



**HAL**  
open science

## Distribution and evolution of the major viruses infecting cucurbitaceous and solanaceous crops in the French Mediterranean area

Cecile Desbiez, Catherine Wipf-Scheibel, Pauline Millot, Karine Berthier, Gregory Girardot, Patrick Gognalons, Judith Hirsch, Benoît Moury, Karine Nozeran, Sylvain Piry, et al.

### ► To cite this version:

Cecile Desbiez, Catherine Wipf-Scheibel, Pauline Millot, Karine Berthier, Gregory Girardot, et al.. Distribution and evolution of the major viruses infecting cucurbitaceous and solanaceous crops in the French Mediterranean area. *Virus Research*, 2020, 286, pp.198042. 10.1016/j.virusres.2020.198042 . hal-02878547

HAL Id: hal-02878547

<https://hal.inrae.fr/hal-02878547v1>

Submitted on 30 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Desbiez, C., Wipf-Scheibel, C., Millot, P., Berthier, K., Girardot, G., Gognalons, P., Hirsch, J., Moury, B., Nozeran, K., Piry, S., Schoeny, A., Verdin, E. (2020). Distribution and evolution of the major viruses infecting cucurbitaceous and solanaceous crops in the French Mediterranean area. *Virus Research*, 286, pp.198042. DOI : 10.1016/j.virusres.2020.198042

Licence CC BY-NC-ND



Cécile Desbiez<sup>1</sup>, Catherine Wipf-Scheibel<sup>1</sup>, Pauline Millot<sup>1</sup>, Karine Berthier<sup>1</sup>, Grégory Girardot<sup>1</sup>, Patrick Gognalons<sup>1</sup>, Judith Hirsch<sup>1</sup>, Benoît Moury<sup>1</sup>, Karine Nozeran<sup>1</sup>, Sylvain Piry<sup>1,2</sup>, Alexandra Schoeny<sup>1</sup>, Eric Verdin<sup>1</sup>

<sup>1</sup>INRAE, Pathologie Végétale, F-84140, Montfavet, France

<sup>2</sup>CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

Corresponding author: C. Desbiez (cecile.desbiez@inrae.fr)

## Abstract

Plant viral diseases represent a significant burden to plant health, and their highest impact in Mediterranean agriculture is on vegetables grown under intensive horticultural practices. In order to understand better virus evolution and emergence, the most prevalent viruses were mapped in the main cucurbitaceous (melon, squashes) and solanaceous (tomato, pepper) crops and in some wild hosts in the French Mediterranean area, and virus diversity, evolution and population structure were studied through molecular epidemiology approaches. Surveys were performed in summer 2016 and 2017, representing a total of 1530 crop samples and 280 weed samples. The plant samples were analysed using serological and molecular approaches, including high-throughput sequencing (HTS). The viral species and their frequency in crops were quite similar to those of surveys conducted ten years before in the same areas. Contrary to other Mediterranean countries, aphid-transmitted viruses remain the most prevalent in France whereas whitefly-transmitted ones have not yet emerged. However, NGS analysis of viral evolution revealed the appearance of undescribed viral variants, especially for watermelon mosaic virus (WMV) in cucurbits, or variants not present in France before, as for cucumber mosaic virus (CMV) in solanaceous crops. Deep sequencing also revealed complex virus populations within individual plants with frequent recombination or reassortment. The spatial genetic structure of cucurbit aphid-borne yellows virus (CABYV) was related to the

landscape structure, whereas in the case of WMV, the recurrence of introduction events and probable human exchanges of plant material resulted in complex spatial pattern of genetic variation.

Keywords: epidemiology; evolution; melon; pepper; tomato; potyvirus; recombination

Highlights (3-5 bullets, 85 characters each, spaces included)

We characterized viruses infecting cucurbitaceous and solanaceous crops in France

Aphid- and thrips-transmitted viruses remain the most prevalent viruses so far

New WMV and CMV strains have been introduced in France in the last decade

Recombination and reassortment are frequent

The frequency of introductions is a major driver of population genetic structure

## 1. Introduction.

Viral diseases are particularly important threats in vegetable crops since they affect crop yield, but also the visual and organoleptic quality of the products. Since no curative methods are available, virus control relies mostly on genetic resistance when available, on prophylactic measures limiting virus introductions and spread and on sanitation (removal of infected hosts). Virus introductions can take place through commercial exchanges of seeds, plants or fruits (Jones, 2009; Lecoq et al., 2003), and/or spread by their biotic vectors, frequently insects. For annual crops, the presence of reservoir hosts constitutes a key component for local viral maintenance when no cultivated hosts are present. To limit epidemics and prevent the emergence of new viruses or new particularly damaging strains, it is important to survey the prevalence and diversity of viruses in reservoirs, and to estimate the genetic exchanges between distant viral populations which can constitute key drivers of viral evolution and emergence.

In Southern France, large-scale surveys (over 2200 samples) were performed from 2004 to 2008 (Desbiez et al., 2009; supplementary material S1) in cucurbits -mostly melon, squashes and cucumber that constitute the main crops in the country (<https://www.franceagrimer.fr/>)- and to a much lesser extent (296 samples in 2005-2008) on tomato, pepper and eggplant (supplementary material S1), with samplings performed based on visual symptoms. The most

common viruses on cucurbits at that time were aphid-transmitted: watermelon mosaic virus (WMV, 79% of the samples), cucurbit-aphid borne yellows virus (CABYV, 40% of the samples), cucumber mosaic virus (CMV, 14%) and zucchini yellow mosaic virus (ZYMV, 13%) (Supplementary material S1), with frequent mixed infections particularly between WMV and CABYV (35% of the samples). Molecular studies revealed a highly dynamic situation for WMV where new strains observed in South-eastern France since 2000 had replaced within a decade the local strains present for at least 30 years (Desbiez et al., 2009), even though their dispersal was limited in the few years following their introduction (Joannon et al., 2010; Piry et al., 2016). In tomato and pepper, the most common viruses were the thrips-transmitted tospovirus tomato spotted wilt virus (TSWV) and the aphid-transmitted CMV and potato virus Y (PVY) (supplementary material S1). Parietaria mottle virus (PMoV), transmitted by pollen, was also frequent in greenhouse tomato. Whitefly-transmitted viruses were found only punctually (Lecoq et al., 2007; <https://gd.eppo.int/taxon/TYLCV0/distribution/FR>).

Whitefly-transmitted viruses are currently emerging in several European and Mediterranean countries, particularly begomoviruses (TYLCV, tomato leaf curl New Delhi virus (ToLCNDV)), and criniviruses (cucurbit yellow stunting disorder virus (CYSDV), cucurbit chlorotic yellows virus (CCYV), tomato infectious chlorosis virus (TICV) and tomato chlorosis virus (ToCV)) (Lecoq and Desbiez, 2012; Lefeuvre et al., 2010; Moriones et al., 2017). Contact- and seed-transmitted viruses from the genera *Tobamovirus* (cucumber green mottle mosaic virus (CGMMV) in cucumber, tomato brown rugose fruit virus (ToBRFV) in tomato) (Dombrovsky et al., 2017; Oladokun et al., 2019) and *Potexvirus* (pepino mosaic virus (PepMV) in tomato) (Hanssen and Thomma, 2010) are also emerging worldwide, particularly in greenhouse conditions, in relation to anthropic exchanges of seeds or material. New strains of “classic” viruses have also been introduced in the last decades: several introductions of cucumber mosaic virus (CMV) group IB, from Asian origin, have taken place in the Mediterranean Basin (Jacquemond, 2012). Recombinant strains of potato virus Y (PVY-NTN) possessing genome parts phylogenetically related to the N and O clades and inducing potato tuber necrosis have spread throughout the world (Glais et al., 2002; Quenouille et al., 2013). These new viruses or strains can represent increased agronomic threats by inducing particularly damaging symptoms on some crops (Desbiez et al., 2009; Glais et al., 2002). They can also contribute to reduce the efficiency or durability of control methods including cross-protection (Hanssen et al., 2010) and genetic resistance (Ben Tamarzizt et al., 2013; Oladokun et al., 2019). Since viral

populations can undergo rapid changes in some areas, it is important to carry out periodic surveys in order to identify the genetic diversity and turnover rates of viruses in agronomically important regions. This allows an updated and accurate estimate of risks for agriculture and the deployment of appropriate control methods.

Ten years after the 2004-2008 surveys, the presence in France of the viruses that are emerging in neighbouring countries, as well as the evolution of populations for local viruses, was not known. In this work, we performed a second run of large-scale surveys of virus prevalence and molecular diversity in cucurbit and solanaceous crops and neighbouring weeds in Southern France, the first area in France for production of these crops. We also applied recently-developed landscape epidemiology approaches to study the spread and intraspecific exchanges of viral populations. This study aims to contribute to a better understanding of the origins of epidemics, the evolution of viral populations and the importance of genetic exchanges, in order to adapt control measures (cultural practices, control of vector populations, use of resistant cultivars) to the current viral threats.

## 2. Materials and methods

### 2.1. Crop and weed samplings

Crop samplings were performed in early July in 2016 and 2017, corresponding to the beginning to middle of the main growing season for cucurbits, tomato, pepper and eggplant in southern France. Sampled cucurbits were melon, different squashes (mostly zucchini and pumpkin), and cucumber, representing the major cucurbit species grown in France. In 2017, surveys were also performed in September, at the end of the growing season. In 2016, samplings were performed in all the Mediterranean arc, from the Spanish border in the west to the Italian border in the east (Figure 1). In 2017, samplings were focused on the three “départements” (counties) Gard, Vaucluse and Bouches-du-Rhône that constitute a major vegetable-growing area in southern France (Figure 1). Six to 12 plots were sampled per département in 2016 (7 départements), and 12 to 46 plots per département in 2017 (3 départements), for a total number of 63 plots in 2016 and 90 plots in 2017. Ten samples, each corresponding to a different plant, were collected in each plot. Samples showing virus-like symptoms of mosaics or yellowing with leaf thickening were preferentially collected; in some plots where virus prevalence appeared very low, particularly in July, plants showing disorders that could be non-viral (necrosis, yellowing) were also collected. This sampling strategy focused on the most

agronomically important viruses –i.e. the ones causing the most obvious symptoms. The presence of insects or mites was noted, as well as the type of crop management (protected or not, conventional, integrated or organic), the cultivar, and the prevalence and distribution of symptoms in the field. Each sample was photographed using a Garmin® Montana GPS recording its geographic coordinates, with a datamatrix® indicating the plant species and a unique code for each sample. The data were imported directly in a dedicated relational database “Virobase” (Wipf-Scheibel et al., 2019).

Weed samples were collected from 24 plots of cucurbitaceous or solanaceous crops in July 2016. Seven species, known as potential reservoirs of crop-infecting viruses, were selected: *Solanum nigrum* (black nightshade), *Senecio vulgaris* (common groundsel), *Capsella bursa-pastoris* (shepherd’s purse), *Datura stramonium* (thorn-apple), *Lamium amplexicaule* (common henbit), *Parietaria officinalis* (wall pellitory), *Chenopodium* sp. (goosefoot). Five weed plants from the same species in one plot were pooled to constitute one sample. Weed samples were collected at random, none of them showing obvious symptoms except a few plants of *Chenopodium* sp. that displayed chloro-necrotic lesions. The samples were geolocalized and their data were imported in “Virobase”.

## 2.2. Serological and molecular diagnostic

All samples were ground in 3 ml cold phosphate solution  $\text{Na}_2\text{HPO}_4$  0.03 M + 0.2% diethyldithiocarbamate (DIECA) in Bioreba® extraction bags to eliminate cross-contaminations. An aliquot of 200  $\mu\text{l}$  was immediately frozen and stored at  $-20^\circ\text{C}$  for RNA and DNA extraction.

The extracts were tested in DAS-ELISA with antisera produced at INRAE Avignon except for the PepMV antiserum (Sediag, France), using a standard protocol used in the laboratory. ELISA was chosen as a cheap and robust method to screen a large number of samples against several viruses at once. Cucurbit samples were tested serologically for the presence of major cucurbit-infecting viruses: CABYV, CMV, WMV and ZYMV that were the most common viruses on melon, squash and cucumber in France in 2004-2008, Moroccan watermelon mosaic virus (MWMV) and papaya ringspot virus (PRSV and MWMV) that used to be very rare in France (supplementary material S1) but are frequent or emerging in several European and Mediterranean countries (Lecoq et al., 2012), melon necrotic spot virus (MNSV) and squash mosaic virus (SqMV) that are seed-transmitted and, as such, present important risks of large-

scale dissemination (Herrera-Vasquez et al., 2009). Since there were no symptoms suggestive of the presence of ToLCNDV, the virus was not tested serologically in cucurbit samples.

Tomato, pepper and eggplant samples were tested with antisera against alfalfa mosaic virus (AMV), CABYV (in order to detect the putative presence of the cross-reacting pepper vein yellows virus (PeVYV)), CMV, eggplant mottled dwarf virus (EMDV), PepMV, pepper mild mottle virus (PMMoV), parietaria mottle virus (PMoV), PVY, tomato mosaic virus (ToMV) and TSWV. The ToMV antiserum cross-reacts with several tobamoviruses, and it would probably cross-react with ToBRFV but this could not be tested due to the lack of positive control. TSWV, CMV and PVY, as well as PMoV in tomato, were the most common viruses in previous surveys. Pepper vein yellows virus is currently emerging in several countries in the Mediterranean Basin (Lotos et al., 2017; Fiallo-Olive et al., 2018) but has not been described in France. The other viruses were found occasionally in previous surveys; their prevalence was low 10 years ago but their current status was not known.

In 2016, the presence of the whitefly-transmitted ToCV and TYLCV was tested in tomato samples by reverse-transcription PCR (RT-PCR) and PCR respectively using specific primers (Supplementary table S2), since antisera against these viruses are either lacking or were not efficient. Cucumber samples were also tested for the presence of beet pseudo-yellows virus (BPYV) and CYSDV by RT-PCR with specific primers (Supplementary table S2).

Since the diversity of virus species present in weeds was not easily predictable, and serological or specific molecular detection in weeds is sometimes inefficient probably due to low virus concentration, irregular distribution in the tissues and/or presence of inhibitors (Lacroix et al., 2016), weeds were tested for the presence of viruses by molecular approaches without *a priori* (deep sequencing).

### 2.3. Nucleic acid extraction and molecular analyses

Total RNAs were extracted from crop and weed samples ground in phosphate solution. Two hundred microliters of frozen sample were extracted with TRI-reagent (Molecular Research Center Inc., Cincinnati, OH) for crop samples (Desbiez et al., 2009). For weeds that often contain more sequencing inhibitors than crop samples, RNA was extracted with PureLink Plant RNA reagent (Thermo Fisher Scientific, Waltham, NA), according to the manufacturers' instructions. Extracted RNAs were resuspended in 20 µl nuclease-free distilled water

(Invitrogen, France) and stored at -20°C until further use. Total DNAs were extracted from tomato samples ground in phosphate solution as described by Gilbertson et al. (1991).

