

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Genetic diversity of Flavescence dorée phytoplasmas at vineyard scale

### **This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1695744> since 2019-03-29T09:54:57Z

*Published version:*

DOI:10.1128/AEM.03123-18

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1                   **Genetic diversity of Flavescence dorée phytoplasmas at vineyard scale**

2

3   **Running title:** FD phytoplasma diversity at vineyard scale

4

5   Marika Rossi<sup>a</sup>, Mattia Pegoraro<sup>a</sup>, Matteo Ripamonti<sup>b</sup>, Simona Abbà<sup>a</sup>, Dylan Beal<sup>b\*</sup>, Alessia Giraud<sup>a,b</sup>,

6   Flavio Veratti<sup>a</sup>, Sylvie Malembic-Maher<sup>c</sup>, Pascal Salar<sup>c</sup>, Domenico Bosco<sup>a,b</sup>, Cristina Marzachi<sup>a#</sup>

7

8

9   <sup>a</sup> Istituto per la Protezione Sostenibile delle Piante, CNR, Strada delle Cacce 73, 10135 Torino, Italy

10   <sup>b</sup> Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco (TO). Italy

11   <sup>c</sup> INRA, Univ. Bordeaux, UMR 1332 de Biologie du Fruit et Pathologie, CS20032, 33882, Villenave

12   d'Ornon Cedex, France

13

14   Corresponding author: cristina.marzachi@ipsp.cnr.it

15   \*Present Address: Department of Environmental Science, Policy, & Management, UC

16   Berkeley, Berkeley, CA 94720

17

18   **Keywords:** *Vitis vinifera*, *Scaphoideus titanus*, *Clematis*, *dnaK*, *vmpA*, *malG*

19

20

21

22

23

24 **ABSTRACT**

25 To study the role of wild areas around the vineyards in the epidemiology of Flavescence dorée (FD)  
26 and track origin of new foci, two phytoplasma genetic markers, *dnaK* and *malG*, were developed for  
27 FD phytoplasma characterization. The two genes and the *vmpA* locus were used to genetically  
28 characterize FDp populations at seven agroecosystems of a wine-growing Italian region. *Vitis vinifera*,  
29 “gone-wild” *V. vinifera* and rootstocks, *Clematis* spp. and *Scaphoideus titanus* adults were sampled  
30 within and outside the vineyards. A range of genotypes infecting the different hosts of the FDp  
31 epidemiological cycle was found. FD-C type isolates were fairly homogeneous compared to FD-D  
32 ones. Most of the FD-D variability was described by *malG* sequence, and a duplication of this locus  
33 was demonstrated for this strain. Co-infection with FD-C and FD-D strains was rare, suggesting  
34 possible competition between the two. Similar levels of FDp genetic variation, recorded for grapevines  
35 or leafhoppers of cultivated and wild areas and co-occurrence of many FDp genotypes inside and  
36 outside the vineyards supported the importance of wild or abandoned *Vitis* spp. plants and associated *S.*  
37 *titanus* in the epidemiology of the disease. Genetic profiles of FDp found in *Clematis* were never found  
38 in the other hosts, indicating that this species does not take part in the disease cycle in the area. Due to  
39 the robustness of *dnaK* for discriminating between FD-C and FD-D strains and the high variability of  
40 *malG* sequence, these are efficient markers to study FDp populations and epidemiology at a small  
41 geographical scale.

42

43 **IMPORTANCE**

44 Flavescence dorée, a threatening disease of grapevine caused by FD phytoplasma (FDp), is distributed  
45 within the most important wine producing areas of Europe and causes severe impacts on both vineyard  
46 productivity and landscape management. FDp is a quarantine pest in Europe, and despite the efforts to

2

47 contain the pathogen, the disease is still spreading. In this work, new genetic markers for the fine genetic  
48 characterization of FDp at local scale are presented. Our findings improve the knowledge of FDp  
49 epidemiological cycle and the possibility to track the route of the FDp infection. In particular, due to its  
50 high genetic variability, one of the newly developed markers could be sufficient to track origin of new  
51 infection foci, either from the wild areas or from nurseries.

52

### 53 INTRODUCTION

54 The causal agent of Flavescence dorée, FD phytoplasma (FDp), is transmitted to grapevines by the  
55 Deltocephalinae leafhopper *Scaphoideus titanus* Ball (1), which is almost monophagous on grapevine  
56 (2). Phytoplasmas are plant-pathogenic bacteria belonging to the class Mollicutes that invade the  
57 phloem sieve tube elements of the host plants and colonize the bodies of insect vectors. Phytoplasmas  
58 are transmitted by leafhoppers, planthoppers and psyllids and by vegetative propagation of infected  
59 plant material. Phytoplasmas are uncultivable and described under the provisional genus “*Candidatus*  
60 *Phytoplasma*” mainly based on 16S rRNA gene phylogeny. Bois Noir phytoplasma (BNp, Stolbur  
61 group, 16SrXII, ‘*Ca. P. solani*’ (3)) and FDp (16SrV) are associated with important phytoplasma  
62 diseases of grapevine (*Vitis vinifera* L.) (4). FDp transmission by the monovoltine leafhopper *S. titanus*  
63 is persistent and propagative (1, 2). The symptoms usually appear the year after the infection and  
64 consist of leaf yellowing or reddening and downward leaf curling, drying of inflorescences and  
65 bunches, and lack of cane lignification (5). Consequently, plant vitality, yields, and wine production are  
66 severely reduced (6). In Piedmont, the presence of high levels of inoculum and abundant vector  
67 populations in vineyards and surrounding areas have made control of the disease especially difficult.  
68 The identification of the ecological components of the FD epidemiological cycle could help in the

