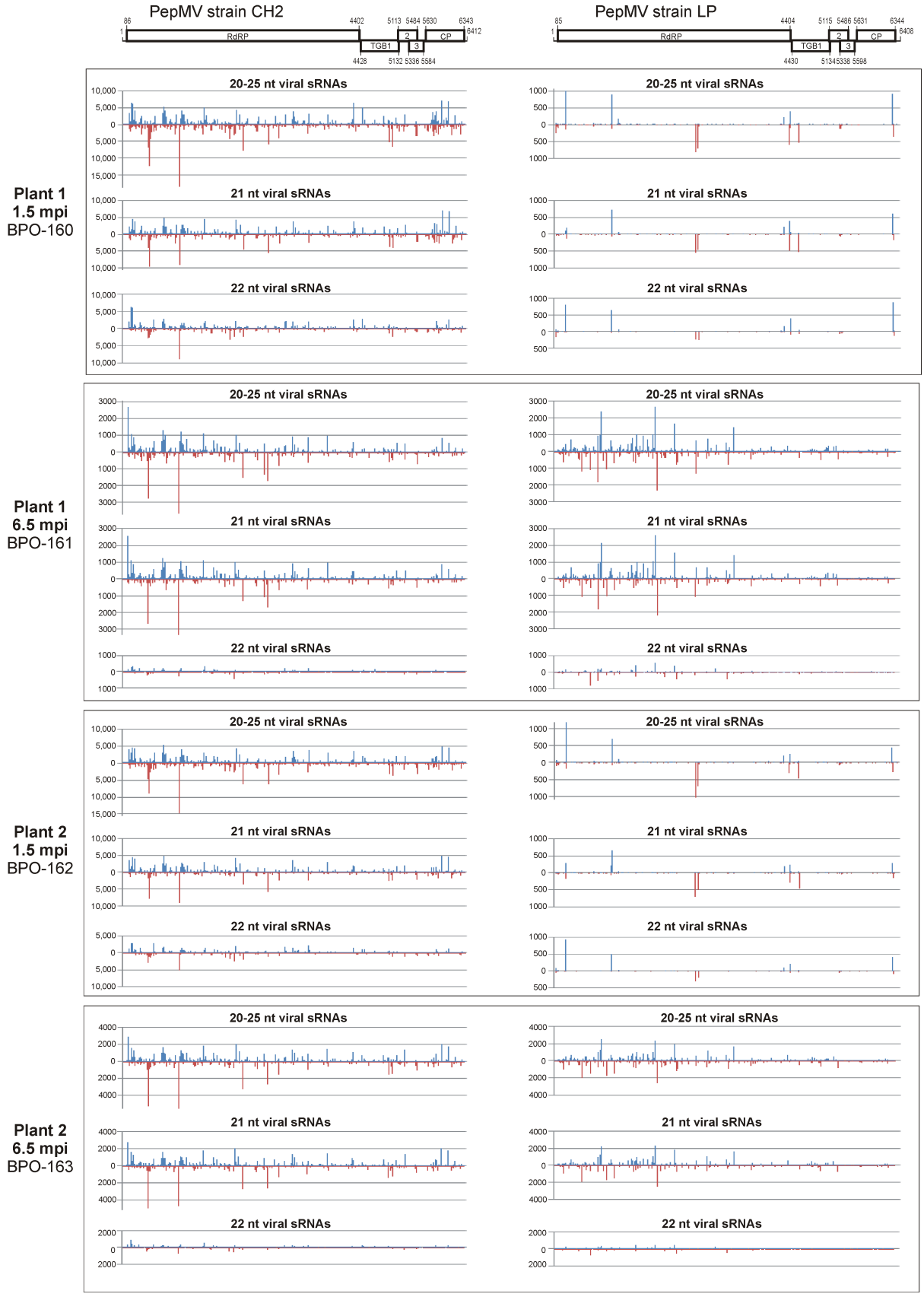
BPO-157

BPO-158

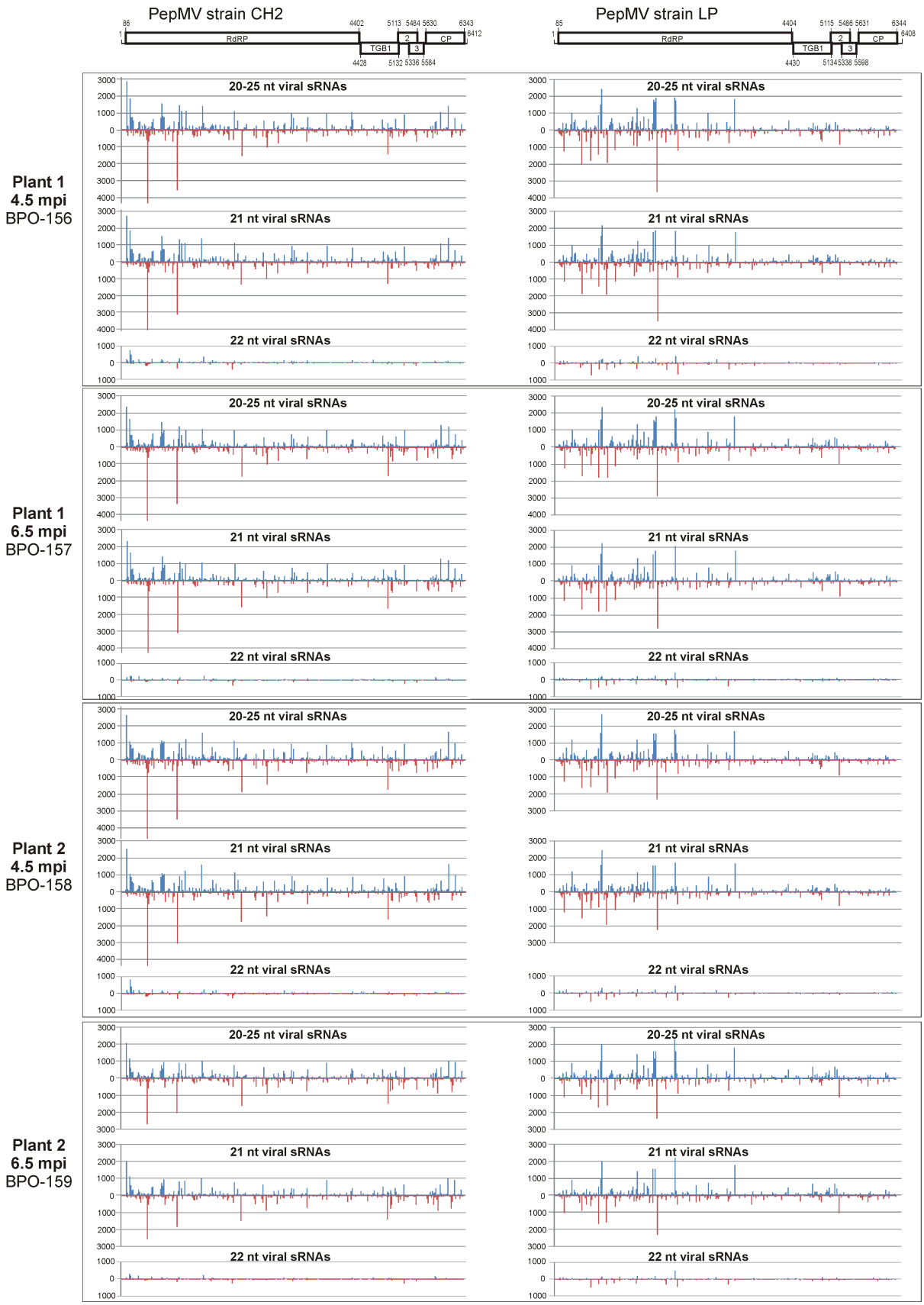
BPO-159

Annex: Supplementary Figure S3

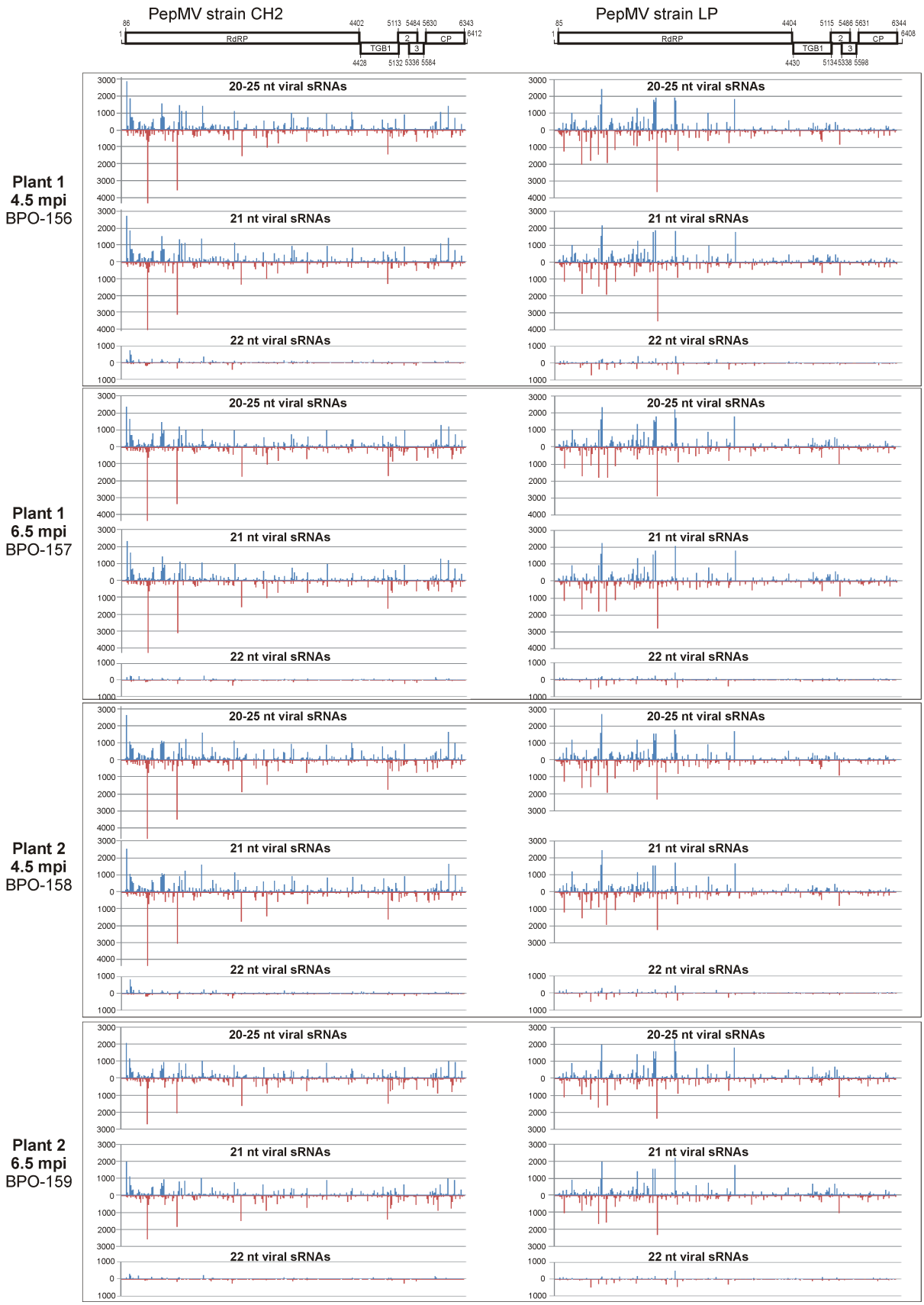
MISIS-generated maps of viral sRNAs aliligned to PepMV CH2 and LP consensus genome sequences with zero mismatches



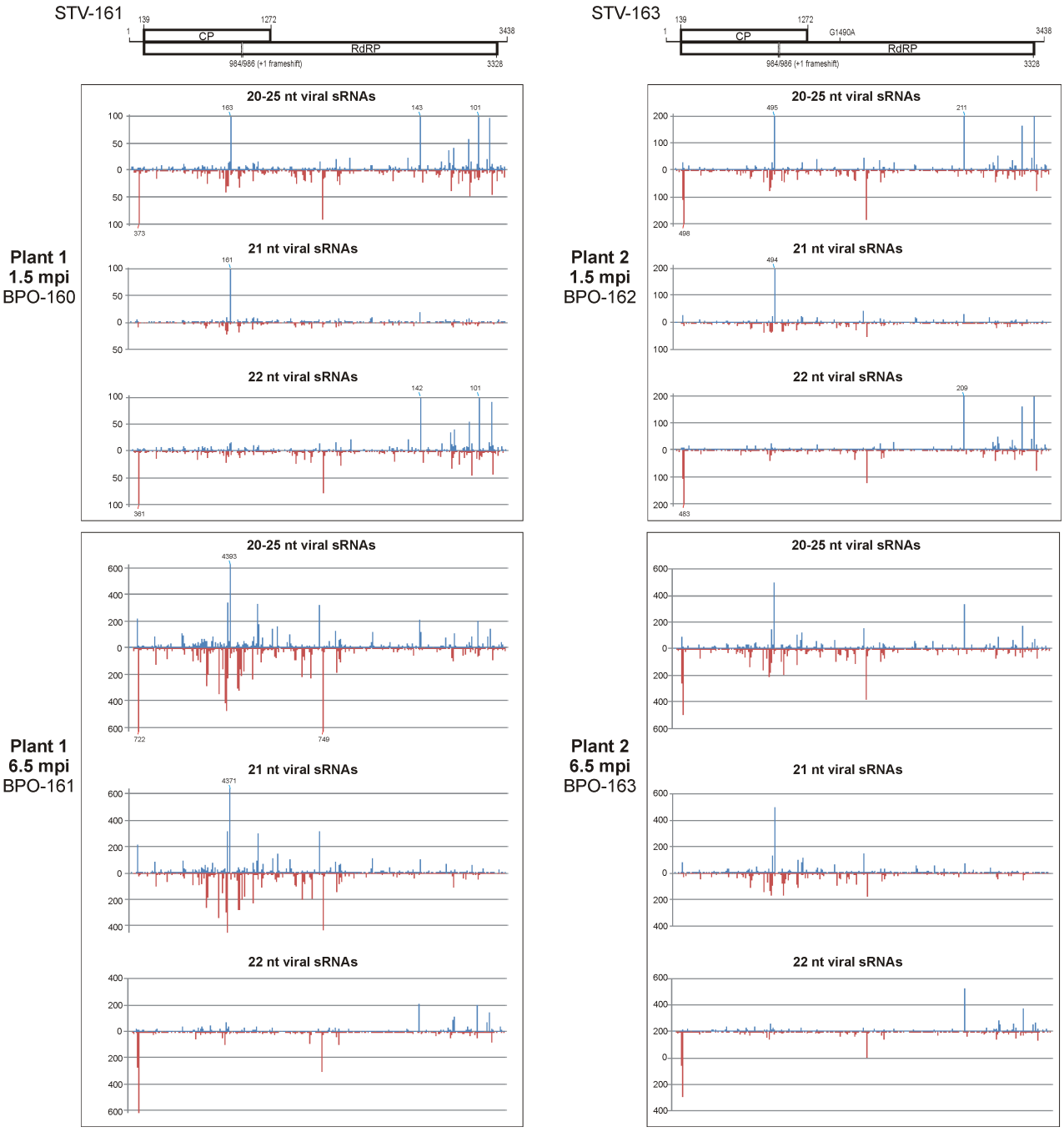
MISIS-generated maps of viral sRNAs aliligned to PepMV CH2 and LP consensus genome sequences with zero mismatches



MISIS-generated maps of viral sRNAs aliligned to PepMV CH2 and LP consensus genome sequences with zero mismatches



MISIS-generated maps of viral sRNAs alilligned to STV-161 and STV-163 consensus genome sequences with zero mismatches



viruses

MDPI

Article

Emergence of a Latent Indian Cassava Mosaic Virus from Cassava Which Recovered from Infection by a Non-Persistent Sri Lankan Cassava Mosaic Virus

Chockalingam Karthikeyan [1,2], Basavaprabhu L. Patil [2,3], Basanta K. Borah [2,4], Thulasi R. Resmi [1,2], Silvia Turco [2], Mikhail M. Pooggin [2], Thomas Hohn [2,*] and Karuppannan Veluthambi [1]

1 Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Madurai-625021, Tamil Nadu, India; karthikbiomdu@gmail.com (C.K.); trresmi@gmail.com (T.R.R.); kveluthambi@rediffmail.com (K.V.)

2 Institute of Botany, University of Basel, Schöenbeinstrasse 6, Basel 4056, Switzerland; basavaprabhu.patil@icar.gov.in (B.L.P.); basantabora@gmail.com (B.K.B.); silvia.turco@unibas.ch (S.T.); Mikhail.Pooggin@unibas.ch (M.M.P.)

3 Present address: ICAR-National Research Centre on Plant Biotechnology, PusaCampus, New Delhi110012, India

4 Present address: Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, India

* Correspondence: hohn@fmi.ch; Tel.: +41-61-701-7502

Academic Editor: Eric O. Freed

Received: 6 August 2016; Accepted: 19 September 2016; Published: 28 September 2016

Abstract: The major threat for cassava cultivation on the Indian subcontinent is cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses which are bipartite begomoviruses with DNA A and DNA B components. Indian cassava mosaic virus (ICMV) and Sri Lankan cassava mosaic virus (SLCMV) cause CMD in India. Two isolates of SLCMV infected the cassava cultivar Sengutchi in the fields near Malappuram and Thiruvananthapuram cities of Kerala State, India. The Malappuram isolate was persistent when maintained in the Madurai Kamaraj University (MKU, Madurai, Tamil Nadu, India) greenhouse, whereas the Thiruvananthapuram isolate did not persist. The recovered cassava plants with the non-persistent SLCMV, which were maintained vegetative in quarantine in the University of Basel (Basel, Switzerland) greenhouse, displayed re-emergence of CMD after a six-month period. Interestingly, these plants did not carry SLCMV but carried ICMV. It is interpreted that the field-collected, SLCMV-infected cassava plants were co-infected with low levels of ICMV. The loss of SLCMV in recovered cassava plants, under greenhouse conditions, then facilitated the re-emergence of ICMV. The partial dimer clones of the persistent and non-persistent isolates of SLCMV and the re-emerged isolate of ICMV were infective in *Nicotiana benthamiana* upon agroinoculation. Studies on pseudo-recombination between SLCMV and ICMV in *N. benthamiana* provided evidence for trans-replication of ICMV DNA B by SLCMV DNA A.

Keywords: cassava; geminivirus; persistent and non-persistent SLCMV; ICMV; pseudo-recombination; *trans*-replication

1. Introduction

Geminiviruses are circular single-stranded DNA (ssDNA) viruses of the family *Geminiviridae* that are encapsidated in paired icosahedral particles [1]. Geminiviruses are classified on the basis of their genome organization and vector transmission into seven genera: *Begomovirus*, *Mastrevirus*,

Curtovirus, *Becurtovirus*, *Eragrovirus*, *Topocuvirus* and *Turncurtovirus* [2,3]. Begomoviruses infect mostly dicotyledonous plants and are transmitted by the whitefly *Bemisia tabaci* [4].

Cassava (*Manihot esculenta* Crantz, Family Euphorbiaceae) is a tropical food crop and one of the major commercial crops of Africa, the Indian sub-continent, Latin America and several Southeast Asian countries. Cassava mosaic disease (CMD), caused by cassava mosaic geminiviruses (CMGs), is the main constraint for cassava cultivation in Africa and Indian sub-continent [5,6]. In India, cassava is mainly grown in Kerala, Tamil Nadu and Andhra Pradesh states.CMD was first reported in India by Abraham [7] and later studied in detail by Alagianagalingam and Ramakrishnan [8]. CMGs are bipartite begomoviruses and the genome is organized into DNA A and DNA B components. DNA A encompasses two genes (*AV1* and *AV2*) in the virion-sense strand and four genes (*AC1*, *AC2*, *AC3* and *AC4*) in the complementary-sense strand. The genes in DNA A encode proteins involved in replication, transcriptional activation, encapsidation and silencing suppression.DNA B encompasses two genes (*BV1* and *BC1*), which assist movement of the virus from the nucleus to the cytoplasm and also from one cell to the other [9]. Indian cassava mosaic virus (ICMV) and Sri Lankan cassava mosaic virus (SLCMV) cause CMD in India [9–11]. Both ICMV and SLCMV also infect *Nicotiana* spp. SLCMV in particular is highly virulent with a broad host-range extending to *Arabidopsis* [12].

Geminiviruses trigger gene silencing which enables the host plants to recover from viral infection [13,14]. Recovery or symptom remission was observed in many geminivirus-infected plants. One such recovery pattern was studied in *Capsicum annum*, which recovered from the infection of Pepper golden mosaic virus (PepGMV) [15,16].

Gene silencing and recovery from CMGs are isolate-dependent [13]. African cassava mosaic virus (ACMV) and SLCMV infections result in higher levels of small interfering RNA (siRNA) accumulation in the recovered plants, whereas isolates that caused non-recovering infection lead to accumulation of lower levels of siRNAs.A correlation between siRNA accumulation and reduction in viral titer was reported in some cucurbits, which showed recovery from Cucurbit leaf crumple virus (CuLCrV) [17]. Elevated temperature, which causes an increase in the siRNA accumulation in plants, also induces recovery of plants from CMD [18,19].

Trans-replication and pseudo-recombination can occur between two different species of begomoviruses in co-infected plants. Pseudo-recombination was reported between Tomato mottle virus (ToMoV) and Bean dwarf mosaic virus (BDMV) [20]. Subsequently, pseudo-recombination between other bipartite begomoviruses has been documented [21,22]. Replication associated protein (Rep) is a multifunctional protein involved mainly in the replication of geminiviral DNA by rolling circle replication (RCR) [23]. Iterative sequences or iterons are the Rep recognition sites present in the common region (CR) of geminiviral DNA A and DNA B, which determine the specificity of Rep binding and facilitate the replication of cognate DNA A and DNA B molecules [24]. ACMV and SLCMV have identical iterons, which can produce infectious pseudo-recombinants, whereas ICMV differs in the iteron sequences, resulting in replicational incompatibility with both ACMV and SLCMV [9].

