

# Metagenomic-based impact study of transgenic grapevine rootstock on its associated virome and soil bacteriome

Jean-Michel Hily<sup>1,†</sup>, Sandrine Demanèche<sup>2,†</sup>, Nils Poulicard<sup>3</sup>, Mélanie Tannières<sup>1,‡</sup>, Samia Djennane<sup>1</sup>, Monique Beuve<sup>1</sup>, Emmanuelle Vigne<sup>1</sup>, Gérard Demangeat<sup>1</sup>, Véronique Komar<sup>1</sup>, Claude Gertz<sup>1</sup>, Aurélie Marmonier<sup>1</sup>, Caroline Hemmer<sup>1</sup>, Sophie Vigneron<sup>1</sup>, Armelle Marais<sup>4</sup>, Thierry Candresse<sup>4</sup>, Pascal Simonet<sup>2</sup> and Olivier Lemaire<sup>1,\*</sup>

<sup>1</sup>INRA, SVQV UMR-A 1131, Université de Strasbourg, Colmar, France

<sup>2</sup>Laboratoire Ampère (CNRS UMR5005), Environmental Microbial Genomics, École Centrale de Lyon, Université de Lyon, Ecully, France

<sup>3</sup>IRD, Cirad, Univ. Montpellier, IPME, Montpellier, France

<sup>4</sup>UMR 1332 Biologie du Fruit et Pathologie, INRA, Université de Bordeaux, Villenave d'Ornon Cedex, France

Received 15 February 2017;

revised 12 May 2017;

accepted 19 May 2017.

\*Correspondence (Tel +33 389 22 49 52;

fax +33 389 22 49 33; email

olivier.lemaire@inra.fr)

‡Present address: European Biological Control Laboratory, USDA-ARS, Campus International de Baillarguet CS 90013 Montferrier-Sur-Lez, 34988, Saint Gely-Du-Fesc Cedex, France.

†These authors contributed equally to this work.

**Keywords:** environmental

microbiology, metagenomics,

grapevine, transgenic rootstock, safety

and regulatory affairs.

## Summary

For some crops, the only possible approach to gain a specific trait requires genome modification. The development of virus-resistant transgenic plants based on the pathogen-derived resistance strategy has been a success story for over three decades. However, potential risks associated with the technology, such as horizontal gene transfer (HGT) of any part of the transgene to an existing gene pool, have been raised. Here, we report no evidence of any undesirable impacts of genetically modified (GM) grapevine rootstock on its biotic environment. Using state of the art metagenomics, we analysed two compartments in depth, the targeted *Grapevine fanleaf virus* (GFLV) populations and nontargeted root-associated microbiota. Our results reveal no statistically significant differences in the genetic diversity of bacteria that can be linked to the GM trait. In addition, no novel virus or bacteria recombinants of biosafety concern can be associated with transgenic grapevine rootstocks cultivated in commercial vineyard soil under greenhouse conditions for over 6 years.

## Introduction

Until the development of genetic transformation tools, genetic improvement of plants relied solely on the selection of the most interesting genotypes, on crossing two individuals and thereby combining their genomes. The potential impact of the new genotype generated through classical breeding has largely been overlooked, as long as the final product obtained displayed the desired characteristics. From the beginning, and especially for plants, the impact of genetically modified (GM) organisms on all components of their environment has been considered as an important biosafety issue. Impacts have been extensively studied (Brookes and Barfoot, 2015; Devos *et al.*, 2014; Scorza *et al.*, 2013; Vázquez-Salat, 2013), and especially the risk of horizontal gene transfer (HGT).

While other transformation methods have been developed (Gordon and Ruddle, 1981; Jaenisch and Mintz, 1974; Klein *et al.*, 1987; Neumann *et al.*, 1982), the fundamentals of genetic engineering are based on a natural mechanism harnessed by *Agrobacterium* species, eventually resulting in a disease known as crown gall (Smith and Townsend, 1907) that affects a wide range of plants worldwide, including grapevine. *In fine*, part of the genetic make-up of the bacteria, the transfer DNA (T-DNA)

(Chilton *et al.*, 1977), is transferred and integrated in the genome of the plant. Recently it has been shown that natural GM crops have been grown and used for consumption for millennia, as all cultivated sweet potato accessions contain and express one or more T-DNA sequences, while close wild relatives do not (Kyndt *et al.*, 2015). This suggests that T-DNA insertion(s) may have contributed to trait(s) selected during domestication. This case of HGT is far from being unique as endogenous viral elements have been described in many animal, fungus and plant species, including grapevine (Bertsch *et al.*, 2009; Feschotte and Gilbert, 2012; Katzourakis and Gifford, 2010; Koonin, 2010). Plants appear to be both donors and recipients of horizontally mobilized genes (Bock, 2010), underlying the key role of HGT in evolution and genetic diversification (Soucy *et al.*, 2015). Also, events of recombination between virus-derived transgene transcripts and infecting viral RNAs have been described under laboratory conditions of high selection pressure (Borja *et al.*, 1999; Gal *et al.*, 1992; Greene and Allison, 1994; Morrioni *et al.*, 2009). Similarly, the transfer of transgenic sequences from plants to bacteria has been observed with specifically selected donor and recipient organisms (Kay *et al.*, 2002). Yet, HGT events between GM plants and virus or bacteria populations should be observed in the field under conditions of natural selection pressure (Badosa

et al., 2004; Capote et al., 2007; Demanèche et al., 2008; Vigne et al., 2004b).

The present investigation is aimed at investigating HGT in the biotic environment of GM perennial plants over a span of at least 6 consecutive years. The work was performed with GM grapevine rootstocks (Vigne et al., 2004b) expressing the coat protein gene of *Grapevine fanleaf virus* (GFLV) strain F13 (*F13-cp*) and the neomycin phosphotransferase (*nptII*) gene, one of the most widely used selectable marker genes in plant genetic engineering (Jelenic, 2003; Turrini et al., 2015). These GM rootstocks were developed in an effort to confer resistance to GFLV (Lomonosoff, 1995; Sanford and Johnston, 1985). This virus is responsible for fanleaf degeneration, the most severe viral disease affecting vineyards worldwide (Andret - Link et al., 2004). GFLV is transmitted by a soil-borne ectoparasitic dagger nematode (*Xiphinema index*) and belongs to the genus *Nepovirus*. Its genome is composed of two single-stranded positive-sense RNAs that are sometimes associated with a satellite RNA (satRNA). In our experimental system, GM and control grapevines were planted in commercial vineyard soil infested with *X. index* and GFLV that was in a confined greenhouse. The virus was detected in all plants within the 4th year of plantation, implying that GFLV replicated within a transgenic background potentially interacting with transgenic *F13-cp* derived transcripts and/or proteins. Our system is, therefore, an excellent model for addressing HGT in a perennial crop, by comparing the structure and genetic diversity of GFLV populations as well as rhizosphere bacterial populations using metagenomics approaches.

## Results

### No transgene-derived sequences were detected in rhizospheric soil bacterial populations

To investigate HGT between GM plants and bacterial populations in the rhizospheric soil, antibiotic resistance in soil bacteria sampled from GM expressing *F13-cp* and *nptII* (Vigne et al., 2004b) and wild-type (WT) grapevine rootstocks was characterized using two different approaches: (i) a classical isolation and culture-based approach and (ii) direct molecular-based methods, such as PCR, qPCR and high-throughput sequencing. The culture-based approach relies on a direct counting of bacteria growing on culture media supplemented with or without kanamycin to estimate the cultivable antibiotic-resistant or total bacteria, respectively. The commercial vineyard soil from Bergheim used in this study contained an average of  $10^5$  cultivable bacteria per gram of dry soil (Figure 1a) of which 0.18% displayed resistance to kanamycin. This percentage represented approximately 300 cultivable kanamycin-resistant bacteria per gram of dry soil. No significant difference between rhizospheric soil of GM and WT grapevines was observed in either total or kanamycin-resistant cultivable bacteria (Mann–Whitney test,  $P = 0.440$  and  $0.930$ , respectively). Kanamycin-resistant bacteria were further tested by PCR for the presence of transgene sequences with two sets of primers targeting *nptII* (*nptII666/668*) and a specific sequence of the transgene construct (TgLarge), respectively (Figure S1). Of the 1,128 individual kanamycin-resistant bacterial cultures tested, no amplification was obtained with either primer set, whereas the *16S rRNA* gene (used as a control) was successfully amplified from all bacteria samples (Figures 1b and S1b). Furthermore, qPCR and DNAseq were applied to total and bacterial DNA from soil samples, bypassing potential limitations associated with the ability to culture bacteria. Total soil DNA samples contained  $1.8 \times 10^2$

to  $2.4 \times 10^3$  copies/ng DNA of *18S rRNA*, as shown by qPCR, whereas no *18S rRNA* (used as a marker for eukaryotic DNA presence) was amplified in any of the total bacteria DNA samples tested, as expected (Figure 1b). From both total and bacterial DNA, no transgene (Tgshort) or *nptII* (*nptII/sens*) sequences (Figure S1a) were amplified. In contrast, plant DNA samples from transgenic grapevines contained  $2.4 \times 10^2$  to  $4.0 \times 10^3$  copies/ng DNA of *nptII* and transgene sequence, while  $3.5 \times 10^5$  to  $2.3 \times 10^6$  copies/ng DNA of *18S rRNA* were identified (Figure 1b). This corresponded to a ratio of 1 copy of transgene for about 100 to 1000 copies of 18S. At the same time, total DNA from six soil samples, three exposed to GM grapevines (GMSo) and three exposed to WT grapevines (WTSO), were submitted to Miseq high-throughput sequencing and direct mapping of reads against the genome of *Agrobacterium tumefaciens* strain LBA4404 (WGS accession number JMKNO1000001-39) and pBin19 (GenBank number U12540.1). Both, the binary plasmid and the *A. tumefaciens* strain, had been used to generate GM rootstocks. The 8.7 million DNA sequences data set was analysed by applying high stringency (0.99 read length and 0.99 sequence similarity) to avoid false positive hits from soil bacteria DNA. Not a single soil DNA read corresponding to *A. tumefaciens* strain LBA4404 or pBin19 was detected. This was still the case even when the stringency was reduced to 0.5/0.8 (read length/similarity). Identical results were obtained with RNAseq of leaf samples from GM grapevines, validating the fact that plant (T0 transformants) and soil samples did not contain, at detectable levels, the recombinant bacterial strain used for transformation.