Total RNAs of 7 to 10 weed samples were pooled by species and sent for sequencing to Fasteris SA, Switzerland. Small RNAs (21-24 nt) were purified from gel before Illumina HiSeq sequencing. The sequences were assembled *de novo* with CLC Genomics Workbench 7 (CLC Bio, Aarhus, Denmark) as described by Verdin et al., 2017. The reads were also aligned with full-length reference sequences of CMV, WMV, PVY, TSWV, CABYV, ToMV and PMoV. Blast analyses were performed on the assembled fragments. When a virus was detected in a pool of samples, the different components of the pool were tested separately by RT-PCR with specific primers (Supplementary table S2). This sequencing strategy allows detection without *a priori* knowledge of the viruses present in weeds.

Total RNA of one WMV-infected melon sample (EM160093) corresponding to Profile 8 (see 3.3.2) was also sent to Fasteris for small RNA sequencing in order to obtain a complete genome sequence (GenBank accession MN914160).

#### 2.4. Amplification and sequencing of viruses in cucurbits, tomato and pepper

In order to design primers against WMV, CMV, CABYV, PVY and TSWV, complete sequences representative of the known molecular diversity of the different viruses were retrieved from GenBank and aligned with ClustalW included in MEGA6 (Tamura et al., 2013). Primers were designed to amplify regions of circa 400 nt in each of the 3 genomic RNAs for CMV (in the helicase (1a), RdRp (2a) and CP genes for RNAs 1, 2 and 3 respectively) and TSWV (in the RdRp, NSm and NSs genes for RNAs L, M and S respectively), and in two genomic areas for the other viruses (Supplementary table S2). For CABYV, the coat protein (CP) and polymerase (RdRp) were targeted since a large number of partial sequences from worldwide isolates were available for these regions. For the two potyviruses WMV and PVY, primers were designed to encompass the variable N-terminal part of the CP-coding region commonly targeted for diversity analyses in potyviruses. For WMV, another primer pair was designed in the CI-coding region that constitute a recombination hotspot for this virus (Desbiez and Lecoq, 2008), whereas for PVY, the VPg-coding region was targeted as involved in host adaptation and resistance-breaking (Ayme et al., 2006). The primers were tested in two-step RT-PCR against a set of 7 isolates and a negative control for each virus. After validation, oligonucleotides with



the same sequence but with an overhang adapter sequence at their 5' extremity were synthesized (supplementary table S2).

After reverse transcription, a two-step PCR strategy was performed on each sample, combined with a single-index (in 2016) or dual-index (in 2017) paired-end sequencing approach (Galan et al., 2018; Martin 2019). RT-PCR were also performed on different negative controls (Piry et al., 2017) (Supplementary material S3). In the first PCR, two independent replicates were performed for each cDNA, to constitute technical replicates and to reduce the risks of false positives (Galan et al., 2018). A second round of PCR (eight cycles) was performed on the products from the first PCR to add multiplexing indices and Illumina sequencing adapters P5 and P7 (supplementary material S3). The tagged samples were pooled, quantified and submitted to pair-end 2x250 (in 2016) or 2x300 (in 2017) Illumina MiSeq sequencing (supplementary material S3).

The MiSeq sequences from amplicons of cucurbit, tomato and pepper-infecting viruses were sorted and analysed with the dada2 package (<http://www.bioconductor.org/packages/release/bioc/html/dada2.html>) (Callahan et al., 2016) in the R software (<https://www.R-project.org>). After demultiplexing and removing the primer sequences, dada2 was used to merge the paired-end reads 1 and 2, with a minimal overlap of 30 nt in 2016 and 70 nt in 2017.

The numbers of different variant sequences (i.e. haplotypes) for each sample, and number of reads for each haplotype, were retrieved for each replicate. Variants present in only one replicate, or presenting less than 30 reads, were removed.

## 2.5. Phylogenetic analyses

Nucleotide sequences of the different haplotypes were aligned with ClustalW and MUSCLE included in MEGA6 (Tamura et al., 2013). Reference sequences from the same viruses were added to the alignments and phylogenetic analyses. Recombination detection analysis was performed with Rdp4 (Martin et al., 2015) and with GARD ([www.datamonkey.org/gard](http://www.datamonkey.org/gard)) on the different fragments separately. For samples that contained only one haplotype for the 2 or 3 fragments sequenced for each virus, concatenated sequences of the different fragments were also analysed for recombination or pseudorecombination detection. For non-recombinant sequences, the best model of multiple substitutions was selected using MODELTEST included in MEGA6. Distance and maximum-likelihood trees were built with

MEGA6 for the aligned sequences using the substitution model previously selected with MODELTEST, with 500 bootstrap replicates.

Molecular groups were defined based on sequence similarity between haplotypes and structure of maximum-likelihood trees. “Profiles” were defined as the combination of groups for the different genetic fragments of each virus. Sequences of one singly-infected isolate corresponding to the major haplotypes for each profile were deposited in GenBank (Supplementary table S4)

## 2.6. Spatial genetic structure of virus populations

Because WMV and CABYV presented the highest prevalence in the survey, displayed some molecular diversity (see 3.3.2 and 3.3.3), and represented contrasted aphid transmission properties (non-persistent vs. persistent for WMV and CABYV, respectively), they were chosen for analysis of spatial genetic structure. Genetic structure of virus populations and its relationship to the landscape structure were analysed for the WMV-CP and CABYV-CP datasets, excluding data from the Pyrénées-Orientales (low number of samples, high geographic distance with the core of the dataset). We used the MAPI (Mapping Averaged Pairwise Information) method (Piry et al., 2016) implemented in the R package “mapi” and random forest (RF) algorithm (Breinan 2001) implemented in the R package “ranger” (Wright and Ziegler, 2017). MAPI is a smoothing procedure for spatial genetic networks for which the connections between georeferenced samples are materialized by ellipses. These ellipses are used to average between-sample genetic distances within the cells of a grid overlaying the study area (in the same way as kernel density can be used to smooth punctual measures). The method provides geographical maps representing the spatial variation in the average level of genetic differentiation between samples (for more details, see Supplementary material S5A). For WMV, the different introduction events that took place since the 1990s (see 3.3.2) are likely to have a strong impact on its spatial genetic structure and to blur the effects of natural (i.e. aphid-mediated) gene flow process. In an attempt to circumvent this difficulty, we performed an additional MAPI analysis considering only the genetic distances computed between samples belonging to the same molecular group (Supplementary material S4 and S5A). Computing the genetic distances independently within each molecular group and then gathering the information from all groups allows keeping enough information to investigate the spatial structure while limiting strong local effects due to the geographic proximity of

highly differentiated variants that do not directly result from natural demographic (e.g. dispersal, genetic drift) and selection processes.

To characterize the landscape structure, we used the level 2 of the European Corine Land Cover typology (supplementary table S5B). The MAPI maps produced for the CABYV and WMV datasets were overlaid on the Corine Land Cover layer and the frequency of each landcover class was computed within each cell of the MAPI grids. These frequencies were used as predictor variables in RF analyses to explain within-cell average genetic distances as computed by MAPI. For each RF analysis, 10,000 trees were built and importance (increase in the mean squared error) and significances were computed for each predictor variable using the permutation (N=1000) approach of Altmann et al. (2010). The shape of the relationship between the average genetic distance and each land cover variable was determined using partial dependence plots computed with the R package “edarf” (Jones and Linder, 2016).

### 3. Results

#### 3.1. Virus prevalence and distribution in crops

Among the 630 crop samples collected in 2016, 63% were diagnosed as infected by at least one of the viruses tested. The proportion of infected plants was similar for cucurbit and solanaceous crops. In 2017, 60% of the 900 plants were infected by at least one virus. The infection rate was higher in cucurbitaceae than in solanaceae (76% vs. 33%) and was also higher in September than in July (Tables 1 and 2). In cucurbits, WMV and CABYV were the most prevalent viruses followed by CMV and ZYMV (Table 1). Mixed infections of WMV and CABYV were common (18 % and 32 % of the total cucurbit samples in 2016 and 2017 respectively). A few plants were infected with up to 4 viruses. On tomato and pepper, the most frequent viruses were PVY, CMV and TSWV (Table 2). Mixed infections of PVY with TSWV or CMV were observed in tomato and pepper. Few eggplants were sampled. Most of them were completely asymptomatic, and a few showed typical symptoms of EMDV, confirmed in ELISA and RT-PCR (Table 2). Other viruses were detected in few fields or greenhouses: MNSV in melon; AMV, EMDV, ToMV, PMoV and PepMV in tomato or pepper. The whitefly-transmitted viruses ToCV and BPYV were found only in greenhouse conditions in a few locations (Tables 1 and 2). TYLCV, TICV, PMMoV and peVYV for solanaceae, as well as CYSDV, MWMV and SqMV for cucurbitaceae, were not detected in any sample.

### 3.2 Viruses in weeds

Sequencing the nine pools of weed samples by HiSeq on small RNAs yielded 46 to 85 million reads (1-50 nt) per pool, more than 97% of which had a quality score above 30. The percentage of reads in the 18-26 nt range, containing the small RNAs, was between 21% and 62% in the different pools, except in the *Datura* pool where the RNA was of poor quality and the 18-26 nt reads represented only 1.1% of the total reads (data not shown). Upon quality control, assembly with CLC and blast analysis, cucurbitaceae- or solanaceae-infecting viruses were found in all pools: TSWV (four pools), CMV (two pools), PVY (two pools), WMV (two pools), TMoV and PMoV (one pool each). Despite the high frequency of CABYV in cucurbits and the fact that several of the tested weeds are hosts of this virus, it was not found in any of the weed pools analysed. The samples from each positive pool were tested separately by RT-PCR for the different viruses, confirming the presence of TSWV, PVY and CMV in 4 samples each, mostly in *S. nigrum* but also in *D. stramonium* for TSWV and *C. bursa-pastoris* for CMV. WMV was detected in three samples (two of *S. vulgaris* and one of *C. bursa-pastoris*), PMoV in one sample of *P. officinalis* and ToMV in one sample of *S. nigrum* (Table 3). Viruses were present in weeds in all sampled areas. The viruses found in the weeds were in some instances also found in the field or greenhouse where the weeds were sampled (Table 3). When the information was available, the virus molecular groups (see below) were also the same in the weeds and neighbouring crops (data not shown). Viruses not infecting the sampled crops but that may infect other crops were also found: turnip mosaic virus detected in shepherd's purse, sowbane mosaic virus in goosefoot (Table 3) but none of the viruses currently emerging in the Mediterranean Basin was detected.

### 3.3. Molecular diversity of viruses in crops

In 2016, 2240 amplicons were pooled and sequenced in one MiSeq run. The amplicons corresponded to two genomic regions for WMV (148 isolates), CABYV (72 isolates) and PVY (53 isolates) and three regions for CMV (72 isolates) and TSWV (74 isolates), with several negative controls. In 2017, the 2 x 2112 amplicons were sequenced in two MiSeq runs (Supplementary figure S3), corresponding to two replicates of amplicons in the same genomic regions as in 2016 for WMV (272 isolates), CABYV (277 isolates) and CMV (172 isolates). The different controls indicated very low contaminations (data not shown). Eighteen to 145 haplotypes –i.e. unique sequences- were obtained for the different virus fragments. Infection

of one plant with several haplotypes of the same virus was frequently observed (data not shown).

### 3.3.1. Cucumber mosaic virus

In 2016, partial sequences of the 3 RNAs were obtained for 67 of the 72 samples tested. Five samples for which at least one of the three amplified fragments did not yield good quality sequences were discarded. Among the three main CMV groups known worldwide –namely groups IA, IB and II (Jacquemond 2012), isolates from group II were observed in two tomato samples in Pyrénées-Orientales, and in 7 tomato samples in Alpes-Maritimes (Supplementary table S4 and data not shown). All other isolates belonged to group IA, with a very low intragroup diversity (Figure 2; Supplementary tables S4). No reassortant or recombinant was observed between groups IA and II. In 2017, among the 172 samples sequenced, the main subgroup remained subgroup IA (Figure 2 for the RNA3 partial sequence). One isolate of group II was found on pepper. Surprisingly, subgroup IB was also detected for the first time in commercial crops in France in two areas about 20 km apart. In one of them, 10 tomato and 3 pepper samples from the same farm were infected with CMV-IB haplotypes closely related to isolates Tfn from Italy and TN9 from Taiwan (Figure 2). The same result was obtained for the 3 fragments corresponding to RNAs 1, 2 and 3. In the other case, the RNA3 fragment from a tomato isolate shared more than 99% identity with IB isolates from Turkey and Iran (Figure 2). RNAs 1 and 2 clustered with group IA, indicating that the RNAs of this isolate must have undergone reassortment (Supplementary table S4).

The CMV isolates detected in weeds by HiSeq sequencing of small RNAs all belonged in their three RNAs to subgroup IA, also found in the nearby crops.

### 3.3.2. Watermelon mosaic virus (WMV)

WMV sequences (CI- and NIb-CP coding regions) were obtained for 148 isolates in 2016 and 272 isolates in 2017. The molecular diversity of WMV was high, and 11 different profiles (i.e. combinations of CI and CP groups) were observed in the dataset (Figure 3 and Supplementary table S4). Profiles 1 to 4 corresponding to the four “emerging” subgroups EM1 to EM4 observed in Southeastern France since 2000-2003 (Desbiez et al., 2009; Joannon et al., 2010) were present in 1 to 154 isolates from 2016 and 2017 (Figure 3 and Supplementary table S4). The “classic” (CL or G1) group present in France at least between 1972 and 2008 (Desbiez et

al., 2009) was not observed anymore, except for one 2016 isolate with a G1/G2 recombinant profile (profile 11) already observed in 2006 in the same area (Alpes-Maritimes) (Desbiez and Lecoq, 2008). Profile 5 corresponding to an apparent CI-CP recombinant between EM isolates, already present in 2008 (Desbiez, unpublished) was found in about 35% of the isolates collected in 2016 and 2017 (Figure 3 and Supplementary table S4). Besides these molecular groups already present in France 10 years ago, three and four new groups for CI and CP respectively were also observed in 2016-2017 (profiles 6 to 10; Figure 3 and supplementary table S4). Some of them shared high similarities with isolates from Asia, America or Europe (data not shown). Apparent discrepancies between the CI and CP clustering suggested recombination event(s) (figure 3), confirmed by RDP4 for profiles 5 and 9 ( $p < 10^{-4}$ ).

Multiple infections of plants with different molecular groups were observed in 2016 in only two samples, whereas in 2017, they concerned 73 samples out of 272 (26.8%). Their frequency was significantly higher in September (59/116) than in July 2017 (14/156) (Chi-square test,  $p < 0.0001$ ). Up to eight CP and CI haplotypes belonging to four different groups were detected in some samples in September 2017. Eighteen samples from five plots, presenting mixed infections with four molecular groups, also displayed recombinant sequences in the CI region (supplementary figure S6). The recombinants represented 0.5% to 12% of the WMV-CI reads per sample. They were always detected at similar frequencies in the two replicates, indicating that they were not PCR artefacts even if RT or sequence mis-pairing artefacts cannot be completely ruled out. Up to three different recombinants in the P3-CI region were present together in the same plant, representing a total of 2% to 9% of the reads.