In our current study, we came across a persistent and a non-persistent SLCMV isolate, infecting the same Sengutchi cultivar of cassava in the cassava fields near the cities Malappuram and Thiruvananthapuram. Cassava plants, which initially recovered from the non-persistent SLCMV infection, later displayed CMD symptoms again in the greenhouse. The symptomatic plants intriguingly carried ICMV DNA A and DNA B. DNA A and DNA B partial dimers of the two SLCMV isolates and the ICMV isolate caused viral disease upon agroinoculation of *Nicotiana benthamiana*. The agroinoculation studies further showed that SLCMV DNA A could trans-replicate ICMV DNA B. This is the first report of pseudo-recombination between SLCMV and ICMV.

2. Material and Methods

2.1. Viral Clones

Field-collected cassava plants (*Manihot esculenta* Crantz, cultivar Sengutchi) collected from Malappuram and Thiruvananthapuram in June 2009 were propagated as stem cuttings initially in

the Madurai Kamaraj University (MKU) greenhouse under natural light (12 h light/12 h darkness, 23 to 33 °C). Details of SLCMV and ICMV isolates, their National Center for Biotechnology Information (NCBI) accession numbers, the sizes of their DNA A and DNA B components and the isolates to which the maximum nucleotide sequence identity is seen are presented in Table 1. SLCMV-Sengutchi Malappuram [SeM] DNA A was cloned by polymerase chain reaction (PCR) using abutting primers. SLCMV-[SeM] DNA B, SLCMV-Sengutchi Thiruvananthapuram [SeT1] DNA A, SLCMV-[SeT1] DNA B, ICMV-[SeT4] DNA A and ICMV-[SeT4] DNA B were cloned by rolling circle amplification (RCA) [25] using *phi*29 DNA polymerase (GE Healthcare UK Ltd., Little Chalfont, UK).

2.2. Diagnostic Multiplex PCR Analysis for Amplification of SLCMV and ICMV

Diagnostic multiplex PCR for amplification of SLCMV and ICMV was carried out by using primers which specifically amplify SLCMV or ICMV DNA A fragments. SLCMV-specific primers 5′-GAAGGGAGACACATATACCTCG-3′ and 5′-CACATATATATTGTCTCCAATTCAC-3′ were used for the amplification of a 615 bp SLCMV DNA A fragment. ICMV-specific primers 5′-AGAAAGGGTTTTGATACGGAG-3′ and 5′-CTCATCTCCACGTGCTCATC-3′ were used for the amplification of a 386 bp ICMV DNA A fragment.

2.3. Construction of SLMV-[SeM] DNA A and DNA B Partial Dimers

SLCMV-[SeM] DNA A (2756 bp) was cloned as a PstI fragment in pBSIIKS$^+$ to yield pBS-SLCMV-Ma-A. A 1.7 kb PstI/HindIII fragment of DNA A from pBS-SLCMV-Ma-A (0.6-mer with CR) was cloned into the corresponding sites of the binary vector pPZP201 [26] to yield pPZP-SLCMV-Ma0.6A. The 2.7 kb full-length DNA A from pBS-SLCMV-Ma-A was taken as a PstI fragment (1-mer) and cloned in the PstI site of pPZP-SLCMV-Ma0.6A to yield the partial dimer binary plasmid pPZP-SLCMV-Ma1.6A.

SLCMV-[SeM] DNA B (2737 bp) was cloned as a BamHI fragment in pBSIIKS$^+$ to yield pBS-SLCMV-Ma-B.A 2.0 kb BamHI/HindIII fragment of DNA B (0.7-mer with CR) from pBS-SLCMV-Ma-B was cloned in the corresponding sites of pBSIIKS$^+$ to yield pBS-SLCMV-Ma0.7B. The 2.7 kb full-length DNA B from pBS-SLCMV-Ma-B was taken as a BamHI fragment (1-mer) and cloned in the BamHI site of pBS-SLCMV-Ma0.7B to yield the partial dimer plasmid pBS-SLCMV-Ma1.7B. A 4.7 kb SacI/SalI fragment which comprises the partial dimer was cloned in the binary vector pPZP201 to yield pPZP-SLCMV-Ma1.7B.

2.4. Construction of SLCMV-[SeT1] DNA A and DNA B Partial Dimers

SLCMV-[SeT1] DNA A (2746 bp) was cloned as a PstI fragment in pBSIIKS$^+$ to yield pBS-SLCMV-Tv-A. A 1.7 kb PstI/HindIII fragment of DNA A (0.6-mer with CR) from pBS-SLCMV-Tv-A was cloned in the corresponding sites of pPZP201 to yield the plasmid pPZP-SLCMV-Tv0.6A. The 2.7 kb full-length DNA A from pBS-SLCMV-Tv-A was cloned as a PstI fragment (1-mer) in the PstI site of pPZP-SLCMV-Tv0.6A to yield the partial dimer binary plasmid pPZP-SLCMV-Tv1.6A.

SLCMV-[SeT1] DNA B (2739 bp) was cloned as a BamHI fragment in pBSIIKS$^+$ to yield pBS-SLCMV-Tv-B. A 1.7 kb BamHI/BglII fragment (0.6-mer with CR) of DNA B from pBS-SLCMV-Tv-B was cloned in the BamHI site of pPZP201 to yield pPZP-SLCMV-Tv0.6B. The 2.7 kb full-length DNA B (1-mer) was taken from pBS-SLCMV-Tv-B as a BamHI fragment and cloned in the BamHI site of pPZP-SLCMV-Tv0.6B to yield the partial dimer binary plasmid pPZP-SLCMV-Tv1.6B.

2.5. Construction of ICMV-[SeT4] DNA A and DNA B Partial Dimers

ICMV-[SeT4] DNA A (2735 bp) was cloned in pBSIIKS$^+$ as a PstI fragment to yield pBS-ICMV-Tv-A. A 2.3 kb XbaI/PstI fragment (0.8-mer with CR) from pBS-ICMV-Tv-A was cloned in the corresponding sites of pBSIIKS$^+$ to yield pBS-ICMV-Tv0.8A. The 2.7 kb full-length DNA A was taken from pBS-ICMV-Tv-A as a PstI fragment (1-mer) and cloned in the PstI site of pBS-ICMV-Tv0.8A to yield the partial dimer pBS-ICMV-Tv1.8A. The 5.0 kb partial dimer was then taken as a SacI/SalI

fragment and cloned in the corresponding sites of pPZP201 to yield the partial dimer binary plasmid pPZP-ICMV-Tv1.8A.

ICMV-[SeT4] DNA B (2716 bp) was cloned as a BamHI fragment in pBSIIKS^{+} to yield pBS-ICMV-Tv-B. A 1.7 kb BamHI/BglII fragment (0.6-mer with CR) from pBS-ICMV-Tv-B was cloned in the BamHI site of pPZP201 to yield pPZP-ICMV-Tv0.6B. The full-length DNA B (1-mer) was taken as a 2.7 kb BamHI fragment from pBS-ICMV-Tv-B and cloned in the BamHI site of pPZP-ICMV-Tv0.6B to yield the partial dimer binary plasmid pPZP-ICMV-Tv1.6B.

All binary plasmids with partial dimers were independently mobilized into the *Agrobacterium tumefaciens* strain Ach5 by triparental mating [27] and the mobilizations were confirmed by Southern blot analysis.

2.6. *Agroinfection of* Nicotiana benthamiana

Three-week-old *N. benthamiana* plants were agroinoculated with the *A. tumefaciens* Ach5 strains harbouring the DNA A and DNA B partial dimers by the two strain method [28]. *A. tumefaciens* Ach5 strain without partial dimers (for mock inoculation) and with partial dimers were grown in AB minimal medium [29] to optical density at 600 nm = 1 and the cells were centrifuged (1100× *g*, Hitachi himac CR 20B2, Rotor 7, Hitachi Koki Co., Ltd., Tokyo, Japan) at 28 °C. The pellets were re-suspended in AB minimal medium (pH 5.6) with 100 µM acetosyringone (Sigma-Aldrich, St. Louis, MO, USA). The cultures were either taken separately for infection or mixed in 1:1 ratio for co-infection of DNA A + DNA B. The stem of *N. benthamiana* plants was pricked three times with a 30G needle above the first fully expanded young leaf and agroinoculation was performed by inoculating 10 µL of the re-suspended *Agrobacterium* cultures at the pricked area [30,31].

2.7. *Southern Blot Analysis*

Total DNA from cassava was extracted using the Nucleon Phytopure Genomic DNA Extraction Kit (GE Healthcare UK Ltd.). DNA from *N. benthamiana* was extracted as described by Rogers and Bendich [32]. DNA was estimated in a fluorimeter (DyNA Quant 200, Hoefer Scientific Instruments, San Francisco, CA, USA) using the Hoechst dye 33258 (Polysciences Inc., Warrington, PA, USA). One microgram plant DNA samples were electrophoresed in 0.8% agarose gels in 1X Tris-sodium acetate-ethylenediaminetetraacetic acid (EDTA) (TNE) buffer [33]. Following ethidium bromide staining and alkali denaturation, DNA was transferred [34] to the Zetaprobe nylon membrane (Bio-Rad Laboratories, Hercules, CA, USA). The probe DNA was labeled with [α-^{32}P] deoxycytidine triphosphate (dCTP) using the Megaprime DNA labeling system (GE Healthcare UK Ltd.). Hybridization was carried out overnight at 65 °C followed by high stringency post hybridization washes [35].

2.8. *siRNA Analysis*

Total RNA from cassava plants was extracted using the Tri Reagent (Sigma-Aldrich, St. Louis, MO, USA). SpeedVac-concentrated total RNA (10 µg) was electrophoresed in a 15% polyacrylamide gel with 8 M urea at 300 V. The separated RNA was blotted onto the nylon membrane (Roche Diagnostics, Indianapolis, IN, USA) in 1X Tris/Borate/EDTA buffer (TBE) using Transblot-SD semidry transfer apparatus (Bio-Rad Laboratories) at 7 V for 45 min. After blotting, the membrane was UV cross-linked twice. A 1.0 kb SLCMV *AC1* fragment was labeled with [α-^{32}P]dCTP using the Megaprime DNA labeling system and used as the probe. Hybridization was carried out for 16–20 h at 37 °C followed by post-hybridization washes for four times with 2X saline sodium citrate (SSC)/0.2% sodium dodecyl sulfate (SDS) for 20 min each at 50 °C [36].

2.9. *Illumina Sequencing and Bioinformatic Analysis of Small RNAs*

Total RNA was isolated from upper leaves of three Sengutchi cassava plants (Figure 1): (i) "SeM-SLCMV", the symptomatic plant from Malappuram which was persistently infected in Basel with SLCMV-[SeM]); (ii) "SeT-recovered", the symptomless plant from Thiruvananthapuram

which had recovered in Basel from infection with SLCMV-[SeT1]; and (iii) "SeT-ICMV", a vegetative progeny of the "SeT-recovered" plant which eventually exhibited disease symptoms and found to be infected with re-emergent ICMV-[SeT4] (see below).cDNA libraries were prepared from 19 to 30 ntRNA fractions following an Illumina Small RNA TruSeq protocol and sequenced on the Illumina Genome Analyzer HiSeq2000 using a TruSeq SBS Kit v3 at Fasteris AG [37]. After trimming the adaptor sequences, 20–25 nt small RNA (sRNA) reads were mapped to the draft genome of *Manihot esculenta* v6.1 [38] and the genome of the respective virus present in each plant (i.e., SLCMV-[SeM], SLCMV-[SeT1], or ICMV-[SeT4]) using the Burrows-Wheeler Alignment Tool (BWA, Version 0.5.9) [39] with up to two mismatches. Percentage of the plant and the viral reads in the total population of 20–25 nt reads was calculated and presented as a bar graph in Figure 2d.

3. Results and Discussion

3.1. Recovery of Cassava Plants from SLCMV Infection in the Greenhouse and Re-Emergence of Symptoms due to Latent ICMV

The vegetative progeny of the infected cassava plants were transferred from MKU as stem-cuttings and established in the greenhouse of the University of Basel in February 2010. The timelines of cassava plant establishment in the Madurai and Basel greenhouses and periods of sample collection are illustrated in Figure 1. DNA was extracted from the two plants (M1 and T1) at the initial stage of greenhouse establishment in MKU and the CMD was confirmed by PCR and Southern blot analysis (data not shown). The full-length viral DNA A was amplified from the plant samples M1 and T1 by either PCR or RCA. The amplified fragments of 2.7 kb were cloned and sequenced. Both Sengutchi plants were found to be infected with SLCMV (Table 1).

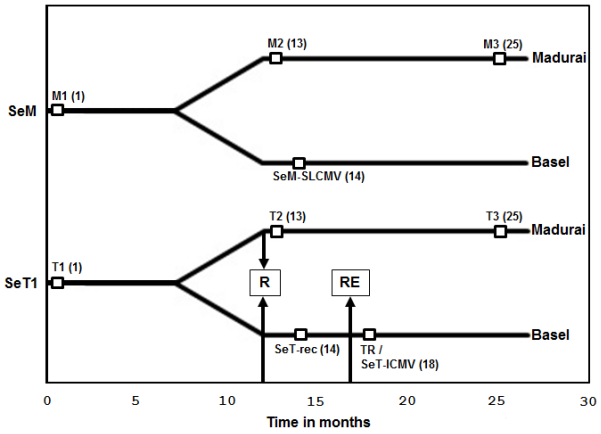


Figure 1. Timelines of cassava plant establishment in Madurai and Basel greenhouses and periods at which samples were taken for DNA and RNA extraction. The periods (in months) following the initial establishment of cassava plants (*Manihot esculenta* Crantz) in Madurai are represented in the *x*-axis. SeM, the cassava plant collected from Malappuram; SeT1, the cassava plant collected from Thiruvananthapuram. The time point 0 indicates the time of initial establishment of field-infected cassava plants in the Madurai greenhouse. The plants were established in the Basel greenhouse eight months after initial establishment in Madurai. R denotes the time of recovery of the SeT 1 plant from cassava mosaic disease (CMD) and RE denotes the time of re-emergence of CMD symptoms. M1, M2 and M3 are the time points at which DNA was extracted from SeM plants. T1, T2 and T3 are the time points at which DNA was extracted from SeT1 plants (both maintained in the Madurai greenhouse). TR is the time point of re-emergence of CMD in SeT1 cassava plants grown in the Basel greenhouse. Time points of RNA extraction in SeM plants (SeM-SLCMV), SeT1 plants at the recovery stage (SeT-rec) and SeT plants at the re-emergent stage (SeT-ICMV) are marked. The number of months from the initial establishment of cassava plants in Madurai is given in brackets.

Table 1. Sri Lankan cassava mosaic virus (SLCMV) and Indian cassava mosaic virus (ICMV) isolates cloned from the Sengutchi variety of cassava plants collected from the fields of Malappuram (SeM) and Thiruvananthapuram (SeT1).

Geographical Region	Cassava Cultivar	Virus	Isolate	NCBI Accession No.	Length(nt)	Identity
Malappuram	Sengutchi	SLCMV	SeM-DNA-A	KR611577	2756	98% to Kerala 17 96% to SeT1
		SLCMV	SeM-DNA-B	KR611578	2737	97% to Kerala 4 98% to SeT1
Thiruvananthapuram	Sengutchi	SLCMV	SeT1-DNA-A	KR611579	2726	99% to Kerala 17 96% to SeM
		SLCMV	SeT1-DNA-B	KR611580	2739	97% to Kerala 4 98% to SeM
Re-emergent in Basel	Sengutchi	ICMV ICMV	SeT4-DNA-A SeT4-DNA-B	KU308385 KU308386	2735 2716	94% to Kerala 2 98% to Kerala 6

The Sengutchi plant collected from Malappuram (infected with SLCMV-[SeM]) showed persistence of CMD symptoms throughout the study period of three years in the MKU greenhouse. However, the Sengutchi plants collected from Thiruvananthapuram (infected with SLCMV-[SeT1]) and maintained in the same facility showed recovery from the CMD symptoms from the second year. Southern blot analysis using a SLCMV DNA A probe in the second (T2) and third years (T3) revealed the absence of viral DNA in the cassava plants originally infected with the SLCMV-[SeT1] isolate (Figure 2a). This shows that SLCMV-[SeT1] is a non-persistent virus isolate. Cassava plants with SLCMV-[SeM] infection maintained viral titer in all the three samples M1, M2 and M3, collected at the beginning of 1st, 2nd and 3rd years, respectively (Figure 2a). Thus, SLCMV-[SeM] is a persistent virus isolate.

The Sengutchi cassava plants, from which the SLCMV-[SeT1] was cloned, displayed initially CMD symptoms in the Basel greenhouse. Recovery from the CMD symptoms and absence of SLCMV was observed in the plants established in both the Basel and Madurai greenhouses. Interestingly, the CMD symptoms re-emerged in those plants during October 2010, after eight months of initial establishment at Basel. DNA was extracted from those plants, the viral DNA was cloned by RCA and the full-length viral DNA A (2735 nt) and DNA B (2716 nt) clones were sequenced (Table 1). The sequences revealed that ICMV was responsible for re-emergence of CMD symptoms.The Basel greenhouse did not contain any other ICMV-infected plants. Cassava plants are not grown in Basel and their viruses are absent. Therefore, the presence of ICMV DNA A and DNA B in the re-emergent infection of cassava plants should be attributed to the re-emergence of ICMV from a latent infection, which was undetectable by PCR at the time of field collection. We therefore carried out an additional experiment to detect the latent ICMV in the original DNA sample of plant T1 (Figure 1) extracted from the Sentgutchi plant (SeT1) at Madurai. Using RCA amplification of circular viral DNA followed by a diagnostic multiplex PCR analysis with the ICMV DNA A and the SLCMV DNA A specific primers, we detected in fact a very low level of ICMV DNA A in addition to abundant SLCMV DNA A in the plant T1 (Figure 2b).

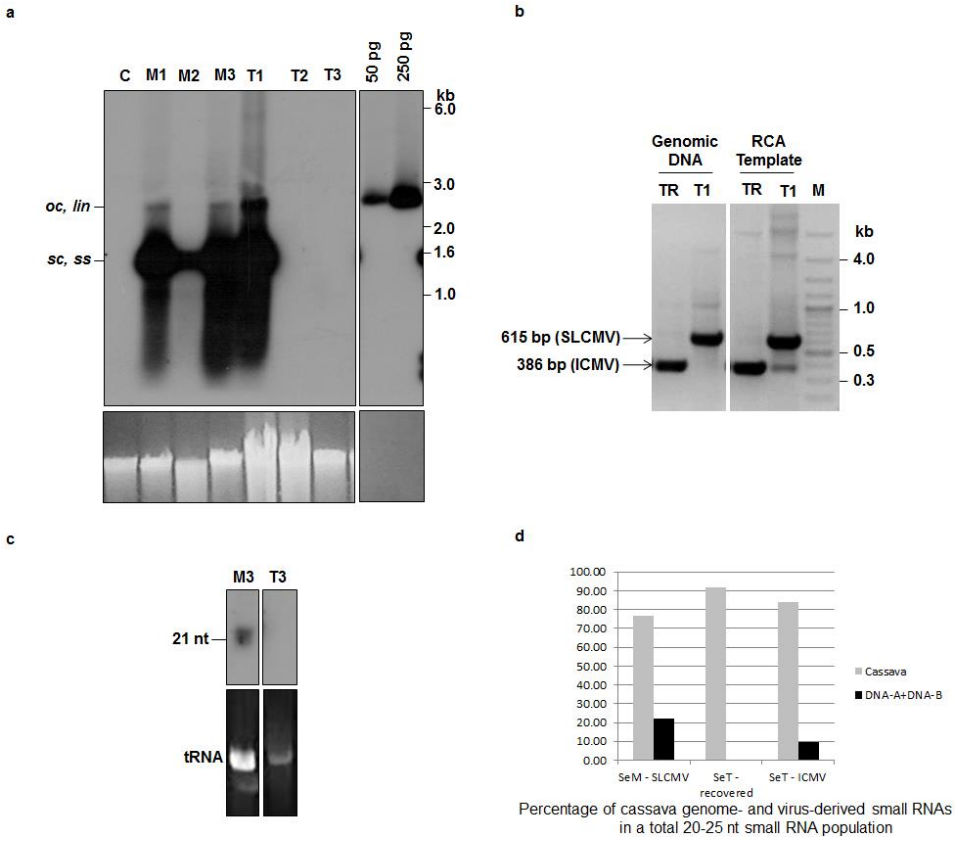


Figure 2. Sri Lankan cassava mosaic virus (SLCMV) DNA A and small interfering RNA (siRNA) analysis in symptomatic and symptom-recovered cassava plants. (**a**) Southern blot analysis of field-infected cassava plants maintained in the greenhouse. DNA (1 µg) samples from a virus-free, axenic (tissue culture-grown) cassava plant (C), and field-infected cassava (cultivar Sengutchi) plants collected from Malappuram (M) and Thiruvananthapuram (T) were analyzed. The field-infected cassava plants were established in the Madurai Kamaraj University (MKU) greenhouse and the DNA was extracted during first year (M1 and T1), second year (M2 and T2) and third year (M3 and T3) of establishment in the greenhouse (Figure 1). pBS-SLCMV-Ma-A plasmid (50 pg and 250 pg) digested with PstI was used as a positive control. [α-^{32}P]dCTP-labeled full-length SLCMV-[SeM] DNA A was used as the probe. Positions of different forms of viral DNA, single stranded (*ss*), super-coiled (*sc*), open circular (*oc*) and linear (*lin*), are marked. Ethidium bromide stained high molecular weight plant DNA is shown as loading control at the bottom panel. (**b**) Diagnostic multiplex PCR using SLCMV- and ICMV-specific primers to analyze mixed infection in the plant sample T1 (field-infected SLCMV-SeT1, symptomatic plant) and TR (plant which initially recovered and subsequently showed re-emergence of CMD symptoms in the Basel greenhouse) (Figure 1). Genomic DNA and RCA-amplified DNA from T1 and TR were used as templates for polymerase chain reaction (PCR) analysis. Amplified fragments of SLCMV (615bp) and Indian cassava mosaic virus (ICMV) (386bp) are marked. M, Molecular weight marker. (**c**) Small RNA Northern blot analysis of field-infected cassava plants. RNA (10 µg) from field-infected cassava plants collected from Malappuram (M3) and Thiruvananthapuram (T3) were analyzed. The field-infected cassava plants were established in the MKU greenhouse and RNA was extracted after 25 months of initial establishment in the greenhouse (M3 and T3). [α-^{32}P]dCTP-labeled SLCMV-[SeM] *AC1* (1 kb) gene fragment was used as the probe. The position of siRNA (21 nt) is marked. The bottom panel shows ethidium bromide stained tRNA, as a loading control. (**d**) The proportion of the plant genome (cassava)- and the viral genome (DNA A + DNA B)-derived small RNAs in a total population of 20–25nt small RNAs accumulating in the Sengutchi cassava plants "SeM-SLCMV", "SeT-recovered" and "SeT-ICMV" (Figure 1).