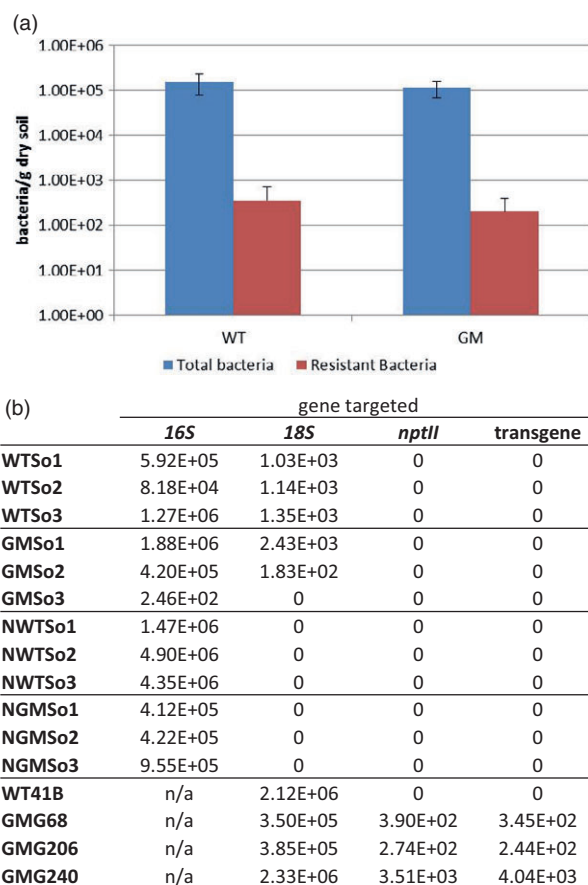
### No transgene-derived sequences translocation to scion or to viral population was observed

GM and WT rootstocks were exposed to a natural GFLV infection. To examine HGT between GM grapevines and GFLV populations, high-throughput sequencing was used, and two data sets were analysed as follows: (i) RNAseq data from 26 samples [corresponding to four sample categories collected from GM grapevine rootstock line G68 (GMR), wild-type grapevine rootstock (WTR), wild-type scion grafted onto GMR (ScGM) and wild-type scion grafted onto WTR (ScWT)] and (ii) high-throughput data from immunocapture-RT-PCR (IC-RT-PCR) products corresponding to encapsidated GFLV RNAs. To detect reads corresponding to transgene transcripts derived from *F13-cp* or *nptII*, high stringent mapping parameters were used to avoid potential cross-mapping with reads derived from the infecting-GFLV population and/or endophytic bacterial communities. No reads obtained by RNAseq matched the transgenic *F13-cp* sequence in any of the samples other than those from the transgenic rootstock (Tables 1 and S1, stringency 0.97/0.99). Similar results were obtained when investigating the *nptII* sequence, with reads being detected only in GMR samples (Tables 1 and S1, stringency 0.97/0.99). These results, therefore, reveal no detectable transfer of transgene-derived transcripts through the graft union as no transgene sequence signature was recovered from any of the ScGM samples (leaf or inflorescence, Table S1). These observations also suggest that no replication or movement-competent GFLV recombinant containing transgene-derived sequences emerged in the transgenic grapevine rootstocks with subsequent translocation into the scion. These particular findings were confirmed by IC-RT-PCR-NGS with no detection of transgenic *F13-cp* reads in any of the samples tested (GM or non-GM, Tables 1 and S1). Such a conclusion was not drawn due to a limitation of the approaches to detect recombination events since many crossover sites were

identified in both GFLV RNA1 and RNA2 [including the coat protein (*cp*) sequence] of the natural infecting-GFLV populations (see section below on natural GFLV recombinants identification). After removal of transgene-derived reads from the data set and relaxing mapping parameters (0.5/0.7), a large number of reads corresponding to GFLV *cp* were detected in all samples (Tables 1 and S1). This result reflected GFLV infection, unveiling the variability of the viral population (see next paragraph). While mapping at low stringency dramatically increased the number of GFLV reads recovered, such an increase in reads was not observed for *nptII* (Tables 1 and S1). This observation suggests that, similar to soil bacterial populations as mentioned above, endophytic bacteria in grapevine samples (see below and Figure S2) do not carry *nptII* sequences, significantly reducing any potential HGT events involving the GM-derived *nptII* gene.

### No specific soil bacterial communities' selection linked to GM grapevine

To characterize the diversity of bacterial communities in soil samples exposed to transgenic (GMSo) and wild-type (WTSO) rootstocks, three metagenomics strategies were used. The first one was a DNaseq approach from which 8.7 million DNA sequences were analysed using MG-RAST (<http://metagenomics.nl.gov/>, last visited 05/2017). No significant differences were observed between GMSo and WTSO soil samples regarding the type of organism and their abundance (ANOVA,  $P > 0.05$ , Figure 2a,b). In addition, a microarray for 16S rRNA detection based on 25 phyla showed some variability between samples (Figure 2c). For example, the *Candidatus/Portibacteria* phylum was not detected in sample WTSO1, whereas it represented 17.3% and 34.6% in samples GMSO3 and WTSO3, respectively. *Spirochaetes* were present at 6.4% in sample WTSO2, but only at 0.2% and 0.1% in samples GMSO1 and GMSO3, respectively, and they were absent in samples WTSO1, WTSO3 and GMSO2. Also, only samples WTSO2 and GMSO3 displayed *Synergistetes* (9.2% and 8.7%, respectively). All these differences were not significant when referring to GM and WT traits (ANOVA,  $P = 0.840$ ; *t*-test WT versus GM,  $P = 0.790$ ). The principal component analysis at the genus level demonstrated that samples WTSO2, GMSO1 and GMSO2 grouped according to the first axis, representing 48% of the observed variability (Figure 2d). Again, no grouping was observed when considering the GM/WT trait. To further describe bacterial communities within soil samples, a 454 sequencing-based approach of 16S rRNA was performed (Beckman Coulter Genomics, V4-V6 region). The dominant bacterial groups across all samples were *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Gemmatimonadetes*, *Planctomycetes* and *Bacteroidetes* (Figure 2e). When considering all samples, a clustering was observed but not according to GM/WT trait (Figure 2f, BGA MC-Test,  $P = 0.001$ ), as most GM samples grouped with WT samples. Interestingly, the sequencing region/primer effect was more discriminating than the GM versus WT grapevine traits (Figure 2f), and sample GMSO1 presented a higher proportion of *Actinobacteria* (Figure 2e,f). Bacterial signatures were also detected from leaf samples (Figure S1). At the genus level, no significant differences were detected when comparing GM/WT grapevines, ( $P > 0.050$ , Welch's two-sided *t*-test, with Benjamini-Hochberg FDR multiple test correction, Figure S2a). When comparing bacterial composition in soil and leaf samples (GMR, ScGM, GMSO, WTR, ScWT and WTSO), 410 significant different components were detected (ANOVA with Tukey-Kramer post hoc test and Benjamini-Hochberg FDR



**Figure 1** Antibiotic resistance of soil bacteria and gene transfer detection. (a) Number of total and Kanamycin-resistant cultivable soil bacteria associated with Genetically Modified (GM) or Wild-Type (WT) grapevine rootstocks. Results are expressed as the mean  $\pm$  95% confidence interval ( $2.31 \times \text{SEM}$ , for  $n = 9$ ). (b) Gene quantification estimated by qPCR. Quantification was expressed as copies per ng of DNA for soil samples (WTSO1 to GMSO3) and for plant samples (WT41B to GMG240) and as copies per g of dry soil for Nycodenz<sup>®</sup> samples (NWTSO1 to NGMSO3). Not applicable, n/a.

multiple test correction). Among these components, endophytic bacteria affiliated to the genus *Staphylococcus* were significantly more abundant in leaf than in soil samples, whereas the *Streptomyces* and *Clavibacter* genera were more abundant in soil than in leaf samples (Figure S2b). While viral signatures were present in all leaf samples, none were detected in soil samples. As expected, eukaryotic sequences were much more abundant in leaf than in soil samples (Figure S2c).

### No evidence of undesired impact of transgenic expression on virome

A comprehensive virological evaluation was performed, and the molecular diversity of the three GFLV RNAs (genomic RNAs 1, 2 and satellite) was assessed for each plant tested. For this, a dual strategy was used as follows: (i) a direct mapping of reads against a collection of reference sequences of grapevine-infecting viruses and viroids species and (ii) a *de novo* assembly of reads followed by BlastN/BlastX annotation of contigs. Remarkably, other than for GFLV-related reads, GM samples did not display any other viral reads (only extremely low contamination level with RPKMs = 0,

**Table 1** Transgenic-derived transcripts detection and sanitary status using RNAseq. Numbers are average of reads directly mapped onto transgenic *F13-cp* and *nptII* sequences, from all WTR (wild-type rootstock), ScWT (scion grafted onto WTR), GMR (transgenic rootstock) and ScGM (scion grafted onto GMR) samples from leaf tissue. Virus and viroid presence are expressed in RPKM: Reads Per Kilobase per Million reads mapped to the reference.