WMV sequences detected in two samples of *Senecio vulgaris* corresponded to profile 5, also present in the cucurbit crops where the weeds were sampled.

### 3.3.3. Cucurbit aphid-borne yellows virus (CABYV)

Partial sequences of CABYV RdRp and CP were obtained for 72 samples in 2016 and 276 samples in 2017. The variability in the CP fragment (maximum sequence divergence: 3.1%) was much lower than in the RdRp region (up to 10% divergence). Five molecular clusters were defined in the RdRp region and two in the CP region, but with a low bootstrap support (data not shown). One major profile (reference isolate EM160196, supplementary table S4) was highly predominant and was present in 75% and 96% of the samples in 2016 and 2017 respectively. Compared to the worldwide diversity of CABYV, all isolates belonged to the

“European-African” clade. They were highly similar to isolates observed in the same area between 2011 and 2015 (Schoeny et al., 2017) and related to isolates from Spain, Egypt and Tunisia (Kassem et al., 2013; Mnari-Hattab et al., 2009). Some isolates clustered differently in the RdRp and CP regions, but the evidence for recombination was not confirmed due to inconsistent phylogenetic signal.

#### 3.3.4. Potato virus Y

In 2016, among the 53 PVY isolates detected by ELISA in pepper and tomato, 49 and 52 sequences were obtained for the VPg and CP coding regions respectively. In 2017, Sanger sequences for the VPg and CP fragments were obtained for 20 isolates from the 5 tomato and pepper plots where PVY was detected that year (Supplementary table S4). PVY isolates clustered in two major clades, without discrepancies between the VPg and CP sequences. Most isolates belonged to the “PVY-C1” clade (Moury et al., 2017), frequent in pepper and tomato in Europe and the Mediterranean Basin (Moury and Verdin, 2012). The other isolates (6 in 2016 and 4 in 2017) clustered in the “N” clade for the VPg-coding region, and were recombinants between the “N” and “O” clades in the CP-coding region. They shared the same recombination breakpoint in the CP as recombinant isolates from the “NTN” type that induce tuber necrosis on potato and can infect other solanaceous hosts (Moury et al., 2017). The NTN isolates were very homogeneous molecularly, whereas C1 isolates clustered in one major group closely related to isolate CAA141 (JQ954317, collected in France in 1999) and several minor variants (supplementary table S4 and Supplementary Figure S7). In the VPg coding fragment, several non-synonymous mutations were localized in the region involved in the interaction with the eukaryotic initiation factor eIF4E and in resistance breakdown (Ayme et al., 2006) (data not shown). However these mutations did not correspond to those previously identified as involved in the breaking of particular resistance alleles (Ayme et al., 2007; Ayme et al., 2006).

#### 3.3.5. Tomato spotted wilt virus

Among the 74 TSWV isolates collected in 2016, 63 and 65 Illumina sequences were obtained for RNAs S and M respectively. Fragment L yielded low amplification for a large number of samples, and Illumina sequences were obtained for only 36 isolates, originating from 12 of the 16 plots where TSWV was detected.

Compared to the global molecular diversity of TSWV described in Tentchev et al. (2011), TSWV isolates from this study fell into two clusters for RNA-L (clusters 3 and 4 according to Tentchev et al. 2011, two clusters for RNA-M (clusters a and b) and three clusters for RNA-S (clusters  $\alpha$ ,  $\delta$  and  $\epsilon$ ) (supplementary table S4). This corresponds to half of the clusters defined worldwide. No new molecular group was detected compared to the work of Tentchev et al., 2011. However, whereas the most common association of the three fragments in France was previously 3-c- $\gamma$  (Tentchev et al., 2011), neither cluster c for RNA-M nor cluster  $\gamma$  for RNA-S were observed in our work. Instead, five different combinations were present, 3-a- $\delta$  being the most common. In 2017, the prevalence of TSWV was low. Among the 18 TSWV samples collected from eight infected fields and characterized by Sanger sequencing, three of the five combinations observed in 2016 were present (supplementary table S4).

#### 3.3.6. Other viruses

The other viruses observed in the survey presented a low prevalence and their molecular diversity was tested only by Sanger sequencing (GenBank accessions MN990975-MN990985). On cucurbits, ZYMV was present in 25 and 6 samples from four and two plots in 2016 and 2017, respectively. Partial CP sequences were obtained and compared to the worldwide diversity of the virus. The 24 isolates sequenced belonged to molecular groups already observed in France in previous surveys, namely group "A1" (1 isolate) present in southeastern France since 1979 (Lecoq et al., 2014), and groups A4 (22 isolates) and A5 (one isolate) that have been introduced in the last 15 years, probably from Asian origin (Lecoq and Desbiez, 2012; Lecoq et al., 2009; Lecoq et al., 2014). Group A4 was the most prevalent in the sampled areas, but the number of plots positive for ZYMV was too low to draw general conclusions. PRSV was detected in only two samples in 2017. Like PRSV isolates previously observed during the rare infections of this virus in France, these two isolates belonged to the "American-European" clade.

PepMV was detected in 12 tomato samples from three plots. Based on sequence comparisons, the isolates detected in France belonged to the mild "Ch2" type already observed in several European countries since 2005 (Hanssen and Lapidot, 2012) and there was no evidence for introduction of severe isolates (Hanssen and Thomma, 2010).

#### 3.4. Spatial genetic structure of CABYV and WMV



Results of the MAPI analysis on the CABYV-CP dataset uncovered a strong spatial structure, with the eastern part of the study area showing a very high level of genetic discontinuity while the central/western part exhibited the lowest level of differentiation (Figure 4A). RF analysis showed that six land cover classes had a significant effect on CABYV spatial genetic structure ( $p$ -value  $< 0.001$ ). “Permanent crops” and “Forest” were the two most important classes (Supplementary figure S8A). “Permanent crops” had a strong negative effect on the average level of genetic differentiation, i.e. favored genetic flow, whereas “Forest” had a strong positive effect, indicating an important barrier to gene flow (Supplementary figure S8B). Among the variables of lower importance, the classes “Heterogeneous agricultural areas” and “Arable land” had almost linear negative and positive relationships with the level of genetic differentiation, respectively.

As could be expected due to the multiple introduction events, the spatial genetic structure retrieved from the WMV dataset was very complex (Supplementary figure S8C). Detection of areas of genetic continuity, reflecting high natural gene flow, is largely impeded by the geographic proximity of introduced strains belonging to different molecular groups with up to 10% molecular divergence (see Supplementary Figure S8C and Figure 3). Therefore, the observed structure could not be related to the landscape structure (RF analysis gave inconsistent results and did not identify any significant variable). When using intra-group genetic distances, MAPI retrieved a more homogeneous spatial structure, with a maximum level of genetic differentiation of 1% (Figure 4B). The method also identified two large significant areas of high genetic continuity, though an area of high genetic discontinuity was still identified in the center of the study area (Figure 4B). The RF analysis identified three land cover variables, with different levels of significance, to explain the variation in the average level of genetic differentiation computed by MAPI (Figure S8D): “Permanent crops” ( $p$ -value  $< 0.0001$ ), “Forest” ( $p$ -value = 0.05) and “Urban fabric” ( $p$ -value  $< 0.001$ ). The shape of the interactions suggested a negative effect of the “Permanent crops” and “Urban fabric” classes while “Forest” mostly had a positive effect on the level of genetic differentiation (Figure S8E).

#### 4. Discussion

In order to characterize viruses infecting cucurbit and solanaceous crops in Southern France, extensive surveys were performed in 2016 and 2017. Surveys in crops and weeds at the beginning/middle of the growing season targeted mostly the primary inocula in reservoirs and

crops, while the survey at the end of the growing season, when virus prevalence and diversity is highest, aimed at studying the global viral diversity in the area and the potential presence of new viruses currently emerging in Europe and the Mediterranean Basin. The most prevalent viruses (WMV, CABYV, CMV and ZYMV in cucurbits; PVY, CMV and TSWV on tomato and pepper) were the same as in surveys conducted ten years ago in the same area (Supplementary material S1), showing that contrary to other Mediterranean areas, the emergence of whitefly-transmitted viruses has not yet taken place in France. This constitutes an important information for virus management, indicating what insect vectors should be particularly targeted by control measures and what genetic resistances should be used in the field. Nearly 40% of the sampled plants were negative for all tested viruses. This can be related in part to limits of the sampling design targeting ten symptomatic plants per plot. In September, almost all sampled plants presented clear virus-like symptoms whereas in July, virus prevalence in some plots appeared very low and plants with ambiguous symptoms that could be related to abiotic stresses or mite proliferation had also been collected. Also, viral symptoms in cucurbits are often unambiguous (mosaic or yellowing with leaf thickening) whereas tomato and pepper often display necrosis that can be related to non-viral causes. No plants with severe begomovirus-like or other atypical symptoms had been observed, and there was no major discrepancy between the symptoms observed and the viruses detected (data not shown). Virus prevalence in weeds was low, but infections by the major crop-infecting viruses, except CABYV, were detected. In several instances, the same viruses –and based on molecular analyses the same strains- were found in weeds and neighbouring crops, suggesting exchanges between crops and weeds. This confirms the importance of careful weeding of fields, greenhouses and edges before and during the growing season. Indeed, host richness –related to weed nature and frequency- in crops and edges can be an important driver of virus prevalence and infection (McLeish et al., 2017). None of the viruses currently emerging in the Mediterranean Basin, begomoviruses and ToBRFV in particular (Moriones et al., 2017; Oladokun et al., 2019) was detected in weeds by non-targeted sequencing of small RNAs, nor observed in the sampled crops. Tobamovirus prevalence appeared low. ToMV was observed in tomato cultivars that did not have the *Tm-2<sup>2</sup>* resistance gene. This confirms the high efficiency of genetic control against “classic” tobamoviruses and the need to use prophylactic measures, including efficient seed control and disinfection, to prevent the introduction in France of ToBRFV that is not controlled by *Tm-2<sup>2</sup>* (Oladokun et al., 2019). The

non-emergence of begomoviruses and criniviruses could be related to low populations of their vector. Viruses transmitted only by *Bemisia tabaci* like TYLCV and CYSDV were not detected, whereas the criniviruses BPYV and ToCV that can be transmitted by the greenhouse whitefly *Trialeurodes vaporariorum* were found in some greenhouses. Insect trapping with yellow water pans performed in fields and in the surroundings of greenhouses in parallel with the plant samplings in July 2016 and 2017 yielded aphids and thrips but no whiteflies (data not shown), even if whiteflies were observed in a few greenhouses in September 2017. Aphid-transmitted viruses currently emerging in southern Europe and the Mediterranean Basin, including the potyvirus MWMV on cucurbits (Lecoq and Desbiez, 2012), and poleroviruses on pepper (Fiallo-Olive et al., 2018; Lotos et al., 2017) were not found in this survey even if their vectors are present in France. They have either not been introduced yet, or have not become established because they did not overwinter in weeds, like MWMV that was observed once in France in 2003 but was not detected again (Lecoq and Desbiez, 2012). Indeed, the growing conditions in southern France, usually presenting an important winter break, must contribute to limit the maintenance of viruses with a narrow host range. Climate change or modifications in agricultural practices could favour their emergence in the future.

Even if no new viruses were observed, analysis of the molecular diversity for the most common viruses revealed in three cases a highly dynamic situation. For WMV, besides the four molecular groups (EM1 to EM4) that have emerged in France in the last decade, new groups were present at a high prevalence, and the “classic” group (G1) observed in southeastern France at least between 1972 and 2013 (Lecoq et al., 2014) was not detected at all. Experimental and modelling approaches based on data from 2000-2008 suggested that G1 isolates had a slightly lower fitness than EM ones (Fabre et al., 2010; Lecoq et al., 2011), and that the EM/G1 ratio should reach an equilibrium (Fabre et al., 2010). This is obviously not the case, maybe due to the new introductions that took place since 2008. Thanks to high-throughput sequencing, complex populations with multiple variants belonging to different molecular groups could be detected at the intra-plant level particularly in September 2017. Evidence for recombination was observed, both between the CI and CP regions and possibly within the P3-6K1-CI region that constitutes a recombination hotspot for WMV (Desbiez et al., 2011; Desbiez and Lecoq, 2008). Even if no increased agronomic impact has been associated so far with the recombinant status of WMV isolates (Desbiez et al., 2011; Desbiez and Lecoq,

2008), intraspecific recombinants are widespread in WMV as well as in other potyviruses (Green et al., 2018; Hajizadeh et al., 2019), and they could be agronomically important. Recombinants are usually detected when they become a major component of a virus population, whereas in this study recombinants were present at low levels in mixed intraplant populations including the parental sequences, suggesting that the recombination took place in the mixed-infected plant itself and that the recombinants did not outcompete their parents. Recombination was also observed in PVY, but in that case it was a single type of recombinant in the CP region corresponding to PVY-NTN strains (Moury et al., 2017). The PVY populations observed in this study, namely PVY-C1 and PVY-NTN, were known in France for more than 10 years even if NTN strains were long considered to be more adapted to potato than to tomato and pepper (Moury et al., 2017), so there is no evidence for a recent change in PVY populations.

For the two tripartite viruses studied here, i.e. CMV and TSWV, sequence comparisons suggested recent introductions as well as reassortment between the different components. CMV-IB was observed for the first time in commercial crops in France, in tomato and pepper but not in cucurbits. Based on sequence comparisons, two different events were observed, one of them associated with IA/IB reassortment. Even if this has not yet been observed in France, CMV-IB can be associated with resistance-breaking in pepper and its introduction can thus be agronomically important. Reassortants between the 3 main CMV groups are common in some countries, and can be related to host adaptation (Ben Tamarzizt et al., 2013). Phylogenetic studies suggested that most current CMV reassortants and recombinants are recent and have emerged in the last 50 years, maybe in relation to the intensification of agriculture and genetic exchanges (Ohshima et al., 2016).