Virus induced post-transcriptional gene silencing (PTGS) in cassava results in recovery from ACMV and SLCMV symptoms, which correlates well with the accumulation of siRNAs [13]. Analysis of viral siRNAs was performed in Sengutchi cassava plants infected with persistent and non-persistent SLCMV isolates. *AC1* was used as the probe, since we detected a high level of siRNAs in field-infected cassava plants with this probe. The Sengutchi cassava plant M3 (Figure 1), which was infected with the persistent isolate SLCMV-[SeM], showed accumulation of siRNAs corresponding to the *AC1* gene (Figure 2c). However, the plant T3 (Figure 1), which was originally infected with the non-persistent isolate SLCMV-[SeT1] and later showed recovery from CMD, did not accumulate virus-derived siRNAs (Figure 2c). The recovery from CMD was expected because of strong silencing (more viral siRNAs). However, siRNA accumulation was not observed in the plant T3. The time point T3 (Figure 1, 25 months) is 13 months past the recovery stage (12 months) and also the insufficient loading of T3 could have resulted in the absence of siRNAs detection. Therefore, a second set of recovered SeT1 samples at the 14-month time point was taken for deep sequencing of small RNAs. We deep-sequenced small RNAs from the vegetative progeny of the Sentgutchi-Thiruvananthapuram plant that had recovered from CMD caused by SLCMV-[SeT1] (designated "SeT-rec") (Figure 1) and later displayed CMD in Basel owing to re-emergence of the latent ICMV (designated "SeT-ICMV") (Figure 1), as well as from the vegetative progeny of the Sentgutchi-Malappuram plant showing persistent CMD (caused by SLCMV-[SeM]; designated "SeM-SLCMV") (Figure 1). Bioinformatics analysis of the deep sequencing data revealed that both symptomatic plants accumulated high proportion of viral siRNAs in a total population of 20–25 nt sRNAs (9.7% in the SeT-ICMV plant and 22% in the SeM-SLCMV plant) (Figure 2d). In contrast, only trace amounts of viral siRNAs were detected in the SeT-rec plant (0.07% of the total 20–25 nt sRNAs). Taken together, our findings agree with the findings of Patil and Fauquet [19], which showed that recovery from symptoms is not always associated with high siRNA accumulation. The re-emergence of CMD symptoms from recovered plants occurred after the complete loss of the SLCMV-[SeT1] titer and near complete loss of viral siRNAs, and this re-emergence of severe disease symptoms was associated with production of viral siRNAs from ICMV-[SeT4] that replicated efficiently in the absence of SLCMV. SLCMV is more widespread in the field in contrast to ICMV in the Indian sub-continent [40] possibly because ICMV does not replicate efficiently in the presence of SLCMV. An alternative hypothesis is that the presence of ICMV in the SeT1 plants caused recovery, whereas the absence of ICMV in the SeM plants did not lead to recovery.

3.2. *Infectivity of Cloned SLCMV DNA A and DNA B in* Nicotiana benthamiana *Plants*

The infectivity of partial dimers of cloned DNA A and DNA B components of SLCMV-[SeM] and SLCMV-[SeT1] were analyzed in three-week-old *N. benthamiana* plants by agroinfection (Figure 3 and Figure S1). Plants were scored for symptoms 14-day-post infection. For both SLCMV isolates, plants agroinoculated with DNA A showed upward leaf rolling symptom, while the plants agroinoculated with DNA B remained symptomless (Table 2). Plants co-agroinoculated with both SLCMV DNAs (A and B) showed more severe typical geminiviral symptoms of stunting, chlorosis and downward leaf curling in all the agroinoculated plants (Figure 3a and Figure S1a). For both SLCMV isolates, Southern blotting revealed much higher DNA A levels in plants agroinoculated with DNA A and DNA B than with DNA A alone (Figure 3b and Figure S1b). Plants agroinoculated with DNA B alone, as expected, did not accumulate the viral DNA. Using a probe lacking the common region (ΔCR), DNA B accumulation was observed only in plants co-agroinoculated with DNA A + DNA B (Figure 3c and Figure S1c).

Table 2. Infectivity analysis by agroinoculation of SLCMV-[SeM], SLCMV-[SeT1] and ICMV-[SeT4] partial dimers in *Nicotiana benthamiana* plants.

Component A	Component B	Symptoms (at 14 days post-inoculation (dpi))	Levels of virus DNA
SLCMV-[SeM]		Upward leaf rolling	Low DNA A
	SLCMV-[SeM]	No symptoms	None
SLCMV-[SeM]	SLCMV-[SeM]	Stunting, chlorosis, downward leaf curling	High DNA A and DNA B
SLCMV-[SeT1]		Upward leaf rolling	Low DNA A
	SLCMV-[SeT1]	No symptoms	None
SLCMV-[SeT1]	SLCMV-[SeT1]	Stunting, chlorosis, downward leaf curling	High DNA A and DNA B
ICMV-[SeT4]		Very mild leaf blade curling	High DNA A
	ICMV-[SeT4]	No symptoms	None
ICMV-[SeT4]	ICMV-[SeT4]	Stunting and mild to moderate leaf blade curling	High DNA A and DNA B
SLCMV-[SeM]	ICMV-[SeT4]	Mild leaf blade curling	High DNA A and low DNA B

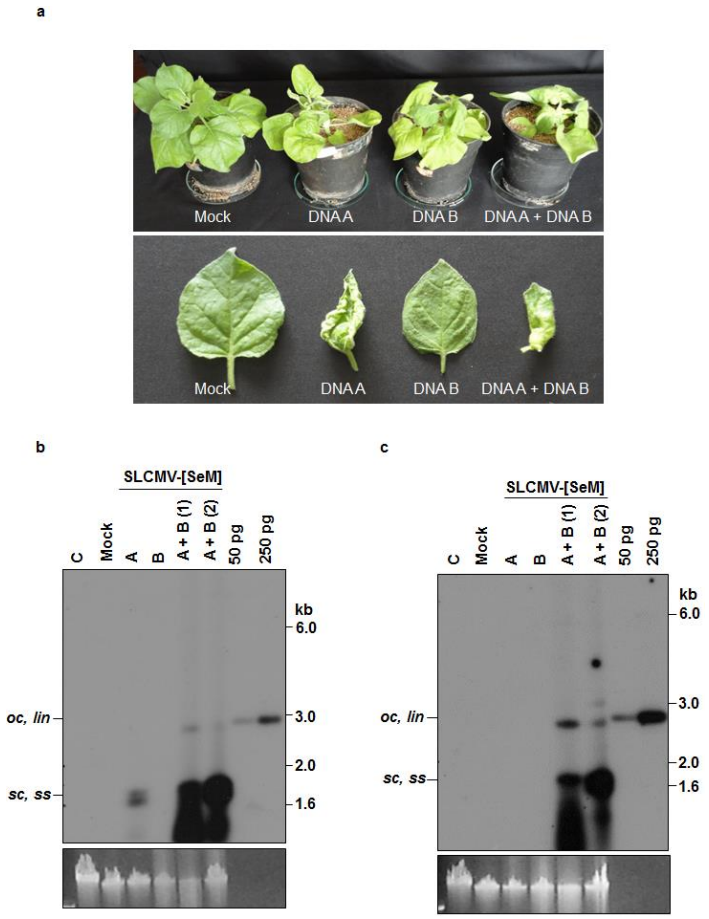


Figure 3. Infectivity analysis of SLCMV-[SeM] partial dimers in *N. benthamiana* plants. (**a**) Symptoms displayed by *N. benthamiana* plants agroinoculated with the SLCMV-[SeM] partial dimers. Bottom half shows individual leaves of the tested plants. (**b**) Southern blot analysis using SLCMV-[SeM] DNA A (without common region, ΔCR) labeled with [α-^{32}P]dCTP as the probe. The plasmid pBS-SLCMV-Ma-A digested with PstI (50 pg and 250 pg) was used as the positive control. (**c**) Southern blot analysis using [α-^{32}P]dCTP-labeled SLCMV-[SeM] DNA B (ΔCR) as the probe. The plasmid pBS-SLCMV-Ma-B digested with BamHI (50 pg and 250 pg) was used as the positive control. (**b**,**c**) DNA (1 µg) from uninfected plant (C), plant mock infected with the *Agrobacterium tumefaciens* strain Ach5 (Mock), plant agroinoculated with partial dimers of DNA A alone (A), DNA B alone (B) and two plants independently co-agroinoculated with the partial dimers of DNA A + DNA B (A + B) were loaded in the respective lanes. Positions of different forms of viral DNA, single stranded (*ss*), super-coiled (*sc*), open circular (*oc*) and linear (*lin*), are marked. Ethidium bromide stained high molecular weight plant DNA is shown as loading control at the bottom.

The upward leaf rolling symptom observed upon agroinoculation with DNA A alone is typical for monopartite geminiviruses, suggesting the evolution of SLCMV from a monopartite geminivirus by the acquisition of DNA B from ICMV [9]. Infectivity analysis of SLCMV-[SeM] and SLCMV-[SeT1] in *N. benthamiana* plants proved that the clones of both isolates from field-infected cassava plants (cultivar Sengutchi) are infectious.

3.3. Infectivity of Cloned ICMV DNA A and DNA B in Nicotiana benthamiana Plants

Similar agroinfection experiments were performed with ICMV partial dimers (Table 2). All plants agroinoculated with ICMV-[TVM4] DNA A alone showed very mild leaf blade curling symptoms (Figure 4a). Interestingly, upward leaf rolling as noted for SLCMV DNA A agroinoculation was not observed. All plants agroinoculated with DNA A + DNA B displayed more severe symptoms (i.e., stunting and leaf blade curling). Again, none of the plants agroinoculated with DNA B alone displayed symptoms.

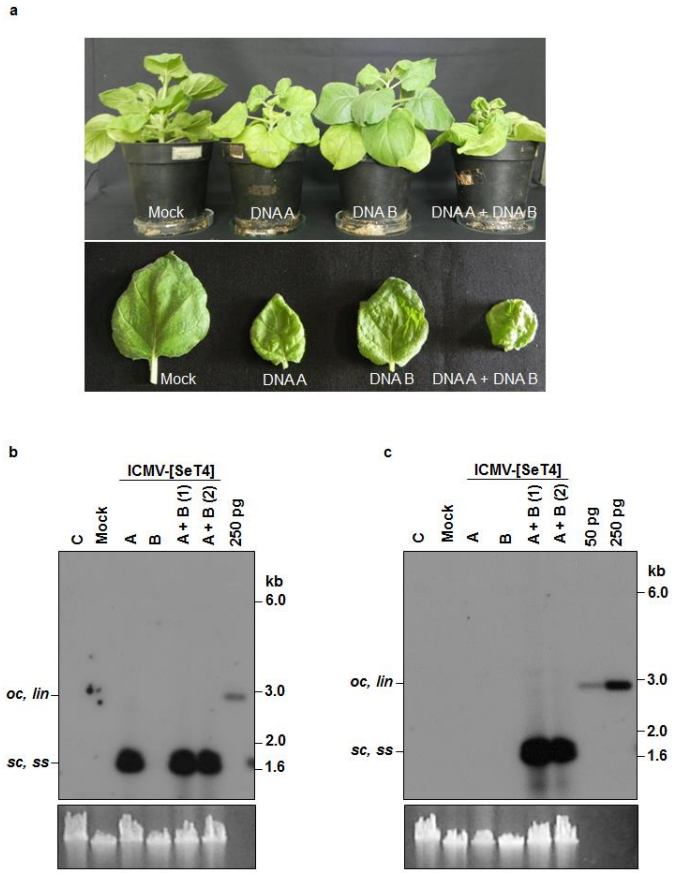


Figure 4. Infectivity analysis of ICMV-[SeT4] in *N. benthamiana* plants. (**a**) Symptoms displayed by *N. benthamiana* plants agroinoculated with the partial dimers of ICMV-[SeT4]. Bottom half shows individual leaves of the tested plants. (**b**) Southern blot analysis using ICMV-[SeT4] DNA A (ΔCR) labeled with [α-^{32}P]dCTP as the probe. The plasmid pBS-ICMV-Tv-A digested with PstI (250 pg) was used as the positive control. (**c**) Southern blot analysis using [α-^{32}P]dCTP-labeled ICMV-[SeT4] DNA B (ΔCR) as the probe. The plasmid pBS-ICMV-Tv-B digested with BamHI (50 pg and 250 pg) was used as the positive control. (**b**,**c**) DNA (1 µg) from uninfected plant (C), plant mock infected with the *A. tumefaciens* strain Ach5 (Mock), plant agroinoculated with the partial dimers of DNA A alone (A), DNA B alone (B) and two plants independently co-agroinoculated with the partial dimers of DNA A + DNA B (A + B) were loaded in the respective lanes. Positions of different forms of viral DNA, single stranded (*ss*), super-coiled (*sc*), open circular (*oc*) and linear (*lin*), are marked. Ethidium bromide stained high molecular weight plant DNA is shown as loading control at the bottom.

In the ICMV case and in contrast to the SLCMV case, Southern blotting revealed comparable levels of DNA A accumulation in the presence and absence of DNA B (Figure 4b). No DNA B accumulation was observed in plants agroinoculated with the DNA B alone (Figure 4c).

The fact that ICMV DNA A accumulated to high levels, regardless whether DNA B was present or not, while symptoms were much more severe upon agroinoculation with both DNAs, revealed that ICMV DNA B plays an important role in symptom development, which validates the earlier reports on symptom determination of bipartite begomoviruses [31,35,41]. The infectivity analysis of ICMV-[SeT4] in *N. benthamiana* plants proved that the isolate cloned from the CMD re-emergent cassava plants is infectious.

3.4. Analysis of Trans-Replication of ICMV-[TVM4] DNA B by SLCMV-[SeM] DNA A

SLCMV-[SeM] DNA A and ICMV-[SeT4] DNA B partial dimers were used to study pseudo-recombination. The infectivity rate in plants co-agroinoculated with SLCMV DNA A + ICMV DNA B was lower than in plants co-agroinocculated with either the SLCMV or ICMV cognate pairs. Two of the four SLCMV DNA A + ICMV DNA B co-agroinoculated plants displayed mild leaf blade curling (Figure 5a) and showed high levels of SLCMV DNA A and low levels of DNA B (Figure 5b,c, lanes 3 and 4).

Figure 5. *Cont.*

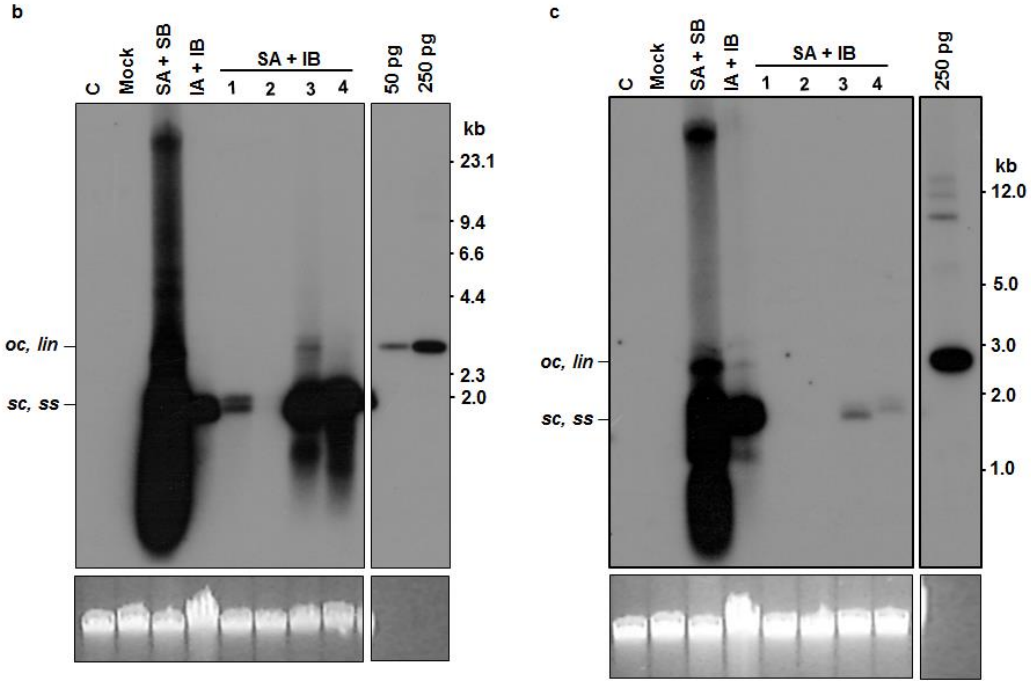


Figure 5. Pseudo-recombination in *N. benthamiana* with SLCMV-[SeM] DNA A and ICMV-[SeT4] DNA B. (**a**) Symptoms in *N. benthamiana* mock inoculated plants (Mock) and plants agroinoculated with the partial dimers of SLCMV-[SeM] DNA A + DNA B, ICMV-[SeT4] DNA A + DNA B and SLCMV-[SeM] DNA A + ICMV-[SeT4] DNA B. (**b**) Southern blot analysis using SLCMV-[SeM] DNA A (ΔCR) labeled with [α-^{32}P]dCTP as the probe. The plasmid pBS-SLCMV-Ma-A digested with PstI (50 pg and 250 pg) was used as the positive control. (**c**) Southern blot analysis using [α-^{32}P]dCTP-labeled SLCMV-[SeM] DNA B (ΔCR) as the probe. The plasmid pBS-SLCMV-Ma-B digested with BamHI (250 pg) was used as the positive control. (**b**,**c**) DNA (1 µg) from uninfected *N. benthamiana* plant (C), plant mock infected with the *A. tumefaciens* strain Ach5 (Mock), plant co-agroinoculated with the partial dimers of SLCMV-[SeM] DNA A + DNA B (SA + SB), plant co-agroinoculated with the partial dimers of ICMV-[SeT4] DNA A + DNA B (IA + IB) and plants co-agroinoculated with the partial dimers of SLCMV-[SeM] DNA A + ICMV-[SeT4] DNA B (SA + IB) were loaded in the respective lanes. Positions of different forms of viral DNA, single stranded (*ss*), super-coiled (*sc*), open circular (*oc*) and linear (*lin*), are marked. Ethidium bromide stained high molecular weight plant DNA is shown as loading control at the bottom.

The leaf blade curling symptom of SLCMV-A/ICMV-B may be associated with the presence of ICMV DNA B, which had elevated the level of SLCMV DNA A accumulation (Figure 5b,c, lanes 3 and 4). The result suggests that SLCMV-[SeM] DNA A trans-replicates ICMV DNA B, while systemic viral movement mediated by ICMV DNA B may be responsible for higher SLCMV DNA A accumulation and enhanced symptoms in comparison to SLCMV DNA A alone infected plants (data not shown). Trans-replication elevated the viral DNA A level possibly by elevating viral movement by DNA B.

Pseudo-recombination was earlier reported between ACMV and SLCMV, which share an identical iteron sequence (5′ AATTGGAGACA 3′) [9]. In the present pseudo-recombination study, ICMV DNA B, which has a non-identical iteron sequence (5′ GGTACTCA 3′) in comparison to that of SLCMV DNA A (5′ AATTGGAGACA 3′), also showed trans-replication. The accumulation of a low level of ICMV DNA B, increased the level of SLCMV DNA A and had an impact on the development of symptoms.

The Rep binding motif (5′ AATCGGTGTC 3′) of the Tomato leaf curl virus (TLCV) is present in the TLCV satellite DNA (sat-DNA) and the replication of this sat-DNA is supported by other taxonomically distinct geminiviruses like Tomato yellow leaf curl virus (TYLCV), ACMV and Beet curly top virus (BCTV) [42]. The infectious pseudo-recombination between two distinct bipartite geminiviruses

BDMV and ToMoV [20] occurred in spite of divergence at one nucleotide position in the Rep binding motif (iteron) [43]. Pseudo-recombination between Tomato leaf curl New Delhi virus (ToLCNDV) and Tomato leaf curl Gujarat virus (ToLCGV) was reported [44]. This pseudo-recombination resulted in enhanced accumulation of ToLCGV DNA B, despite having a different iteron sequence. Hence, pseudo-recombination can occur between two geminivirus species even with non-identical iterons. Our results show that SLCMV DNA A trans-replicates ICMV DNA B at low levels, in spite of differences in the iteron sequences.

In the same Sengutchi cultivar of cassava, the isolate SLCMV-[SeT1] was non-persistent, whereas the isolate SLCMV-[SeM] was persistent. In cassava plants with mixed infections, recovery from SLCMV facilitated the re-emergence of latent ICMV. This shows that SLCMV suppresses ICMV, which results in the prevalence of SLCMV over ICMV in the cassava fields of the Indian sub-continent. The finding of trans-replication of ICMV DNA B by SLCMV DNA A shows that the dynamics of mixed infections in CMD is very complex.

Elucidation of the basis of recovery of cassava plants from the infection caused by the non-persistent SLCMV-[SeT1] isolate can help in designing a strategy to develop CMD resistance in cassava. The alterations of relative titers of SLCMV and ICMV in cassava under different growth conditions and trans-replication of ICMV DNA B by SLCMV DNA A suggest that both ICMV and SLCMV should be simultaneously targeted in any effort to develop CMD resistance in the Indian subcontinent.