Sample name		WTR	ScWT	GMR	ScGM	
Total trimmed reads		51 379 136 ± 4 580 482 <sup>b</sup>	39 203 579 ± 7 454 413 <sup>ab</sup>	30 368 183 ± 5 195 118 <sup>a</sup>	36 263 138 ± 6 431 238 <sup>ab</sup>	
References	size (nt)	Stringency*				
Transgene						
<i>F13-CP</i>	1 515	0.97/0.99	0.17 ± 0.17 <sup>a</sup>	0.67 ± 0.42 <sup>a</sup>	204.6 ± 48.99 <sup>b</sup>	0.00 ± 0.00 <sup>a</sup>
		0.5/0.7 <sup>†</sup>	416 922 ± 42 119 <sup>b</sup>	197 349 ± 44 058 <sup>a</sup>	166 419 ± 34 989 <sup>a</sup>	126 110 ± 15 537 <sup>a</sup>
<i>nptII</i>	796	0.97/0.99	0.00 ± 0.00 <sup>a</sup>	0.50 ± 0.50 <sup>a</sup>	648.40 ± 139.76 <sup>b</sup>	0.00 ± 0.00 <sup>a</sup>
		0.5/0.7 <sup>†</sup>	0.00 ± 0.00 <sup>a</sup>	0.17 ± 0.17 <sup>a</sup>	355.80 ± 72.27 <sup>b</sup>	0.00 ± 0.00 <sup>a</sup>
Virus/viroid						
GFLV RNA1	6 855	0.5/0.7 <sup>†</sup>	2 142.46 ± 210.04 <sup>c</sup>	1 095.74 ± 137.37 <sup>ab</sup>	1 349.96 ± 197.86 <sup>b</sup>	818.78 ± 83.30 <sup>a</sup>
GFLV RNA2	3 333	0.5/0.7 <sup>†</sup>	4 687.45 ± 654.53 <sup>b</sup>	2 765.81 ± 360.86 <sup>a</sup>	2 956.36 ± 424.20 <sup>a</sup>	1 995.55 ± 230.72 <sup>a</sup>
GFLV	1 004	0.5/0.7 <sup>†</sup>	1 682.54 ± 364.25 <sup>b</sup>	981.05 ± 166.56 <sup>a</sup>	1 030.70 ± 150.24 <sup>ab</sup>	628.84 ± 186.02 <sup>a</sup>
satRNA						
GRSPaV	8 725	0.5/0.7 <sup>†</sup>	82.50 ± 21.24 <sup>ab</sup>	318.10 ± 80.81 <sup>b</sup>	0.12 ± 0.04 <sup>a</sup>	149.97 ± 83.98 <sup>ab</sup>
HSVd	298	0.5/0.7 <sup>†</sup>	75.00 ± 9.13 <sup>b</sup>	142.06 ± 20.22 <sup>c</sup>	0.24 ± 0.13 <sup>a</sup>	176.35 ± 10.29 <sup>c</sup>
GYSVd	366	0.5/0.7 <sup>†</sup>	7.98 ± 7.98 <sup>b</sup>	59.01 ± 11.61 <sup>c</sup>	0.06 ± 0.06 <sup>a</sup>	47.90 ± 6.37 <sup>c</sup>

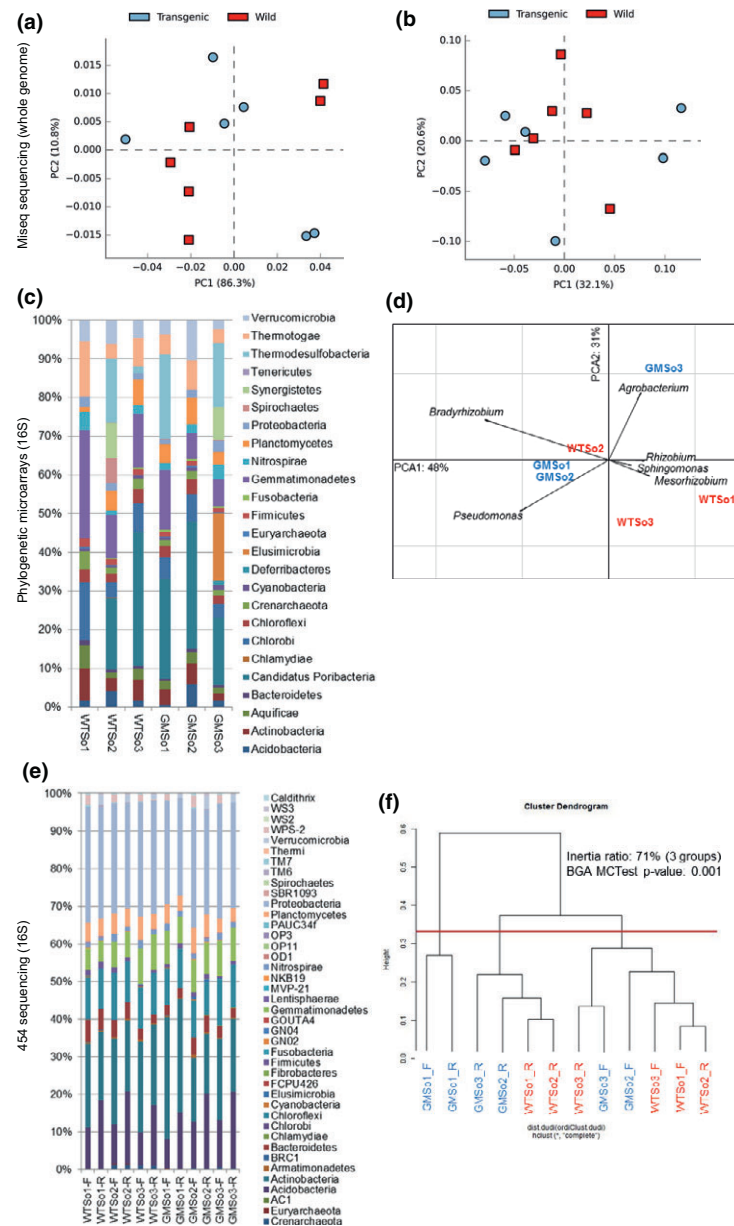
a, b, cSymbolize the post hoc analysis (Fisher's least significant difference, LSD).

\*Denotes the stringency (minimum length of the read/ homology with the sequence) for mapping reads to sequence of interest.

<sup>†</sup>Indicates that transgenic *cp* and *nptII* reads were removed prior to performing analysis at a lower stringency.

Tables 1 and S2). On the other hand, in all nontransgenic samples (e.g., WTR, ScWT and ScGM), reads matching *Grapevine rupestris stem pitting-associated virus* (GRSPaV), *Hop stunt viroid* (HSVd) and *Grapevine yellow speckle viroid* (GYSVd) were readily detected in addition to GFLV (RNA1, RNA2 and satRNA) reads (Tables 1 and S2). The absence in GM samples of GRSPaV, HSVd and GYSVd, which are extremely widespread grapevine pathogens, likely results from their elimination through somatic embryogenesis, which is the cornerstone of the grapevine transformation process (Gambino *et al.*, 2006; Panattoni *et al.*, 2013). While GFLV genomic RNAs and satRNA were detected in all samples, rootstock samples presented greater RPKM values (e.g., virus titre) than scions grafted onto them, regardless of the viral RNA (ANOVA test,  $P \leq 0.0421$ , Table S2). Even though infected with GFLV, GM rootstocks accumulated less virus than WTR samples (ANOVA test,  $P = 0.024$ , 0.064 and 0.159, for RNAs 1, 2 and satellite, respectively, Table 1). From the direct mapping analyses, a single-nucleotide polymorphism (SNP) study of GFLV genomic RNAs and satellite RNA was performed. Various SNP detection levels were tested in order to check different depths of the variability spectrum: from very deep (with detection limit being set at 0.5%), up to the detection of SNPs being the most represented (80%), with two levels in between (20% and 51%). Quantitatively, while SNP numbers fluctuated among samples (Table S3), no statistical differences (ANOVA test,  $P \geq 0.0744$ ) were found between sampling categories in each detection level tested, within any of the GFLV RNAs or the *cp*. To further investigate GFLV population diversity, a *de novo* assembly was performed. From the RNAseq data set, seventy complete GFLV RNA1 sequences (GenBank accessions KX011072-KX011075 and KX034840-KX034905) were obtained. A phylogenetic analysis of the RNA1 sequences grouped them into three distinct clades (Figure 3a) in which all sampling categories [GM-related samples (GMR and ScGM) or WT-related samples (WTR