69 control of disease spread. A possible approach to evaluate disease dispersal patterns over spatial scales  
70 is through analyses of pathogen genetic markers (7). Based on sequence and restriction fragment length  
71 polymorphism (RFLP) analysis of the 16S rRNA and 16S-23S intergenic spacers, two FDp taxonomic  
72 groups were described: 16SV-C and 16SV-D (8). Moreover, sequencing of two non-ribosomal loci,  
73 *secY* and *rpsC*, allowed for the identification of three genetic clusters within FDp populations sampled  
74 in France and Italy (9). Arnaud and co-workers (2007) analyzed the sequences of two other genetic  
75 loci, *map* and *deg*, and confirmed the existence of three genetic clusters of FDp characterized by  
76 different geographical distribution and genetic variability: strain Cluster FD1, characterized by low  
77 genetic variability and high incidence in Southwestern France; strain Cluster FD2, including isolates  
78 FD-92 and FD-D, with no genetic variability and present both in France and Italy; and strain cluster  
79 FD3 comprising FD-C, showing high variability and found only in Italy (10). Both FD-C and FD-D  
80 phytoplasma are found in Piedmont (9). This genetic classification was not sufficiently accurate to  
81 describe the genetic variability present at a single agroecosystem. In fact, the low variability of the  
82 genetic loci considered until now represented a limit in the study of populations from a small  
83 geographical area. In the present study, two new genetic markers for FDp characterization, *dnaK* and  
84 *malG*, were developed. These two genes, together with the *vmpA* locus, already described by Renaudin  
85 and co-workers (11), were used to genetically characterize the FDp populations at seven  
86 agroecosystems in Piedmont. To identify the components of the FD epidemiological cycle and to study  
87 the epidemic flow of the disease, *V. vinifera*, *V. vinifera* and “wild” rootstocks (hybrids of *V. riparia*, *V.*  
88 *rupestris* and *V. berlandieri*) from abandoned vineyards, *Clematis* spp. (12) and *S. titanus* were  
89 sampled both within and outside the vineyards. Due to the robustness of *dnaK* for discriminating  
90 between FD-C and FD-D isolates as well as the high variability of *malG* sequence, these two markers  
91 provide a new tool to study FDp populations and epidemiology on a small geographical scale.

92

93 **RESULTS**94 **FDp diagnosis**

95 In order to study the role of wild areas around the vineyards in the epidemiology of Flavescence dorée  
96 in Piedmont (Italy), the following seven vineyards were selected within the main wine areas of the  
97 region: Cisterna d’Asti (CI), Castel Rocchero (CR), CREA-Asti (AS), La Morra (LM), Montà (MO),  
98 Paderna (PA), and Portacomaro (PC) (Figure S1). Samples of five categories, cultivated grapevines  
99 (VV), *S. titanus* from inside each vineyard (ST\_IN) and the wild areas bordering the vineyard edge  
100 (ST\_OUT), *Clematis* (CL) and wild or abandoned grapevines (WG) from the wild areas surrounding  
101 each vineyard were collected and tested for the presence of FDp by nested PCR driven by a 16SrV  
102 phytoplasma-specific primer pair. The distribution of FDp positive samples for each category at each  
103 site is reported in Table 1. In the case of CR, all the collected 19 wild or abandoned grapevines  
104 growing nearby were negative for FD presence. Therefore, three cultivated grapevines from  
105 neighboring vineyards were sampled as potential external sources of FD (VV\_OUT). Forty one WG  
106 (including the three cultivated grapevines sampled outside of the CR vineyard) out of 192 and 7 CL  
107 samples out of 32 were positive to FDp diagnosis. All FDp-infected samples were negative to the  
108 nested PCR assay aimed at detecting ‘*Ca. P. solani*’ (Bois noir) and ‘*Ca. P. asteris*’, phytoplasmas  
109 known to infect grapevine in the Piedmont region. At least six FDp positive samples for each category  
110 at each site were further characterized for their genetic variability on the selected target genes.

111 **Selection of candidate genes to characterize the genetic diversity of FDp**

112 The following 17 genes were selected on the basis of their difference in sequence identity (ranging  
113 from 87 to 100%) between FD-C (13) and FD-D (14) isolates: *dnaK*, *mntA*, *nrdF*, *malG*, *malF*, *map*,  
114 *rpoC*, *rpsE*, *rsmA*, *htmp1*, *htmp2*, *htmp3*, *htmp4*, *htmp5*, *loiD*, *glyA*, *vmpA*. PCR amplification of the

5

115 17 genes from total DNA of periwinkle-maintained FD-C (13) and FD-D (14) isolates provided  
116 amplicons of the expected sizes for 15 and 17 targets of FD-C and -D, respectively. In particular,  
117 primers designed on *htmp1* and 4 failed to amplify their targets from FD-C (Table 2). These two genes  
118 were excluded from further analyses. In preliminary experiments on FD-infected grapevine samples  
119 collected in 2013, amplification with specific primers designed on the remaining 15 genes provided  
120 amplicons of the expected sizes for most of the genes analyzed. Amplicons from *htmp1*, 2, 3, and 4  
121 genes were not efficiently amplified from most of the field samples and were, consequently, excluded  
122 from successive analysis (Table 2). All the obtained amplicons were sequenced in both directions. Upon  
123 sequencing, *rpoC* and *rsmA* amplicons provided a single genotype and due to their low sequence  
124 variability, they were excluded from further analysis. Among the tested target genes, only *malG*  
125 provided more than two genotypes and, due to this high sequence variability, it was selected for FDp  
126 characterization. Sequencing of the remaining target genes always provided 2 genotypes,  
127 corresponding to FD-C (13) and FD-D (14) reference isolates. Among these, *dnaK* was arbitrary  
128 selected for further studies. Gene *vmpA* was also chosen due the adhesion role of VmpA that could be  
129 essential in the colonization of the insect by FDp (15). These three genes were used for subsequent  
130 characterization of FDp from the different geographical sites.