Supplementary Materials: The following are available online at www.mdpi.com/1999-4915/8/10/264/s1, Figure S1: Infectivity analysis of Sri Lankan cassava mosaic virus (SLCMV)-[SeT1] partial dimers in *Nicotiana benthamiana* plants.

Acknowledgments: This work was funded by Indo-Swiss Collaboration in Biotechnology (ISCB # CV1.2) which is jointly supported by the Department of Biotechnology (DBT), Govt. of India and Swiss Agency for Development and Co-operation (SDC). University Grants Commission (UGC), Government of India is thanked for the UGC-Basic Scientific Research (BSR) Faculty Fellowship (No. F. 18-1 (61)/2014 (BSR)) to K. Veluthambi. B. Abirami is acknowledged for her technical support.

Author Contributions: K.V. planned and guided *N. benthamiana* agroinfection experiments; T.H. made the critical interpretation on the recovery of cassava from SLCMV and re-emergence of ICMV; M.M.P. carried out deep sequencing and provided the data on small RNAs in cassava; C.K. made partial dimers of both SLCMV and ICMV, performed agroinfection experiments and wrote the manuscript; B.L.P. provided evidence for re-emergence of ICMV; T.R.R. and B.K.B. cloned full-length SLCMV-[SeM] and ICMV-[SeT4] isolates; and S.T. interpreted the data of small RNA deep sequencing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fauquet, C.M.; Briddon, R.W.; Brown, J.K.; Moriones, E.; Stanley, J.; Zerbini, M.; Zhou, X. Geminivirus strain demarcation and nomenclature. *Arch. Virol.* **2008**, *153*, 783–821. [CrossRef] [PubMed]
2. Hanley-Bowdoin, L.; Bejarano, E.R.; Robertson, D.; Mansoor, S. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* **2013**, *11*, 777–788. [CrossRef] [PubMed]
3. Varsani, A.; Navas-Castillo, J.; Moriones, E.; Hernández-Zepeda, C.; Idris, A.; Brown, J.K.; Zerbini, F.M.; Martin, D.P. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Arch. Virol.* **2014**, *159*, 2193–2203. [CrossRef] [PubMed]
4. Idris, A.M.; Brown, J.K. Molecular analysis of Cotton leaf curl virus-Sudan reveals an evolutionary history of recombination. *Virus Genes* **2002**, *24*, 249–256. [CrossRef] [PubMed]
5. Legg, J.P.; Fauquet, C.M. Cassava mosaic geminiviruses in Africa. *Plant Mol. Biol.* **2004**, *56*, 585–599. [CrossRef] [PubMed]
6. Legg, J.P.; Owor, B.; Sseruwagi, P.; Ndunguru, J. Cassava mosaic virus disease in East and Central Africa: Epidemiology and management of a regional pandemic. *Adv. Virus Res.* **2006**, *67*, 355–418. [PubMed]
7. Abraham, A. Tapioca cultivation in India. In *Farm Bulletin No. 17*; Indian Council of Agricultural Research: New Delhi, India, 1956; p. 20.
8. Alagianagalingam, M.N.; Ramakrishnan, K. Cassava mosaic in India. *South Indian Hortic.* **1966**, *14*, 441–448.

9. Saunders, K.; Nazeera, S.; Mali, V.R.; Malathi, V.G.; Briddon, R.W.; Markham, P.G.; Stanley, J. Characterisation of *Sri Lankan cassava mosaic virus* and *Indian cassava mosaic virus*: Evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* **2002**, *293*, 63–74. [CrossRef] [PubMed]
10. Hong, Y.G.; Robinson, D.J.; Harrison, B.D. Nucleotide sequence evidence for the occurrence of three distinct whitefly transmitted geminiviruses in cassava. *J. Gen. Virol.* **1993**, *74*, 2437–2443. [CrossRef] [PubMed]
11. Patil, B.L.; Rajasubramaniam, S.; Bagchi, C.; Dasgupta, I. Both Indian cassava mosaic virus and Sri Lankan cassava mosaic virusare found in India and exhibit high variability as assessed by PCR-RFLP. *Arch. Virol.* **2005**, *150*, 389–397. [CrossRef] [PubMed]
12. Mittal, D.; Borah, B.K.; Dasgupta, I. Agroinfection of cloned Sri Lankan cassava mosaic virus DNA to *Arabidopsis thaliana*, *Nicotianatabacum* and cassava. *Arch. Virol.* **2008**, *153*, 2149–2155. [CrossRef] [PubMed]
13. Chellappan, P.; Vanitharani, R.; Fauquet, C.M. Short interfering RNA accumulation correlates with host recovery in DNA virus infected hosts and gene silencing targets specific viral sequences. *J. Virol.* **2004**, *78*, 7465–7477. [CrossRef] [PubMed]
14. Vanitharani, R.; Chellappan, P.; Fauquet, C.M. Geminiviruses and RNA silencing. *Trends Plant Sci.* **2005**, *10*, 144–151. [CrossRef] [PubMed]
15. Rodríguez-Negrete, E.A.; Carrillo-Tripp, J.; Rivera-Bustamante, R.F. RNA silencing against geminivirus: Complementary action of posttranscriptional gene silencing and transcriptional gene silencing in host recovery. *J. Virol.* **2009**, *83*, 1332–1340. [CrossRef] [PubMed]
16. Góngora-Castillo, E.; Ibarra-Laclette, E.; Trejo-Saavedra1, D.L.; Rivera-Bustamante, R.F. Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). *Virol. J.* **2012**, *9*, 295. [CrossRef] [PubMed]
17. Hagen, C.; Rojas, M.R.; Kon, T.; Gilbertson, R.L. Recovery from *Cucurbit leaf crumple virus* (Family *Geminiviridae*, Genus *Begomovirus*) infection is an adaptive antiviral response associated with changes in viral small RNAs. *Virology* **2008**, *98*, 1029–1037.
18. Chellappan, P.; Vanitharani, R.; Ogbe, F.; Fauquet, C.M. Effect of temperature on geminivirus-induced RNA silencing in plants. *Plant Physiol.* **2005**, *138*, 1828–1841. [CrossRef] [PubMed]
19. Patil, B.L.; Fauquet, C.M. Studies on differential behavior of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes* **2015**, *50*, 474–486. [CrossRef] [PubMed]
20. Gilbertson, R.L.; Hidayat, S.H.; Paplomatas, E.J.; Rojas, M.R.; Hou, Y.; Maxwell, D.P. Pseudo-recombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses. *J. Gen. Virol.* **1993**, *74*, 23–31. [CrossRef] [PubMed]
21. Unseld, S.; Ringel, M.; Höfer, P.; Höhnle, M.; Jeske, H.; Bedford, I.D.; Markham, P.G.; Frischmuth, T. Host range and symptom variation of pseudorecombinant virus produced by two distinct bipartite geminiviruses. *Arch. Virol.* **2000**, *145*, 1449–1454. [CrossRef] [PubMed]
22. Unseld, S.; Ringel, M.; Konrad, A.; Lauster, S.; Frischmuth, T. Virus-specific adaptations for the production of a pseudorecombinant virus formed by two distinct bipartite geminiviruses from Central America. *Virology* **2000**, *274*, 179–188. [CrossRef] [PubMed]
23. Fondong, V.N. Geminivirus protein structure and function. *Mol. Plant Pathol.* **2013**, *14*, 635–649. [CrossRef] [PubMed]
24. Argüello-Astorga, G.R.; Guevara-González, R.G.; Herrera-Estrella, L.R.; Rivera-Bustamante, R.F. Geminivirus replication origins have a group-specific organization of iterative elements: A model for replication. *Virology* **1994**, *203*, 90–100. [CrossRef] [PubMed]
25. Fujii, R.; Kitaoka, M.; Hayashi, K. Error-prone rolling circle amplification: The simplest random mutagenesis protocol. *Nat. Protoc.* **2006**, *1*, 2493–2497. [CrossRef] [PubMed]
26. Hajdukiewicz, P.; Svab, Z.; Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **1994**, *25*, 989–994. [CrossRef] [PubMed]
27. Ditta, G.; Stanfield, S.; Corbin, D.; Helinski, D.R. Broad host-range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7347–7351. [CrossRef] [PubMed]
28. Jacob, S.S.; Vanitharani, R.; Karthikeyan, A.S.; Chinchore, Y.; Thillaichidambaram, P.; Veluthambi, K. *Mungbean yellow mosaic virus*-Vi agroinfection by codelivery of DNA A and DNA B from one *Agrobacterium* strain. *Plant Dis.* **2003**, *87*, 247–251. [CrossRef]

29. Chilton, M.; Currier, T.C.; Farrand, S.K.; Bendich, A.J.; Gordon, M.P.; Nester, E.W. *Agrobacterium tumefaciens* DNA and PS8 bacteriophage DNA not detected in crown gall tumours. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3672–3676. [CrossRef] [PubMed]
30. Grimsley, N.; Hohn, B.; Hohn, T.; Walden, R. "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3282–3286. [CrossRef] [PubMed]
31. Mahajan, N.; Parameswari, C.; Veluthambi, K. Severe stunting in blackgram caused by the *Mungbean yellow mosaic virus* (MYMV) KA27 DNA B component is ameliorated by co-infection or post-infection with the KA22 DNA B: MYMV nuclear shuttle protein is the symptom determinant. *Virus Res.* **2011**, *157*, 25–34. [CrossRef] [PubMed]
32. Rogers, S.O.; Bendich, A.J. Extraction of total cellular DNA from plants, algae and fungi. In *Plant Molecular Biology Manual*, 2nd ed.; Gelvin, S.B., Schilperoort, R.A., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; pp. 1–8.
33. Hong, Y.; Stanley, J. Virus resistance in *Nicotiana benthamiana* conferred by African cassava mosaic virus replication associated protein (AC1) transgene. *Mol. Plant-Microbe Interact.* **1996**, *9*, 219–225. [CrossRef]
34. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **1975**, *98*, 503–517. [CrossRef]
35. Balaji, V.; Vanitharani, R.; Karthikeyan, A.S.; Anbalagan, S.; Veluthambi, K. Infectivity analysis of two variable DNA B components of *Mungbean yellow mosaic virus-Vigna* in *Vigna mungo* and *Vignaradiata*. *J. Biosci.* **2004**, *29*, 297–308. [CrossRef] [PubMed]
36. Sunitha, S.; Shanmugapriya, G.; Balamani, V.; Veluthambi, K. *Mungbean yellow mosaic virus* (MYMV) *AC4* suppresses post-transcriptional gene silencing and an *AC4* hairpin RNA gene reduces MYMV DNA accumulation in transgenic tobacco. *Virus Genes* **2013**, *46*, 496–504. [CrossRef] [PubMed]
37. Fasteris—DNA Sequencing Service. Available online: https://www.fasteris.com/dna/ (accessed on 23 September 2016).
38. Phytozome v11.0. *Manihot esculenta* v6.1 (Cassava). Available online: https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mesculenta (accessed on 23 September 2016).
39. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler Transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef] [PubMed]
40. Rothenstein, D.; Haible, D.; Dasgupta, I.; Dutt, N.; Patil, B.L.; Jeske, H. Biodiversity and recombination of cassava-infecting begomoviruses from southern India. *Arch. Virol.* **2006**, *151*, 55–69. [CrossRef] [PubMed]
41. Von Arnim, A.; Stanley, J. Determinants of *Tomato golden mosaic virus* symptom development located on DNA B. *Virology* **1992**, *186*, 286–293. [CrossRef]
42. Dry, I.B.; Krake, L.R.; Rigden, J.E.; Rezaian, M.A. A novel subviral agent associated with a geminivirus: The first report of a DNA satellite. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7088–7093. [CrossRef] [PubMed]
43. Fontes, E.P.; Gladfelter, H.J.; Schaffer, R.L.; Petty, I.T.; Hanley-Bowdoin, L. Geminivirus replication origins have a modular organization. *Plant Cell* **1994**, *6*, 405–416. [CrossRef] [PubMed]
44. Chakraborty, S.; Vanitharani, R.; Chattopadhyay, B.; Fauquet, C.M. Supervirulent pseudo-recombination and asymmetric synergism between genomic components of two distinct species of begomovirus associated with severe tomato leaf curl disease in India. *J. Gen. Virol.* **2008**, *89*, 818–828. [CrossRef] [PubMed]

New Phytologist

Viral protein suppresses oxidative burst and salicylic acid-dependent autophagy and facilitates bacterial growth on virus-infected plants

Anna S. Zvereva[1], Victor Golyaev[1], Silvia Turco[1], Ekaterina G. Gubaeva[1], Rajendran Rajeswaran[1], Mikhail V. Schepetilnikov[2], Ola Srour[2], Lyubov A. Ryabova[2], Thomas Boller[1] and Mikhail M. Pooggin[1]

[1]Department of Environmental Sciences, Botany, University of Basel, Hebelstrasse 1, Basel 4056, Switzerland; [2]Institut de Biologie Moléculaire des Plantes du CNRS, Université de Strasbourg, Strasbourg Cedex 67084, France

Author for correspondence:
Mikhail Pooggin
Tel: +41 61 2672314
Email: pooggin@yahoo.com

Received: *1 February 2016*
Accepted: *11 March 2016*

New Phytologist (2016) **211:** 1020–1034
doi: 10.1111/nph.13967

Key words: autophagy, *Cauliflower mosaic virus*, effector protein, innate immunity, oxidative burst, RNA silencing, salicylic acid (SA), target-of-rapamycin.

Summary

- Virus interactions with plant silencing and innate immunity pathways can potentially alter the susceptibility of virus-infected plants to secondary infections with nonviral pathogens.
- We found that *Arabidopsis* plants infected with *Cauliflower mosaic virus* (CaMV) or transgenic for CaMV silencing suppressor P6 exhibit increased susceptibility to *Pseudomonas syringae* pv. *tomato* (*Pst*) and allow robust growth of the *Pst* mutant *hrcC*-, which cannot deploy effectors to suppress innate immunity. The impaired antibacterial defense correlated with the suppressed oxidative burst, reduced accumulation of the defense hormone salicylic acid (SA) and diminished SA-dependent autophagy.
- The viral protein domain required for suppression of these plant defense responses is dispensable for silencing suppression but essential for binding and activation of the plant target-of-rapamycin (TOR) kinase which, in its active state, blocks cellular autophagy and promotes CaMV translation.
- Our findings imply that CaMV P6 is a versatile viral effector suppressing both silencing and innate immunity. P6-mediated suppression of oxidative burst and SA-dependent autophagy may predispose CaMV-infected plants to bacterial infection.

Introduction

Plants have evolved potent defense mechanisms to repel pathogenic agents such as viruses, bacteria, fungi and oomycetes, as well as invertebrate animals that damage plants directly by feeding and indirectly by transmitting viruses (Hull, 2014).

Innate immunity plays a major role in plant defense against nonviral pathogens. The frontline of this defense, called pattern-triggered immunity (PTI), is mediated by cell surface receptors that detect pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) and trigger a wide array of responses, including rapid production of reactive oxygen species (ROS burst), phosphorylation of mitogen-activated protein kinases (MAPKs), synthesis of defense hormones such as salicylic acid (SA) and ethylene, and induction of defense genes (Boller & Felix, 2009). PTI against bacteria is elicited through recognition of the flagellin-derived MAMP flg22 by receptor kinase FLS2 and its coreceptor kinase BAK1. BAK1 is also required for PTI responses to another bacterial MAMP, the translation elongation factor-derived peptide elf18, and to plant damage-associated molecular patterns such as the *Arabidopsis* peptides AtPeps (Bartels *et al.*, 2013). The second line of immune defense is mediated by the intracellular receptors, largely from the nucleotide-binding leucine-rich repeat (NB-LRR) family, which detect protein effectors deployed by the pathogens to suppress PTI. Effector-triggered immunity (ETI) is an amplified version of PTI, often associated with a hypersensitive response (HR) involving programmed cell death (Coll *et al.*, 2011; Bonardi & Dangl, 2012) as well as pro-death and/or pro-survival autophagy regulated by SA and ROS (Yoshimoto *et al.*, 2009; Hofius *et al.*, 2011; Zhou *et al.*, 2014).

Plant PTI responses, including ROS burst, cell wall callose deposition and defense gene induction, are also involved in deterring aphids, which transmit a large number of viruses. The elicitor(s) of these responses are not well characterized as yet, but BAK1 is required for the elicitation process in *Arabidopsis* (Prince *et al.*, 2014).

The primary plant defense against viruses is mediated by RNA silencing. In virus-infected plants, the host Dicer-like (DCL) enzymes detect and process viral double-stranded RNA (dsRNA) into short interfering RNAs (siRNAs), and the host Argonaute (AGO) proteins bind the resulting viral siRNAs to form RNA-induced silencing complexes that repress viral gene expression (Blevins *et al.*, 2006; Carbonell *et al.*, 2012; Pooggin, 2013). RNA silencing is reinforced by RNA-dependent RNA polymerase (RDR) enzymes that generate dsRNA precursors of

secondary siRNAs (Wang *et al.*, 2011). Besides its antiviral function in plants and animals (Bronkhorst *et al.*, 2013; Li *et al.*, 2013), the RNA silencing machinery generates endogenous miRNAs and siRNAs that regulate gene expression at both post-transcriptional and transcriptional levels. Certain endogenous miRNAs and siRNAs are involved in regulation of plant defenses against nonviral pathogens. Thus, a subset of NB-LRR genes in dicots and monocots are repressed by miRNAs and RDR6-dependent siRNAs (Zhai *et al.*, 2011; Li *et al.*, 2012; Shivaprasad *et al.*, 2012; Boccara *et al.*, 2014), and successful nonviral pathogens deploy a battery of effector proteins that suppress not only PTI or ETI but also RNA silencing, as shown for bacteria and oomycetes (Navarro *et al.*, 2008; Qiao *et al.*, 2013).

Besides the effectors of nonviral pathogens, plant NB-LRRs can recognize viral avirulence (Avr) proteins. This recognition triggers ETI-like responses, often resulting in HR, which restrict viral infection in certain host plants expressing virus-specific NB-LRRs. In susceptible hosts, the viral Avr proteins have virulence functions and, in some cases, suppress RNA silencing, thus resembling effectors of nonviral pathogens (Zvereva & Pooggin, 2012). This similarity led us to hypothesize that viral effector proteins could also suppress PTI. PTI suppression might offer an advantage to the virus to establish robust systemic infection and to be transmitted by invertebrate vectors. PTI suppression may also predispose virus-infected plants to nonviral pathogens. Interestingly, transgenic *Arabidopsis* plants expressing P6 protein from the *Cauliflower mosaic virus* (CaMV) exhibited enhanced growth of virulent and avirulent *Pseudomonas syringae*, which correlated with delayed HR and reduced accumulation of SA (Love *et al.*, 2012). This finding indicated that CaMV P6 suppresses plant ETI responses and implicated this viral protein in PTI suppression.

Cauliflower mosaic virus belongs to the family *Caulimoviridae*, comprising viruses with circular double-stranded DNA genome, which replicate via reverse transcription of pregenomic RNA (pgRNA) (Hohn, 2013; Hohn & Rothnie, 2013). Aphids transmit CaMV disease among host plants mostly from the family Brassicaceae, which reduces yields and quality of brassica crops worldwide, but some strains of CaMV infect hosts from the family Solanaceae. P6 is a multifuctional protein essential for CaMV virulence, because it is a main component of the viroplasm (viral replication and assembly factory), a transactivator of pgRNA translation, and a suppressor of RNA silencing (Hohn, 2013; Hohn & Rothnie, 2013). P6 has also been implicated in viral movement (Harries *et al.*, 2009; Angel *et al.*, 2013; Rodriguez *et al.*, 2014). Moreover, P6 is a main determinant of the host range and pathogenicity and an Avr factor that triggers HR in resistant hosts (Daubert *et al.*, 1984; Schoelz *et al.*, 1986). The Avr domain was mapped within a variable N-terminus of P6, and this domain is also required for CaMV virulence in susceptible hosts (Agama *et al.*, 2002; Palanichelvam & Schoelz, 2002; Kobayashi & Hohn, 2004; Hapiak *et al.*, 2008). Downstream of the Avr domain, P6 contains a mini-TAV domain with a so-called dsRNA-binding motif implicated in binding and activation of target of rapamycin (TOR) kinase, which is required for translation transactivation (Schepetilnikov *et al.*, 2011). TOR is an evolutionary conserved protein kinase that regulates cell growth and proliferation in response to cellular energy status, growth factors, hormones and nutrient abundance (Ma & Blenis, 2009). In yeast and animal cells, it is engaged in two large complexes, TORC1 and TORC2. TORC1 mediates temporal control of cell growth by activating anabolic processes such as ribosome biogenesis, protein synthesis, transcription, and nutrient uptake, and by inhibiting catabolic processes such as ubiquitin-dependent proteolysis and autophagy. Similar to mammalian TOR, plant TOR is a negative regulator of autophagy (Liu & Bassham, 2010). As animal viruses suppress antiviral autophagy (Levine *et al.*, 2011; Liang *et al.*, 2015), it is conceivable that CaMV P6-mediated activation of TOR may also be required to suppress antiviral autophagy.

In silencing suppression, CaMV P6 interferes with DCL4-mediated processing of RDR6-dependent dsRNA precursors of siRNAs (Love *et al.*, 2007a; Shivaprasad *et al.*, 2008). The antisilencing activity of P6 does not appear to determine the host range or pathogenicity of CaMV strains. Thus, the strain CM1841 infecting Brassicaceae family hosts causes severe disease in *Arabidopsis thaliana* Col-0 plants (Blevins *et al.*, 2006), while the strain D4 infecting Solanaceae family hosts exhibits very mild symptoms in Col-0 plants (Yu *et al.*, 2003), despite the fact that the P6 proteins from these strains have comparable antisilencing activities in Col-0 plants (Shivaprasad *et al.*, 2008).