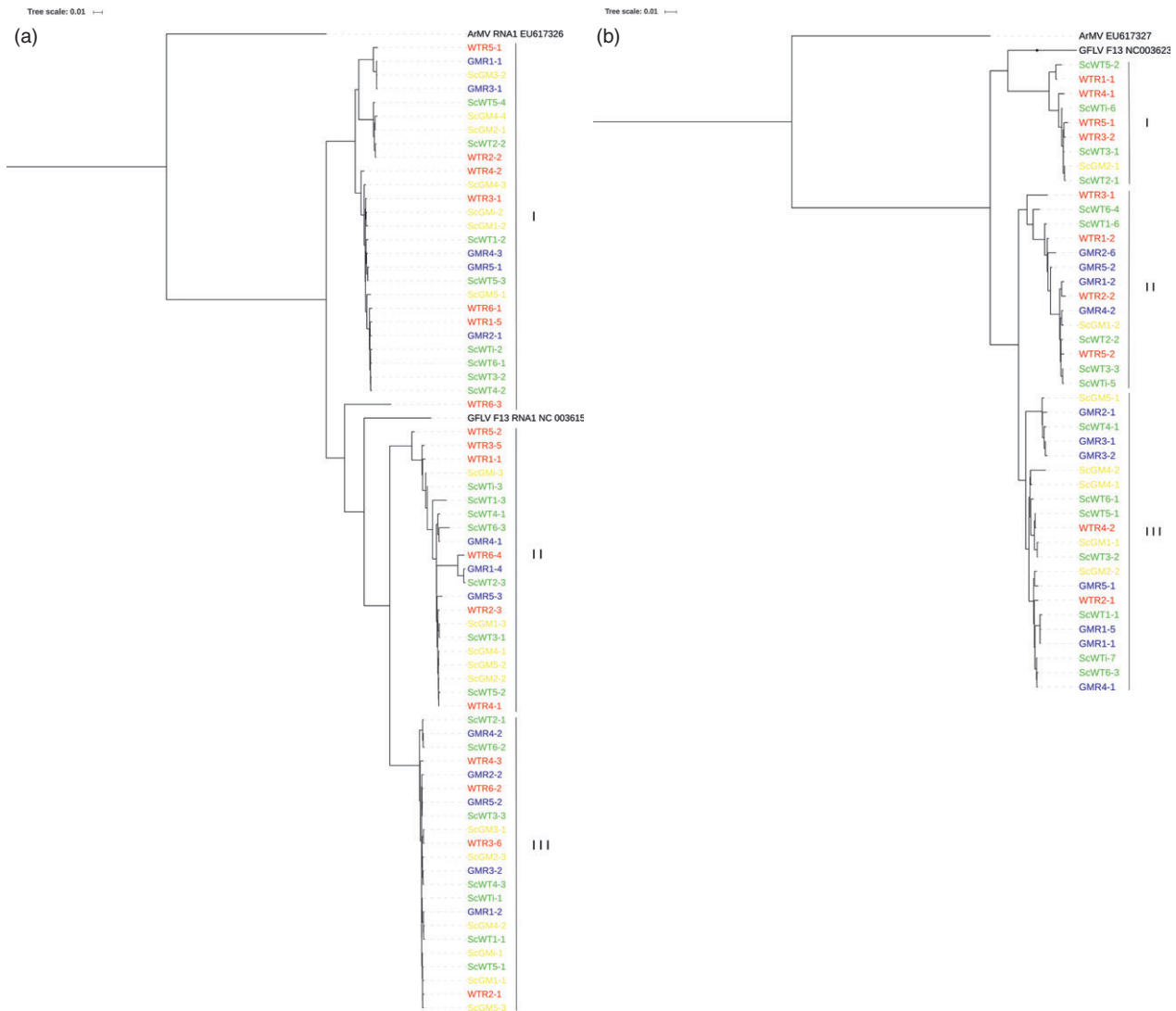
and ScWT)] were represented. This suggests no selection for a particular virus population among the different sampling categories (Chi-squared test,  $P = 0.976$ ). In addition, the genetic diversity ( $\pi = 0.0966 \pm 0.0035$  and  $0.0973 \pm 0.0036$  for WT- and GM-related populations, respectively), the ratio of synonymous and nonsynonymous substitution ( $d_N-d_S = -0.2166 \pm 0.0097$  and  $-0.2157 \pm 0.0098$ , respectively) and the Tajima's D values ( $D_T = 1.353$  and 1.832, respectively) followed the same pattern between WT and GM populations, suggesting no particular effect of the GM grapevine rootstocks on the GFLV RNA1 diversification and selection. This was supported by the very low value of the fixation index  $F_{ST}$  between the WT and GM populations ( $F_{ST} = -0.022$ ;  $P = 0.921$ ), supported no genetic differentiation between these two populations (Figure 4a). Similarly, 31 full-length sequences of GFLV satRNA were obtained (GenBank accessions KX034950-KX034980). While all sequences clustered together (clade I, Figure S3a) compared to the known satellite diversity (Čepin *et al.*, 2016; Gottula *et al.*, 2013), no effect of the GM plants on the satRNA diversity and selection was observed, as confirmed by the absence of genetic structure between GM and WT populations (Figure S4a,  $F_{ST} = -0.013$ ;  $P = 0.645$ ). As for GFLV RNA2, 44 complete sequences (GenBank accessions KX034906-KX034949) were assembled, alongside a total of 75 partial sequences spanning the 1515 bp of the *cp*. Phylogenetic analyses of the RNA2 and *cp* sequences grouped them into three clades (Figures 3b and S3b, respectively). Interestingly, the three clades did not display members from each sampling category as GMR sample sequences were not accounted for in clade I (Figures 3b and S3b). This observation indicated a potential population imbalance, but the existence of a selective bias linked to the presence of the *F13-cp* transgene from the GMR category was not statistically supported (Chi-squared test,  $P = 0.208$  and  $P = 0.135$  for the RNA2 and *cp* data sets, respectively). Nonetheless, the genetic diversity was drastically



**Figure 2** Microbial diversity comparison using Miseq whole-genome sequencing (a and b), phylogenetic microarrays (c and d) and 454 16S rRNA gene sequencing (e and f). (a) Principal component analysis (PCA) on MG-RAST organism abundance with representative hits classification at 80% cut-off. (b) PCA on MG-RAST functional abundance with hierarchical classification at 80% cut-off. (c) Sample comparison among the 25 phyla represented on the microarray. (d) Sample comparison at the genus level by PCA. (e) Sample comparison at the phylum level (%). (f) Groups generated by clustering. WTS01 to WTS03: rhizospheric soil samples of wild-type plants; GMS01 to GMS03: rhizospheric soil samples of genetically modified plants; \_F : forward reads related to the V4 region of the 16S RNA gene; \_R : reverse reads related to the V6 region of the 16S RNA gene.

lower in GM than WT populations (Figure 4b;  $0.0385 \pm 0.0021$  and  $\pi = 0.0670 \pm 0.0036$ , respectively). In addition, Tajima's D test was significantly negative for the GM population (Figure 4b;  $D_T = -0.742$ ), underlying a recent selective sweep of the *cp* in particular, while WT population followed a balancing selection ( $D_T = 1.521$ ). Furthermore, the  $F_{ST}$  values indicated a significant genetic differentiation between GM and WT populations for both RNA2 and *cp* (Figure 4b,c,d;  $F_{ST} = 0.077$ ;  $P = 0.033$  for RNA2 data set;  $F_{ST} = 0.094$ ,  $P = 0.038$  and  $F_{ST} = 0.078$ ,  $P = 0.019$  for *cp* data sets). More precisely, when testing sampling classes in a pairwise comparison, a number of

$F_{ST}$  with statistical significance were observed, in each case involving the GMR category (Table S4). These results clearly demonstrate that the RNA2 and especially the *cp* gene are distinct between WT- and GM-related populations (Figure 4b,d). Remarkably, this difference in *cp* structuration was only due to the clade I missing from GMR class, as removal of clade I sequences from the data set abolished the genetic differentiation between populations (Figure S4b;  $F_{ST} = -0.020$ ;  $P = 0.474$ ). These results were confirmed by performing the same analyses on the IC-RT-PCR data set (Figures S3c and S4c and Table S4).



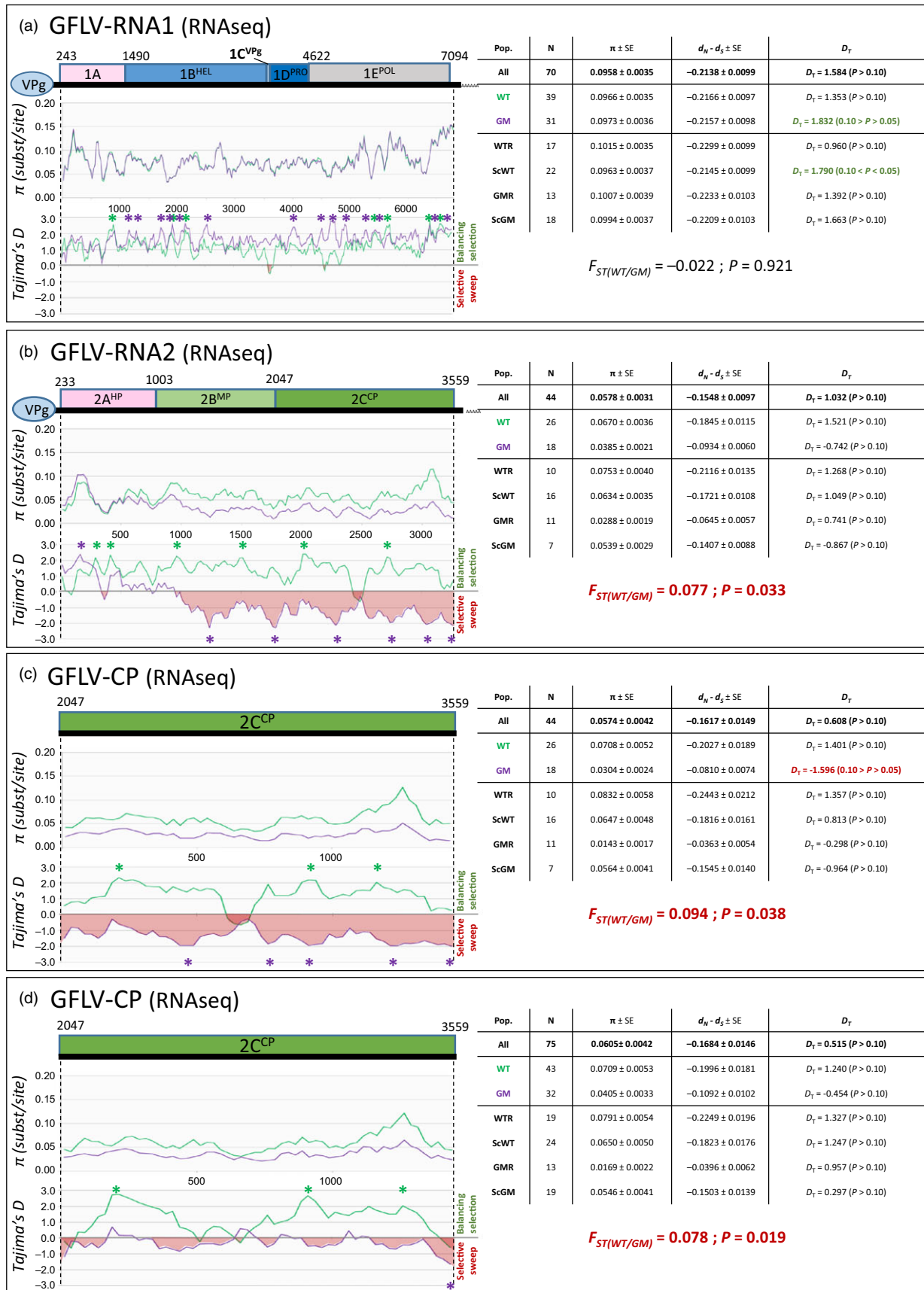
**Figure 3** Phylogenetic relationships of GFLV RNA1 and RNA2 genomes obtained from leaf samples from GM rootstock (GMR, in blue), non-GM rootstock (WTR, in red), scion grafted onto GMR (ScGM, in yellow) and scion grafted onto WTR (ScWT, in green) samples assembled with CLC Workbench 8.5.1 software. Phylogenetic tree based on the Maximum likelihood of (a) 70 full-length sequences of GFLV RNA1 and (b) 44 full-length sequences of GFLV RNA2. Bootstrap values are shown.