For TSWV, the molecular groups observed were among those described by Tentchev et al. 2011 for the three RNA fragments. Clade 4 for RNA-L that had not been detected in France before although it was present in Spain (Tentchev et al., 2011) was observed in this work, suggesting a recent introduction. Several clades highly prevalent in Spain, Italy or the USA were not observed here, showing that long-distance exchanges have not been responsible for the emergence of these clades in France in the last decade. As in the work of Tentchev et al 2011, reassortants appeared common in this study, but the combinations of the different clades for the three RNAs were different from those observed in France before 2011. This shows that TSWV in Southeastern France undergoes extensive reassortment, and that the

situation is highly dynamic. Whether some reassortants present a higher fitness in some environmental conditions is not known. Reassortment could contribute to a rapid adaptation to host resistance, some determinants for resistance breaking in tomato and pepper being related to NSs and NSm proteins encoded by RNAs S and M respectively (Zhu et al., 2019)

Spatial structure analyses yielded very different results for CABYV and WMV. For CABYV, two types of land covers including vegetable crops (“Permanent crops” and “Heterogeneous agricultural areas” – table S4B) corresponded to the highest level of gene flow. The spatial organisation of these two types of agricultural land use provide evident corridors for virus dispersal in the center and western parts of the study area (Figure 1). In contrast, large areas mostly covered by forests were associated to high levels of genetic differentiation. This situation mostly corresponds to the eastern part of the study area (figures 1 and 4). These results suggest that CABYV transmission at the landscape scale mainly occurs through successive short/middle distance aphid flights, the absence of host plant patches over long distance being a limiting factor to virus spread even under a persistent mode of transmission. Due to the multiple introduction events that create local spots of high diversity, the spatial genetic structure of WMV was very complex to decipher and to relate to natural demographic processes. When accounting for phylogenetic information (division in molecular profiles), results were partly consistent with those on CABYV: gene flow would be enhanced by the presence of permanent crops and limited by the presence of large forested areas. These results are, however, to be taken with caution as the range of intra-group genetic differentiation is very low (<1%) and detected effects rely on very small variations in the genetic pattern. The complexity of WMV genetic structure, with multiple molecular groups including newly-introduced ones, suggests that human-assisted plant material exchanges are likely the major cause for WMV population structure differentiation whereas natural aphid-mediated spread plays only a minor role in medium- or long-distance transmission. For instance, the same molecular group (profile 5) was present in the center and eastern part of the study area, which are separated by landscape elements identified as impeding dispersal for CABYV. The contrast between CABYV and WMV could be related to a hypothetical seed transmission in WMV (never observed so far), and/or to the fact that some prophylactic methods, including insecticide sprays in nurseries before plantlets are transported to different

areas, can be efficient against persistently transmitted viruses like CABYV but not against non-persistently transmitted ones including WMV.

This study confirms that, even if no new viruses have been observed in this study, virus populations can evolve quickly in relation to intrinsic mechanisms (mutations and recombination/reassortment) but also to long-distance introductions probably mediated by human activity. Prophylactic and cultural practices including control and limitation of plant material exchanges remain key factors to reduce virus epidemics and emergence risks.

### **Acknowledgements**

This work was funded through the EMERAMB project in the ARIMNet2 2015 Call. ARIMNet2 (2014-2017) is an ERA-NET project coordinated by INRA (France). It has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 618127.

We thank people from the 20 extension services (Technical Institutes, Chambres d'Agriculture) who helped us for the surveys, and particularly Anne Terrentroy (Chambre d'Agriculture des Bouches-du-Rhône) and Catherine Taussig (APREL, Association Provençale de Recherche et d'Expérimentation Légumière). We particularly thank the 80 growers involved in this survey.

The authors declare that they have no conflict of interest

### **References**

- Altmann, A., Tolosi, L., Sander, O. and Lengauer, T. (2010) Permutation importance: a corrected feature importance measure. *Bioinformatics* 26(10), 1340-1347.
- Ayme, V., Petit-Pierre, J., Souche, S., Palloix, A. and Moury, B. (2007) Molecular dissection of the potato virus Y VPg virulence factor reveals complex adaptations to the pvr2 resistance allelic series in pepper. *J. Gen. Virol.* 88, 1594-1601.
- Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadoeuf, J., Palloix, A. and Moury, B. (2006) Different Mutations in the Genome-Linked Protein VPg of Potato virus Y Confer Virulence on the pvr2<sup>3</sup> Resistance in Pepper. *Mol. Plant-Microbe Interact.* 19, 557-563.

- Ben Tamarzizt, H., Montarry, J., Girardot, G., Fakhfakh, H., Tepfer, M. and Jacquemond, M. (2013) Cucumber mosaic virus populations in Tunisia peppers are mainly composed of virus reassortants with resistance-breaking properties. *Plant Pathol.* 62, 1415-1428.
- Breinan, L. (2001) Random forests. *Machine Learning* 45, 5–32.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Meth.* 13, 581.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581.
- Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C. and Lecoq, H. (2009) Emergence of new strains of *Watermelon mosaic virus* in South-Eastern France: evidence for limited spread but rapid local population shift. *Virus Res.* 141, 201-208.
- Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C. and Lecoq, H. (2011) Recombination in natural populations of watermelon mosaic virus: new agronomic threat or damp squib? *J. Gen. Virol.* 92, 1939-1948.
- Desbiez, C. and Lecoq, H. (2008) Evidence for multiple intraspecific recombinants in natural populations of Watermelon mosaic virus (WMV). *Arch. Virol.* 153, 1749-1754.
- Dombrovsky, A., Tran-Nguyen, L.T.T. and Jones, R.A.C. (2017) Cucumber green mottle mosaic virus: Rapidly Increasing Global Distribution, Etiology, Epidemiology, and Management. *Ann. Rev. Phytopathol.* 55, 231-256.
- Fabre, F., Chadœuf, J., Costa, C., Lecoq, H. and Desbiez, C. (2010) Asymmetrical over-infection as a process of plant virus emergence. *J. Theor. Biol.* 265, 377-388.
- Fiallo-Olive, E., Navas-Hermosilla, E., Ferro, C.G., Zerbini, F.M. and Navas-Castillo, J. (2018) Evidence for a complex of emergent poleroviruses affecting pepper worldwide. *Arch. Virol.* 163, 1171-1178.
- Galan, M., Pons, J.-B., Tournayre, O., Pierre, E., Leuchtman, M., Pontier, D. and Charbonnel, N. (2018) Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Mol. Ecol.* 18, 474–489.
- Gilbertson, R.L., Rojas, M.R., Russell, D.R. and Maxwell, D.P. (1991) Use of the asymmetric polymerase chain reaction and DNA sequencing to determine genetic variability of bean golden mosaic geminivirus in the Dominican Republic. *J. Gen. Virol.* 72, 2843-2848.
- Glais, L., Tribodet, M. and Kerlan, C. (2002) Genetic variability in *Potato potyvirus Y* (PVY): evidence that PVY<sup>NW</sup> and PVY<sup>NTN</sup> variants are single to multiple recombinants between PVY<sup>O</sup> and PVY<sup>N</sup> isolates. *Arch. Virol.* 147, 363-378.

- Green, K.J., Brown, C.J. and Karasev, A.V. (2018) Genetic diversity of potato virus Y (PVY): sequence analyses reveal ten novel PVY recombinant structures. *Arch. Virol.* 163, 23-32.
- Hajizadeh, M., Gibbs, A.J., Amirnia, F. and Glasa, M. (2019) The global phylogeny of Plum pox virus is emerging. *J. Gen. Virol.* 100, 1457-1468.
- Hanssen, I.M., Gutierrez-Aguirre, I., Paeleman, A., Goen, K., Wittemans, L., Lievens, B., Vanachter, A., Ravnika, M. and Thomma, B. (2010) Cross-protection or enhanced symptom display in greenhouse tomato co-infected with different Pepino mosaic virus isolates. *Plant Pathol.* 59, 13-21.
- Hanssen, I.M. and Lapidot, M. (2012) Major tomato viruses in the Mediterranean Basin. *Adv. Virus Res.* 84, 31-66.
- Hanssen, I.M. and Thomma, B.P.H.J. (2010) Pepino mosaic virus: a successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Mol. Plant Pathol.* 11, 179-189.
- Herrera-Vasquez, J.A., Cordoba-Selles, M.C., Cebrian, M.C., Alfaro-Fernandez, A. and Jorda, C. (2009) Seed transmission of Melon necrotic spot virus and efficacy of seed-disinfection treatments. *Plant Pathol.* 58, 436-442.
- Jacquemond, M. (2012) Cucumber mosaic virus. *Adv. Virus Res.* 84, 440-504.
- Joannon, B., Lavigne, C., Lecoq, H. and Desbiez, C. (2010) Barriers to gene flow between emerging populations of Watermelon mosaic virus in south-eastern France. *Phytopathol.* 100, 1373-1379.
- Jones, R.A.C. (2009) Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* 141, 113-130.
- Jones, Z.M. and Linder, F.J. (2016) edarf: Exploratory Data Analysis using Random Forests. *The Journal of Open Source Software* 1(6), 92.
- Kassem, M.A., Juarez, M., Gómez, P., Mengual, C.M., Sempere, R.N., Plaza, M., Elena, S.F., Moreno, A., Fereres, A. and Aranda, M.A. (2013) Genetic Diversity and Potential Vectors and Reservoirs of Cucurbit aphid-borne yellows virus in Southeastern Spain. *Phytopathol.* 103, 1188-1197.
- Lacroix, C., Renner, K., Cole, E., Seabloom, E.W., Borer, E.T. and Malmstrom, C.M. (2016) Methodological Guidelines for Accurate Detection of Viruses in Wild Plant Species. *Appl. Env. Microb.* 82, 1966-1975.
- Lecoq, H. and Desbiez, C. (2012) Viruses of cucurbit crops in the mediterranean region: an ever-changing picture. *Adv. Virus Res.* 84, 67-126.



- Lecoq, H., Desbiez, C., Wipf-Scheibel, C. and Girard, M. (2003) Potential involvement of melon fruit in the long distance dissemination of cucurbit potyviruses. *Plant Dis.* 87, 955-959.
- Lecoq, H., Fabre, F., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., Schoeny, A. and Desbiez, C. (2011) Search for factors involved in the rapid shift in watermelon mosaic virus (WMV) populations in south-eastern France. *Virus Res.* 159, 115-123.
- Lecoq, H., Wipf-Scheibel, C., Chandeysson, C., Le Van, A., Fabre, F. and Desbiez, C. (2009) Molecular epidemiology of *Zucchini yellow mosaic virus* in France: an historical overview. *Virus Res.* 141, 190-200.
- Lecoq, H., Wipf-Scheibel, C., Nozeran, K., Millot, P. and Desbiez, C. (2014) Comparative molecular epidemiology provides new insights into Zucchini yellow mosaic virus occurrence in France. *Virus Res.* 186, 135-143.
- Lefevre, P., Martin, D.P., Harkins, G., Lemey, P., Gray, A.J.A., Meredith, S., Lakay, F., Monjane, A., Lett, J.-M., Varsani, A. and Heydarnejad, J. (2010) The Spread of Tomato Yellow Leaf Curl Virus from the Middle East to the World. *PLoS Pathog.* 6(10), e1001164.
- Lotos, L., Olmos, A., Orfanidou, C., Efthimiou, K., Avgelis, A., Katis, N.I. and Maliogka, V.I. (2017) Insights Into the Etiology of Polerovirus-Induced Pepper Yellows Disease. *Phytopathol.* 107(12), 1567-1576.
- Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1(1), 1-5.
- Martin, J.-F. (2019) Creating error-proof indexes for high throughput sequencing. DOI : 10.5281/zenodo.33502.
- McLeish, M., Sacristan, S., Fraile, A. and Garcia-Arenal, F. (2017) Scale dependencies and generalism in host use shape virus prevalence. *Proc. R. Soc. B-Biol. Sci.* 284(1869).
- Mnari-Hattab, M., Gauthier, N. and Zouba, A. (2009) Biological and Molecular Characterization of the Cucurbit aphid-borne yellows virus Affecting Cucurbits in Tunisia. *Plant Dis.* 93, 1065-1072.
- Moriones, E., Praveen, S. and Chakraborty, S. (2017) Tomato Leaf Curl New Delhi Virus: An Emerging Virus Complex Threatening Vegetable and Fiber Crops. *Viruses* 9(10), 264.
- Moury, B., Simon, V., Faure, C., Svanella-Dumas, F., Marais, A. and Candresse, T. (2017) Host groups to Potato virus Y: vanishing barriers. In: C. Lacomme, L. Glais, D.U. Bellsted, B. Dpui, A.V. Karasev and E. Jacquot (Eds), *Potato virus Y: biodiversity, pathogenicity, epidemiology and management*, pp. 234-261. Springer.
- Moury, B. and Verdin, E. (2012) Viruses of pepper crops in the Mediterranean Basin: a remarkable stasis. *Adv. Virus Res.* 84, 127-162.

- Ohshima, K., Matsumoto, K., Yasaka, R., Nishiyama, M., Soejima, K., Korkmaz, S., Ho, S.Y.W., Gibbs, A.J. and Takeshita, M. (2016) Temporal analysis of reassortment and molecular evolution of Cucumber mosaic virus: Extra clues from its segmented genome. *Virology* 487, 188-197.
- Oladokun, J.O., Halabi, M.H., Barua, P. and Nath, P.D. (2019) Tomato brown rugose fruit disease: current distribution, knowledge and future prospects. *Plant Pathol.* 68, 1579-1586.
- Piry, S., Chapuis, M.P., Gauffre, B., Papaix, J., Cruaud, A. and Berthier, K. (2016) Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure. *Meth. Ecol. Evol.* 7, 1463-1475.
- Piry, S., Wipf-Scheibel, C., Martin, J.-F., Galan, M. and Berthier, K. (2017) High throughput amplicon sequencing to assess within- and between-host genetic diversity in plant viruses. *bioRxiv*, 168773.
- Quenouille, J., Vassilakos, N. and Moury, B. (2013) Potato virus Y: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Mol. Plant Pathol.* 14, 439-452.
- Schoeny, A., Desbiez, C., Millot, P., Wipf-Scheibel, C., Nozeran, K., Gognalons, P., Lecoq, H. and Boissot, N. (2017) Impact of Vat resistance in melon on viral epidemics and genetic structure of virus populations. *Virus Res.* 241, 105-115.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725-2729.
- Tentchev, D., Verdin, E., Marchal, C., Jacquet, M., Aguilar, J.M. and Moury, B. (2011) Evolution and structure of tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. *J. Gen. Virol.* 92, 961-973.
- Verdin, E., Wipf-Scheibel, C., Gognalons, P., Aller, F., Jacquemond, M. and Tepfer, M. (2017) Sequencing viral siRNAs to identify previously undescribed viruses and viroids in a panel of ornamental plant samples structured as a matrix of pools. *Virus Research* 241, 19-28.
- Wipf-Scheibel, C., Millot, P., Berthier, K., and Piry, S. (2019). ViroBase: An ad hoc relational database for the management and traceability of biological data. In: 17èmes Rencontres de Virologie Végétale : RVV 2019 (p. 65). Aussois, France.
- Wright, M.N. and Ziegler, A. (2017) ranger: A Fast Implementation of Random Forests for High Dimensional Data in C plus plus and R. *J. Stat. Software* 77(1), 1-17.
- Zhu, M., van Grinsven, I.L., Kormelink, R. and Tao, X.R. (2019) Paving the Way to Tospovirus Infection: Multilined Interplays with Plant Innate Immunity. *Ann. Rev. Phytopathol.* 57, 41-62.