### 131 **Genetic diversity of the selected target genes**

132 *sec-map*. Representative samples from the diverse host categories of each location were also  
133 characterized on the sequence of the *sec-map* locus to link the obtained results to existing literature. To  
134 provide a starting point for mapping FDp diversity in Piedmont, the genetic diversity of the *sec-map*  
135 locus was measured for the representative samples of each sampling site. Twenty nine samples  
136 representative of the seven vineyards were characterized for this locus following the protocol described  
137 by Arnaud *et al.* (2007) (10): 7 VV, 6 WG, 8 ST\_IN, 2 ST\_OUT, and 6 CL. Twenty one samples of

138 three categories (VV, WG, and ST) had M54 (AM384886; FD-D), and two samples (VV) showed M12  
139 (AM384896; FD-C) genotypes (Figure S2). *Clematis* samples had M50 (LT221945) and M51  
140 (LT221946) genotypes, and a third one identical to the FDp found in *Clematis* in Serbia (KJ911219).

141 ***dnaK* and *vmpA***. Amplification with primers *dnaK\_F/R* produced a specific amplicon of the expected  
142 size from 46 cultivated vines, 26 *Vitis* spp. plants from outside the vineyards, 42 *S. titanus* collected  
143 inside the vineyards, and 38 *S. titanus* from outside the vineyards, as well as the six *Clematis* spp from  
144 wild vegetation. One hundred and fifty-eight *dnaK* sequences were analyzed, and three *dnaK*  
145 genotypes were identified as *dnaK1*, 2 and 3. A *dnaK1* genotype was found in 116 isolates (73 %) and  
146 the reference isolate FD92 (FD-D), *dnaK2* genotype was obtained from 36 isolates (23 %) and the  
147 reference isolate FD-C, *dnaK3* genotype was obtained from the six *Clematis* spp. (Figure 1). Genotype  
148 *dnaK2* differed by three SNPs from *dnaK1* at positions 624, 888 and 969, and by one SNP from *dnaK3*  
149 at position 789. All the mutations were synonymous (Table 3). Mixed infections were evident from the  
150 chromatograms of seven samples (3 VV, 2 WG, and 2 ST, not shown), from AS, PA, and PC, so PCR  
151 amplicons were cloned in the plasmid vector pGEM-T (Promega, Madison, WI) for further  
152 investigations. Sequencing of five clones for each of the seven samples confirmed double infection  
153 with *dnaK1* and *dnaK2* isolates for all samples. The incidence of *dnaK1* was higher than that of *dnaK2*  
154 for all analyzed sample categories ( $\chi^2$ ,  $p=0.025$ ). Interestingly, the *dnaK1/dnaK2* frequency ratio for  
155 VV (1.4) was lower than for WG (3.8), ST\_IN (3.6) and ST\_OUT (8.3) (not shown).

156 The phytoplasma variable membrane protein *VmpA* gene is characterized by a stretch of 234 nt  
157 repeated sequences (R) (11, 16). Insertion/deletion of one repeat sequence determines size variability of  
158 the gene. Amplification with primers *VmpAF3/VmpAR* yielded two possible amplicons of 1.488 bp  
159 (A) and 1.254 bp (B) respectively. Sequencing of R1 repeats from 158 samples identified two R1  
160 genotypes (R1\_1, R1\_2) from all cultivated *V. vinifera*, wild *Vitis* spp. plants, and *S. titanus* from

161 inside and outside the vineyards (Figure 1). Sequencing of the R1 repeat from the *Clematis* spp.  
162 samples yielded two profiles, named R1\_3 and R1\_4 (Figure 1). In summary, taking together  
163 information from size polymorphism of the amplicon (A vs B) and R1 sequence profiles (R1\_1 to  
164 R1\_4), two and 108 of the 158 sequenced samples showed R1\_1A and \_1B profiles, 34 and two  
165 showed R1\_2A and \_2B profiles. As for R1\_3 and \_4 profiles, these only showed amplicons of the A  
166 type (Table 4). Isolates from two samples showed mixed profiles, and four were not amplified under  
167 our experimental conditions. The *vmpA* R1 profiles of the reference isolates FD-C and FD-D were  
168 R1\_2A and \_1B, respectively. The four R1 genotypes differed in their sequences at 26 sites (Figure  
169 S3A) some of which corresponded to non-silent mutations (Figure S3B). Profiles corresponding to  
170 multiple infections were absent upon analyses of the R1 repeats of the *vmpA* gene except for a WG and  
171 a ST\_OUT both collected in PA. This plant showed a mixed profile also for the *vmpA* amplicon size (A  
172 and B) and *dnaK* genotype (*dnaK1* and *dnaK2*). The ST\_OUT showing both *vmpA* amplicon types had  
173 a R1\_2 profile associated with *dnaK* mixed profile. When the length of *vmpA* gene was considered  
174 together with R1 sequence and *dnaK* genotypes, 7 types of *dnaK-vmpA* profiles were detected and  
175 listed in Table 4.