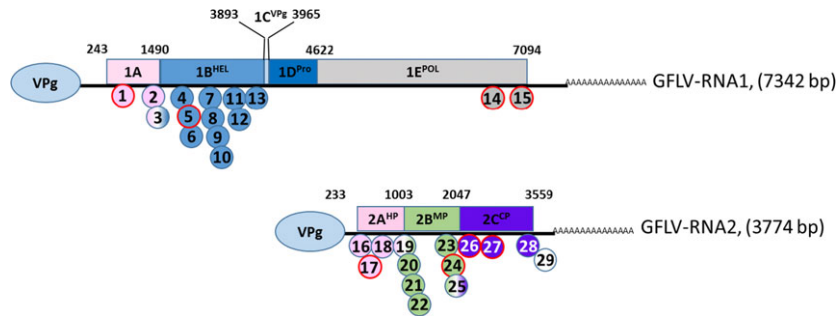
#### Detection of viral recombination events, none involving transgenic sequence

In previous work (Vigne *et al.*, 2004b), despite the fact that many natural recombination events were detected (Vigne *et al.*, 2004a, 2005), no HGT between *F13-cp* transgene-derived transcripts and

GFLV populations was identified. In this study, of the 70 complete RNA1 sequences (mentioned above), 15 recombination sites were identified (Figure 5 and Table S5, with corrected *P*-value  $\leq 5.59 \times 10^{e-12}$ ). While no recombination events were detected in the *Vpg* or *protease* sequences, between two and 11 breakpoints were identified in the *1A*, *helicase* and *polymerase*

**Figure 4** Genetic diversity analyses of GFLV RNA1 sequences (a), GFLV RNA2 sequences (b), GFLV *cp* gene sequences (c) and all GFLV sequences spanning the *cp* gene (d) assembled with CLC Workbench 8.5.1 software. Sequences were obtained from GM rootstock (GMR), non-GM rootstock (WTR), scion grafted onto GMR (ScGM) and scion grafted onto WTR (ScWT) samples. Graphics represent  $\pi$  (substitution per site) along each sequence and Tajima's *D* ( $D_T$ ) for evolution study, comparing GM grapevine-related sequences (GMR + ScGM, in aquamarine) and WT grapevine-related samples (WTR + ScWT, in purple). Overall genetic diversity is composed of *N*: number of sequences,  $\pi \pm SE$ : overall genetic diversity (substitution/site)  $\pm$  standard error,  $d_S$ ,  $d_N$ : diversity of synonymous and nonsynonymous substitutions, respectively ( $d_N - d_S < 0$ : negative/purifying selection;  $d_N - d_S = 0$ : neutral selection;  $d_N - d_S > 0$ : positive/diversifying selection), Tajima's *D* ( $D_T$ ):  $D_T = 0$  correspond to a mutation-drift equilibrium,  $D_T > 0$  indicates balancing selection, sudden population contraction and  $D_T < 0$  distinguish a recent selective sweep, population expansion after a recent bottleneck, (\*  $P < 0.05$ ). Genetic differentiation expressed as the Fixation Index ( $F_{ST}$ ) either overall or pairwise, yielding the genotypic frequencies in the entire population, with associated *p*-value. Significant *p*-values are in red.





**Figure 5** Localization of recombination sites detected in GFLV RNA1 and RNA2 sequences obtained from ‘*de novo*’ assembly using CLC Genomics Workbench software (positions based on GFLV-F13 RNA1 and RNA2 sequences, NC003615 and NC003623, respectively). Colour code indicates gene sequences that have undergone recombination in each viral genome. In red are shown the sites that were analysed and confirmed to be biologically present in the viral population. Numbers correspond to those in Table S6.

sequences. All sites were confirmed in the RNASeq data set when recombinants were used as references for a direct mapping analysis using very stringent parameters (Table S5). Of these, four randomly selected crossover sites were confirmed as being biologically genuine in the viral population following RT-PCR and Sanger sequencing. The same investigation was performed using the complete set of full-length RNA2 sequences, and 14 recombination sites (corrected  $P$ -value  $\leq 4.10 \times 10^{-09}$ ) were identified encompassing all three RNA2 coding regions (Figure 5 and Table S5). Using the larger *cp* data set, four recombination events were identified, two of which, randomly selected, were confirmed to be biologically present in the viral populations. As previously documented (Vigne *et al.*, 2004b), none of the crossover events identified within the *cp* were associated with the *F13-cp* transgene sequence.

## Discussion

It is now widely accepted that horizontal gene transfer (HGT) is shaping the web of life (Soucy *et al.*, 2015). The selectable marker *nptII* that confers resistance to kanamycin is present in a large number of GM crops and has been extensively studied (Dantas *et al.*, 2008; Jelenic, 2003). Two decades ago it was concluded that *nptII* poses no risk to humans, animals or the environment (Fuchs *et al.*, 1993; Nap *et al.*, 1992). The results presented here demonstrate that the presence of *nptII* in GM grapevine rootstocks did not increase the level of kanamycin-resistant bacteria in the rhizospheric soil because similar amounts of kanamycin-resistant bacteria were detected in soil samples of WT rootstocks. This confirms previous results on the occurrence of bacteria resistant to kanamycin in various ecosystems (D’Costa *et al.*, 2007; Dantas *et al.*, 2008; Li *et al.*, 2009). Also, *nptII* was not detected in DNA from soil samples by qPCR or NGS even after a six-year period of continuous GM grapevine cultivation. In addition to the transgene flanked by the left and right T-DNA borders, fragments from the tumour-inducing plasmid and larger fragments, up to 18 kb, of *Agrobacterium* chromosomal DNA can occasionally be integrated into the plant genome during *Agrobacterium tumefaciens*-mediated transformation (Ulker *et al.*, 2008), potentially increasing the risk of HGT. Within our different NGS data sets, sequences of the *A. tumefaciens* strain used for grapevine transformation were never detected in DNA recovered from rhizospheric soil sampled from around GM rootstocks, in leaf samples or within the genome of GM grapevine rootstocks. Additionally, while differences in microbial

populations were observed among soil samples subjected to various conditions, those were due to biological sample variations, and not to the presence of GM rootstocks. More importantly, similar dominant bacterial groups have been observed in American vineyard soils (Burns *et al.*, 2015; Zarraonaindia *et al.*, 2015) as well as in our study, although differences in abundance were noticed. For example, *Proteobacteria*, *Actinobacteria* and *Acidobacteria* were the three most abundant bacteria in our traditional vineyard soil and in Californian soils (Burns *et al.*, 2015), whereas *Bacteroidetes* were in the top three in New York soil (Zarraonaindia *et al.*, 2015) but were less abundant in our study (7th most abundant). No specific bacterial groups seemed to be selected for in either WT or GM rhizospheric soils. These data support the role of soil properties and crop management practices in shaping the composition of microbial and fungal communities and their diversity (Campisano *et al.*, 2014; Pancher *et al.*, 2012), with no detectable effect of GM rootstocks.

As expected, transgene-derived transcripts were detected only in GM samples and were not found in any WT scions grafted onto GM rootstocks. Furthermore, as for bacteria, no HGT events between the *F13-cp* transgene sequence and GFLV populations were identified, while a large number of natural recombination events were readily detected *in silico* and *in vivo*, confirming previous studies (Vigne *et al.*, 2004a, 2005). When added to previous investigations (Tepfer *et al.*, 2015), these findings, using state of the art high-throughput sequencing technology, confirm that GM grapevine did not result in the emergence of novel recombinant viral isolates. Nonetheless, the GM grapevine line investigated in this study impacted the GFLV population diversity. Such a specific effect on the viral RNA2/*cp* population, observed only in the GMR category, could potentially be explained by a transgene-mediated silencing mechanism specifically targeting homologous *cp* sequences. Two additional elements strengthen this hypothesis: (i) GMR, here G68 line, is a GM rootstock presenting multiple transgene inserts (Vigne *et al.*, 2004b), potentially leading to the production of *F13-cp* dsRNA, a major trigger of silencing (Meister and Tuschl, 2004; Wesley *et al.*, 2001), and (ii) the counter-selected RNA2 population from clade I (Figure 3b) is the closest to the *F13-cp* transgene sequence (Figures 3b and S3b), suggesting a homology-dependent exclusion by gene silencing mechanism. This hypothesis is re-enforced by the fact that while the level of substitutions per site ( $\pi$ ) is lower in GM than WT population, the ratio of nonsynonymous/synonymous substitutions ( $d_N$ - $d_S$ ) is similar in both populations,



suggesting that the selection does not occur at the protein level. These results were consistent not only in the GM rootstock but also in most (4/5) of the scions grafted onto it, increasing the potential usefulness of the rootstock-mediated resistance technology against a telluric viral agent such as *Grapevine fanleaf virus*.