Table 1 : viruses detected in melon, squashes and cucumber in the 2016-2017 surveys in Southern France

	Host	Nb positive <sup>a</sup> / nb samples	BPYV	CABYV	CMV	EMDV	MNSV	PRSV	WMV	ZYMV
July 2016	Cucumber	21/60	17	4	0	nt	0	0	0	0
	Squash	81/100	nt	35	3	nt	0	0	77	10
	Melon	104/170	nt	34	47	nt	4	0	72	15
	Total 2016	206/330	17	73	50	nt	4	0	149	25
July 2017	Cucumber	11/20	nt	2	5	5	0	0	0	0
	Squash	113/120	nt	92	10	nt	0	0	89	0
	Melon	172/280	nt	93	87	nt	6	0	67	0
Sept. 2017	Cucumber	10/10	nt	7	9	0	0	0	9	0
	Squash	50/50	nt	26	6	nt	0	2	50	3
	Melon	65/70	nt	57	29	nt	0	0	57	3
Total 2017	Cucumber	21/30	nt	9	14	5	0	0	9	0
	Squash	163/170	nt	118	16	nt	0	2	139	3
	Melon	237/350	nt	150	116	nt	6	0	124	3
	Total 2017	421/550	nt	277	146	5	6	2	272	6

nt : not tested

<sup>a</sup> : plants positive for at least one virus

CYSDV, MWMV and SqMV were not detected in the survey.

Table 2 : Viruses detected in tomato, pepper and eggplant crops in the 2016-2017 surveys in Southern France

	Host	Nb positive/ nb samples	AMV	CMV	EMDV	PepMV	PMoV	PVY	ToMV	ToCV	TSWV
July 2016	Tomato	131/210	3	15	2	11	2	49	18	22	34
	Pepper	58/80	0	7	6	0	3	4	0	nt	40
	eggplant	3/10	0	0	3	0	0	0	0	nt	0
	Total 2016	192/300	3	22	11	11	5	53	18	22	74
July 2017	Tomato	65/220	19	3	14	1	0	10	11	nt	13
	Pepper	7/40	0	1	6	0	0	0	0	nt	0
	eggplant	0	0	0	0	0	0	0	0	0	0
Sept. 2017	Tomato	24/45	9	15	0	0	0	10	1	nt	5
	Pepper	19/44	0	14	1	0	0	8	0	0	0
	eggplant	1/1	0	0	1	0	0	0	0	nt	0
Total 2017	Tomato	89/265	28	18	14	1	0	20	12	nt	18
	Pepper	26/84	0	15	7	0	0	8	0	nt	0
	eggplant	1/1	0	0	1	0	0	0	0	nt	0
	Total 2017	116/350	28	33	22	1	0	28	12	nt	18

PeVYV, PMMoV, TYLCV and TICV were not detected in the survey

Table 3 : Viruses detected in weed samples and in the crops where the weeds were collected

Location	Weed plant	Viruses in weed	Crop where the weed was sampled	Viruses in crop
Gard	<i>Solanum nigrum</i>	PVY	tomato	TSWV
Hérault	<i>Solanum nigrum</i>	PVY	tomato	PepMV
Var	<i>Solanum nigrum</i>	PVY	pepper	TSWV
Vaucluse	<i>Solanum nigrum</i>	<u>CMV</u>	melon	CABYV, <u>CMV</u> , WMV, ZYMV
Vaucluse	<i>Solanum nigrum</i>	<u>CMV</u>	melon	<u>CMV</u>
Vaucluse	<i>Solanum nigrum</i>	<u>TSWV</u>	tomato	<u>TSWV</u>
Gard	<i>Solanum nigrum</i>	<u>PVY</u>	tomato	<u>PVY</u>
Gard	<i>Solanum nigrum</i>	<u>TSWV</u>	tomato	PVY, <u>CMV</u> , <u>TSWV</u>
PO	<i>Solanum nigrum</i>	<u>TSWV, ToMV</u>	tomato	<u>TSWV, ToMV</u>
AM	<i>Senecio vulgaris</i>	<u>WMV</u>	squash	<u>WMV</u> , CABYV
AM	<i>Senecio vulgaris</i>	<u>WMV</u>	melon	<u>WMV</u> , CABYV
AM	<i>Capsella bursa-pastoris</i>	<u>CMV</u>	melon	<u>CMV</u>
PO	<i>Capsella bursa-pastoris</i>	CMV	pepper	TSWV
AM	<i>Capsella bursa-pastoris</i>	TuMV	melon	CABYV, CMV
AM	<i>Chenopodium sp.</i>	SoMV	melon	WMV
AM	<i>Parietaria officinalis</i>	<u>PMoV</u>	pepper	TSWV, <u>PMoV</u>
PO	<i>Datura stramonium</i>	<u>TSWV</u>	pepper	<u>TSWV</u>

Viruses detected both in weeds and crops in one sampling site are underlined. PO = Pyrénées-Orientales ; AM = Alpes-Maritimes. SoMV : sowbane mosaic virus ; TuMV : turnip mosaic virus

## Figure legends

Figure 1: Study area for the 2016-2017 samplings. Sampling sites (fields or greenhouses) are indicated by grey circles with a level of transparency informing on the geographic proximity of several sites (circles are overlapping). The landscape layer is adapted from the level 2 of the European Corine Land Cover typology (see Supplementary Material S5B for details).

Figure 2: Maximum-likelihood tree obtained for CMV RNA 3 partial sequences. Bootstrap values (n=500 bootstrap) above 50% are indicated for each node. Sequences corresponding to isolates from the survey are boxed.

Figure 3: Maximum-likelihood tree obtained for WMV P3-CI (left) and NIb-CP (right) partial sequences, excluding the 6 recombinant haplotypes in the P3-CI region. Bootstrap values above 50% are indicated for each node. Branches with bootstrap values below 40% were collapsed. The highly divergent isolate WMV-Pg was used as an outgroup. Profile numbers (1 to 11) and the name of the reference isolate for each profile, as defined in Supplementary table S4, are indicated for each branch tip. Profiles with an asterisk were detected as recombinant with RDP4.

Figure 4: MAPI analysis for A) CABYV and B) WMV considering only the genetic distances computed within the molecular groups. For both viruses the sampling sites are represented with circles whose size is proportional to the number of samples, grey circles for CABYV and pie charts indicating the proportion of each molecular group in the sampling site for WMV. The color gradation indicates genetic distances between 0.0001 and 0.211. Hatched areas indicate areas of significant genetic discontinuity, whereas dotted areas correspond to areas of significant genetic continuity.

Supplementary material

Supplementary material S1: Viruses detected during surveys performed on vegetable crops in Southern France in 2004-2008.

Supplementary table S2 (A) Primers used for amplification before Illumina sequencing; (B) other primers used for diagnostic and/or Sanger sequencing

Supplementary material S3: Amplification and pooling strategy for MiSeq sequencing of viral amplicons from the 2016-2017 surveys.

Supplementary table S4: Molecular profiles and their frequency in the 2016-2017 survey for CMV, WMV, PVY, TSWV and CABYV

Supplementary material S5 (A): Method of geographic structure analysis performed with MAPI on the CABYV and WMV datasets; (B): Corine Land Cover nomenclature and proportion of the surface of the MAPI grids covered by the different classes.

Supplementary figure S6: Alignment of **polymorphic sites only** (23 nt out of 412 nt) in the WMV-CI fragment of reference and potentially recombinant haplotypes.

Supplementary figure S7: Maximum-likelihood tree obtained for PVY partial CP sequences. Bootstrap values above 50% (n=500 bootstraps) are indicated for each node. Reference isolates for the different molecular groups are in bold.

Supplementary material S8: (A): Random forest (RF) ranking of the importance of land cover variables to explain the variation in the average level of genetic differentiation as computed by MAPI; (B): Partial dependence plots from the RF analysis for the CABYV dataset assessing the shape of the relationships between the land cover variables and the level of genetic differentiation as computed by MAPI; (C): Result of the MAPI analysis for the WMV-CP dataset. The sampling sites are represented with pie charts indicating the proportion of each molecular profile (as defined in Figure 3 and Supplementary table S4) within the sampling sites. The color

gradation indicates genetic distances between 0.0001 and 0.211. Hatched areas indicate areas of significant genetic discontinuity, whereas dotted areas correspond to areas of significant genetic continuity.; (D): WMV (considering intra-group connections): RF ranking of the importance of land cover variables to explain the variation in the average level of genetic differentiation as computed by MAPI; (E) WMV (considering intra-group connections): partial dependence plots from the RF assessing the shape of the relationships between the land cover variables and the level of genetic differentiation as computed by MAPI. Only the three classes identified as significant are presented.