176 **malG.** Partial sequencing of *malG* gene detected a mix of divergent sequences in 108 of the 158  
177 analyzed samples. The *malG* PCR products of these isolates were cloned and 3 to 5 clones for each  
178 sample were sequenced. Four hundred and forty-seven sequences were then analyzed and 183  
179 genotypes were detected (Figure S4). To simplify the successive analyses at each location, the 183  
180 identified genotypes were manually checked and genotypes with SNPs that were present less than three  
181 times were grouped into the closest node (Figures 2, 3, S3, and S5). This procedure did not alter the  
182 overall picture of the Median-joining networks (Figure 3 and S3), and provided enough sensitivity to  
183 cope with the sampling size strategy of the experiment. After this procedure, 50 *malG* genotypes were

184 left in the list, *malG1* identical to *malG* of the FD-D isolate (*dnaK1* profile), *malG2* identical to *malG*  
185 of the FD-C isolate (*dnaK2* profile) and *malG3* (*dnaK1* profile) being the predominant ones (28,2 %, 30,4 %, 9,4 % of the 447 analyzed sequences, respectively). *malG38* and *malG39* were identified only  
186 in *Clematis* spp. (*dnaK3*, *sec-map*: M50/M51/ KJ911219; Figure 2); *malG38* was always associated  
187 with *vmpA\_R1\_3A*, and *malG39* with *vmpA\_R1\_4A* (Figure 2). Most of the samples (78), showed  
188 multiple profiles, predominantly *malG1* and *malG3*, and among these, 43 samples had more than two  
189 *malG* types. Profiles *malG1* and *malG3* were found 7 and 11 times as pure profiles, respectively. Type  
190 *malG2* was found with other *malG* genotypes in four samples, of which three (1 VV, 1 WG, and 1  
191 St\_OUT) had FD-C/FD-D mixed infections and one (WG) had just the *dnaK1* (FD-D) profile.  
192 Overall, the frequencies of *malG1* and *malG3*, both associated to *dnaK1* profile, were similar (around  
193 30 %) suggesting a possible gene duplication of this gene locus. Southern blot confirmed the  
194 duplication of *malG* gene at least in the FD92 chromosome (Figure 4, Panel A). Copy-specific PCR  
195 amplification confirmed the duplication (Figure 4, Panel B), as primer pair *malGtestF* / *malGtestR2* and  
196 *malGtestF* / *malGtestR5* were able to amplify both *malG* copies from total DNA of FD-D reference  
197 isolate. These primers were designed on the sequences of the contigs 002 and 005 of FD92 draft  
198 genome, which included two identical *malG1* sequences. Primer *malGtestF* was designed on the *malG*  
199 coding sequence whereas *malGtestR2* and *malGtestR5* were designed to align downstream of the  
200 identical region, so that they could amplify specifically contig 002 and 005 sequences, respectively. A  
201 single copy of *malG* (in the context of contig 002) was detected in the chromosome of the FD-C isolate  
202 (Figure 4).  
203 Overall, 32 *malG* types were found in insects, 25 of which were present with less than 2 % frequency  
204 (Figure 2). Thirteen *malG* genotypes were detected in cultivated grapevines and wild growing *Vitis* spp.  
205

206 plants, suggesting that the genetic variability of the phytoplasma was lower in the plants than in  
207 vectors.

208 The AS and PC sites, with 15 *malG* types each, showed the highest genetic variability, which was  
209 mainly determined by insect and wild grapevine isolates (Figure 2). Six genotypes were identified at  
210 PA, which was the site with the lowest variability. This vineyard was also characterized by the  
211 prevalence of FD-C (*dnaK2*, *malG2*, *vmpA2A*) both inside and outside the vineyard.

212 The DNA sequence identity among the *malG* genotypes ranged between 99.7 % and 97.3 %. Among  
213 the three most frequent genotypes, only the SNP at position 380 (G to A) determined a non-  
214 synonymous substitution (V127I) (Figure S6). This was the most frequent mutation and determined a  
215 clear distinction between *malG1* and *malG3* clusters (Figure S6). A second mutation at position 332,  
216 determining a valine to leucine substitution, characterized *malG38* and explained its relationship with  
217 *malG103* and *malG126* genotypes (Figure 3 and Figure S6) found only in WG category at PC. The  
218 silent mutation T/C at position 629 was found only in isolates from CI.

219 The Median-joining network analysis of the 50 *malG* genotypes (Figure 4) identified seven main nodes  
220 (based on sequence and frequency): *malG1*, *malG2*, *malG3*, *malG6*, *malG16*, *malG18*, *malG34*. In  
221 particular, most of the genetic diversity was linked to *malG1* and 3, especially *malG1*. Eighteen minor  
222 genotypes were linked to *malG1*, 15 of which were found only in insects. Eleven minor genotypes were  
223 linked to *malG3*, and seven of those were found only in insects. Two genotypes were directly linked to  
224 *malG2* (141 and 146). Genotypes *malG103* and *malG126* were found only in the WG category. These  
225 two genotypes were linked to *malG38* of CL, with no direct link to *malG* genotypes of other sample  
226 categories.