Here we demonstrate that CaMV P6 from strains CM1841 and JI, but not D4, can suppress ROS burst and SA-dependent autophagy and make *A. thaliana* plants more susceptible to infection with *P. syringae*. Moreover, CaMV-infected and P6-transgenic plants could sustain robust growth of the *P. syringae* mutant *hrcC*-, which cannot deploy effectors, suggesting that P6 functionally substitutes for the bacterial effectors in suppression of PTI. Careful investigation of P6 functional domains revealed that the TOR-binding domain is responsible for suppression of PTI, but dispensable for silencing suppression. Our data indicate that P6, via its TOR-binding domain, can down-regulate autophagy by activating TOR kinase. P6 from the strain D4, which is defective in TOR activation in *A. thaliana*, failed to suppress autophagy and PTI-mediated antibacterial defense.

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh wild-type (Col-0) and mutant/transgenic lines were germinated in sterile soil or half-strength Murashige and Skoog (½ MS) salt media (Phyto-technology Laboratories, Shawnee Mission, KS, USA) supplied with 1% sucrose and 0.7% phytoagar and grown in phytochambers (Sanyo, Gunma, Japan) at 20–22°C and 10, 12 or 24 h photoperiod. *Nicotiana benthamiana* was grown in soil in an open glasshouse at 25°C. The *A. thaliana* homozygous mutant lines (all in Col-0 background) used in our study have been genotyped and characterized previously: *atg5-1* (NASC SAIL_129_B07) (Thompson *et al.*, 2005), *atg18a-1* (Lenz *et al.*, 2011), *fls2* (SAIL_691_C04) (Zipfel *et al.*, 2004), *efr1* (SALK_N544334) (Zipfel *et al.*, 2006), *bak1-3* (SALK_034523) and *bak1-4*

(SALK_116202) (Chinchilla *et al.*, 2007); *bak1-5* (Schwessinger *et al.*, 2011), *adr1* triple (Bonardi *et al.*, 2011), *sid2-1* (Nawrath & Metraux, 1999), *pepr1/2* (Flury *et al.*, 2013), *nik1* (Fontes *et al.*, 2004), *rps4* (NASC SALK_012799C) (Alonso *et al.*, 2003) and *sag101 (*NASC SALK_022911C*)* (Alonso *et al.*, 2003). The transgenic lines (all in Col-0 background) have been described previously: P6-CM1841 (as lines CM-6 and CM-8) and P6-D4 (as D4-2) (Yu *et al.*, 2003); P6-JI (as AT7) and P6-JIΔdsR (as AT7ΔdsR) (Schepetilnikov *et al.*, 2011); and GTOR (as G548) and TORi (as 35-7) (Deprost *et al.*, 2007). The transgenic lines P6-CM-HA and P6-CMΔ23-HA were generated as follows. Coding sequences of P6-CM and P6-CMd23 were subcloned from the CaMV strain CM1841 infectious clone pCa122 and its d23 deletion derivative (Kobayashi & Hohn, 2004), respectively, into pEarlyGate p201 vector (Earley *et al.*, 2006) using PCR with primers AttB1_TAV_s and AttB2_TAV_as in Supporting Information Table S1 and following a standard Gateway protocol. The resulting plasmids carrying N-terminal HA-tag fusion-protein expression cassettes under the control of CaMV 35S promoter were mobilized to *Agrobacterium tumefaciens* strain GV3101 for transformation of Col-0 plants by the floral dip method following the protocol described in Clough & Bent (1998). Note that all the P6-transgenic *A. thaliana* (Col-0) lines studied here stably express different P6 variants under the control of the constitutive CaMV 35S promoter.

Inoculation with CaMV and its mutants

Following germination in soil, 4-wk-old Arabidopsis was inoculated by biolistic delivery of 1 μg plasmid pCa122 (strain CM1841) (Kobayashi *et al.*, 2002) or its derivatives (CM-d3 and CM-dsR), as described previously (Akbergenov *et al.*, 2006). The CM-dsR mutant virus was constructed by subcloning an artificially synthesized fragment *Sac*I-*Pvu*II (Eurofins, Ebersberg, Germany), lacking the dsR domain of P6 (amino acids 136–182), into pCa122 plasmid. The resulting viral clone was sequenced to confirm the introduced deletion. One month after inoculation, unless otherwise stated, virus-infected plants were used for pathotests or harvested in pools and ground in liquid nitrogen for total RNA, DNA and protein preparations as described later. Viral DNA titers were measured by real-time PCR using CaMV-specific and plant 18S rDNA-specific primers (Table S1).

P6 transient expression in *N. benthamiana*

The CM1841 P6 coding sequence was subcloned from plasmid pCa122 to pEarlyGate vector p100 (Earley *et al.*, 2006) as described earlier. The resulting plasmid, which carries the CaMV 35S promoter-driven wild-type P6-CM protein expression cassette, was mobilized to *A. tumefaciens* strain C58C1 for agroinfiltration assays. At 3 wk postgermination, the *N. benthamiana* leaves were syringe-infiltrated with the agrobacterial suspension at $OD_{600} = 0.3$ in agroinfiltration buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$, 100 μM acetosyringone). At 2 dpi, the P6- and the empty agroinfiltrated leaves were treated with the bacterial MAMP, and the ROS burst was measured as described later.

Bacterial infection assays

The *Pst* DC3000 (Katagiri *et al.*, 2002) and its 'effector-deficient' (*Pst hrcC*- defective in type III effector secretion system) (Roine *et al.*, 1997; Boch *et al.*, 2002) and avirulent (*Pst* AvrRpt2) (Kunkel *et al.*, 1993) strains were grown to $OD_{600} = 0.6$–1 in LB media supplemented with 10 mM $MgSO_4 \times 7H_2O$ with the appropriate antibiotics. Cells were harvested by centrifugation and the pellets were resuspended in 10 mM $MgCl_2$ to $OD_{600} = 0.001$, corresponding to 5×10^5 colony-forming units (CFU) ml^{-1} for bacterial growth assays, or to $OD_{600} = 0.2$, corresponding to 10^8 CFU ml^{-1} for SA assay. Before inoculation, Silwet L-77 was added to a final concentration 0.04% (v/v). The *A. thaliana* plants were dipped in bacterial suspension and exposed for 10 s to a vacuum and then left in the suspension for another 5 min. Plants were grown further under the transparent plastic covers in the Sanyo phytochamber at 20°C and 10 h photoperiod. Leaf disc samples were taken from two leaves per plant, and four plants per genotype. The samples were ground in 10 mM $MgCl_2$ with a pestle and plated as serial dilutions on the LB media supplemented with 10 mM $MgSO_4 \times 7H_2O$. Plates were incubated at 28°C for 2 d and the number of colonies was counted. The bacteria were genotyped by PCR (see Table S1 for primer sequences) to confirm the presence or absence of the *hrcC* locus.

Measurement of SA

Free SA and glucose-conjugated SA (SAG) measurements were performed as previously described (Bonardi *et al.*, 2011) with some modifications, as follows. Leaves (100 mg) were collected at 24 h postinfiltration with *Pst* AvrRpt2 at 10^8 CFU ml^{-1} and frozen in liquid nitrogen. Samples were ground and the tissues homogenized in 200 μl 0.1 M acetate buffer (pH 5.6). Samples were then centrifuged for 15 min at 16 000 ***g*** at 4°C. Supernatant (100 μl) was transferred to a new tube for free SA measurement, and 10 μl was incubated with 1 μl of 0.5 U μl^{-1} β-glucosidase from almonds (Sigma-Aldrich) for 90 min at 37°C for total SA measurement. After incubation, plant extracts were diluted fivefold with 44 μl acetate buffer for total SA measurement. A total of 60 μl of LB media, 5 μl of plant extract (treated or not with β-glucosidase), and 50 μl of *Acinetobacter* sp. *ADPWH-lux* (OD = 0.4) were added to each well of a LIA white 96-well plate (Greiner Bio-One, Frickenhausen, Germany). The plate was incubated at 37°C for 60 min, and luminescence was read with the MicroLumat LB96P plate reader (Berthold Technologies, Bad Wildbad, Germany). For the standard curve, 1 μl of the known amount of SA stock (0–1000 μg ml^{-1}) was diluted 10-fold in acetate buffer, and 5 μl of each standard (undiluted for free SA measurement, or fivefold diluted for total SA) was added to the wells of the plate containing 60 μl of LB and 50 μl of *Acinetobacter* sp. *ADPWHlux* (OD = 0.4). SA standards were read in parallel with the experimental samples.

ROS burst measurement

Reactive oxygen species burst was measured as described by Bartels *et al.* (2013). Briefly, leaf discs of diameter 4 mm were floated

overnight in sterile water in the darkness at room temperature. The next day water was replaced with a solution 10 μg ml^{-1} horseradish peroxidase (Sigma-Aldrich) and 100 μM luminol (Sigma-Aldrich). Luminescence was measured in a MicroLumat LB96P plate reader (Berthold Technologies) for 30 min immediately after addition of elicitor flg22 or elf18 to a final concentration of 1 μM, and plotted in all figures as the peak of luminescence achieved during the 30 min of measurements.

Ethylene measurement

Ethylene was measured as described by Bartels *et al.* (2013). Briefly, leaf discs of diameter 4 mm of 4-wk-old plants were floated overnight in sterile water. The following day the leaves were transferred in 6 ml glass vials (two discs per vial) containing 200 μl of 1 μM aqueous solution of flg22 or elf18. The tubes were closed with rubber septa, and ethylene accumulating in the free air space was measured by GC (GC-14A Shimadzu, Tokyo, Japan) after 3–4 h incubation.

Western blot analysis

Ten milligrams of plant material were ground in liquid nitrogen and solubilized in 100 μl of 6 × sodium dodecyl sulfate (SDS) sample buffer (0.35M Tris, pH 6.8, 22.4% glycerol, 10% SDS, 0. 6M DTT, bromophenol blue), heated to 95°C for 5 min, and centrifuged at 12 000 ***g*** for 5 min, and the supernatant was loaded onto 9% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) for separation of P6, FLS2, and BAK1, and 12% for MAPKs and EDS1. The proteins were blotted onto polyvinylidene fluoride membrane (GE Healthcare, Europe GmbH, Glattbrugg, Switzerland), and blocked with 2% BSA w/v (Sigma) in TBS-tween (0.1%) for 1–2 h. The primary antibodies were diluted to the following concentrations in TBS-tween (0.1%), if not indicated otherwise, and incubated overnight at +4°C: anti-P6 (Schepetilnikov *et al.*, 2011), 1 : 20 000 in TBS-tween (1%); anti-FLS2 (Chinchilla *et al.*, 2007), 1 : 300; anti-BAK1 (Schulze *et al.*, 2010), 1 : 300; anti-phospho-p44/p42-MAPK (Cell Signalling Technology, Danvers, MA, USA), 1 : 1000 in 5% BSA w/v; anti-AtMPK3 (Gonzalez Besteiro *et al.*, 2011), 1 : 2000 in 2% BSA w/v; anti-TOR (Schepetilnikov *et al.*, 2011), 1 : 1000 in 2% BSA w/v; phospho-TOR-S2448P (Cell Signalling Technology), 1 : 500 in 2% BSA w/v. Secondary goat anti-rabbit antibodies (Agrisera, Vännäs, Sweden) were diluted 1 : 25 000 in TBS-tween (0.1%) and incubated for 1 h at room temperature. Signals were visualized using ECL Prime Western Blot Detection Reagent (GE Healthcare). The membranes were then stained with amidoblack or ponceau for loading control.

MAPK phosphorylation assay

The assay was conducted as described by Bartels *et al.* (2013) with some modifications. Two-week-old seedlings grown in soil or CaMV-infected plants at 1 month postinoculation were pooled as follows: five seedlings or four leaf discs (6 mm in diameter) per tube, respectively, floated in 300 μl of sterile water overnight. The following day, flg22 or elf18 was added to a final concentration of 1 μM and, after 5 and 15 min incubation, the plant material was frozen in liquid nitrogen. We then proceeded with western blot analysis as described earlier.

Induction of autophagy by SA analogue benzothiadiazole (BTH)

The roots of 9-d-old seedlings grown on ½ MS media solidified with 0.7% phytoagar at 20°C with a 24 h photoperiod were excised and incubated in liquid ½ MS media supplemented with 10 μM E-64d (Enzo), in the presence or absence of 100 μM BTH (Sigma-Aldrich). After 4–5 h incubation, the roots were transferred in water and observed by conventional transmission light microscopy (Nikon Eclipse 80i) under ×100 magnification. An autophagy-deficient *atg5-1* mutant was used as a negative control. Induction was calculated as $\log_{10}$ of the ratio between the number of vesicles per cell in BTH-treated roots and buffer-treated roots of each plant line. The experiment was repeated three times and at least 100 cells per root sample were inspected for the characteristic vesicles representing autophagosomes (Yoshimoto *et al.*, 2009).

RNA isolation and blot hybridization

Total RNA preparation and high-resolution blot hybridization were performed as described in detail previously (Akbergenov *et al.*, 2006; Rajeswaran & Pooggin, 2012). Briefly, total RNA was extracted from leaf tissues ground in liquid nitrogen using TRIzol reagent (Sigma-Aldrich), following the manufacturer's instructions. For analysis of small RNAs, 10 μg samples of total RNA were separated on 15% polyacrylamide gel. Ethidium bromide (EtBr) staining of the gels was used for loading control. RNAs were transferred to Hybond N+ membrane (Amersham) and the membrane was sequentially hybridized with ^{32}P-ATP-labeled DNA oligonucleotide probes specific for plant or viral small RNAs (Table S1).

DNA isolation, quantitative PCR (qPCR) and reverse transcription qPCR (RT-qPCR)

Total DNA was isolated as described previously (Pooggin *et al.*, 1998). SybrGreen master mix (Roche) was used for qPCR reactions. Accumulation of CaMV DNA was monitored by qPCR analysis, normalized to the internal reference gene 18S rDNA (At2g01010) and plotted relative to the Col-0 infected with CaMV CM1841. Gene expression of *PR-1* (At2g14610) was monitored by RT-qPCR analysis. cDNA was synthesized from 2.5 μg total RNA using SuperScript III reverse transcriptase (Invitrogen) and oligo-dT. The expression was normalized to the internal reference gene *ACT2* (At3g18780) and plotted relative to the Col-0 or mock steady-state expression level. The PCR primers are given in Table S1.

Transient expression assay in plant protoplasts

Cauliflower mosaic virus P6 translation transactivation activity was measured as described in detail previously (Schepetilnikov *et al.*,

New Phytologist

2011). Briefly, mesophyll protoplasts were prepared from 3- to 4-wk-old seedlings of *A. thaliana* (Col-0) and cotransfected with plasmid DNA of a bicistronic β-glucuronidase (GUS) construct (pbiGUS), a monocistronic green fluorescent protein (GFP) construct (pGFP) and a CaMV 35S promoter-driven P6 expression construct (pTAV, or its derivative with different P6 variants, or an empty vector). After 20–24 h of incubation at 27°C, protoplasts were harvested and protein extract was prepared for measuring activity of GUS (fluorimetric assay) and accumulation of P6 and GFP proteins (western blotting). Subcloning of P6 variants into *Sal*I and *Kpn*I sites of pTAV, in place of the P6 (strain JI) coding sequence, was performed by PCR from the corresponding CaMV strains or mutants (for PCR primers, see Table S1).

Results

CaMV infection facilitates bacterial growth in *Arabidopsis* via P6 protein

To test if the viral infection can predispose plants to bacterial pathogens, we examined growth rates of *P. syringae* on *A. thaliana* Col-0 plants infected with CaMV strain CM1841. After appearance of CaMV disease symptoms, the plants were inoculated with the virulent *Pst* DC3000 or its 'effector-deficient' derivative *Pst hrcC*- defective in the type III secretion system and therefore incapable of deploying effectors (Roine *et al.*, 1997; Boch *et al.*, 2002). Bacterial growth was estimated by measuring bacterial cell number (CFU) at 1 and 3 d postinoculation (dpi). Both wild-type and effector-deficient bacteria exhibited increased growth rates on CaMV-infected plants compared with control virus-free plants (Fig. 1a). Notably, the effector-deficient bacteria grew on virus-infected plants almost as fast as the virulent bacteria on virus-free plants. This indicated that one or more viral components accumulating in CaMV-infected plants could functionally substitute the bacterial effectors in suppression of PTI.

As transgenic expression of P6 from CaMV strain JI was shown to promote growth of *P. syringae* (Love *et al.*, 2012), we tested whether the P6 protein from strain CM1841 (P6-CM) is responsible for CaMV-mediated enhancement of bacterial growth. To this end, two independent *A. thaliana* Col-0 transgenic lines expressing comparable levels of P6-CM protein (lines CM6 and CM8; Yu *et al.*, 2003) were used. Both lines

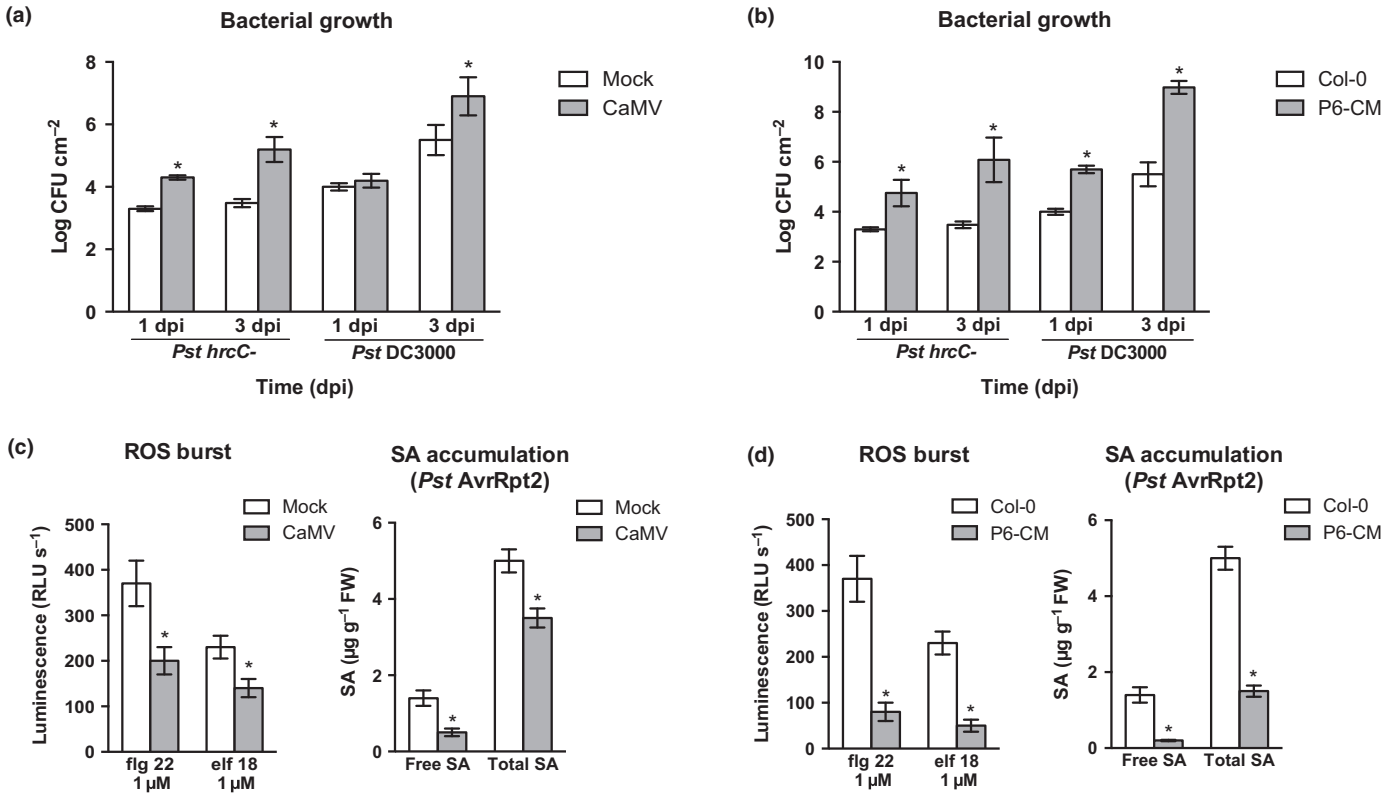


Fig. 1 *Cauliflower mosaic virus* (CaMV) infection and P6 expression facilitate growth of virulent (*Pseudomonas syringae* pv. *tomato* (*Pst* DC3000)) and effector-deficient (*Pst hrcC*-) bacteria on *Arabidopsis thaliana* by suppressing reactive oxygen species (ROS) burst and salicylic acid (SA) production. (a, b) Number of bacteria on CaMV-infected (CaMV), compared with virus-free mock-inoculated Col-0 plants (mock) (a), and noninfected P6-CM transgenic lines (P6-CM) compared with nontreated Col-0 plants, at 1 and 3 d postinoculation (dpi) with *Pst* DC3000 and *Pst hrcC*- (b), plotted as $\log_{10}$ of colony-forming units (CFU) cm^{-2} of the leaf. (c, d, left plots) ROS burst triggered by the bacterial microbe-associated molecular patterns (MAMPs) flg22 and elf18 on CaMV-infected (c) and P6-transgenic plants (d), plotted as peak of relative luminescence units (RLU) s^{-1} during 30 min of measurements following the addition of 1 μM MAMP peptide. (c, d, right plots) SA production triggered by *Pst* AvrRpt2 on CaMV-infected (c) and P6-transgenic plants (d), plotted as μg of free and total SA g^{-1} of leaf tissue (FW). Data in all the panels represent means ± SE; *n* (number of biological replicates) is as follows: *Pst* DC3000 ($n = 7$), *hrcC*- ($n = 7$), ROS ($n = 7$), SA ($n = 5$). Significant differences between means are indicated (*, $P < 0.05$; multiple *t*-test – one per row; www.graphpad.com/guides/prism/6/statistics/index.htm?stat) in each pairwise comparison (i.e. CaMV vs Mock, or P6-CM vs Col-0).

demonstrated the ability to promote growth of virulent *Pst* DC3000 and to allow robust growth of effector-deficient *Pst hrcC*-; the latter was comparable to the growth of virulent *Pst* on control nontransgenic Col-0 plants (Fig. 1b; Fig. S1a; note that 'P6-CM' stands for the line CM6, unless otherwise stated). These results indicate that CaMV protein P6 is responsible for the virus-mediated suppression of plant immunity and can thus compensate for the *P. syringae* effectors in suppression of PTI.