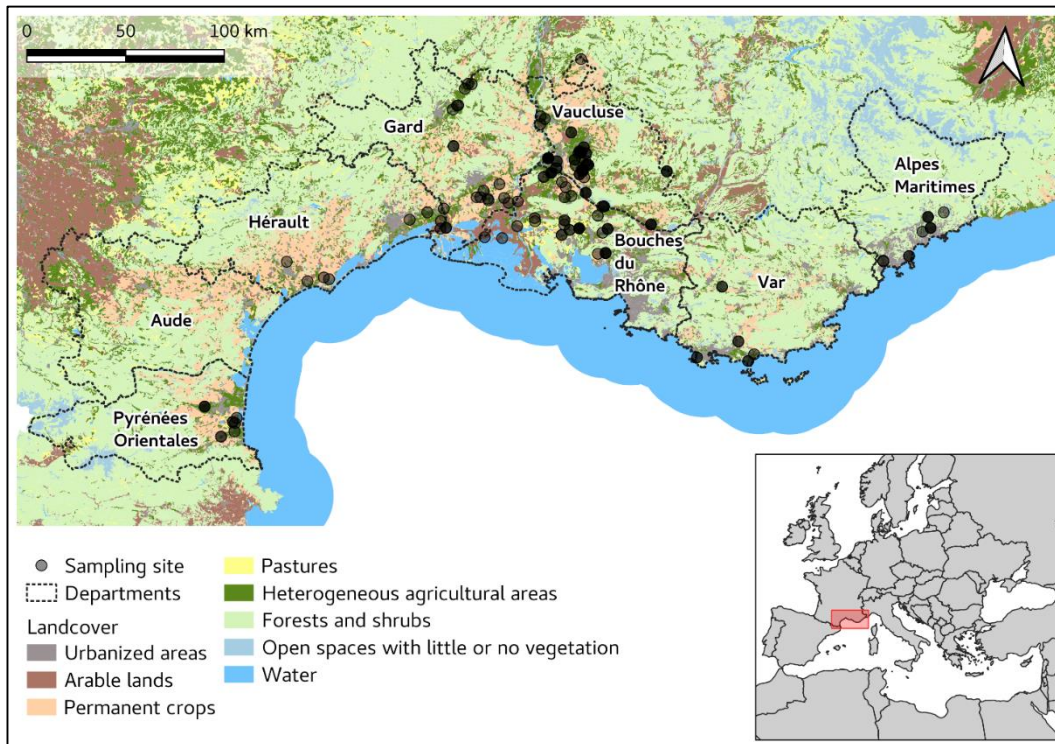


Figure 1



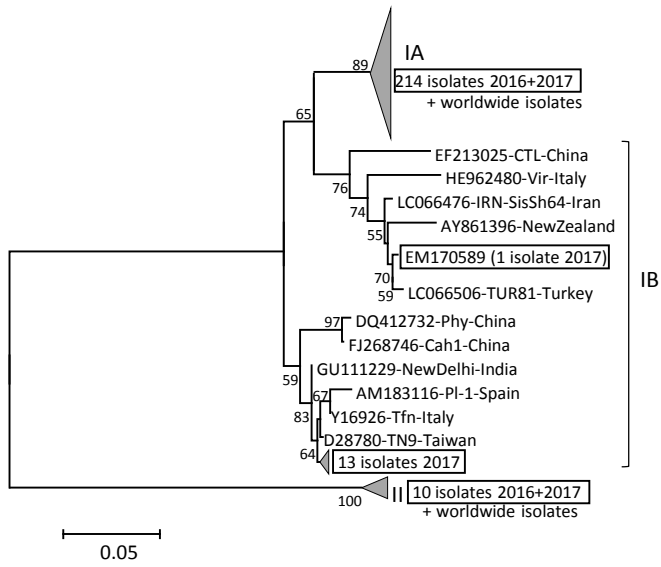


Figure 2

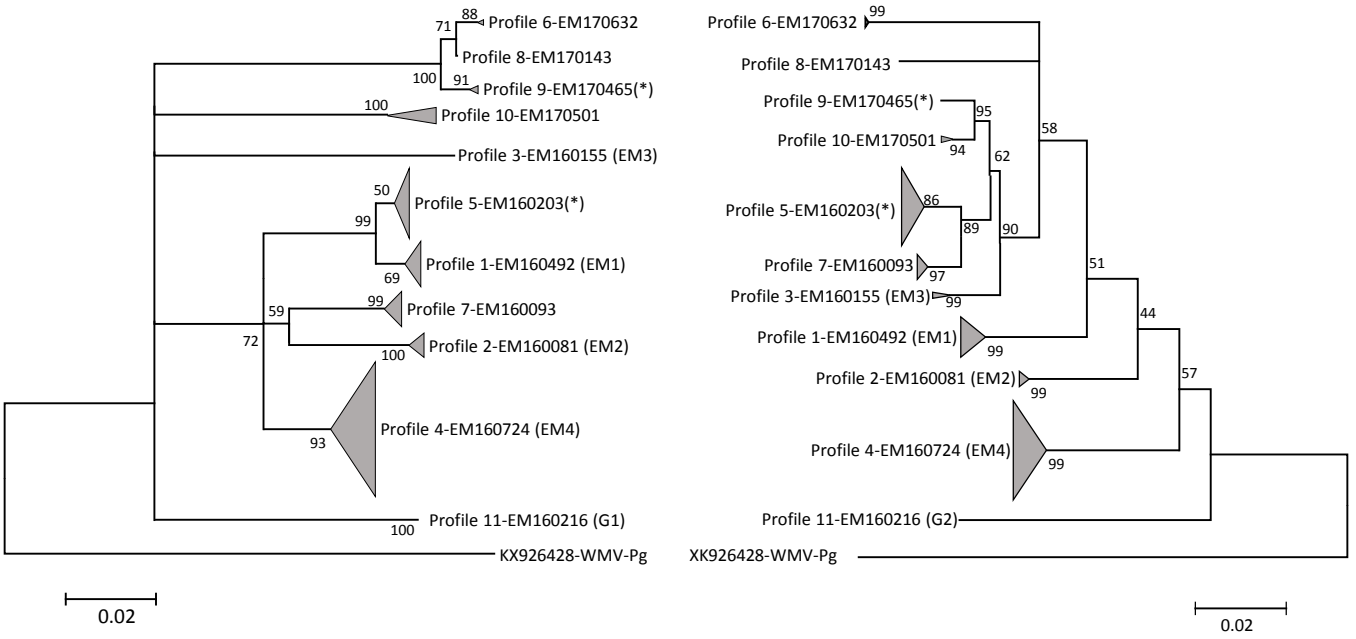


Figure 3

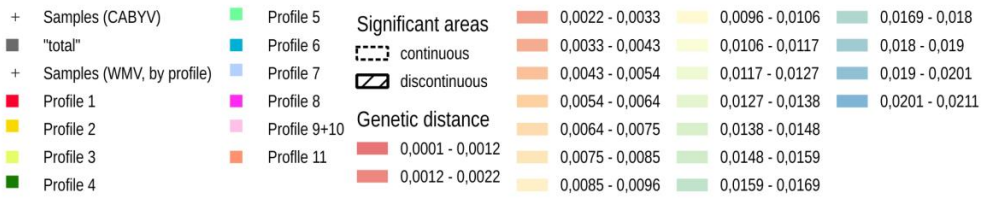
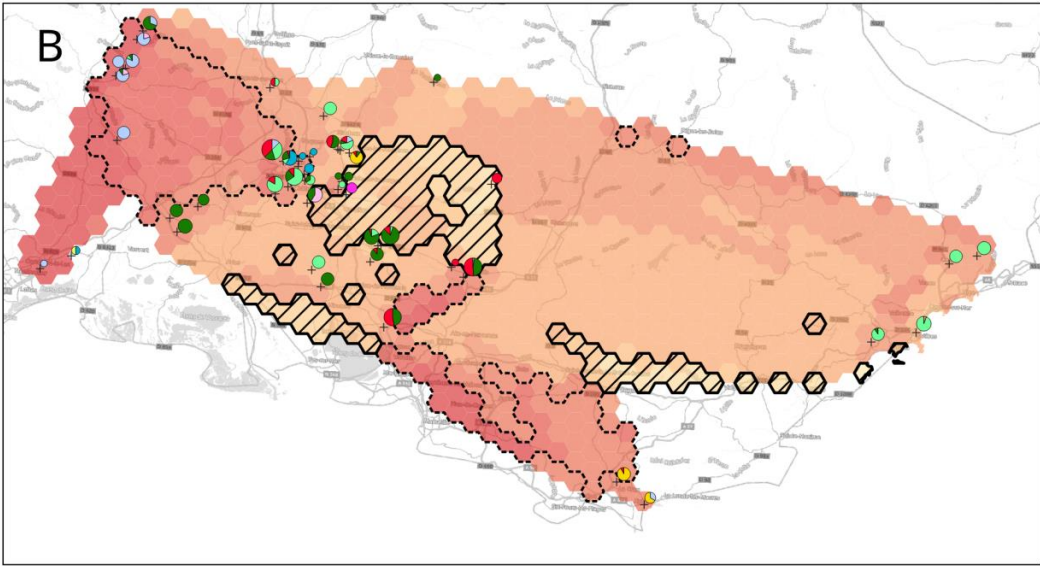
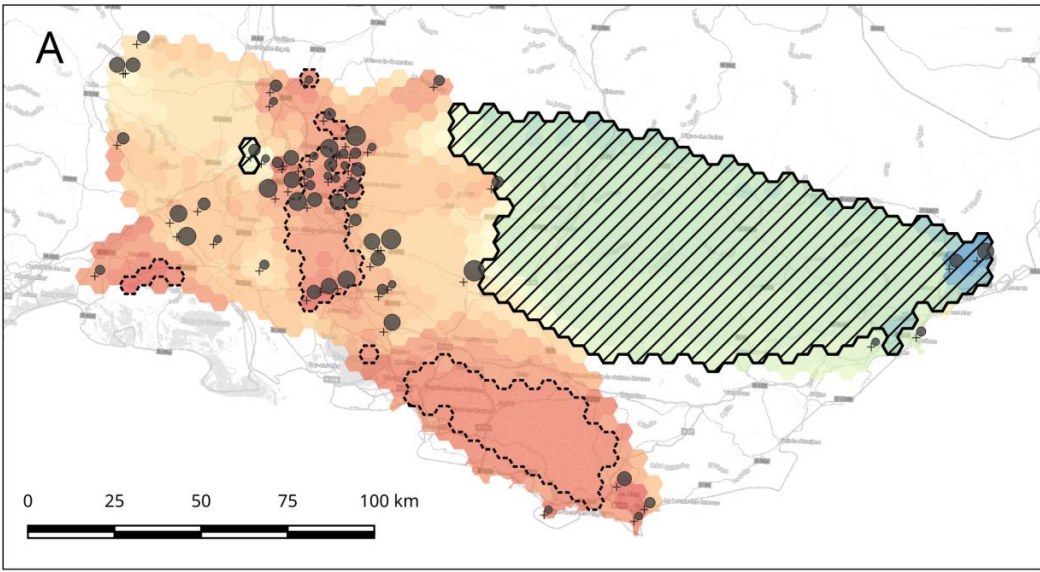


Figure 4 – Result of the MAPI analysis for A) CABYV and B) WMV considering only the genetic distances computed within the molecular groups.

Supplementary material S1: viruses detected during surveys performed on vegetable crops in Southern France in 2004-2008. The samples were sent by growers and extension services based on virus-like symptoms (mosaics, deformation, yellowings, necrosis). The samples were tested as in 2016-2017 (see materials and methods), except that no tests were performed against CYSDV, PeVYV and PMMoV.

Table S1A : viruses detected in melon, squashes and cucumber in the 2004-2008 surveys in Southern France

Host	Nb positive <sup>a</sup> / nb samples	BPYV	CABYV	CMV	EMDV	MNSV	PRSV	WMV	ZYMV
Cucumber	62/110	7	19	26	6	0	0	15	4
Squash	916/960	nt	452	36	0	0	3	870	164
Melon	1129/1153	nt	402	249	0	84	0	868	119
Total 2004-2008	2007/2223	7	873	311	6	84	3	1753	287

MWMV and SqMV were not detected in the survey.

Table S1B : Viruses detected in tomato, pepper and eggplant crops in the 2005-2008 surveys in Southern France

Host	Nb positive/ nb samples	AMV	CMV	EMDV	PepMV	PMoV	PVY	TMV + ToMV	ToCV	TSWV	TYLCV
Tomato	111/213	4	26	0	12	5	28	1	5	41	2
Pepper	39/59	0	6	1	0	0	3	0	nt	29	nt
eggplant	11/24	0	1	10	0	0	0	0	nt	0	nt
Total 2005-2008	162/296	4	33	11	12	5	31	1	5	70	2

TICV was not detected in the survey



Supplementary table S2A : primers used for amplification before Illumina sequencing

Virus	Amplified region	Primer position	Primer name	Primer sequence	Size of amplified fragment (nt)
WMV	P3-CI	3715	WMV3715F	AWTGAGGATGAGCAAGCWGT	
WMV	P3-CI	4166	WMV4166R	TGCAACTGCCTRTGCCACCA	452
WMV	Nib-CP	8689	WMV8689F	ACTGAGGCAATTTGTGCAGC	
WMV	Nib-CP	9144	WMV9144R	TATCTTYTGYAGTCGTGGGAC	450 or 456
CABYV	RdRp	2316	CABYV2316F	CTCCTTCCGATATTGGCTCG	
CABYV	RdRp	2730	CABYV2730R	CTCCAGTCAAARCCRGAGCAGTC	419
CABYV	CP	3617	CABYV3617F	CGGAAGACGACAACGAAGAA	
CABYV	CP	4040	CABYV4040R	CCGTTCCCCTTGTAGAGGAT	423
CMV	RNA-1	703	1CMV703F	CGAYGGYGCKATGATGTTTGAC	
CMV	RNA-1	1076	1CMV1076R	AGAGGGGAACCARATRCAATG	376
CMV	RNA-2	1063	2CMV1063F	ACCGGGAGYGGTCACMAGAG	
CMV	RNA-2	1510	2CMV1510R	TCYCGAAGGCATCTCTGGAA	446
CMV	RNA-3	1468	3CMV1468F	CCTTTGCCGAAATTYGATTC	
CMV	RNA-3	1889	3CMV1889R	TGGAYGGACAACCCGTTT	434 or 451
PVY	CP	8907	PVY8907F	GGGYTTATGGTTYGGTGCAT	
PVY	CP	9300	PVY9300R	TGGTSTGYCTCTCTGTGTT	412
PVY	VPg	5946	PVY5946F	GATCCRCTCACTGGDGDCAAAT	
PVY	VPg	6322	PVY6322R	GCRAYTGGRRTGAARTCYCT	377
TSWV	RNA-S	238	STSWV238F	CCTATTTTGTATGATATTGAYTTYAGC	
TSWV	RNA-S	669	STSWV669R	GAGCAGAGAYTTCACTGCAAG	429
TSWV	RNA-M	43	MTSWV43F	TTTAGTTTCAYTWGCTAAACRTAATGG	
TSWV	RNA-M	465	MTSWV465R	TTGYTTTTCWGRTGGCATGTT	427
TSWV	RNA-L	280	LTSWV280F	YCAATCYACAGAGAACTTGCTTTR	
TSWV	RNA-L	717	LTSWV717R	CCATGATTARACCTGTTMCAAAA	440
MeCMV <sup>a</sup>	DNA-B	1808	MeCMVB-F	GAACTTGCGGAAATGCGTACC	
MeCMV	DNA-B	2212	MeCMVB-R	TCTAAATTGGCTGATCCTC	405

<sup>a</sup> : MeCMV was used as an « alien » sequence to estimate the frequency of contaminations during second PCR and Illumina sequencing

For NGS sequencing, the primers were « tailed » in 5' with the following sequences :

Tail primers F-2016	CTTCCCTACACGACGCTCTTCCGATCT
Tail primers R-2016	GGAGTTCAGACGTGTGCTCTTCCGATCT
Tail primers F-2017	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Tail primers R-2017	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Supplementary table S2B : other primers used for diagnostic and/or Sanger sequencing

Virus	Amplified region	Primer name	Primer sequence	Reference
ZYMV	Nib	ZYMV-CP-5'	GGTTCATGTCCCACCAAGC	Lecoq & Desbiez, 2012
ZYMV	CP	ZYMV-CP-3'	ATGTCGAGTATCACATTTCC	
PRSV	CP	PRSV-milCP-5'	TCTAACTCGTGCCACTCA	Lecoq & Desbiez, 2012
PRSV	CP	PRSV-finCP-3'	YARTTGCGCATACCAGGAG	
CYSDV	CP	CYSDV-CP-5'	GCGAGTTCGAGTGAGAATAA	Lecoq & Desbiez, 2012
CYSDV	CP	CYSDV-CP-3'	TCAATTACCACAGCCACCTG	
PepMV	Pol	potex-1	CAYCARCARGCNAARGAYSA	Gibbs et al., 1998 <sup>a</sup>
PepMV	Pol	potex-2	TCDGTRTTDGCRTCRAADGT	
TICV	CP	TICP1 (+)	CCCCTCGAGATGGAAAACCTTATCTGG	Jacquemond et al., 2008 <sup>b</sup>
TICV	CP	TICP2 (-)	GGTTCCGCACCTCCAGTAC	
ToCV	CP	ToCP1 (-)	CCCCTCGAGTTAGCAACCAGTTATCGATGC	Jacquemond et al., 2008

ToCV	CP	ToCP2 (+)	CCTCAGTTAAAGCAGCCGG	
BPYV	CP	BPYV-CP-5	CTGACATATGGGAGATAATGATGATGG	Lecoq & Desbiez, 2012
BPYV	CP	BPYV-CP-3	CTGACTCGAGTCAGTTTCCATAAGAAGC	
Begomovirus	CP (DNA-A)	GemCP-V-5'	GCCYATRTAYAGRAAGCCMAG	Lecoq & Desbiez, 2012
Begomovirus	CP (DNA-A)	GemCP-C-3'	GGRTTNGANGCRTGHGTACAYG	

<sup>a</sup>Gibbs, A., Armstrong, J., Mackenzie, A.M. and Weiller, G.F. (1998) The GPRIME package: computer programs for identifying the best regions of aligned genes to target in nucleic acid hybridisation-based diagnostic tests, and their use with plant viruses. *Journal of Virological Methods* 74, 67-76.

<sup>b</sup>Jacquemond, M., Verdin, E., Dalmon, A., Guilbaud, L. and Gognalons, P. (2009) Serological and molecular detection of Tomato chlorosis virus and Tomato infectious chlorosis virus in tomato. *Plant Pathology* 58(2), 210-220.

Supplementary material S3: Amplification, pooling and sequencing strategy for MiSeq sequencing of the 2016-2017 samples.

After validation of the RT-PCR primers, oligonucleotides with the same sequence as the validated primers but with an overhang adapter sequence at their 5' extremity were synthesized. In 2017, "pads" of 1 to 4 bases were added between the adapter and the specific primer sequences to introduce some additional sequence complexity. The "tailed" primers were tested against the same virus isolates as the untailed ones and with the same conditions, before being used to amplify the positive samples of the survey.

Reverse transcription was performed for ELISA-positive samples using the untailed reverse primer for each fragment. Two  $\mu\text{l}$  of extracted RNA and 8  $\mu\text{l}$  ddH<sub>2</sub>O were heated at 80°C for 3 mn, immediately cooled on ice and added to 10  $\mu\text{l}$  of a 2x reverse transcription mix for a final concentration of 1.25  $\mu\text{M}$  primer, 1 mM dNTP, 4U AMV reverse transcriptase (Promega, France) in 1x RT buffer (Promega). The samples were incubated for 1 hr at 42°C and stored at -20°C before PCR amplification.