227

228 **Genetic variability of FDp at different sampling sites**

229 The distribution and frequency of the most represented *malG* genotypes are reported in Figure 5. At all  
230 locations, several FDp variants were detected, with the exception of PA, where the FDp infection  
231 showed low variability. FDp from *Clematis* spp. were never found in other plant or insect hosts. At  
232 most locations, more than three genetic FDp variants were detected within each sample category, but at  
233 PA only one and two variants were detected from cultivated and wild grapevines, respectively. At each  
234 location, except CR and PC, most genetic variants of FDp were detected in the vectors, both within the  
235 vineyard and the nearby wilderness. At each site, the most frequent FDp genotypes were present in all  
236 categories except *Clematis*, although the frequency could differ. Indeed, at each location where its  
237 presence was recorded, *malG2* genotypes were more frequently detected from plant hosts than vectors,  
238 and from cultivated grapevines than wild *Vitis* spp. plants. Due to the lack of *dnaK2* profiles at CI and  
239 CR, only *malG* analysis was able to describe FDp variability at these locations which were comparable  
240 to the variability of the other sampling sites (Figure 2). At the remaining locations, *dnaK1* and *dnaK2*  
241 profiles were present both inside and outside the vineyards, irrespective of the sample categories, with  
242 the exception of FDp isolates from the *Clematis* samples, which all showed *dnaK3* profiles. *Clematis*,  
243 instead, differed in terms of *VmpA* profile: R1\_3A was obtained from all *Clematis* spp samples from  
244 the LM site, and R1\_4A was obtained from the CR and PC samples.

245

## 246 **DISCUSSION**

247 FD presence in Northwestern Italy dates back to 1998 (14), and despite intense control efforts, the  
248 disease has spread to the most important viticultural areas of Piedmont. Since the beginning of the  
249 epidemics, FD-C was the prevalent strain, whereas the incidence of FD-D was occasional (9). In this  
250 study, a protocol was developed to decipher the genetic variability of FDp strains involved in the  
251 epidemics of disease in Northwestern Italy. A previously developed genotyping protocol, based on the

252 sequence of *sec-map* locus, clearly identified three lineages among FDp isolates from Italy and France  
253 (10). The same gene was used also in this study. Genotype M54 was predominant in the seven selected  
254 vineyards of the Piedmont region and, together with M12, represented most of the genetic diversity of  
255 FDp isolates from vines and insects at the seven sampling sites. Therefore, the genetic variability  
256 associated to this locus was not enough to provide a detailed molecular typing of FDp at the vineyard  
257 scale. Genotyping based on the three newly selected genes identified different degrees of genetic  
258 variabilities of FDp in the seven vineyards. In particular, FDp was less variable at PA than at the other  
259 sites. Interestingly, this vineyard is located in the area where the first epidemics of FD were spotted in  
260 1998 in the Region, and it is geographically isolated from the other six, which, in contrast, form a  
261 continuous vineyard landscape. In addition, although characterized by different genetic resolution  
262 powers, *dnaK*, *vmpA*, and *malG* always provided consistent results about the presence of mixed  
263 infections. Like *sec-map*, each of them, could, in fact, detect the presence of both FD-C and -D in some  
264 of the analyzed samples.

265 According to our genotype analysis, FD-C type isolates were fairly homogeneous while FD-D types  
266 were highly variable. Most of the observed FD-D type variability was described by *malG* sequence.  
267 Indeed a duplication of this locus has occurred, as demonstrated by the analysis of the FD92 draft  
268 genome (14) as well as by the results of the PCR and Southern blot with contig-specific reagents. Even  
269 if *malG2* type of FD-C showed no variability at all, our results cannot exclude the presence of a *malG*  
270 duplication also in the FD-C genome, as possible mismatches on the sequence of the contig 5-specific  
271 reverse primer could also explain failure of amplification of the *malG* copy in this context. Poor quality  
272 of FD-C draft genome (13) in that region does not support any of the hypotheses. Yet, *malG* operon is  
273 present as a single copy in most bacterial genomes, and, in particular, in many phytoplasma genomes  
274 such as ‘*Ca. P. asteris*’, ‘*Ca. P. mali*’, and ‘*Ca. P. australiense*’. The *malG1* and *malG3* genotypes

275 showed the highest variability; interestingly, 15 and 7 of the 18 *malG1* and 11 *malG3* were found only  
276 in insects, indicating that most of the variability was detected in this host. All *malG1* and 3-derived  
277 types were associated to a *dnaK1* profile, and again, this was more abundant in insects, compared to  
278 vines, especially cultivated ones. In contrast, the unique *malG2* profile was associated only with the  
279 *dnaK2* type, which is more abundant in plants, in particular in cultivated grapevines, than in insects.  
280 Infections with different *dnaK* profiles were rare, especially in vectors, indicating a possible  
281 antagonism between the two *dnaK* genotypes, which is an issue currently under investigation. Such  
282 large genetic variability of FDp described on a very restricted geographical scale, supports the  
283 hypothesis of a European origin of the phytoplasma (10). Interestingly, similar levels of FDp genetic  
284 variation were recorded in the cultivated and wild areas for either grapevines or leafhoppers. This  
285 finding, together with the co-occurrence of many FDp genotypes inside and outside the vineyards,  
286 confirms the importance of wild or abandoned *Vitis* spp. plants and associated *S. titanus* in the  
287 epidemiology of the disease. The large overlap of FDp genotypes in the two environments is confirmed  
288 at the single site level, with the partial exception of the PA site. In this site, very few *S. titanus* from the  
289 wild area were found infected, so epidemiology of the disease at this site, should be mainly explained  
290 by within-vineyard spread (“secondary infection”). At the other sites, FDp genetic diversity was  
291 consistent with the hypothesis of “primary infection” by incoming vectors from outside the vineyard.  
292 This hypothesis is further supported by the “edge” effect recorded for FDp-infected grapevines (18–  
293 20). Primary infections are likely to occur late in the season, when cultivated grapevines are no longer  
294 protected by insecticides, due to the need of respecting a safety period before grape harvest (21). Data  
295 on *S. titanus* dispersal capability (22) indicate that 80% of adults do not fly beyond 30 m, although few  
296 can move up to 300 m. In our study, wild or abandoned grapevines and associated *S. titanus* adults  
297 were always collected within this distance.