Environmental safety concerns have been raised on the release of GM plants. These risks could be particularly pertinent in the case of a perennial crop, such as grapevine, as vineyards are worked for decades. Our study strongly suggests that GM grapevine rootstock cultivation does not favour the development of recombinant viruses and/or endophytes of biosafety concern nor disturb the composition of nontargeted rhizospheric bacterial communities. Also, transgenic-derived transcripts recovery was limited to GM tissues and was not detected in WT scion grafted onto GM rootstocks. Taken together, all these results could potentially guide policy makers when deciding about GM rootstock regulation, likely shaping the agriculture of tomorrow. Such comprehensive multiscale environmental impact study of a GM crop should be performed not only in a mesocosm environment (represented by a confined greenhouse with high-density planting) but also in open-field trials where HGT events have yet to be observed and where abiotic conditions might distinctly modify the selection and evolution of virus and/or bacteria communities.

## Experimental procedure

### Plant material, soil samples and conditions

In this work, five genetically modified (GM) grapevine rootstock lines (G68, G77, G206, G219 and G240) (Vigne *et al.*, 2004b) were used. For the bacterial genomic analysis, soil surrounding all GM line rhizospheres was tested; however, in the viral genomic study, only the G68 GM line was fully investigated. While presenting different insertion numbers and sites (Vigne *et al.*, 2004b), all GM lines were transformed with the *coat protein (cp)* gene of *Grapevine fanleaf virus* (GFLV) strain F13 (F13-*cp*) in sense under the 35S promoter and the *npTll* gene under nos promoter, which was used as the selective marker gene (Figure S1a). To test potential movement of transgenic molecules intraplant through conduit-vessels, 'Pinot meunier' grape variety was grafted onto GM and non-GM rootstocks (41B rootstocks, same variety as the GM lines). Five repeats of each GM line and six repeats of untransformed controls, grafted or not, were randomly planted (Figure S5) in September 2006 in a homogenized commercial vineyard soil (Bergheim, France, 48.199405 lat., 7.349196 long.), infested with nematode *Xiphinema index* and GFLV. The soil was transported to the INRA experimental station in Colmar (The Local Monitoring Committee *et al.*, 2010), and was kept in a greenhouse (48.064457 lat., 7.334899 long.) within a 6 × 2 × 0.5 m confined arena made of concrete. Emergence and spread of the disease were monitored every year. In this setting, we studied the impact of GM plants on telluric microbiota and on the natural GFLV population *in planta*, using metagenomics sequencing analyses.

For the study on the metavirome, two different leaf samplings were performed; all completed in spring/early summer, when the virus titre is believed to be at its peak (Čepin *et al.*, 2010; Walter and Etienne, 1987; Walter *et al.*, 1984). In 2013, leaf samples were collected from G68-GM Rootstock (GMR), non-GM Rootstock (WTR), Scion grafted onto GMR (ScGM) and Scion grafted

onto WTR (ScWT) from which IC-RT-PCR-based NGS analyses were performed (Figure S5). In 2015, RNAseq-based analyses were performed on the same plants.

For the soil microbiota genomic study, three samples of 1.5 kg of soil were collected in October 2012 along WT rootstocks rooting system (WTS01, WTS02 and WTS03) as well as along roots of three GM rootstocks (GMS01, GMS02 and GMS03) (Figure S5). Samples were immediately sieved through a 2 mm mesh, providing for about 700 g soil for each sample, that were sent overnight to École Centrale de Lyon, Ecully (France) and stored at 22 °C in sterile plastic 2-L flasks. For a detailed experimental procedure, see Appendix S1.

### Bacterial genomics and soil study

#### Isolation of antibiotic-resistant bacteria

Total cultivable and antibiotic-resistant bacteria were isolated from the six soil samples described above. Soil suspensions were serially diluted for total cultivable bacteria counts. For antibiotic-resistant bacteria counts, soil suspensions were spread on medium supplemented with 50 µg/mL kanamycin and 50 µg/mL neomycin. Three Petri dishes per dilution were used, and soil dispersion was spread in triplicate for each of the six soil samples.

#### Bacterial cells recovery

Telluric bacteria were extracted from the six soil samples (see above, and named NWTSo1, NWTSo2, NWTSo3, NGMS01, NGMS02 and NGMS03) using a Nycodenz<sup>®</sup> gradient (Axis-shield, Oslo, Norway).

#### Soil DNA extraction

Total DNA was extracted from the six soil samples (Total RNA Isolation Kit with DNA elution accessory kit, Mo Bio Laboratories, Carlsbad, CA) as per manufacturer's instructions. Concentrations were determined with Qubit<sup>®</sup> 2.0 fluorometer (London, UK) and were stored at -20 °C.

#### PCR amplifications

The different targeted genes, primers set, annealing temperature, time of elongation and PCR product sizes are shown in Figure S1b. PCR amplifications were performed using the Titanium Taq Polymerase (Clontech Laboratories Inc., Palo Alto, CA). PCR amplifications to prepare standards for qPCR were performed using the Invitrogen Taq Polymerase (Invitrogen, Carlsbad, CA) with appropriate primers (Edwards *et al.*, 1989; Fierer *et al.*, 2005) (Figure S1b).

#### Quantitative PCR assays

Quantitative PCR was performed using the SYBR Green chemistry. The different targeted genes and other information are described in Figures 1 and S1. PCR amplifications were performed with SensiMix SYBR No-ROX Kit (Bioline, Reagents Ltd, London, UK). Amplifications were carried out using the Rotor-gene 6000<sup>™</sup> (QIAGEN, Courtaboeuf, France). The real-time PCR (qPCR) reaction mixture yielded a final volume of 20 µL including 1X SensiMix<sup>™</sup> SYBR qPCR Kit, 0.8 µL of each primer at a concentration of 10 µM, 2 µL of DNA (samples and standards), Nycodenz<sup>®</sup> samples or water (*q.s.p.* 20 µL). Gene copy numbers of samples were determined according to standard curves with  $R^2 > 0.99$ , obtained from serial dilutions of purified PCR products (10-fold dilutions from 10<sup>9</sup> to 10<sup>2</sup> copies were used). Data were analysed with Rotor-Gene 6000 software. The threshold limits

were manually positioned at the beginning of the exponential phase (linear scale). The dynamic tube normalization method was selected.

#### Metagenome sequencing (Miseq)

For high-throughput sequencing, 50 ng of DNA of each of the three WT and GM soil were fragmented by 'tagmentation' according to manufacturer's protocol (Nextera DNA sample prep, Illumina) in a final reaction volume of 50  $\mu$ L. The indexes were then added by PCR according to the Illumina protocol (Nextera Index kit). Libraries (1  $\mu$ L) were checked on the Bioanalyzer (Agilent, Santa Clara, CA) with High Sensitivity kit. The sequencing was then performed with the Miseq sequencer 2  $\times$  250 bp (v2 chemistry) at École Centrale de Lyon. Sequences were used as singleton, and the 12 data sets were then trimmed (Table S6) with CLC Genomics Workbench 8.5.1 (CLC bio Genomics, Qiagen, Aarhus, Denmark) and mapped against reference sequences (see NGS data analysis section).

#### Phylogenetic Microarrays

The 16S rRNA genes of soil DNA were amplified using the primers pA and T7-pH. The forward primer contained a T7 promoter site at the 5' end (5'-TAATACGACTCACTATAG-3'), which enabled T7 RNA polymerase mediated *in vitro* transcription using PCR products as templates. Data analysis of microarrays was performed as previously described (Bodrossy *et al.*, 2003; Sanguin *et al.*, 2006). Probe sequences are available on the website: <http://www.genomenviron.org/Research/Microarrays.html>, last visited 05/2017).

#### 16S rRNA sequencing (454)

The Beckman recommended primers (16S-0515F and 16S-1061R with a MID tag per sample on both primers) targeting the V4-V6 region were used to obtain 560 bp products from extracted soil DNA and sequenced (Beckman Coulter Genomics). Data were trimmed (Table S6) and analysed according to QIIME pipeline (Caporaso *et al.*, 2010). Data were submitted to MG-RAST website (Bergheim project, <http://metagenomics.anl.gov/linkin.cgi?project=mgp12555>).

#### Metagenomics analyses

Miseq and Hiseq data (28 metagenomes) were submitted to MG-RAST website and made publicly available (IMAGMO project, <http://metagenomics.anl.gov/linkin.cgi?project=mgp18015>). Sample name correspondence is provided in Figure S5. The representative hit classification table with a cut-off at 80% was downloaded at the genus level and further analysed with STAMP software (Parks *et al.*, 2014) version 2.1.3 and MEGAN version 4.70.4 (Huson *et al.*, 2011).

#### Viral genomic study

##### RNA extraction, cDNA amplification and high-throughput sequencing

The NGS method of choice was an RNAseq-based experiment. This method allows the sequencing of any poly-A-tailed RNA in the sample. From this, a sanitary status inspection of our testing system was performed. We also checked for potential movement of the transgenic molecule intraplant. As previously mentioned, four categories of samples (GMR, ScGM, WTR and ScWT) were tested. We focused on the GMR transgenic line G68 and ScGM scion grafted onto this line. Total RNA was extracted from

100 mg of leaf tissue using the RNeasy Plant mini kit (Qiagen), as per manufacturer's recommendations, from 6 different WTR, 6 ScWT as well as 5 GMR and 5 ScGM (Figure S5). Inflorescences were also sampled and extracted separately and then samples from the same category were mixed at a 1:1 ratio and that mix was sequenced and renamed WTRi, ScWTi, GMRi and ScGMi. The cDNA libraries were then made at the GeT-Genotoul platform facility (INRA-Toulouse, France), using TruSeq Stranded mRNA sample prep kit with in-house modifications. Experiments were performed on an Illumina Hiseq 3000 (Illumina, San Diego, CA) using a paired-end read length of 2 $\times$ 150bp with the Illumina Hiseq3000/4000 SBS sequencing kits.