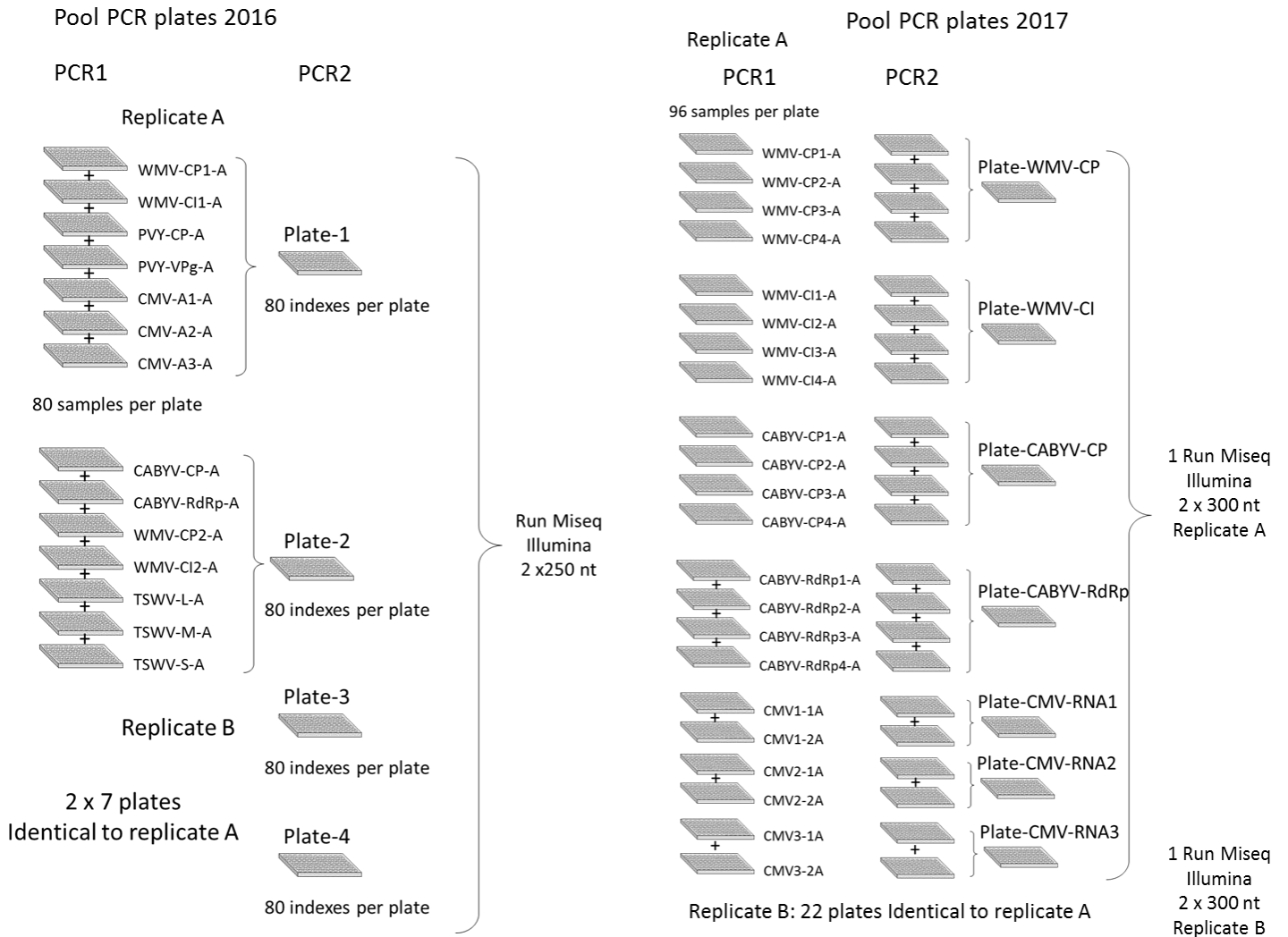
RT-PCR were also performed on different negative controls: extracts from healthy plant and from water, RT mix and PCR mix without RNA added. As a control for cross-contamination during the sequencing reaction, amplicons from an "alien" virus –here, a 405-nt fragment from DNA-B of the begomovirus melon chlorotic mosaic virus, originating from Venezuela and totally absent from Europe,- was added in one well before the second PCR.

In 2016, amplicons from different virus fragments were pooled before PCR2 in two plates for each replicate. A single index was added to each amplicon, using the 320 (4 x 80) i5 indexes developed by Illumina® available at that time. Second PCR and library preparation (320 libraries) were performed on the Genotoul platform (Toulouse, France). After the second PCR, all samples from one 96-well plate were pooled. The four super-pools of amplicon libraries were quantified using the KAPA library quantification kit (KAPA Biosystems) before High-throughput sequencing was performed on a MiSeq sequencer using v3 technology for pair-end, 2x250 nt sequencing.

In 2017, the second PCR was performed with a double-indexing as described by Galan et al., 2018, using 384 9-bp i5 and 384 i7 indices defined by Martin, 2019. Thus, each of the 9-bp i5 and i7 dual-index was used only for one PCR sample, eliminating the problem of 'leak' due to false index-pairing (Martin, 2019).



After the second PCR, all samples from one replicate corresponding to the same amplified region were pooled and purified on 1% agarose gel (7 pools per replicate). The 14-pools of amplicon libraries were quantified using the KAPA library quantification kit (KAPA Biosystems). For each replicate, the seven pools were then combined into a super-pool at a final concentration of 4 nM before being loaded on a MiSeq flow cell for pair-end 2 x 300 nt sequencing.



Supplementary table S4 : molecular profiles and their frequency in the 2016-2017 survey for CMV, WMV, PVY, TSWV and CABYV

Virus	Profile	Molecular group/reference	Reference isolate in the survey	Accessions <sup>a</sup>	Samples 2016 <sup>b</sup>	Plots 2016 <sup>c</sup>	Samples 2017	Plots 2017	Total samples
CMV	1	IA-IA-IA	EM160191	MN990924, MN990928, MN990932	58	13	157	27	215
	2	IB-IB-IB	EM170683	MN990925, MN990929, MN990933	0	0	13	2	13
	3	IA-IA-IB	EM170589	MN990926, MN990930, MN990934	0	0	1	1	1
	4	II	EM160348	MN990927, MN990931, MN990935	9	2	1	1	10
WMV	1	EM1	EM160492	MN990906, MN990916	24	3	65	12	89
	2	EM2	EM160451	MN990905, MN990915	14	2	8	1	22
	3	EM3	EM160155	MN990911, MN990921	1	1	0	0	1
	4	EM4	EM160724	MN990912, MN990922	18	4	136	19	154
	5	C08-451	EM160203	MN990913, MN990923	56	6	84	12	140
	6	Cg15-5055	EM170632	MN990910, MN990920	1	1	1	1	2
	7	New	EM160093	MN914160	33	5	46	8	79
	8	New	EM170143	MN990907, MN990917	0	0	5	1	5
	9	New	EM170465	MN990908, MN990918	0	0	9	1	9
	10	New	EM170501	MN990909, MN990919	0	0	4	2	4
	11	C06-188	EM160216	MN990904, MN990914	1	1	0	0	1
PVY	1	C1	EM160661	MN990952, MN990956	32	5	16	4	48
	2	C1	EM160700	MN990954, MN990958	3	2	0	0	4
	3	C1	EM160634	MN990953, MN990957	8	1	0	0	8
	4	NTN	EM160001	MN990955, MN990959	6	4	4	1	10
TSWV <sup>d</sup>	1	3-a- $\delta$	EM160608	MN990961, MN990966, MN990971	14	4	16	7	29
	2	3-a- $\varepsilon$	EM160039	MN990960, MN990965, MN990970	7	1	1	1	8
	3	3-b- $\alpha$	EM160251	MN990962, MN990967, MN990972	3	1	0	0	3
	4	3-b- $\varepsilon$	EM160319	MN9909631, MN990968, MN990973	5	2	0	0	5
	5	4-b- $\varepsilon$	EM160322	MN990964, MN9909696, MN990974	6	2	1	1	7
CABYV	1		EM160196	MN990936, MN990944	54	16	261	40	
	nd <sup>e</sup>		EM160730	MN990942, MN990950	nd				
			EM160394	MN990943, MN990951					
			EM160201	MN990937, MN990945					
			EM160202	MN990938, MN990946					
			EM170427	MN990939, MN990947					
			EM170257	MN990940, MN990948					
		EM170195	MN990941, MN990949						

<sup>a</sup> : GenBank accession number of the different fragments for the reference isolate from the survey

<sup>b</sup> : total number of samples in the survey (including mixed-infected ones) belonging to the molecular profile. For WMV, in cases of mixed infections, the components of the mix were considered as belonging to the same groups in the whole genome, without recombination between the different fragments. For CABYV it was not always possible to associate the RdRp and CP components from mixed infections

<sup>c</sup> : total number of plots where samples showing the profile were observed

<sup>d</sup> : TSWV groups were as defined by Tentchev et al., 2011

°: except for the major Profile 1, CABYV profiles and their relative frequencies could not be determined unambiguously because of the low phylogenetic signal in the sequences, the lack of full-length reference isolate and the high frequency of mixed infections in the dataset. The sequences of the main variants observed at least once without mixed infections were deposited in the GenBank.

Supplementary material S5A: Analysis of geographic structure performed with MAPI on the CABYV and WMV datasets.

For the CABYV-CP and WMV-CP datasets, pairwise genetic distances between all pairs of geolocalized samples were calculated and attributed to ellipses materializing the connections between the samples. Data from Pyrénées Orientales were excluded since they involved few samples and had a very high geographic distance with the rest of the dataset. A grid of hexagonal cells with a half-width of 3km was superimposed on the study area and each cell received the arithmetic mean of the ellipses intercepting its geographical extent. The mean was weighted using the inverse of the ellipse areas to limit long distance effects (i.e. Isolation by distance effect). The setting of the MAPI analysis can be controlled through several parameters (see Piry et al. 2016 for details on these parameters and their effect on the output of the method) : 1) the ellipse eccentricity, which controls the smoothing intensity (here set to the default value of 0.975) , 2) the radius of the error circle that controls for uncertainty on sample coordinates (here set to 10m according to GPS error measure), 3) the range of between-sample distances to consider in the analysis (here we set the minimum distance to 500m to exclude intra-field connections). We used the permutation procedure (1000 permutations), along with the FDR approach as described in Piry *et al.* (2016), to identify major areas of high genetic continuity and discontinuity. The cells of the MAPI grid for which the average value of genetic distance was computed using a small number of ellipses, and/or only long-distance ellipses, were discarded.

For WMV, we performed an additional MAPI analysis for which we only considered the genetic distances between the samples belonging to the same molecular group. Then, we used the ellipses materializing the genetic distances computed within each group altogether to compute the average level of genetic differentiation within the cells of the MAPI grid. The rationale behind this approach is that after their introduction, variants from the different molecular groups are transmitted by the same vector, which should lead, if there is no strong differential selection between the groups, to similar patterns of gene flow though the landscape (e.g. same dispersal corridors and barriers). Computing the genetic distances independently within each molecular group and then gathering the information from all groups to run MAPI allow to get a stronger dataset to investigate the spatial structure while limiting strong local effects due to the geographic proximity of highly differentiated variants

(belonging to different molecular groups) that do not directly result from natural demographic process (e.g. dispersal).

Supplementary material S5B – Corine Land Cover nomenclature (<https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover>). The land cover classes based on the level 2 of the Corine Land Cover nomenclature are presented in the first column. The proportion of the surface of the MAPI grid covered by these classes is indicated in brackets. The classes that represented more than 2% of the grid (indicated in bold) were retained in the RF analysis. For these classes, the detailed composition (level 3 of the Corine Land Cover typology) is provided along with their description (the proportion of each level 3 sub-classes is indicated in brackets in column 2). For the remaining classes that covered 2% or less of the spatial extent of the MAPI grid (excluded from the RF analysis), the description corresponds to the level 2 Corine Land Cover nomenclature.

Level 2 CLC	Level 3 CLC	Description
<b>Urban fabric (0.080)</b>	• Continuous urban fabric (0.02)	Urban structures and networks > 80% surface
	• Discontinuous urban fabric (0.98)	Urban structures and networks 30% to 80% surface. Includes villages with scattered blocks of residential buildings with numerous non-sealed spaces and complex cultivation pattern areas with scattered houses > 30 % surface.
Industrial, commercial and transport units (0.017)		Areas mainly occupied by industrial activities, services, transport infrastructures, airport, river and sea port installations.
Mine, dump and construction sites (0.002)		Areas mainly occupied by extractive activities, construction sites, man-made waste dump sites and their associated lands.
Artificial, non-agricultural vegetated areas (0.003)		Areas voluntarily created for recreational use. Includes green or recreational and leisure urban parks, sport and leisure facilities.
<b>Arable land (0.077)</b>	• Non irrigated arable land (0.88)	Cultivated land parcels under rainfed agricultural use for annually harvested non-permanent crops.
	• Rice fields (0.12)	Cultivated land parcels prepared for rice production, consisting of periodically flooded flat surfaces with irrigation channels.
<b>Permanent crops (0.11)</b>	• Vineyards (0.77)	Vineyard parcels covering >50% surface. Includes patches of orchards or annual crops < 50% surface
	• Fruit trees and berry (0.19)	Cultivated parcels planted with fruit trees and shrubs. Includes complex cultivation pattern mosaics where fruit parcels cover at least 50 % of the area.

	<ul style="list-style-type: none"> <li>Olive groves (0.3)</li> </ul>	Cultivated areas planted with olive trees. Includes interspersed annual crops occupying < 50% surface.
Pastures (0.018)		Lands permanently used for fodder production. Includes natural or sown herbaceous species, unimproved or lightly improved meadows and grazed or mechanically harvested meadows.
<b>Heterogeneous agricultural areas (0.13)</b>	<ul style="list-style-type: none"> <li>Complex cultivation patterns (0.78)</li> </ul>	Mosaic of small cultivated land parcels with annual crops, pasture and/or permanent crops, eventually with scattered houses or gardens. Includes agricultural mosaics with scattered houses, or garden huts (< 30% surface) in proximity of rural or urban settlements and used for growing agricultural crops, fruit, and vegetable for own consumption.
	<ul style="list-style-type: none"> <li>Land principally occupied by agriculture with significant areas of natural vegetation (0.22)</li> </ul>	Mosaic of <25 ha parcels of agricultural land (arable crops, pasture, permanent crops). Includes hortillonnage (vegetable crops and canals) and city gardens primarily for agricultural production use
<b>Forest (0.36)</b>	<ul style="list-style-type: none"> <li>Broad leaved forest (0.47)</li> </ul>	Vegetation formation composed principally of trees, including shrub and bush understorey, where broad-leaved species predominate (>75% formation).
	<ul style="list-style-type: none"> <li>Coniferous forest (0.28)</li> </ul>	Vegetation formation composed principally of trees, including shrub and bush understorey, where coniferous species predominate (>75% formation).
	<ul style="list-style-type: none"> <li>Mixed forest (0.25)</li> </ul>	Vegetation formation composed principally of trees, including shrub and bush understorey, where neither broad-leaved nor coniferous species predominate.
<b>Shrubs (0.17)</b>	<ul style="list-style-type: none"> <li>Natural grasslands (0.24)</li> </ul>	Low productivity grasslands under no or moderate human influence.
	<ul style="list-style-type: none"> <li>Moors and heathland (0.06)</li> </ul>	Vegetation with low and closed cover, dominated by bushes, shrubs, dwarf shrubs and herbaceous plants, forming a climax stage of development.
	<ul style="list-style-type: none"> <li>Sclerophyllous vegetation (0.48)</li> </ul>	Bushy sclerophyllous vegetation in a climax stage of development, including maquis, matorral and garrigue.
	<ul style="list-style-type: none"> <li>Transitional woodland shrub (0.22)</li> </ul>	Transitional bushy and herbaceous vegetation with occasional scattered trees. Can represent woodland degradation, forest regeneration / recolonization or natural succession.
Bareground (0.020)		Natural areas covered with little or no vegetation.
Inland waters (0.010)		Lakes, ponds and pools of natural origin containing fresh water and running waters made of all rivers and streams. Man-made fresh water bodies including reservoirs and canals.
Marine waters (0.001)		Oceanic and continental shelf waters, bays and narrow channels Saline or brackish coastal waters often formed from sea inlets by sitting and cut-off from the sea by sand or mud banks.

```

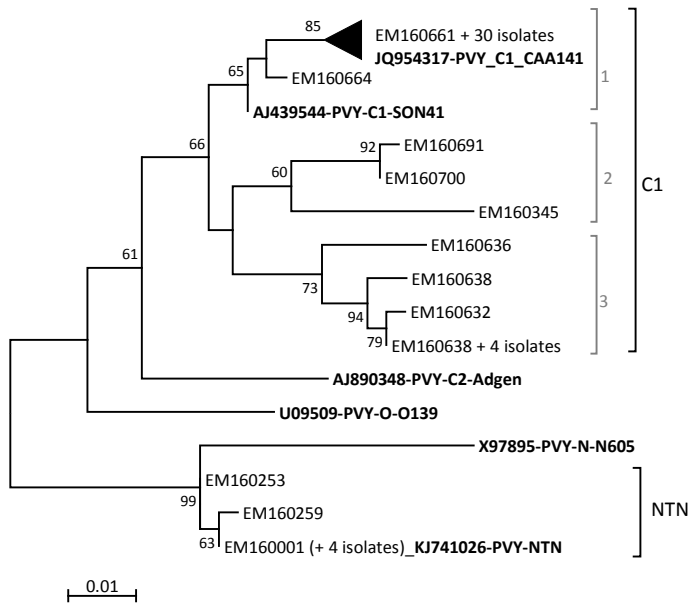
>EM160093 (Profile 7) AGAGGTACAAAGGAAGAGCATCC
>WMV_CI_3030          AGAGGTACAATAAGGAGATGCTT
>EM160724 (Profile 4) GAGAACGTGGTAAGGAGATGCTT

>EM160724 (Profile 4) GAGAACGTGGTAAGGAGATGCTT
>WMV_CI_3065          GAGAACGTGGAGGAAGAGCATCC
>WMV_CI_3036          GAGAACGCAAAGGAAGAGCATCC
>EM160093 (Profile 7) AGAGGTACAAAGGAAGAGCATCC

>EM160203 (Profile 5) GGCAACGTATCAAACGCAGCACC
>WMV_CI_3026          GGCAACGTATCAAACGCATGTT
>EM160724 (Profile 4) AATGGTAGGCTGGGTATGATGTT

```

Supplementary figure S6 : alignment of **polymorphic sites only** (23 nt out of 412 nt) in the WMV-CI fragment of reference and potentially recombinant haplotypes. Nucleotides corresponding to reference profile 4 are highlighted in grey, profile 5 is highlighted in black and profile 7 is not highlighted. Nucleotide positions of polymorphic sites are indicated above each alignment.