298 Lack of transovarial transmission of FDp in *S. titanus*, implies that insects must acquire FDp from  
299 plants. The presence of about 50 % FDp types only in the insects can be explained by latent infection of  
300 FDp types in the plant that are able to multiply efficiently only in the vector body, and /or by high  
301 variation rate of the FDp population within the vector that, being persistently infected and hosting an  
302 active multiplication of FDp, might act a strong selection pressure towards these phytoplasmas. Lack of  
303 identification of some FDp types in the plants could be due to insufficient plant sample sizes, to  
304 inefficient multiplication of some FDp types in the plant host, and/or inefficient transmission of some  
305 FDp genotypes. Although other plant species, besides *Vitis* spp. ones, are known reservoir hosts of  
306 FDp, the genotypes of the phytoplasma identified in *Clematis* sp. in the investigated areas, were  
307 different from those infecting grapevine and *S. titanus*. Actually, *sec-map* types identified in FDp from  
308 *Clematis* in our study were consistent with those described in Malembic-Maher et al. (23), M50 and  
309 M51, and these latter were never found in grapevines and vectors. Therefore, we can conclude that,  
310 even if *Dictyophara europea* plant hoppers can occasionally transmit *Clematis* phytoplasmas to  
311 grapevine (24), the frequency of such transfer is negligible in the investigated areas. Alder (*Alnus* spp.)  
312 and *Ailanthus altissima* are also known hosts, and potential reservoirs of FDp (25, but they were absent  
313 in the vicinity of the analyzed vineyards.

314 The newly developed protocol, based on the analyses of three loci of the FDp chromosome, provided  
315 enough sensitivity to describe the genetic population structure at the vineyard level and assess the  
316 composition of FDp population within the cultivated and wild areas of seven geographic locations.  
317 These results also highlight the importance of both *Vitis* spp. plants and *S. titanus* populations of the  
318 uncultivated areas nearby productive vineyards in the epidemiology of the disease, in the analyzed  
319 areas. In particular, a direct consequence of these results would support the urgent need of an effort  
320 aimed at controlling both vectors and *Vitis* spp. plants of areas surrounding productive vineyards, at

321 least by plant eradication to reduce FDp reservoir within vector flying distance from cultivated  
322 grapevines. Moreover, due to its high genetic variability, *malG* can be applied to track origin of new  
323 infection foci, either from the wild area or from nursery. Actually it is worth noting that, in the  
324 presence of the vector, spread of the disease in previously uninfected areas, can be either due to the  
325 introduction of infected plant material, or to the transfer of FDp phytoplasmas already present in the  
326 wild area into the vineyard. According to EFSA (2016) (25), it is likely that emergence of FDp from  
327 the wild reservoir has occurred in some European region.

328

## 329 MATERIALS AND METHODS

### 330 Vineyard selection

331 For the selection of the sampling sites the following criteria were adopted: 1) presence of actively  
332 cultivated *Vitis vinifera* with plants positive for FDp; 2) presence of the FDp vector *Scaphoideus*  
333 *titanus* 3) presence of potential alternative host plants (*Vitis vinifera*, *V. riparia* and hybrids of different  
334 *Vitis* species, and *Clematis vitalba*) for FDp along plot edges and surrounding landscape. Following  
335 these guidelines, seven sites across the Piedmont region were selected. The sites were named after the  
336 villages closest to them: CR, AS, CI, LM, MO, PA, and PC (Figure S1).

### 337 Plants, insects and phytoplasma reference isolates

338 Total DNA extracts from FD-infected grapevines sampled in 2013 at representative sites in Piedmont  
339 were used for the initial selection of the best candidate genes to characterize the genetic diversity of  
340 FDp. For the detailed study of FDp diversity at selected sites, grapevines showing FD symptoms (VV)  
341 were collected at each vineyard described above during July and August of 2014 and 2015.  
342 Representative samples from wild grapevines (including *V. vinifera* and rootstocks, hybrids of *V.*  
343 *riparia*, *V. rupestris*, and *V. berlandieri* from abandoned vineyards, and hybrids from different *Vitis*

344 species, WG), as well as *C.vitalba* (CL) were collected in the wild areas around each vineyard site,  
345 whenever present. At each site, both asymptomatic and yellows-showing samples were collected,  
346 aiming at testing all potential sources of FDp, irrespective of the expressed symptomatology. Adult *S.*  
347 *titanus* individuals (ST) were detached from the yellow sticky traps placed inside (ST\_IN) and outside  
348 (ST\_OUT) each vineyard. At one of the sites (CR), no FDp-infected wild grapevines were found, so  
349 symptomatic, cultivated *V. vinifera* from adjacent/neighboring vineyards (VV\_OUT) were collected  
350 instead.

351 Phytoplasma reference isolates FD92 (FD-D) (26) and FD Piedmont (FD-C) (13) were maintained in  
352 *Catharanthus roseus* by grafting of infected scions at the Institute of Sustainable Plant Protection  
353 collection (Torino, Italy).

#### 354 **Total DNA extraction and FDp diagnosis**

355 Total nucleic acids were extracted from 1 g of leaf midribs and petioles and from single leafhoppers  
356 according to the method of Pelletier (27, 28). Total DNA extracts from plants and insects were then  
357 suspended in 100  $\mu$ L or 75  $\mu$ L of Tris-HCl 10mM pH 8, respectively. DNA concentration was  
358 measured with NanoDrop 2000<sup>TM</sup> Spectrophotometer (Thermo Scientific, Waltham, MA), and all  
359 samples were then diluted to 20 ng/ $\mu$ L.