CaMV infection and P6 protein expression suppress oxidative burst and SA production

To study the mechanism of PTI suppression we measured MAMP-triggered immune responses including ROS burst, ethylene accumulation and MAPK phosphorylation. The *A. thaliana* NADPH-oxidase AtRbohD produces extracellular ROS burst when activated upon MAMP recognition (Torres *et al.*, 1998; Galletti *et al.*, 2008). In both CaMV-infected and P6-CM transgenic plants, examined at a time point corresponding to the time point before inoculation in the bacterial growth assays, the extracellular ROS burst in response to the bacterial MAMPs flg22 and elf18 was significantly suppressed (Fig. 1c,d), and the degree of suppression was well correlated with the bacterial growth rates (Fig. 1a,b). We also measured MAMP-triggered ROS burst at the earlier time points postinoculation with CaMV. Interestingly, at 5 dpi the ROS burst in response to the bacterial MAMPs was not suppressed but rather enhanced (Fig. 2a), while during the later stages of CaMV infection (at 27 and 33 dpi), when the viral protein P6 accumulated at the highest levels (Fig. 2b), the ROS burst was suppressed (Fig. 2a). This indicates that the MAMP-triggered oxidative burst, apparently promoted by certain viral component(s) during the early stages of CaMV infection, is eventually suppressed by the viral protein P6. We corroborated this finding by examining the flg22-triggered ROS burst in *N. benthamiana*: *Agrobacterium*-mediated transient expression of P6-CM protein in leaves of *N. benthamiana* seedlings significantly suppressed the ROS burst (Fig. 2c), similar to transgenic expression of P6-CM in *A. thaliana*. We conclude that P6 is responsible for virus-mediated suppression of the ROS burst.

Induction of ethylene production is among the hallmarks of plant PTI (Boller & Felix, 2009). We tested whether accumulation of ethylene was compromised upon CaMV infection and P6 expression. The basal concentration of ethylene in P6-transgenic plants was slightly higher than in control Col-0 or CaMV-infected Col-0 plants, while the levels of elf18- and flg22-induced ethylene accumulation in both CaMV-infected and P6 transgenic plants were slightly lower than in the Col-0 control (Fig. 2d). Although our statistical analysis revealed that these differences within each treatment were not significant ($P > 0.05$; multiple *t*-tests – one per row (GraphPad Prism), $n = 3$), comparison between treatments (i.e. mock vs each MAMP) showed that ethylene production was strongly induced in control plants (as expected), less markedly induced in CaMV-infected plants, and not induced in P6-CM transgenic plants. Thus, the inducibility of ethylene production upon MAMP treatment was somewhat compromised by CaMV infection and strongly compromised by P6 expression.

It has been reported that accumulation of the FLS2 receptor is compromised in ethylene-insensitive mutants, thereby reducing their ability to produce a ROS burst in response to MAMPs (Boutrot *et al.*, 2010; Mersmann *et al.*, 2010). However, FLS2 protein accumulation in P6-transgenic or CaMV-infected plants was comparable to that in control Col-0 plants (Fig. 2e). Likewise, BAK1 protein accumulation was not reduced (Fig. 2e).

To examine if CaMV P6 interferes with MAMP-triggered MAPK activation, young leaves of P6-transgenic plants and systemic leaves of CaMV-infected plants (taken at a time point corresponding to the time point before inoculation in the bacterial growth assays) were treated with flg22 and elf18 and the kinetics of MAPK phosphorylation was examined by western blotting using antibodies specific to the phosphorylated MAPKs AtMPK3, AtMPK4 and AtMPK6. At time zero, no phosphorylation was observed, indicating that neither CaMV P6 or any other viral component at the late infection stage (30 dpi) preactivate these MAPKs. At 5 and 15 min posttreatment with either flg22 or elf18, all three AtMPKs were found to be phosphorylated in both CaMV-infected and P6 transgenic plants and the degree of their phosphorylation was higher than in control plants (Fig. 2f). Using antibody specific to nonphosphorylated AtMPK3, this kinase was found to overaccumulate in CaMV-infected plants (Fig. 2g). Hence, the increased PAMP-dependent phosphorylation of AtMPK3 (and possibly AtMPKs 4 and 6) in CaMV-infected plants accumulating high levels of P6 (Fig. S1e) is correlated with the increased MAPK protein accumulation. A similar increase in accumulation of nonphosphorylated AtMPKs 3 has been observed upon treatment with the SA analog BTH, which in turn resulted in up-regulation of the defense gene *PR-1* (Beckers *et al.*, 2009). We therefore examined if *PR-1* gene was up-regulated in CaMV-infected or P6-transgenic plants. Despite the increased accumulation of AtMPK3, the levels of *PR-1* mRNA were slightly decreased in CaMV-infected plants (at 30 dpi) and more drastically decreased in P6-transgenic plants (Fig. S1d). Notably, the more pronounced decrease in *PR-1* mRNA accumulation correlates with higher bacterial growth rates in P6 transgenic plants (Fig. 1a,b). Taken together, our findings indicate that CaMV P6 suppresses plant immunity at a step downstream of MAPK activation. Alternatively, P6-mediated suppression of PTI is independent of the flg22 or elf18 signal transduction pathways involving the three AtMPKs.

It has been proposed that an oxidative burst during the early phase of immune response induces SA synthesis at the site of infection and that this increased SA concentration potentiates further ROS production in a self-amplifying feedback loop (Vlot *et al.*, 2009). As CaMV infection and P6 expression suppress the MAMP-triggered oxidative burst, SA synthesis might in turn be also impaired. We therefore tested the SA accumulation at 1 dpi with *Pst* carrying an AvrRpt2 effector, which is known to be a potent trigger of ROS burst and SA synthesis (Tsuda & Katagiri, 2010). In both CaMV-infected and P6-transgenic plants, examined at a time point corresponding to the time point before

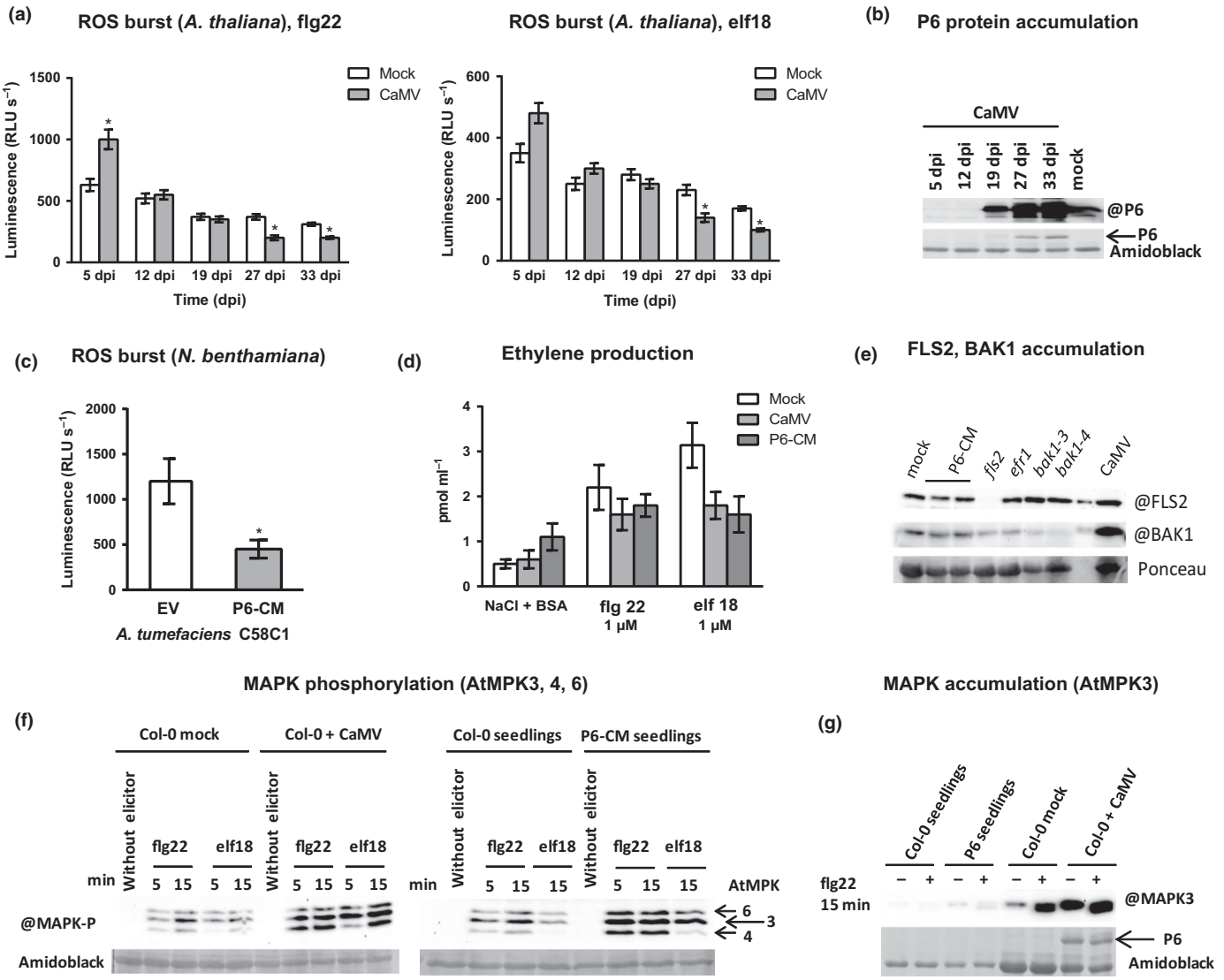


Fig. 2 Plant defense responses modulated by *Cauliflower mosaic virus* (CaMV) infection and P6 expression. (a) Reactive oxygen species (ROS) burst triggered by flg22 and elf18 on *Arabidopsis thaliana* Col-0 plants mock-inoculated (mock) or infected with CaMV strain CM1841 (CaMV) at 5, 12, 19, 27 and 33 d postinoculation (dpi), plotted as peak of relative luminescence units (RLU) s^{-1} during 30 min of measurements following the addition of 1 μM microbe-associated molecular pattern (MAMP) peptide (flg22, or elf18). (b) Western blot analysis of P6 protein accumulation in the CaMV-infected plants from (a). The mock sample corresponds to 33 dpi. Amidoblack staining of the blot membrane is shown as loading control, with an arrow indicating P6. (c) ROS burst triggered by flg22 on *Nicotiana benthamiana* leaves infiltrated with an empty *Agrobacterium tumefaciens* strain C58C1 (EV) or the C58C1 carrying a binary vector with the 35S promoter-driven P6-CM expression cassette (P6-CM), plotted as RLU s^{-1} following the addition of 1 μM flg22 peptide. (d) Ethylene production triggered by flg22 or elf18 on CaMV-infected Col-0 plants at 30 dpi, plotted as pmol of ethylene ml^{-1} of air, following the addition of 1 μM MAMP peptide or buffer without MAMP (NaCl + BSA). (e) Western blot analysis of FLS2 and BAK1 protein accumulation in P6-CM transgenic plants, CaMV-infected Col-0 plants (CaMV) and the Col-0 mutant lines *fls2*, *efr1*, *bak1-3*, and *bak1-4*. Ponceau staining is shown as loading control. (f) Western blot analysis of accumulation of phoshorylated mitogen-activated protein kinases (MAPKs) in P6-CM transgenic and Col-0 control seedlings as well as CaMV-infected and mock-inoculated plants treated with flg22 or elf18 or buffer without elicitor (NaCl + BSA) at two time points (5 and 15 min) posttreatment, using anti-phospho-p44/42-MAPK antibodies (AtMPK3, AtMPK3, AtMPK3 proteins are indicated by arrows). Amidoblack staining is shown as loading control. (g) Western blot analysis of nonphosphorylated AtMPK3 protein in P6-CM transgenic and Col-0 control seedlings as well as CaMV-infected and mock-inoculated Col-0 plants treated with 1 μM flg22 (+) or buffer without elicitor (−) at 15 min posttreatment. Data in (a, c, d) show means ± SE; $n = 3$. Significant differences between means are indicated (*, $P < 0.05$; multiple *t*-test – one per row) in each pairwise comparison (i.e. CaMV vs mock, or P6 vs respective control).

inoculation in the bacterial growth assays, SA production in response to AvrRpt2 was reduced (Fig. 1c,d) and the degree of SA reduction was correlated with the growth rates of virulent and effector-deficient bacteria (Fig. 1a,b). Our finding that P6 from strain CM1841 suppresses SA accumulation in response to avirulent bacteria is corroborated by similar findings in transgenic *A. thaliana* La-er plants expressing P6 from strain JI (Love *et al.*, 2012).

P6 from strain D4 failed to suppress antibacterial defense, but not silencing, in *Arabidopsis*

Cauliflower mosaic virus strain D4 develops severe systemic symptoms in Solanaceous hosts (Schoelz *et al.*, 1986). However, this strain exhibits only very mild symptoms in *A. thaliana* Col-0, compared with the severe disease symptoms of strain CM1841 (Yu *et al.*, 2003), even though its P6 protein (P6-D4) exhibited strong antisilencing activity when expressed in Col-0 transgenic plants (Shivaprasad *et al.*, 2008). We therefore investigated whether P6-D4 is able to suppress PTI. We hypothesized that the pathogenicity and the host range of different CaMV strains might be determined by P6-mediated suppression of innate immunity, in addition to its antisilencing activity. We tested transgenic Col-0 plants expressing P6-D4 for bacterial growth and immune responses as described earlier. Unlike P6-CM, P6-D4 expression did not facilitate the growth of *Pst* DC3000 or *Pst hrcC-* (Fig. 3a; note that the growth rates are presented as an increment of bacterial numbers between days 1 and 3; for the original data, see Fig. S1b). Furthermore, the MAMP-triggered ROS burst and the AvrRpt2-triggered SA production in P6-D4 plants were similar to those in Col-0 control plants (Fig. 3b,c). Consistent with the previous findings (Shivaprasad *et al.*, 2008), both P6-D4 and P6-CM proteins exhibited strong antisilencing activities, as highly abundant dsRNA precursors of plant siRNAs were detected in both P6-D4 and P6-CM plants (Fig. 3d). The precursor–siRNA ratio was six times higher in the case of P6-D4, probably because this protein accumulated at higher levels than P6-CM (Fig. 3e). These results indicate that P6-D4 is not able to interfere with PTI-based responses in *A. thaliana*. Furthermore, the antisilencing activity of CaMV P6 is not sufficient for its effector function in PTI suppression.

Identification of the P6 protein domain required for suppression of antibacterial defense in *Arabidopsis*

The mini-TAV domain of P6 (amino acids 113–242) has been implicated in P6-mediated transactivation of viral pgRNA translation (Ryabova *et al.*, 2006; Hohn & Rothnie, 2013) (Fig. 4d). This domain contains an RNase H motif (amino acids 140–182), which was sufficient for binding RNA-DNA hybrids and dsRNA *in vitro* (Cerritelli *et al.*, 1998). To evaluate if this dsRNA-binding (dsR) motif is required for P6-mediated suppression of plant defenses, we tested *A. thaliana* Col-0 transgenic lines that express P6 protein from CaMV strain JI (P6-JI) or its dsR motif (amino acids 136–182) deletion mutant (P6-JIΔdsR) for their susceptibility to *P. syringae* and their ability to suppress immune responses and silencing. Note that, similar to the strain CM1841 (but unlike D4), CaMV strain JI causes severe disease symptoms in *A. thaliana* Col-0 plants (Schepetilnikov *et al.*, 2011). We found that similar to P6-CM, transgenic expression of P6-JI allowed robust growth of the effector-deficient bacteria and facilitated growth of the virulent bacteria, which correlated well with suppression of the MAMP-triggered ROS burst and AvrRpt2-triggered SA production. By contrast, transgenic expression of the P6-JIΔdsR mutant protein did not promote the growth of virulent or suppressor-deficient bacteria, and did not suppress ROS burst or SA production (Figs 3a–c, S1c). However, both P6-JI and P6-JIΔdsR were equally able to interfere with siRNA processing in the transgenic plants (Fig. 3d). We conclude that the dsR motif, being dispensable for P6 antisilencing activity, is required for P6-mediated suppression of innate immunity.

P6 protein may suppress SA-dependent authophagy via TOR binding and activation

Schepetilnikov *et al.* (2011) demonstrated that the dsR motif of P6 from CaMV strain JI is essential for binding and activation of *A. thaliana* TOR kinase and for P6-mediated transactivation of translation, and that TOR-deficient (TORi) *A. thaliana* plants are resistant to JI-virus infection. P6-mediated activation of TOR is manifested by its phosphorylation at Ser 2424 (Schepetilnikov *et al.*, 2013). It is well documented that plant TOR inactivation strongly up-regulates autophagy (Liu & Bassham, 2010), and SA signaling might be involved in this regulation (Yoshimoto *et al.*, 2009). We therefore hypothesized that P6-mediated activation of TOR contributes to suppression of antiviral autophagy, in addition to transactivation of viral pgRNA translation. The ability of CaMV P6 variants to suppress antibacterial defense may also depend on the efficiency of autophagy suppression. To validate this hypothesis we first compared TOR phosphorylation in the P6-CM and P6-JI plants, which exhibit suppressed PTI responses and compromised antibacterial defense, with that in the P6-D4 and P6-JIΔdsR plants, which exhibit normal PTI responses and antibacterial defense. Consistent with our hypothesis, high amounts of phosphorylated TOR accumulated in seedlings expressing P6-JI and P6-CM, whereas a substantially lower amount of phosphorylated TOR was detected in seedlings expressing P6-D4 and only a trace amount of phosphorylated TOR was detectable in seedlings expressing P6-JIΔdsR (Fig. 4a). Hence, both P6-JI and P6-CM proteins are able to activate TOR, which correlates with their ability to facilitate bacterial growth. By contrast, P6-JI protein lacking the TOR-binding dsR domain and P6-D4 being defective in TOR activation both fail to facilitate bacterial growth (Fig. 3a). It should be noted that P6-D4 does activate TOR, albeit much less pronouncedly than P6-JI or P6-CM, and such low amounts of phosphorylated TOR may not be sufficient for suppression of antibacterial defense. To test if P6-D4 is also defective in transactivation of TOR-dependent translation in *A. thaliana*, we made use of a well-established assay in protoplasts from *A. thaliana* Col-0 leaves, in which transient expression of P6-JI transactivates bicistronic translation to the levels of *c.* 15–20% of control monocistronic translation (Schepetilnikov *et al.*, 2011). In this assay, P6-CM transactivated translation 2.2 times more efficiently than P6-JI (Fig. 4b). By contrast, P6-D4 transactivated translation four times less efficiently than P6-JI, while P6-JI lacking the dsR motif failed to transactivate translation (Fig. 4b). Thus, the relative abilities of P6 variants to transactivate translation are correlated with their abilities to activate TOR (Fig. 4a). P6-D4, being derived from the viral strain adapted to Solanaceae family plants, is somewhat compromised

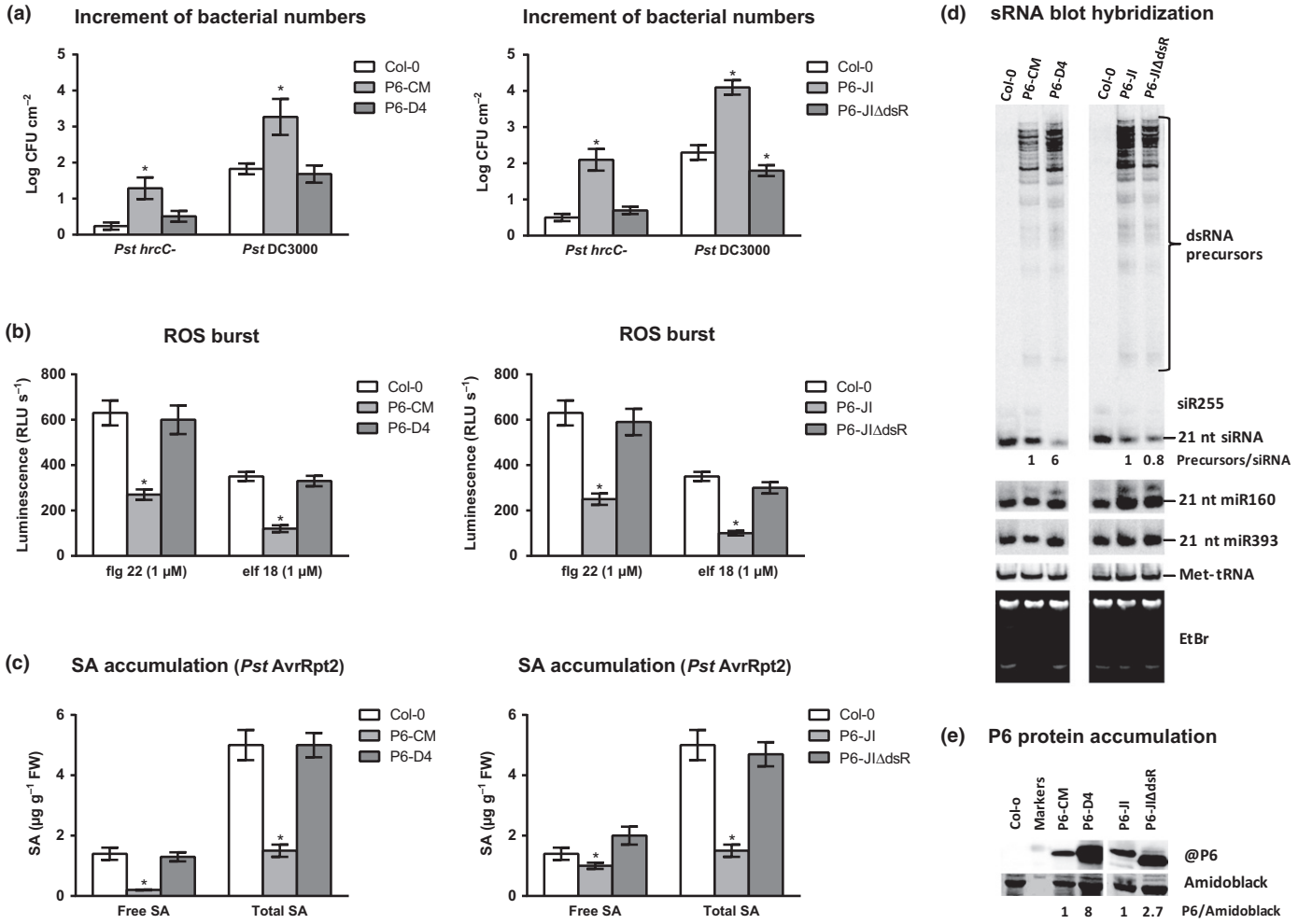


Fig. 3 Effects of P6 homologs from *Cauliflower mosaic virus* (CaMV) strains CM1841, D4 and JI and the P6-JI dsR motif-deletion mutant on bacterial growth, reactive oxygen species (ROS) burst, salicylic acid (SA) production and double-stranded (ds)RNA processing in *Arabidopsis thaliana*. (a) Increment of bacterial numbers on nontransgenic (Col-0) and P6-transgenic (P6-CM, P6-D4, P6-JI, P6-JIΔdsR) plants between 1 and 3 d postinoculation (dpi) with *Pseudomonas syringae* pv. *tomato* (*Pst*) *hrcC*- and *Pst* DC3000, plotted as $\log_{10}$ colony-forming units (CFU) cm^{-2} of the leaf. (b) ROS burst triggered by the bacterial microbe-associated molecular patterns (MAMPs) flg22 and elf18 on the P6-transgenic and control plants, plotted as the peak of relative luminescence units (RLU) s^{-1} during 30 min of measurements following the addition of 1 μM MAMP peptide. (c) SA accumulation triggered by *Pst* AvrRpt2 on the P6-transgenic and control plants, plotted as μg of free and total SA g^{-1} of leaf tissue. (d) Blot hybridization analysis of total RNA from the P6-transgenic and control plants. The blot membranes were successively hybridized with short DNA probes specific for plant 21 nt transacting short interfering RNA (siRNA; siR255), micro-RNAs (miRNAs; miR160, miR393) and Methyonine transfer-RNA (Met-tRNA) (see Supporting Information Table S1 for probe details). Positions of the siRNA and its long dsRNA precursors are indicated and the precursor/siRNA relative ratios are shown under the respective scan, with the ratio for P6-CM or P6-JI set to 1. (e) Western blot analysis of P6 protein accumulation in the P6-transgenic and control plants using anti-P6 antibody. Amidoblack staining of the blot membranes is shown as loading control. The normalized densities (P6/amidoblack) are shown under the scans, with the value for P6-CM or P6-JI set to 1. Bars in (a–c) show means ± SE ($n = 5$). Significant differences between means are indicated (*, $P < 0.05$; multiple *t*-test – one per row) in each pairwise comparison (i.e. each P6 line vs Col-0).

in its transactivation function in *A. thaliana*, probably owing to its weak ability to activate TOR in this Brassicaceous plant.