On a second set of samples, a method focusing solely on encapsidated GFLV sequences IC step (Immuno Capture) followed by an RT-PCR step (Reverse Transcription-Polymerase Chain reaction) was performed. Polyclonal antibody @GFLV (from our lab) was used at 1:1000 dilution as previously described (Vigne *et al.*, 2004b). RT-PCR was performed using the Long-Range (2 step) RT-PCR kit (Qiagen), amplifying most of the RNA2 molecule. Post-IC and RNA extraction, some samples from the same category were mixed at a 1:1 ratio (Figure S5) prior to sequencing being performed at 2  $\times$  250 bp on a MiSeq250.

#### NGS data analyses

Analyses of data sets were performed using CLC Genomics Workbench 8.5.1 software (Qiagen). After trimming procedure and quality check, only reads above 70 nucleotides were kept (see Table S6). Commonly, for transgenic reads detection, very stringent parameters (nonetheless allowing for potential PCR and sequencing errors) were used with length fraction of 0.97 and similarity of 0.99. In a second phase, for sanitary status examination [Table S7 (Martelli, 2014)] and variant detection, less stringent parameters were used to detect a maximum diversity of the viral population using length fraction of 0.5 with similarity of 0.7. This was performed after removal of reads corresponding to the transgenic *cp* sequence from each GMR samples.

For the genetic diversity study, 'de novo assembly tool' from the Workbench 8.5.1 software was used, and the list of contigs obtained was then mapped against RNA1, RNA2 and satRNA complete GFLV consensus genomes, allowing the identification of GFLV contigs.

#### Sequences analyses, genetic diversity and recombination detection

Alignment analysis and tentative maximum likelihood-based phylogenetic trees were performed using MUSCLE (Edgar, 2004) and MEGA7 (Kumar *et al.*, 2016) software. The best ML-fitted model for each sequence alignment was used, and bootstrapping analyses of 100 replicates were conducted. Genetic diversity ( $\pi$ ) of the viral populations was estimated using the Kimura 2-parameter model, with standard errors of each measure based on 100 replicate bootstraps, as implemented in MEGA6 (Tamura *et al.*, 2013). The variation of  $\pi$  along the GFLV genome was evaluated by sliding window analyses using DnaSP v. 5.10 (Librado and Rozas, 2009). The difference between non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions over the coding sequences from GFLV populations was estimated by the Pamilo-Bianchi-Li method as implemented in MEGA6. Differences in nucleotide diversity of the viral populations between modalities were tested by analysis of molecular variance (AMOVA), as implemented in Arlequin v. 5.3.1.2 (Excoffier *et al.*, 2005). AMOVA calculates the  $F_{ST}$  index explaining the between-groups fraction of

total genetic diversity. Significance of these differences was obtained by performing 1000 permutations. Tajima's  $D$  ( $D_T$ ) and sliding window analyses were conducted using DnaSP v. 5.10 (Librado and Rozas, 2009) in order to distinguish the viral populations evolving randomly (per mutation-drift equilibrium;  $D_T = 0$ ) to those evolving under a nonrandom process ( $D_T > 0$ : balancing selection, sudden population contraction;  $D_T < 0$ : recent selective sweep, population expansion after a recent bottleneck).

All sequences obtained from 'de novo assembly' analyses were submitted to Recombination Detection Program (RDP v.4.46) (Martin et al., 2015) for recombination detection. To confirm the biological occurrence of the recombination, eight sites were tested by RT-PCR on RNA samples that had been used for RNAseq library construction. Resulting PCR products were Sanger-sequenced to confirm sequences (Table S5).

### Statistical analyses

Statistical analyses were performed using the statistical software package Statgraphics Centurion version 15.1.02 (StatPoint technologies Inc., Warrenton, VA) and XLSTAT (v2016-03-30882, Addinsoft, Paris, France). When data were not following a normal distribution and/or homoscedastic, both parametric (ANOVA, t-test) and nonparametric (Kruskal–Wallis, Mann–Whitney) tests were used. While both tests gave similar results and as ANOVA is robust to the partial violation of its assumptions, for simplicity, only ANOVA analyses are presented. *Chi-squared* statistical tests were performed using website: <http://www.socscistatistics.com/tests> (last visited 05/2017). Principal component analysis (PCA) and Clustering of PCA (OrdiClust) were computed using Ade4TkGUI package of R software (version 3.2.2) (Thioulouse and Dray, 2007). Statistical significance of groups was evaluated by Monte Carlo test (MC-Test) with 10 000 iterations.

### Acknowledgements

This work was supported by the Institut National de la Recherche Agronomique and the Agence National de la Recherche project IMA-GMO Grant ANR-2011-CESA-00501 granted to OL and SDE, after discussion initiated within the Local Monitoring Committee of the GMO experiment. The authors acknowledge Cécillon S. for microarray analysis, Rustenholz C. for helpful bioinformatics discussion, the GeT-Genotoul platform for NGS and the greenhouse team for technical support, as well as Fuchs M. and Mitchell E. for helpful discussion and valuable comments on the manuscript. In memory of Valerie Goldschmidt for initiating the soil microbiome study.

### Conflict of interest

The authors declare no conflict of interest.

### References

- Andret - Link, P., Laporte, C., Valat, L., Ritzenthaler, C., Demangeat, G., Vigne, E., Laval, V. et al. (2004) Grapevine fanleaf virus: Still a major threat to the grapevine industry. *J. Plant Pathol.* **86**, 183–195.
- Badosa, E., Moreno, C. and Montesinos, E. (2004) Lack of detection of ampicillin resistance gene transfer from Bt176 transgenic corn to culturable bacteria under field conditions. *FEMS Microbiol. Ecol.* **48**, 169–178.
- Bertsch, C., Beuve, M., Dolja, V.V., Wirth, M., Pelsy, F., Herrbach, E. and Lemaire, O. (2009) Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biol. Direct* **4**, 21–21.
- Bock, R. (2010) The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* **15**, 11–22.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A. and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ. Microbiol.* **5**, 566–582.
- Borja, M., Rubio, T., Scholthof, H.B. and Jackson, A.O. (1999) Restoration of wild-type virus by double recombination of tombusvirus mutants with a host transgene. *Mol. Plant Microbe Interact.* **12**, 153–162.
- Brookes, G. and Barfoot, P. (2015) Environmental impacts of genetically modified (GM) crop use 1996–2013: Impacts on pesticide use and carbon emissions. *GM Crops Food* **6**, 103–133.
- Burns, K.N., Kluepfel, D.A., Strauss, S.L., Bokulich, N.A., Cantu, D. and Steenwerth, K.L. (2015) Vineyard soil bacterial diversity and composition revealed by 16S rRNA genes: Differentiation by geographic features. *Soil Biol. Biochem.* **91**, 232–247.
- Campisano, A., Antonielli, L., Pancher, M., Yousaf, S., Pindo, M. and Pertot, I. (2014) Bacterial endophytic communities in the grapevine depend on pest management. *PLoS ONE* **9**, e112763.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336.
- Capote, N., Pérez-Panadés, J., Monzó, C., Carbonell, E., Urbaneja, A., Scorza, R., Ravelonandro, M. et al. (2007) Assessment of the diversity and dynamics of Plum pox virus and aphid populations in transgenic European plums under Mediterranean conditions. *Transgenic Res.* **17**, 367–377.
- Čepin, U., Gutiérrez-Aguirre, I., Balažič, L., Pompe-Novak, M., Gruden, K. and Ravnikar, M. (2010) A one-step reverse transcription real-time PCR assay for the detection and quantitation of Grapevine fanleaf virus. *J. Virol. Methods* **170**, 47–56.
- Čepin, U., Gutiérrez-Aguirre, I., Ravnikar, M. and Pompe-Novak, M. (2016) Frequency of occurrence and genetic variability of Grapevine fanleaf virus satellite RNA. *Plant. Pathol.* **65**, 510–520.
- Chilton, M.-D., Drummond, M.H., Merlo, D.J., Sciaky, D., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell*, **11**, 263–271.
- Dantas, G., Sommer, M.O.A., Oluwasegun, R.D. and Church, G.M. (2008) Bacteria subsisting on antibiotics. *Science*, **320**, 100–103.
- D'Costa, V.M., Griffiths, E. and Wright, G.D. (2007) Expanding the soil antibiotic resistome: exploring environmental diversity. *Curr. Opin. Microbiol.* **10**, 481–489.
- Demanèche, S., Sanguin, H., Poté, J., Navarro, E., Bernillon, D., Mavingui, P., Wildi, W. et al. (2008) Antibiotic-resistant soil bacteria in transgenic plant fields. *Proc. Natl Acad. Sci. USA* **105**, 3957–3962.
- Devos, Y., Aguilera, J., Diveki, Z., Gomes, A., Liu, Y., Paoletti, C., du Jardin, P. et al. (2014) EFSA's scientific activities and achievements on the risk assessment of genetically modified organisms (GMOs) during its first decade of existence: looking back and ahead. *Transgenic Res.* **23**, 1–25.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Edwards, U., Rogall, T., Blöcker, H., Emde, M. and Böttger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17**, 7843–7853.
- Excoffier, L., Laval, G. and Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinf. Online* **1**, 47–50.
- Feschotte, C. and Gilbert, C. (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**, 283–296.
- Fierer, N., Jackson, J.A., Vilgalys, R. and Jackson, R.B. (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl. Environ. Microbiol.* **71**, 4117–4120.
- Fuchs, R.L., Ream, J.E., Hammond, B.G., Naylor, M.W., Leimgruber, R.M. and Berberich, S.A. (1993) Safety assessment of the neomycin phosphotransferase II (NPTII) Protein. *Nat. Biotech.* **11**, 1543–1547.
- Gal, S., Pisan, B., Hohn, T., Grimsley, N. and Hohn, B. (1992) Agroinfection of transgenic plants leads to viable cauliflower mosaic virus by intermolecular recombination. *Virology* **187**, 525–533.