Supplementary figure S7



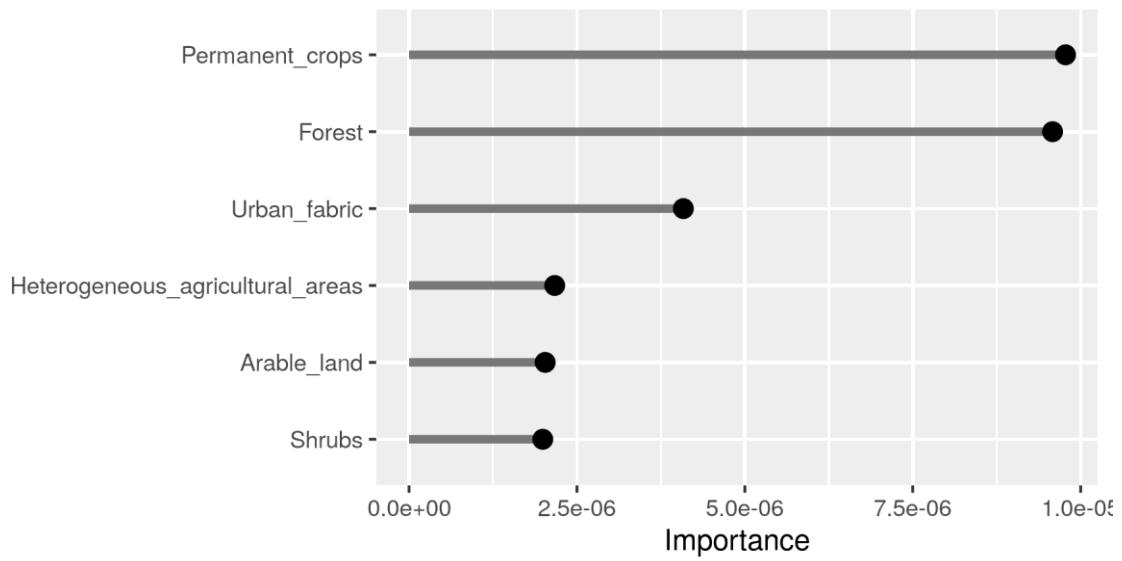


Figure S8A – CABYV: Random forest ranking of the importance of land cover variables to explain the variation in the average level of genetic differentiation as computed by MAPI.

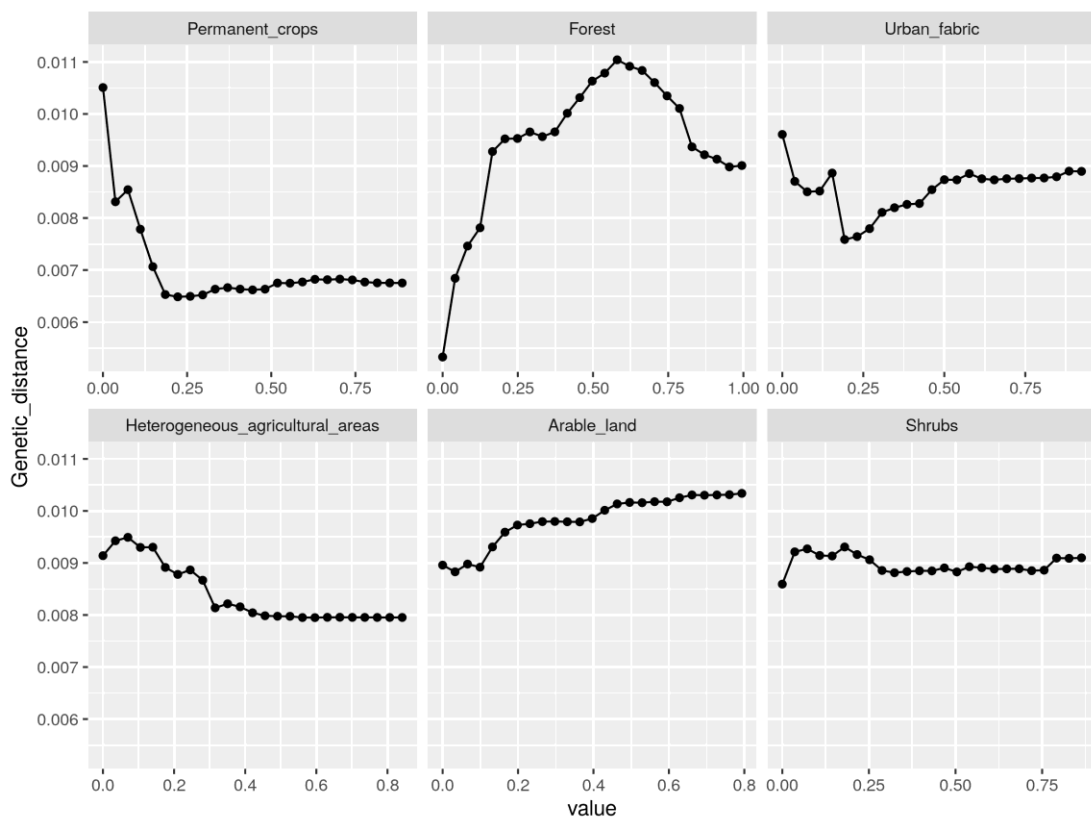


Figure S8B – CABYV: partial dependence plots from the RF assessing the shape of the relationships between the land cover variables and the level of genetic differentiation as computed by MAPI.

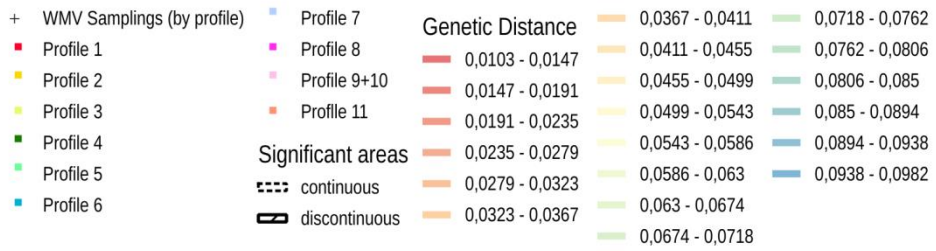
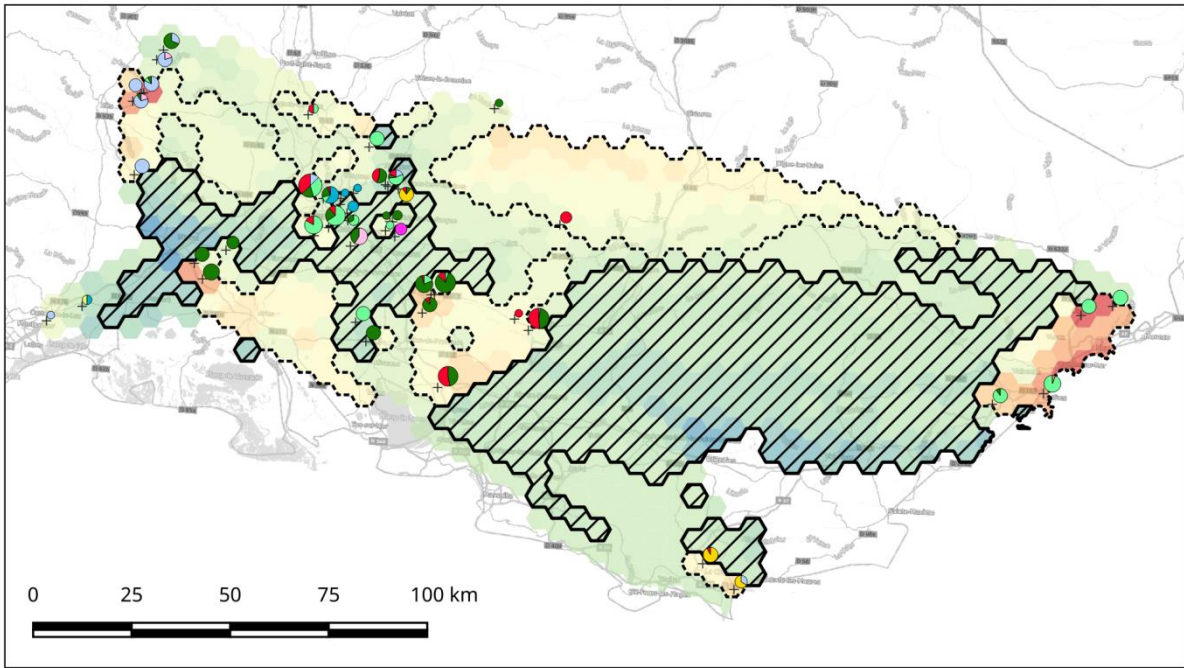


Figure S8C – Result of the MAPI analysis for the WMV-CP dataset. The sampling sites are represented with pie charts indicating the proportion of each molecular profile (as defined in Figure 3 and Supplementary table S4) within the sampling sites.

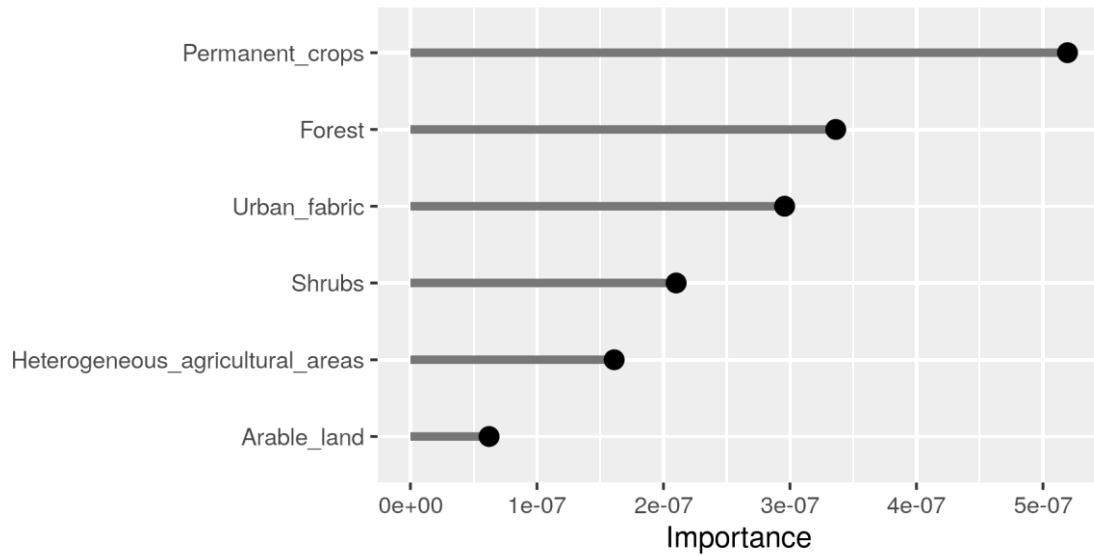


Figure S8D – WMV (considering intra-group connections): random forest ranking of the importance of land cover variables to explain the variation in the average level of genetic differentiation as computed by MAPI.

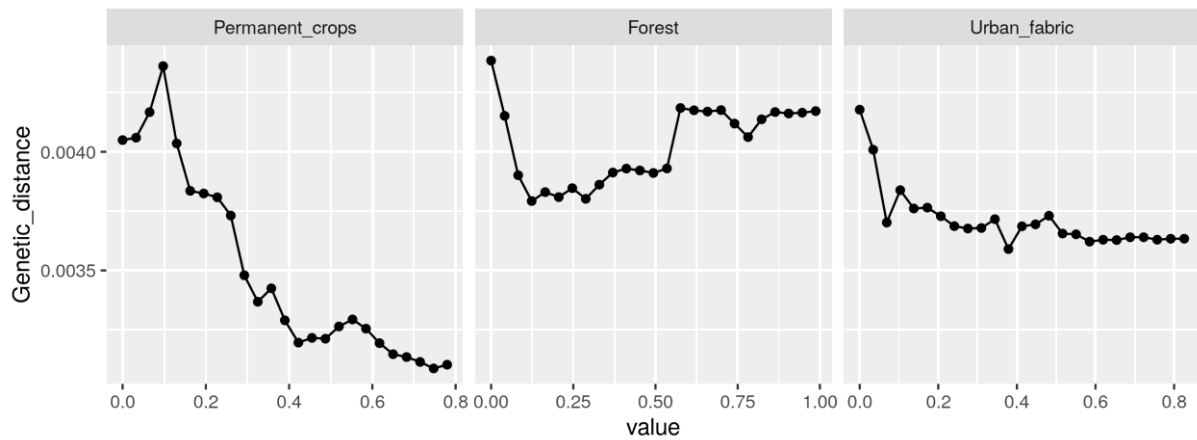


Figure S8E – WMV (considering intra-group connections): partial dependence plots from the RF assessing the shape of the relationships between the land cover variables and the level of genetic differentiation as computed by MAPI. Only the three classes identified as significant are presented.