Following up our hypothesis, as P6-D4 and P6-JIΔdsR proteins are not able to activate TOR as efficiently as P6-JI and P6-CM, they may not be able to suppress autophagy. To examine SA-dependent autophagy in the P6-transgenic lines, we monitored formation of autophagosome vesicles in their roots upon treatment with the SA analog BTH as described previously (Yoshimoto *et al.*, 2009). The induction of autophagy was estimated as the ratio between the numbers of vesicles per cell in the roots treated with BTH to mock-treated roots of the same line. Roots of the *A. thaliana atg5* mutant line deficient in autophagosome formation were used as a negative control. After 4 h of BTH treatment, the induction of autophagy was substantially compromised in *atg5*, P6-CM and P6-JI roots compared with the Col-0 control. By contrast, the induction of autophagy in P6-D4 and P6-JIΔdsR roots was similar to that in Col-0 (Fig. 4c). We conclude that the TOR-binding dsR motif is essential for P6-mediated suppression of SA-dependent autophagy. Furthermore, P6 protein from strain D4 infecting Solanaceae family hosts is defective in activation of TOR and suppression of SA-dependent autophagy in *A. thaliana* that belongs to the family Brassicaceae.

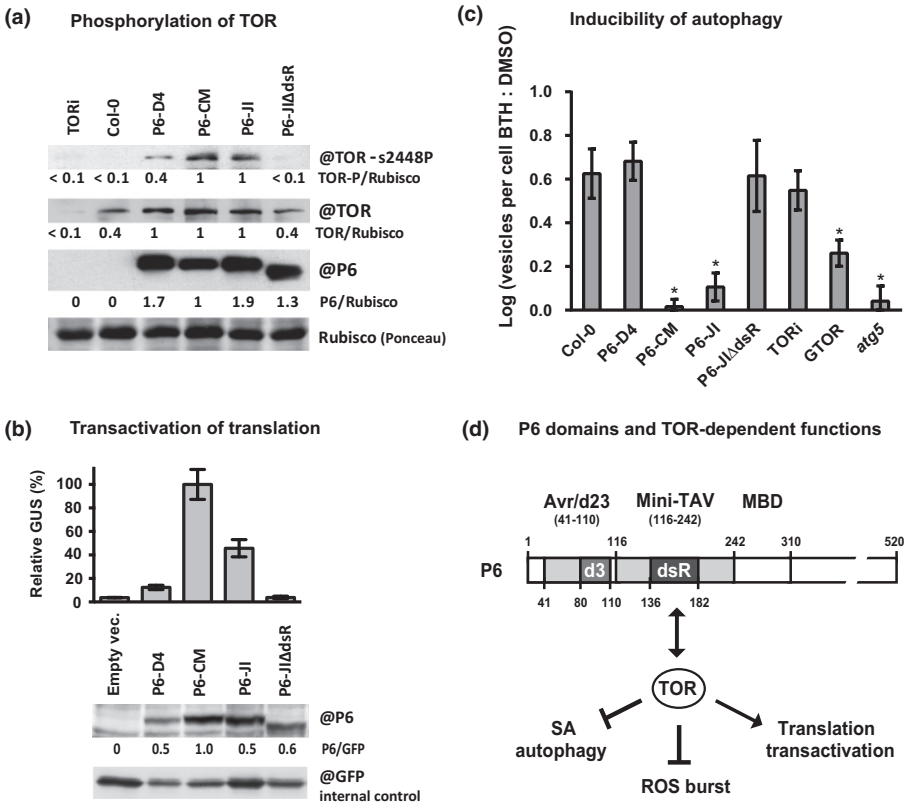


Fig. 4 Effects of P6 homologs from *Cauliflower mosaic virus* (CaMV) strains CM1841, D4 and JI and the P6-JI dsR motif-deletion mutant on phosphorylation of plant target-of-rapamycin (TOR) kinase, transactivation of TOR-dependent translation and induction of salicylic acid (SA)-dependent autophagy in *Arabidopsis thaliana*. (a) Western blot analysis of total protein from nontransgenic (Col-0), P6-transgenic (P6-D4, P6-CM, P6-JI, P6-JIΔdsR) and TOR-knockdown (TORi) seedlings. The membrane was cut in two parts. The upper part was first hybridized with anti-TOR-S2448-P and then anti-TOR antibody (@), while the lower part was hybridized with anti-P6 antibody and then stained with ponceau for loading control. (b) Transactivation of translation of a bicistronic β-glucuronidase (GUS) reporter construct in *A. thaliana* Col-0 protoplasts coexpressing the P6 variants P6-D4, P6-CM, P6-JI, P6-JIΔdsR, or an empty vector as a negative control. GUS transactivation by P6-CM is set to 100%. Western blot analysis of P6 protein accumulation in protoplasts is shown under the GUS bar graph. The blot membrane was cut in two parts. The upper part was hybridized with anti-P6 antibody, while the lower part was hybridized with anti-green fluorescent protein (anti-GFP) antibody). GFP served as an internal control for protein transient expression efficiency. (c) Inducibility of autophagy in roots of wild-type (Col-0), P6-transgenic (P6-D4, P6-CM, P6-JI, P6-JIΔdsR), TOR-knockdown (TORi), TOR-overexpression (GTOR) and autophagy-deficient (*atg5*) lines, plotted as $\log_{10}$ ratios between the numbers of autophagy vesicles per cell in benzothiadiazole (BTH)-treated roots and buffer (dimethyl sulfoxide (DMSO))-treated roots for each line. Data show means ± SE ($n = 3$). Significant differences between means are indicated (*, $P < 0.05$; multiple *t*-test – one per row) in each pairwise comparison (i.e. each transgenic/mutant line vs Col-0). (d) Schematic representation of P6 protein domains including Avr/d23, mini-TAV with TOR-binding dsR motif, and multiple binding domain (MBD; Ryabova *et al.*, 2006). The domains are boxed with the deletions within Avr/d23 (d3) and mini-TAV (dsR) highlighted in dark gray. Borders of the domains and the deletions are indicated with amino acid numbering. Effects of TOR binding to the dsR domain on SA-dependent autophagy, reactive oxygen species (ROS) burst and translation are depicted under the scheme.

To further confirm that SA-dependent autophagy is controlled by plant TOR, we compared BTH-triggered autophagosome formation in roots of the transgenic *A. thaliana* lines deficient in TOR (TORi) or overexpressing TOR (GTOR). Consistent with our hypothesis, the induction of autophagy was compromised only in GTOR (not TORi) roots (Fig. 4c).

The P6 dsR domain is essential for CaMV infectivity in *Arabidopsis*

To test the importance of P6-mediated suppression of ROS burst and SA-dependent authophagy for CaMV virulence, we generated the CaMV CM1841 mutant virus lacking the dsR motif of P6 (amino acids 136–182) and tested it for infectivity in *A. thaliana* Col-0 plants. None of the 50 plants in several rounds of inoculation were infected with the mutant virus. This indicates that the dsR motif is essential for CaMV infectivity. Consistent with this finding, CaMV P6 deletion mutants lacking amino acids 137–166 or 167–200 did not replicate in turnip protoplasts (Kobayashi & Hohn, 2003).

We then tested if replication and systemic spread of the dsR-deletion mutant virus can be restored in *A. thaliana* lines deficient in immune responses (*bak1-4*, *bak1-5*, *adr1* triple, *sid2-1*, *rps4*, *sag101*) and autophagy (*atg5*, *atg18a-1* and GTOR). None of these plant lines could support systemic infection of the dsR-deletion mutant virus. At this stage, we are not sure that inability of the P6ΔdsR mutant protein to suppress plant immune defenses can, on its own, explain the loss of infectivity of the

dsR-deletion mutant virus, because the dsR motif is required not only for P6-mediated suppression of oxidative burst and authophagy but also for TOR-dependent transactivation of viral pgRNA translation.

The P6 Avr domain modulates immune responses but not silencing suppression

The P6 Avr domain (amino acids 41–110) contributes to CaMV pathogenicity and host range (Agama *et al.*, 2002; Palanichelvam & Schoelz, 2002; Kobayashi & Hohn, 2004; Hapiak *et al.*, 2008). Deletions of (or within) the Avr domain drastically affect infectivity of the CaMV strain CM1841 and disease symptom development in turnip plants (Kobayashi & Hohn, 2004). However, the CaMV CM1841 mutant virus CM-d3, carrying a deletion within the P6 Avr domain (amino acids 80–110), was shown to accumulate at high titers in systemic leaves of *A. thaliana* Col-0 plants (as measured by enzyme-linked immunosorbent assay), albeit developing no disease symptoms. Furthermore, CM-d3 infection was not able to suppress transgene-induced silencing in *A. thaliana*, which led to the conclusion that the deleted domain is responsible for silencing suppression (Laird *et al.*, 2013). To examine if this domain contributes to P6-mediated suppression of plant immune responses, we tested the same mutant virus CM-d3. Although we reproduced the finding that CM-d3 develops attenuated and mild disease symptoms in turnip Just Right plants (Kobayashi & Hohn, 2004), this mutant virus was noninfectious in *A. thaliana* Col-0 plants under our conditions. We also tested CM-d3 for infectivity in several immune-deficient *A. thaliana* lines, including *bak1-3*, *bak1-4*, *fls2*, *efr1*, *pepr1/2* and *nik1*. All these lines were resistant to the mutant virus, and most of the plants accumulated no viral DNA, with a notable exception for a small proportion of *bak1-3* plants (two out of 50 inoculated plants) accumulating the mutant virus DNA in systemic leaves (Fig. 5a). However, CM-d3 DNA titer did not exceed 1.2% of the wild-type virus DNA titer on Col-0 plants, as measured by real-time PCR (Fig. 5b). Notably, the degree of virus-mediated interference with siRNA processing correlated well with the titers of the wild-type and the mutant virus, suggesting that CM-d3 is not deficient in silencing suppression (Fig. 5c).

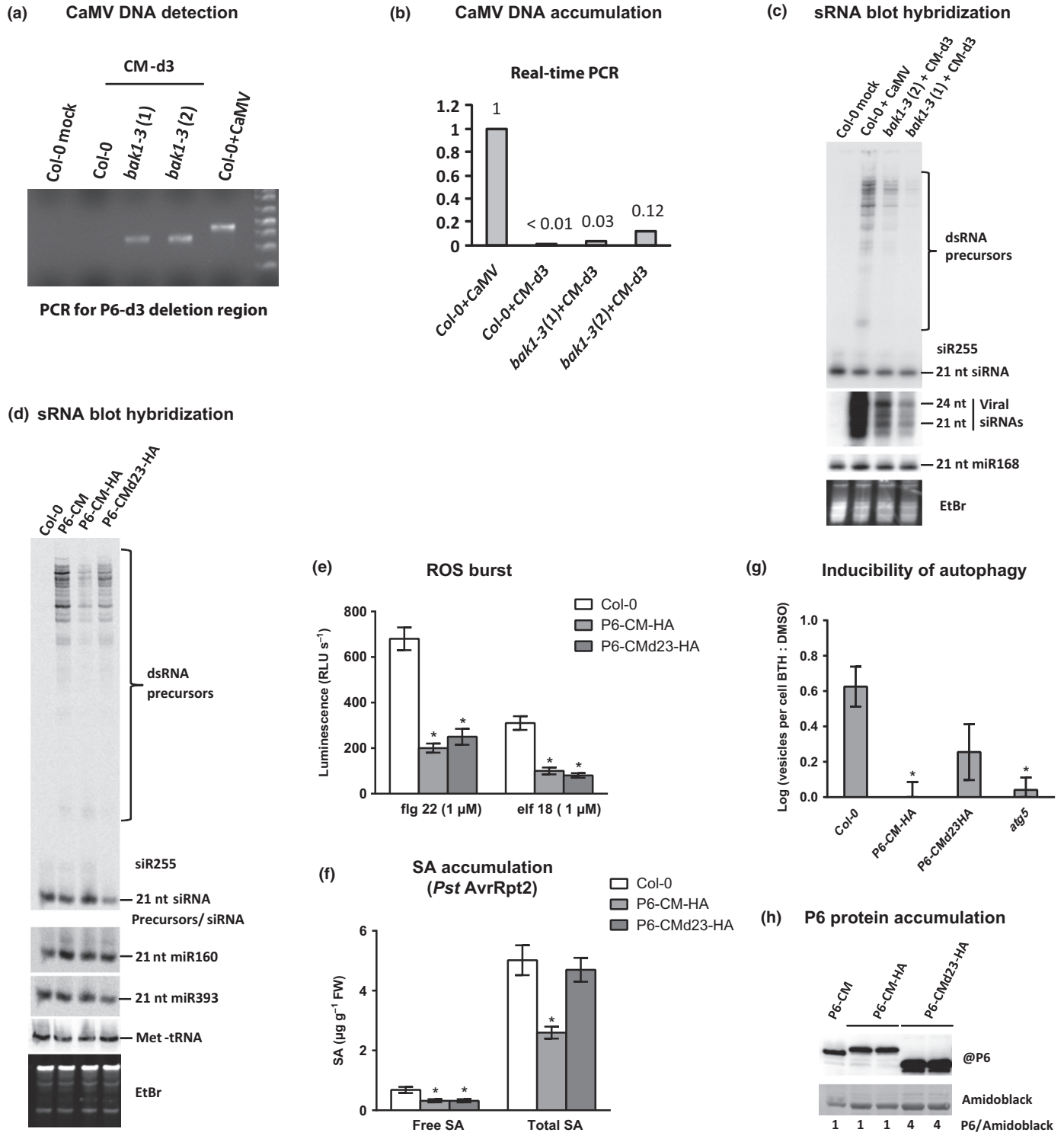
To further clarify whether or not the Avr domain is required for P6-mediated silencing suppression, we generated transgenic Col-0 lines that express HA-tagged P6 protein from strain CM1841 (P6-CM-HA) or its deletion variant d23 (amino acids 41–110) lacking the entire Avr domain (P6-CMd23-HA). Although the addition of HA tag reduced P6-mediated interference with plant siRNA processing, the deletion of Avr domain did not have any negative effect, but rather increased accumulation of the unprocessed siRNA precursors (Fig. 5d; compare P6-CMd23-HA and P6-CM-HA); the latter positive effect was correlated with higher accumulation of the P6-CMd23-HA protein compared with the P6-CM-HA protein (Fig. 5h). Taking our findings together, the Avr domain is not essential for P6 antisilencing activity, albeit this domain is crucial for CaMV virulence in *A. thaliana*. Note that we investigated here a role of the Avr domain in P6-mediated interference with DCL4 processing of plant siRNA precursors, whereas Laird *et al.* (2013) investigated its role in P6-mediated suppression of transgene-induced silencing, which may explain the conflicting findings.

We then tested responsiveness of P6-CMd23 transgenic plants to the bacterial MAMPs and AvrRpt2 and to BTH treatment. The results revealed that the P6 Avr domain is not required for suppression of ROS burst (Fig. 5e) but contributes to suppression of SA production (Fig. 5f) and SA-dependent autophagy (Fig. 5g). We note that the inducibility of autophagy was somewhat compromised in P6-CMd23 plants, compared with the Col-0 control (albeit this difference was not found to be statistically significant; $P > 0.05$ in three biological replicates), while in P6-CM plants the inducibility of autophagy was abolished ($P < 0.05$; $n = 3$), similar to the autophagy-deficient *atg5* plants (Fig. 5g). We speculate the Avr domain modulates P6-mediated suppression of SA-dependent immune responses, although it is not absolutely essential for this TOR-dependent function of P6 (Fig. 4d).

Fig. 5 Effect of deletions within the P6 Avr domain on *Cauliflower mosaic virus* (CaMV) infectivity and P6-mediated suppression of silencing in *Arabidopsis thaliana*. (a) PCR analysis of total DNA extracted from Col-0 and *bak1-3* plants inoculated with the Avr domain deletion-mutant virus (CM-d3) and Col-0 plants infected with the wild-type virus (CaMV). The deletion in CM-d3 virus was verified by PCR amplification of the N-terminal part of the P6 gene. Of 50 inoculated *bak1-3* plants, only two plants, designated (1) and (2), were found to be PCR-positive. (b) Real-time PCR analysis of viral DNA accumulation in the CM-d3 infected *bak1-3* plants (1) and (2), compared with CaMV-infected Col-0 plant (set to 1). (c) Blot hybridization analysis of total RNA from the CM-d3-infected *bak1-3* plants (1) and (2), compared with the CaMV-infected Col-0 plant. The blot membranes were successively hybridized with probes specific for plant 21 nt transacting siRNA (siR255), 21 nt miRNA (miR168) and viral 21–24 nt siRNAs (see Supporting Information Table S1 for probe details). Positions of the small RNAs and long dsRNA precursors of siR255 are indicated. Ethidium bromide (EtBr) staining is shown as loading control. (d) Blot hybridization analysis of total RNA from transgenic plants expressing wild-type P6 (P6-CM), its HA-tagged version (P6-CM-HA) and Avr-deletion version (P6-CMd23-HA). The blot membrane was successively hybridized with probes specific for siR255, miR160, miR393 and Met-tRNA. (e) Reactive oxygen species (ROS) burst triggered by flg22 and elf18 on Col-0 control and P6-transgenic (P6-CM-HA and P6-CMd23-HA) plants, plotted as relative luminescence units (RLU) s^{-1} following the addition of 1 μM flg22 or elf18. (f) Salicylic acid (SA) production triggered by *Pseudomonas syringae* pv. *tomato* (*Pst*) AvrRpt2 on Col-0 control and P6-transgenic (P6-CM-HA and P6-CMd23-HA) plants, plotted as μg of free and total SA g^{-1} of leaf tissue. (g) Inducibility of autophagy in roots of P6-transgenic lines (P6-CM-HA and P6-CMd23-HA) vs wild-type (Col-0) and autophagy-deficient (*atg5*) lines, plotted as log ratios between the numbers of autophagy vesicles per cell in BTH-treated roots and dimethyl sulfoxide (DMSO)-treated roots for each line. Data in (e–g) are means ± SE; $n = 3$. Significant differences between means are indicated (*, $P < 0.05$; multiple *t*-test – one per row) in each pairwise comparison (i.e. each P6 line vs Col-0). (h) Western blot analysis of P6 protein accumulation in P6-transgenic lines (P6-CM, P6-CM-HA and P6-CMd23-HA). Amidoblack staining is shown as loading control. The normalized densities (P6/amidoblack) are shown under the scans, with the value for P6-CM set to 1. siRNA, short interfering RNA; dsRNA, double-stranded RNA; miRNA, micro-RNA; Met-tRNA, methyonine transfer-RNA.