- Gambino, G., Bondaz, J. and Gribaudo, I. (2006) Detection and elimination of viruses in callus, somatic embryos and regenerated plantlets of grapevine. *Eur. J. Plant Pathol.* **114**, 397–404.
- Gordon, J. and Ruddle, F. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science*, **214**, 1244–1246.
- Gottula, J., Lapato, D., Cantilina, K., Saito, S., Bartlett, B. and Fuchs, M. (2013) Genetic variability, evolution, and biological effects of grapevine fanleaf virus satellite RNAs. *Phytopathology*, **103**, 1180–1187.
- Greene, A. and Allison, R. (1994) Recombination between viral RNA and transgenic plant transcripts. *Science*, **263**, 1423–1425.
- Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N. and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560.
- Jaenisch, R. and Mintz, B. (1974) Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc. Natl Acad. Sci. USA* **71**, 1250–1254.
- Jelenic, S. (2003) Controversy associated with the common component of most transgenic plants - Kanamycin resistance marker gene. *Food Technol. Biotechnol.* **41**, 183–190.
- Katzourakis, A. and Gifford, R.J. (2010) Endogenous viral elements in animal genomes. *PLoS Genet.* **6**, e1001191.
- Kay, E., Vogel, T.M., Bertolla, F., Nalin, R. and Simonet, P. (2002) In situ transfer of antibiotic resistance genes from transgenic (Transplastomic) Tobacco Plants to Bacteria. *Appl. Environ. Microbiol.* **68**, 3345–3351.
- Klein, T.M., Wolf, E.D., Wu, R. and Sanford, J.C. (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* **327**, 70–73.
- Koonin, E.V. (2010) Taming of the shrewd: novel eukaryotic genes from RNA viruses. *BMC Biol.* **8**, 1–4.
- Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.
- Kyndt, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., Gheysen, G. et al. (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: an example of a naturally transgenic food crop. *Proc. Natl Acad. Sci. USA* **112**, 5844–5849.
- Li, D., Yang, M., Hu, J., Zhang, J., Liu, R., Gu, X., Zhang, Y. et al. (2009) Antibiotic-resistance profile in environmental bacteria isolated from penicillin production wastewater treatment plant and the receiving river. *Environ. Microbiol.* **11**, 1506–1517.
- Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Lomonosoff, G.P. (1995) Pathogen-derived resistance to plant viruses. *Annu. Rev. Phytopathol.* **33**, 323–343.
- Martelli, G.P. (2014) Grapevine-infecting viruses. *J. Plant Pathol.* **96**, 7–8.
- 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**, 1–5.
- Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349.
- Morrioni, M., Thompson, J.R. and Tepfer, M. (2009) Analysis of recombination between viral RNAs and transgene mRNA under conditions of high selection pressure in favour of recombinants. *J. Gen. Virol.* **90**, 2798–2807.
- Nap, J.-P., Bijvoet, J. and Stiekema, W.J. (1992) Biosafety of kanamycin-resistant transgenic plants. *Transgenic Res.* **1**, 239–249.
- Neumann, E., Schaefer-Ridder, M., Wang, Y. and Hofschneider, P.H. (1982) Gene transfer into mouse lymphoma cells by electroporation in high electric fields. *EMBO J.* **1**, 841–845.
- Panattoni, A., Luvisi, A. and Triolo, E. (2013) Review. Elimination of viruses in plants: twenty years of progress. *Spanish Journal of Agricultural Research* **11**, 173–188.
- Panther, M., Ceol, M., Corneo, P.E., Longa, C.M.O., Yousaf, S., Pertot, I. and Campisano, A. (2012) Fungal endophytic communities in grapevines (*Vitis vinifera* L.) respond to crop management. *Appl. Environ. Microbiol.* **78**, 4308–4317.
- Parks, D.H., Tyson, G.W., Hugenholtz, P. and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, **30**, 3123–3124.
- Sanford, J.C. and Johnston, S.A. (1985) The concept of parasite-derived resistance—Deriving resistance genes from the parasite's own genome. *J. Theor. Biol.* **113**, 395–405.
- Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T.M., Nesme, X., Moëgne-Loccoz, Y. et al. (2006) Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. *Appl. Environ. Microbiol.* **72**, 4302–4312.
- Scorza, R., Kriss, A.B., Callahan, A.M., Webb, K., Demuth, M. and Gottwald, T. (2013) Spatial and temporal assessment of pollen- and seed-mediated gene flow from genetically engineered plum *Prunus domestica*. *PLoS ONE* **8**, e75291.
- Smith, E.F. and Townsend, C.O. (1907) A plant-tumor of bacterial origin. *Science*, **25**, 671–673.
- Soucy, S.M., Huang, J. and Gogarten, J.P. (2015) Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482.
- Tamura, K., Stecher, G., Peterson, D., Filipinski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.
- The Local Monitoring, C., Lemaire, O., Moneyron, A. and Masson, J. E. (2010) “Interactive Technology Assessment” and Beyond: the Field Trial of Genetically Modified Grapevines at INRA-Colmar. *PLoS Biol* **8**, e1000551.
- Tepfer, M., Jacquemond, M. and García-Arenal, F. (2015) A critical evaluation of whether recombination in virus-resistant transgenic plants will lead to the emergence of novel viral diseases. *New Phytol.* **207**, 536–541.
- Thioulouse, J. and Dray, S. (2007) Interactive multivariate data analysis in R with the ade4 and ade4TKGUI Packages. *Journal of Statistical Software* **22**, 1–14.
- Turrini, A., Sbrana, C. and Giovannetti, M. (2015) Belowground environmental effects of transgenic crops: a soil microbial perspective. *Res. Microbiol.* **166**, 121–131.
- Ulker, B., Li, Y., Rosso, M.G., Logemann, E., Somssich, I.E. and Weisshaar, B. (2008) T-DNA-mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants. *Nat. Biotech.* **26**, 1015–1017.
- Vázquez-Salat, N. (2013) Are good ideas enough?: The impact of socio-economic and regulatory factors on GMO commercialisation. *Biol. Res.* **46**, 317–322.
- Vigne, E., Bergdoll, M., Guyader, S. and Fuchs, M. (2004a) Population structure and genetic variability within isolates of Grapevine fanleaf virus from a naturally infected vineyard in France: evidence for mixed infection and recombination. *J. Gen. Virol.* **85**, 2435–2445.
- Vigne, E., Komar, V. and Fuchs, M. (2004b) Field safety assessment of recombination in transgenic Grapevines Expressing the coat protein gene of Grapevine fanleaf virus. *Transgenic Res.* **13**, 165–179.
- Vigne, E., Demangeat, G., Komar, V. and Fuchs, M. (2005) Characterization of a naturally occurring recombinant isolate of Grapevine fanleaf virus. *Adv. Virol.* **150**, 2241–2255.
- Walter, B. and Etienne, L. (1987) Detection of the Grapevine Fanleaf viruses away from the period of vegetation. *J. Phytopathol.* **120**, 355–364.
- Walter, B., Vuittenez, A., Kuszala, J., Stocky, G., Burckard, J. and van Regenmortel, M.H.V. (1984) Détection sérologique des virus du court-noué de la vigne par le test ELISA. *Agronomie*, **4**, 527–534.
- Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M., Rouse, D.T., Liu, Q., Gooding, P.S. et al. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590.
- Zarraonaindia, I., Owens, S.M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N.A. et al. (2015) The soil microbiome influences grapevine-associated microbiota. *mBio*, **6**, e02527–14.

## Supporting information

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

**Figure S1** Genes targeted for detection by PCR and qPCR experiments.

**Figure S2** Microbial diversity comparison between leaf (RNAseq 2 × 150) and soil samples (Miseq sequencing).

**Figure S3** Phylogenetic relationships of satRNA genomes, GFLV *cp* gene sequences from RNAseq and from IC-RT-PCR NGS-based dataset obtained from GM rootstock (GMR, in blue), non-GM rootstock (WTR, in red), scion grafted onto GMR (ScGM, in yellow) and scion grafted onto WTR (ScWT, in green) samples assembled with CLC Workbench 8.5.1 software. Phylogenetic tree based on the Maximum likelihood of (a) 31 full-length satRNA sequences obtained from RNAseq dataset; (b) 75 full-length GFLV *cp* gene sequences obtained from RNAseq dataset (T is for the transgenic sequence) and (c) 50 full-length of GFLV *cp* gene sequences obtained from IC-RT-PCR dataset. Bootstrap values are shown.

**Figure S4** Genetic diversity analyses of satRNA sequences (a), GFLV *cp* gene sequences (b) and GFLV *cp* gene sequences without clade I sequences (c) all from RNAseq dataset assembled with CLC Workbench 8.5.1 software.

**Figure S5** Map of the greenhouse assay with location of each sample.

**Table S1** Detection of transgenic plant-derived transcripts from RNAseq (total RNA) and IC-RT-PCR-NGS (virus encapsidated RNA)-based techniques.

**Table S2** Sanitary status.

**Table S3** SNPs quantification.

**Table S4**  $F_{ST}$  values for each sampling category.

**Table S5** (part A and B): Crossover sites.

**Table S6** Quality control of DNAseq and RNAseq.

**Table S7** List of reference of grapevine-infecting viruses and viroids tested for sanitary status investigation with accession number, based on the directory of virus and virus-like diseases of the grapevine and their agents<sup>62</sup>.

**Appendix S1** Bacterial genomics and soil study. Viral genomic study.