360 To confirm the presence of FDp in single infection, 40 ng of each DNA extract was used in direct PCR  
361 assays with the universal ribosomal primers P1/P7 (29, 30), followed by nested PCRs with the group  
362 specific ribosomal primers R16(V)F1/R1 (30), as well as R16(I)F1/R1 (30). Samples with FD and BN  
363 mixed infections were excluded from the analysis. PCR conditions were as described by Lee et al.  
364 (1994) (30). Taq DNA polymerase (1 U) (Polymed) was used in each assay. PCR products were  
365 analyzed by electrophoresis through 1% agarose gel in 1 $\times$  Tris-borate-EDTA (TBE) buffer along with a

366 1-kb-plus DNA size marker (Gibco BRL). Gels were stained with ethidium bromide and visualized on  
367 a UV transilluminator.

368 **Selection of candidate FDp genes, primer design, cloning, transformation and sequencing**

369 Seventeen genes (Table 2) were selected on the basis of their difference in sequence identity (ranging  
370 from 87 to 100%) between FD-C (13) and FD-D (14) isolates determined by Blastn. Specific primers  
371 able to amplify both FD-C and FD-D genes were designed. PCR was carried out in 30  $\mu$ l reactions.  
372 Each reaction contained 0.3 U of the proofreading DyNAzyme EXT DNA Polymerase (Thermo  
373 Scientific, Waltham, MA). The cycling conditions were set as follow: 2 min at 94°C and 35 cycles with  
374 1 cycle consisting of 30 sec at 94°C, 30 sec at 55°C and 40 sec at 72°C followed by a final extension of  
375 5 min at 72°C. Obtained amplicons were sequenced in both directions as detailed below. For  
376 sequencing purposes, portions of the genes were amplified by PCR with the corresponding primers as  
377 listed in (Table 2) and sequenced as detailed below. To determine *vmpA* gene size, PCRs were  
378 performed using primers vmpAF3/R (Table 2) in a 30  $\mu$ l reaction solution at the following cycling  
379 conditions: 2 min at 94°C and 35 cycles with 1 cycle consisting of 30 sec at 94°C, 30 sec at 52°C and 1  
380 min and 30 sec at 68°C followed by a final extension of 5 min at 68°C. PCR products (5  $\mu$ l) were  
381 loaded on a 1% agarose gel in TBE buffer using FD-C and FD-D *vmpA* amplicons as size reference. To  
382 determine the sequence of *vmpA* R1 repeat, a PCR with the primers VmpAF5/R3 (Table 2) was  
383 performed at the following conditions: 2 min at 94°C and 35 cycles with 1 cycle consisting of 30 sec at  
384 94°C, 30 sec at 56°C and 30 sec at 66°C followed by a final extension of 5 min at 66°C. Direct PCR  
385 products (1  $\mu$ l) were used as templates for nested PCR with primers VmpAF8/R9 at the following  
386 cycling conditions: 2 min at 94°C and 35 cycles with 1 cycle consisting of 30 sec at 94°C, 30 sec at  
387 50°C and 30 sec at 66°C followed by a final extension of 5 min at 66°C. Nested PCR products were

388 purified using the DNA Clean and Concentrator kit<sup>TM</sup> (Zymo Research, Irvine, CA) and sequenced as  
389 detailed below with primer VmpAF3.

390 In case of mixed infections (presence of double peaks in the analyzed pherograms from sequencing of  
391 the original PCR amplicon), purified PCR products were ligated into pGEM-T easy cloning vector  
392 following the manufacturer's instructions (pGEM-T- clone kit, Promega, Madison, WI) and  
393 transformed into *E. coli* DH5 $\alpha$  competent cells by heat shock. Positive colonies were selected by  
394 blue/white screening followed by colony PCR using M13F/R primers at the following conditions: 5  
395 min at 95°C and 35 cycles with 1 cycle consisting of 60 sec at 95°C, 60 sec at 51°C and 1min and 20  
396 sec at 72°C followed by a final extension of 5 min at 72°C. Recombinant plasmids were extracted  
397 using the Wizard SV Plus Minipreps DNA Purification System (Promega, Madison, WI). Purified  
398 plasmids were sent for sequencing (Macrogen, Seoul, South Korea) with appropriate primers for each  
399 target gene (Table 2). Each sequence had a 2x coverage. The partial sequences of *dnaK*, *vmpA* (R1  
400 repeat), and *malG* were deposited in NCBI with the following submission ID: 2121943. Partial  
401 sequences of the remaining genes were deposited in NCBI with the following submission ID: 216155.

#### 402 **Sequence analysis**

403 Raw sequences were trimmed of the unwanted 5' and 3' fragments generally characterized by low  
404 sequence quality using BioEdit (31), before further analyses. The reading frames of the sequences were  
405 maintained. Sequences from the same gene were aligned with MEGA7 (32), and the MUSCLE  
406 algorithm (33) was used for sequence alignments. In the case of *malG*, only the Parsimony Informative  
407 Sites present at least three times upon sequencing of all samples and cloned plasmids were considered  
408 as significant to define a new genotype. The other mutations were at first, analyzed with a Median  
409 joining network analysis, as detailed below, then they were corrected according to closest node. This  
410 procedure underestimated *malG* variability, but did not hamper characterization of FD variability at the

411 required geographical scale (Figure S1). To study the correlations among the different genotypes,  
412 Median joining network analysis was used. The median joining method to construct networks is more  
413 used in case of intraspecific data than other phylogenetic analysis (34).