Discussion

Here we hypothesized that, in order to establish systemic and vector-transmissible infection, viruses must express effector proteins that suppress both silencing- and innate immunity-based host defenses. As a result of virus interactions with the host defense systems, virus-infected plants can potentially become more susceptible to nonviral pathogens. To provide evidence in support of this hypothesis, we used the model plant *A. thaliana* with its vast genetic resources, the biotrophic bacterial pathogen *P. syringae* with its virulent, avirulent and effector-deficient strains, and the aphid-transmissible pararetrovirus CaMV, the extensive studies of which had a major impact in plant molecular pathology (Scholthof *et al.*, 2011). Our study demonstrates that systemic infection of *A. thaliana* plants with CaMV leads to increased susceptibility to *P. syringae*, owing to suppression of the



plant antibacterial defense responses by the viral effector protein P6. Notably, the antisilencing activity of P6 was dispensable for this effector function, whereas the ability of P6 to activate plant TOR kinase correlated well with suppression of ROS burst and SA-dependent autophagy. Furthermore, the P6 Avr domain, which, in addition to its TOR-binding domain, contributes to suppression of SA-dependent autophagy is also dispensable for silencing suppression. Thus, CaMV P6 protein appears to be a multidomain effector protein with uncoupled activities in suppression of RNA silencing and innate immunity. It is worth noting that the TOR-binding and dsR motif-containing domain of CaMV P6 has also been reported to physically interact with the ribosomal proteins L13 and L18, both implicated in viral translation (Ryabova *et al.*, 2006). Furthermore, a larger mini-TAV sequence containing the TOR-binding domain physically interacts with the chloroplast outer membrane protein CHUP1 and the plasma membrane (plasmodesmata-associated) C2-calcium-dependent protein AtSRC2.2, both implicated in viral movement in *Arabidopsis* (Angel *et al.*, 2013; Rodriguez *et al.*, 2014). Interestingly, a pepper homolog of AtSRC2.2 is also involved in plant defenses against viral and nonviral pathogens (Kim *et al.*, 2008). Taken together, these findings imply a complex regulation of P6 effector functions in viral translation, movement and defense suppression via multiple host proteins competing for one binding domain in the viral effector protein.

We observed that at the early time point postinoculation with CaMV (5 dpi), when P6 is not detectable in systemic leaves, the oxidative burst in response to bacterial MAMPs was not suppressed but rather enhanced (Fig. 2a,b), suggesting that other viral components may potentiate immune responses at the early stage of viral infection. Love *et al.* (2005) have reported increased accumulation of ROS (H_2O_2) in systemic leaves at 2 h postinoculation with CaMV virions (but not naked viral DNA), thus implicating the CaMV coat protein in elicitation of plant immune responses. Interestingly, ethylene signaling has been implicated in this elicitation process (Love *et al.*, 2005). However, the plant mutant lines compromised in ethylene signaling exhibit enhanced resistance to CaMV (Love *et al.*, 2007b). We observed here that the plants systemically infected with CaMV or transgenic for P6-CM are both compromised in inducibility of ethylene production in response to bacterial PAMPs, suggesting that P6 interferes with ethylene biosynthesis. Earlier, the transgenic plants expressing P6-JI were found to be insensitive to ethylene (Geri *et al.*, 2004). Taken together, P6-mediated interference with ethylene biosynthesis and signaling may contribute to weakening the plant antibacterial defense.

Our findings imply that virus-infected plants can generally be predisposed to secondary infections with bacterial (and perhaps other nonviral) pathogens and raise several outstanding questions for future research. It remains to be examined by comprehensive field surveys if plants systemically infected with CaMV and other viruses are more susceptible to biotrophic nonviral pathogens under natural conditions. In the CaMV–*Arabidopsis* model system, it remains to be addressed if suppression of ROS burst and SA-dependent authophagy by the viral protein P6 is solely required for the virus to overcome the plant defense and thereby establish systemic infection, or if it is, additionally or alternatively, required to facilitate virus transmission from plant to plant by aphids. We speculate that the impaired ROS burst at the late stages of CaMV infection might make virus-infected plants more susceptible to aphids. Indeed, BAK1-dependent ROS burst is one of the potent immune responses elicited by aphids on *A. thaliana*, reducing aphids' survival rates (Prince *et al.*, 2014). Autophagy has been implicated in plant defense against bacterial and viral pathogens (Liu *et al.*, 2005; Hofius *et al.*, 2009; Yoshimoto *et al.*, 2009; reviewed in Zhou *et al.*, 2014). Our findings implicate SA-dependent autophagy in plant–CaMV interactions. As a well-studied mammalian TOR kinase blocks autophagy in its active phosphorylated state (Kim & Guan, 2015), it is conceivable that phosphorylation of plant TOR mediated by CaMV P6 might lead to suppression of antiviral authophagy.

Acknowledgements

We thank Jeff Dangl for providing *Acinetobacter* sp. *ADPWH-lux*, *Pst* AvrRpt2, and *adr1* triple, *eds1-2* and *sid2-1* mutant lines; Dagmar Hann for *Pst* DC3000 and *Pst hrcC-*; Delphine Chinchilla for anti-BAK1 and anti-FLS2 antibodies, *bak1-3*, *bak1-4*, *fls2*, *efr1* mutant lines and flg22 and elf18 peptides; Sebastian Bartels for anti-AtMPK3 and *pepr1/2* mutant lines; Jim Schoelz for P6-CM6, P6-CM8 and P6-D4 transgenic lines; Cyril Zipfel for the *bak1-5* mutant line; Elizabeth Fontes for the *nik1* mutant line; and Andrea Gust for the *atg18a-1* mutant line. The work was supported by the Swiss National Science Foundation grant 31003A_143882 to M.M.P. and the Swiss Government Excellence Scholarship to V.G.

Author contributions

A.S.Z., L.A.R., T.B. and M.M.P. planned and designed the research. A.S.Z., V.G., S.T., E.G.G., R.R., M.V.S. and O.S. performed experiments. A.S.Z., L.A.R. and M.M.P. analyzed data. A.S.Z. and M.M.P wrote the manuscript.

References

Agama K, Beach S, Schoelz J, Leisner SM. 2002. The 5′ third of cauliflower mosaic virus gene VI conditions resistance breakage in Arabidopsis ecotype Tsu-0. *Phytopathology* **92**: 190–196.

Akbergenov R, Si-Ammour A, Blevins T, Amin I, Kutter C, Vanderschuren H, Zhang P, Gruissem W, Meins F Jr, Hohn T *et al.* 2006. Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Research* **34**: 462–471.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al.* 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.

Angel CA, Lutz L, Yang X, Rodriguez A, Adair A, Zhang Y, Leisner SM, Nelson RS, Schoelz JE. 2013. The P6 protein of cauliflower mosaic virus interacts with CHUP1, a plant protein which moves chloroplasts on actin microfilaments. *Virology* **443**: 363–374.

Bartels S, Lori M, Mbengue M, van Verk M, Klauser D, Hander T, Boni R, Robatzek S, Boller T. 2013. The family of Peps and their precursors in Arabidopsis: differential expression and localization but similar induction of pattern-triggered immune responses. *Journal of Experimental Botany* **64**: 5309–5321.

Beckers GJ, Jaskiewicz M, Liu Y, Underwood WR, He SY, Zhang S, Conrath U. 2009. Mitogen-activated protein kinases 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana*. *Plant Cell* **21**: 944–953.

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F Jr, Hohn T *et al.* 2006. Four plant dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research* **34**: 6233–6246.

Boccara M, Sarazin A, Thiebeauld O, Jay F, Voinnet O, Navarro L, Colot V. 2014. The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathogens* **10**: e1003883.

Boch J, Joardar V, Gao L, Robertson TL, Lim M, Kunkel BN. 2002. Identification of *Pseudomonas syringae* pv. tomato genes induced during infection of *Arabidopsis thaliana*. *Molecular Microbiology* **44**: 73–88.

Boller T, Felix G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* **60**: 379–406.

Bonardi V, Dangl JL. 2012. How complex are intracellular immune receptor signaling complexes? *Frontiers in Plant Science* **3**: 237.

Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. 2011. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proceedings of the National Academy of Sciences, USA* **108**: 16463–16468.

Boutrot F, Segonzac C, Chang KN, Qiao H, Ecker JR, Zipfel C, Rathjen JP. 2010. Direct transcriptional control of the Arabidopsis immune receptor FLS2 by the ethylene-dependent transcription factors EIN3 and EIL1. *Proceedings of the National Academy of Sciences, USA* **107**: 14502–14507.

Bronkhorst AW, Miesen P, van Rij RP. 2013. Small RNAs tackle large viruses: RNA interference-based antiviral defense against DNA viruses in insects. *Fly (Austin)* **7**: 216–223.

Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT, Carrington JC. 2012. Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell* **24**: 3613–3629.

Cerritelli SM, Fedoroff OY, Reid BR, Crouch RJ. 1998. A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucleic Acids Research* **26**: 1834–1840.

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones JD, Felix G, Boller T. 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**: 497–500.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant Journal* **16**: 735–743.

Coll NS, Epple P, Dangl JL. 2011. Programmed cell death in the plant immune system. *Cell Death and Differentiation* **18**: 1247–1256.

Daubert SD, Schoelz J, Debao L, Shepherd RJ. 1984. Expression of disease symptoms in cauliflower mosaic virus genomic hybrids. *Journal of Molecular and Applied Genetics* **2**: 537–547.

Deprost D, Yao L, Sormani R, Moreau M, Leterreux G, Nicolai M, Bedu M, Robaglia C, Meyer C. 2007. The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. *EMBO Reports* **8**: 864–870.

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant Journal* **45**: 616–629.

Flury P, Klauser D, Schulze B, Boller T, Bartels S. 2013. The anticipation of danger: microbe-associated molecular pattern perception enhances AtPep-triggered oxidative burst. *Plant Physiology* **161**: 2023–2035.

Fontes EP, Santos AA, Luz DF, Waclawovsky AJ, Chory J. 2004. The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes & Development* **18**: 2545–2556.

Galletti R, Denoux C, Gambetta S, Dewdney J, Ausubel FM, De Lorenzo G, Ferrari S. 2008. The AtrbohD-mediated oxidative burst elicited by oligogalacturonides in Arabidopsis is dispensable for the activation of defense responses effective against *Botrytis cinerea*. *Plant Physiology* **148**: 1695–1706.

Geri C, Love AJ, Cecchini E, Barrett SJ, Laird J, Covey SN, Milner JJ. 2004. Arabidopsis mutants that suppress the phenotype induced by transgene-mediated expression of cauliflower mosaic virus (CaMV) gene VI are less susceptible to CaMV-infection and show reduced ethylene sensitivity. *Plant Molecular Biology* **56**: 111–124.

Gonzalez Besteiro MA, Bartels S, Albert A, Ulm R. 2011. Arabidopsis MAP kinase phosphatase 1 and its target MAP kinases 3 and 6 antagonistically determine UV-B stress tolerance, independent of the UVR8 photoreceptor pathway. *Plant Journal* **68**: 727–737.

Hapiak M, Li Y, Agama K, Swade S, Okenka G, Falk J, Khandekar S, Raikhy G, Anderson A, Pollock J *et al.* 2008. Cauliflower mosaic virus gene VI product N-terminus contains regions involved in resistance-breakage, self-association and interactions with movement protein. *Virus Research* **138**: 119–129.

Harries PA, Palanichelvam K, Yu W, Schoelz JE, Nelson RS. 2009. The cauliflower mosaic virus protein P6 forms motile inclusions that traffic along actin microfilaments and stabilize microtubules. *Plant Physiology* **149**: 1005–1016.

Hofius D, Munch D, Bressendorff S, Mundy J, Petersen M. 2011. Role of autophagy in disease resistance and hypersensitive response-associated cell death. *Cell Death and Differentiation* **18**: 1257–1262.

Hofius D, Schultz-Larsen T, Joensen J, Tsitsigiannis DI, Petersen NH, Mattsson O, Jorgensen LB, Jones JD, Mundy J, Petersen M. 2009. Autophagic components contribute to hypersensitive cell death in Arabidopsis. *Cell* **137**: 773–783.

Hohn T. 2013. Plant pararetroviruses: interactions of cauliflower mosaic virus with plants and insects. *Current Opinion in Virology* **3**: 629–638.

Hohn T, Rothnie H. 2013. Plant pararetroviruses: replication and expression. *Current Opinion in Virology* **3**: 621–628.

Hull R. 2014. *Plant virology*, 5th edn. San Diego, CA, USA: Academic Press.

Katagiri F, Thilmony R, He SY. 2002. The *Arabidopsis thaliana–Pseudomonas syringae* interaction. *Arabidopsis Book* **1**: e0039.

Kim YC, Guan KL. 2015. mTOR: a pharmacologic target for autophagy regulation. *Journal of Clinical Investigation* **125**: 25–32.

Kim YC, Kim SY, Choi D, Ryu CM, Park JM. 2008. Molecular characterization of a pepper C2 domain-containing SRC2 protein implicated in resistance against host and non-host pathogens and abiotic stresses. *Planta* **227**: 1169–1179.

Kobayashi K, Hohn T. 2003. Dissection of cauliflower mosaic virus transactivator/viroplasmin reveals distinct essential functions in basic virus replication. *Journal of Virology* **77**: 8577–8583.

Kobayashi K, Hohn T. 2004. The avirulence domain of Cauliflower mosaic virus transactivator/viroplasmin is a determinant of viral virulence in susceptible hosts. *Molecular Plant-Microbe Interactions* **17**: 475–483.

Kobayashi K, Tsuge S, Stavolone L, Hohn T. 2002. The cauliflower mosaic virus virion-associated protein is dispensable for viral replication in single cells. *Journal of Virology* **76**: 9457–9464.

Kunkel BN, Bent AF, Dahlbeck D, Innes RW, Staskawicz BJ. 1993. RPS2, an Arabidopsis disease resistance locus specifying recognition of *Pseudomonas syringae* strains expressing the avirulence gene avrRpt2. *Plant Cell* **5**: 865–875.

Laird J, McInally C, Carr C, Doddiah S, Yates G, Chrysanthou E, Khattab A, Love AJ, Geri C, Sadanandom A *et al.* 2013. Identification of the domains of cauliflower mosaic virus protein P6 responsible for suppression of RNA silencing and salicylic acid signalling. *Journal of General Virology* **94**: 2777–2789.

Lenz HD, Haller E, Melzer E, Kober K, Wurster K, Stahl M, Bassham DC, Vierstra RD, Parker JE, Bautor J *et al.* 2011. Autophagy differentially controls plant basal immunity to biotrophic and necrotrophic pathogens. *Plant Journal* **66**: 818–830.

Levine B, Mizushima N, Virgin HW. 2011. Autophagy in immunity and inflammation. *Nature* **469**: 323–335.

Li Y, Lu J, Han Y, Fan X, Ding SW. 2013. RNA interference functions as an antiviral immunity mechanism in mammals. *Science* **342**: 231–234.

Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B. 2012. MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences, USA* **109**: 1790–1795.

Liang C, Oh BH, Jung JU. 2015. Novel functions of viral anti-apoptotic factors. *Nature Reviews Microbiology* **13**: 7–12.

Liu YM, Bassham DC. 2010. TOR is a negative regulator of autophagy in *Arabidopsis thaliana*. *PLoS ONE* **5**: e11883.

Liu Y, Schiff M, Czymmek K, Tallóczy Z, Levine B, Dinesh-Kumar SP. 2005. Autophagy regulates programmed cell death during the plant innate immune response. *Cell* **121**: 567–577.

Love AJ, Geri C, Laird J, Carr C, Yun BW, Loake GJ, Tada Y, Sadanandom A, Milner JJ. 2012. Cauliflower mosaic virus protein P6 inhibits signaling responses to salicylic acid and regulates innate immunity. *PLoS ONE* **7**: e47535.

Love AJ, Laird J, Holt J, Hamilton AJ, Sadanandom A, Milner JJ. 2007a. Cauliflower mosaic virus protein P6 is a suppressor of RNA silencing. *Journal of General Virology* **88**: 3439–3444.

Love AJ, Laval V, Geri C, Laird J, Tomos AD, Hooks MA, Milner JJ. 2007b. Components of Arabidopsis defense- and ethylene-signaling pathways regulate susceptibility to Cauliflower mosaic virus by restricting long-distance movement. *Molecular Plant–Microbe Interactions* **20**: 659–670.

Love AJ, Yun BW, Laval V, Loake GJ, Milner JJ. 2005. Cauliflower mosaic virus, a compatible pathogen of Arabidopsis, engages three distinct defense-signaling pathways and activates rapid systemic generation of reactive oxygen species. *Plant Physiology* **139**: 935–948.

Ma XM, Blenis J. 2009. Molecular mechanisms of mTOR-mediated translational control. *Nature Reviews Molecular Cell Biology* **10**: 307–318.

Mersmann S, Bourdais G, Rietz S, Robatzek S. 2010. Ethylene signaling regulates accumulation of the FLS2 receptor and is required for the oxidative burst contributing to plant immunity. *Plant Physiology* **154**: 391–400.

Navarro L, Jay F, Nomura K, He SY, Voinnet O. 2008. Suppression of the microRNA pathway by bacterial effector proteins. *Science* **321**: 964–967.

Nawrath C, Metraux JP. 1999. Salicylic acid induction-deficient mutants of Arabidopsis express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell* **11**: 1393–1404.

Palanichelvam K, Schoelz JE. 2002. A comparative analysis of the avirulence and translational transactivator functions of gene VI of Cauliflower mosaic virus. *Virology* **293**: 225–233.

Pooggin MM. 2013. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *International Journal of Molecular Sciences* **14**: 15233–15259.

Pooggin MM, Hohn T, Futterer J. 1998. Forced evolution reveals the importance of short open reading frame A and secondary structure in the cauliflower mosaic virus 35S RNA leader. *Journal of Virology* **72**: 4157–4169.

Prince DC, Drurey C, Zipfel C, Hogenhout SA. 2014. The leucine-rich repeat receptor-like kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the cytochrome P450 PHYTOALEXIN DEFICIENT3 contribute to innate immunity to aphids in Arabidopsis. *Plant Physiology* **164**: 2207–2219.

Qiao Y, Liu L, Xiong Q, Flores C, Wong J, Shi J, Wang X, Liu X, Xiang Q, Jiang S *et al.* 2013. Oomycete pathogens encode RNA silencing suppressors. *Nature Genetics* **45**: 330–333.

Rajeswaran R, Pooggin MM. 2012. RDR6-mediated synthesis of complementary RNA is terminated by miRNA stably bound to template RNA. *Nucleic Acids Research* **40**: 594–599.

Rodriguez A, Angel CA, Lutz L, Leisner SM, Nelson RS, Schoelz JE. 2014. Association of the P6 protein of Cauliflower mosaic virus with plasmodesmata and plasmodesmal proteins. *Plant Physiology* **166**: 1345–1358.

Roine E, Wei W, Yuan J, Nurmiaho-Lassila EL, Kalkkinen N, Romantschuk M, He SY. 1997. Hrp pilus: an hrp-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. tomato DC3000. *Proceedings of the National Academy of Sciences, USA* **94**: 3459–3464.

Ryabova LA, Pooggin MM, Hohn T. 2006. Translation reinitiation and leaky scanning in plant viruses. *Virus Research* **119**: 52–62.

Schepetilnikov M, Dimitrova M, Mancera-Martínez E, Geldreich A, Keller M, Ryabova LA. 2013. TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3 h. *EMBO Journal* **32**: 1087–1102.

Schepetilnikov M, Kobayashi K, Geldreich A, Caranta C, Robaglia C, Keller M, Ryabova LA. 2011. Viral factor TAV recruits TOR/S6K1 signalling to activate reinitiation after long ORF translation. *EMBO Journal* **30**: 1343–1356.

Schoelz J, Shepherd RJ, Daubert S. 1986. Region VI of cauliflower mosaic virus encodes a host range determinant. *Molecular and Cellular Biology* **6**: 2632–2637.

Scholthof KB, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P *et al.* 2011. Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology* **12**: 938–954.

Schulze B, Mentzel T, Jehle AK, Mueller K, Beeler S, Boller T, Felix G, Chinchilla D. 2010. Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. *Journal of Biological Chemistry* **285**: 9444–9451.

Schwessinger B, Roux M, Kadota Y, Ntoukakis V, Sklenar J, Jones A, Zipfel C. 2011. Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genetics* **7**: e1002046.

Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC. 2012. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* **24**: 859–874.

Shivaprasad PV, Rajeswaran R, Blevins T, Schoelz J, Meins F Jr, Hohn T, Pooggin MM. 2008. The CaMV transactivator/viroplasmin interferes with RDR6-dependent trans-acting and secondary siRNA pathways in Arabidopsis. *Nucleic Acids Research* **36**: 5896–5909.

Thompson AR, Doelling JH, Suttangkakul A, Vierstra RD. 2005. Autophagic nutrient recycling in Arabidopsis directed by the ATG8 and ATG12 conjugation pathways. *Plant Physiology* **138**: 2097–2110.

Torres MA, Onouchi H, Hamada S, Machida C, Hammond-Kosack KE, Jones JDG. 1998. Six *Arabidopsis thaliana* homologues of the human respiratory burst oxidase (gp91(phox)). *Plant Journal* **14**: 365–370.

Tsuda K, Katagiri F. 2010. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology* **13**: 459–465.

Vlot AC, Dempsey DA, Klessig DF. 2009. Salicylic Acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology* **47**: 177–206.

Wang XB, Jovel J, Udomporn P, Wang Y, Wu Q, Li WX, Gasciolli V, Vaucheret H, Ding SW. 2011. The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in *Arabidopsis thaliana*. *Plant Cell* **23**: 1625–1638.

Yoshimoto K, Jikumaru Y, Kamiya Y, Kusano M, Consonni C, Panstruga R, Ohsumi Y, Shirasu K. 2009. Autophagy negatively regulates cell death by controlling NPR1-dependent salicylic acid signaling during senescence and the innate immune response in Arabidopsis. *Plant Cell* **21**: 2914–2927.

Yu W, Murfett J, Schoelz JE. 2003. Differential induction of symptoms in Arabidopsis by P6 of Cauliflower mosaic virus. *Molecular Plant–Microbe Interactions* **16**: 35–42.

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA *et al.* 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development* **25**: 2540–2553.

Zhou J, Yu JQ, Chen Z. 2014. The perplexing role of autophagy in plant innate immune responses. *Molecular Plant Pathology* **15**: 637–645.

Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JD, Boller T, Felix G. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* **125**: 749–760.

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JD, Felix G, Boller T. 2004. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **428**: 764–767.

Zvereva AS, Pooggin MM. 2012. Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* **4**: 2578–2597.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1 Bacterial growth on transgenic *Arabidopsis* lines expressing different variants/mutants of CaMV P6 protein.

Table S1 DNA oligonucleotide probes for RNA blot hybridization and PCR primers

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.