#### 414 **Southern hybridization**

415 Southern hybridization was performed following standard procedures (35) using the DIG labeling and  
416 detection system (Roche, Basel, Switzerland). Briefly, genomic DNA from *C. roseus* infected with FD-  
417 C and FD-D reference strains were digested with 30U *EcoRI* (Invitrogen, Carlsbad, CA) for overnight  
418 at 37 °C, then electrophoresed through 1 % (w/v) agarose gel, depurinated and denatured in denaturing  
419 solution for 30 min. The gel was then neutralized in neutralizing solution for 30 min, and DNA was  
420 transferred to a positively charged nylon membrane (Roche, Basel, Switzerland) by capillary action  
421 overnight using 10x SSC solution. The transferred DNA was fixed to the membrane by UV-irradiation.  
422 The membrane was pre-hybridized in 10 ml hybridization buffer (5x SSC, 0.1 % N-lauroylsarcosine  
423 (w/v), 0.02 % SDS (w/v), 1 % blocking solution (Roche, Basel, Switzerland), 600 µg salmon sperm  
424 DNA) for 4 h at 65 °C. The DIG probe was synthesized using the PCR DIG Probe Synthesis Kit  
425 following the manufacturer's instructions (Roche, Basel, Switzerland). pGEM-T-*malG1* plasmid was  
426 used as template for the amplification with the *MalG\_F/R* primers pair. The DIG probe was diluted at  
427 25ng/ml in 5ml of hybridization buffer, denatured by boiling (10 min) and incubated with the  
428 membrane overnight at 65 °C. The membrane was washed twice with 2x SSC, 0.1 % SDS for 5 mins at  
429 RT, and twice in 0.5x SSC, 0.1 % SDS for 15 min at 65 °C. The hybridized probe was then detected  
430 using anti-Digoxigenin antibody (Roche, Basel, Switzerland) using CSPD as the chemiluminescent  
431 substrate according to the manufacturer's instructions. The blot was then visualized by exposing an  
432 autoradiographic film to chemiluminescence.

#### 433 **Accession numbers**

434 Sequence data were submitted to GenBank and the following accession numbers were provided:  
435 MH547710 - 12 (*dnaK1* - 3); MH547713 - 47 (*malG1* - 35); MH547748 - 93 (*malG38* - 183);  
436 MH547894 - 96 (*vmpA\_R1\_1* - 3); MK091396-97 (*mntA1*-2); MK091398-99 (*nrdF1*-2); MK091400-  
437 01 (*malF1*-2); MK091402-3 (*map1*-2); MK091404 (*rpoC1*); MK091405-6 (*rpsE1*-2); MK091407  
438 (*rsmA1*); MK091408 (*htmp1\_1*); MK091409-10 (*htmp2\_1-2htmp*); MK091411-12 (*htmp3\_1-2*)  
439 MK091413 (*htmp4\_1*); MK091414-15 (*htmp5\_1-2*); MK091416-17 (*lolD1*-2); MK091418-19 (*glyA1*-  
440 2) .

441

#### 442 **ACKNOWLEDGEMENTS**

443 This work was part of the ‘INTEFLAVI’ project funded by Fondazione Cassa di Risparmio di Cuneo,  
444 Fondazione Cassa di Risparmio di Torino, and Fondazione Cassa di Risparmio di Asti. MR and MP  
445 were supported by fellowships funded by the above cited grant-making foundations. CM and DB were  
446 partially supported by CNR grants on the Short Term Mobility 2015 and 2016 programs, respectively.  
447 A special thank to Dr. Loretta Panero (CREA-VE) for her support in sample collection. The funders  
448 had no role in study design, data collection and interpretation, or the decision to submit the work for  
449 publication.

450

#### 451 **REFERENCES**

- 452 1. Schvester D, Carle P, Moutous G. 1967. Testing susceptibility of vine varieties to (Flavescence  
453 doree) by means of inoculation by *Scaphoideus littoralis* Ball. *Annales Des Epiphyties* 18:143.
- 454 2. Chuche J, Thiery D. 2014. Biology and ecology of the Flavescence doree vector *Scaphoideus*  
455 *titanus*: a review. *Agron Sustain Dev* 34:381–403.

- 456 3. Quaglino F, Zhao Y, Casati P, Bulgari D, Quaglino F, Wei W, Quaglino F. 2013. “Candidatus  
457 Phytoplasma solani”, a novel taxon associated with stolbur- and bois noir-related diseases of plants. *Int*  
458 *J Syst Evol Microbiol* 63:2879–2894.
- 459 4. Osler R, Carraro L, Loi N, Refatti E. 1993. Symptom expression and disease occurrence of a  
460 yellows disease of grapevine in Northeastern Italy. *Plant Dis* 77:496–498.
- 461 5. Caudwell A. 1990. Epidemiology and characterization of Flavescence-doree (FD) And other  
462 grapevine yellows. *Agronomie* 10:655–663.
- 463 6. Morone C, Boveri M, Giosue S, Gotta P, Rossi V, Scapin I, Marzachi C. 2007. Epidemiology  
464 of flavescence doree in vineyards in Northwestern Italy. *Phytopathology* 97:1422–1427.
- 465 7. Coletta HD, Francisco CS, Almeida RPP. 2014. Temporal and Spatial Scaling of the Genetic  
466 Structure of a Vector-Borne Plant Pathogen. *Phytopathology* 104:120–125.
- 467 8. Martini M, Murari E, Mori N, Bertaccini A. 1999. Identification and epidemic distribution of  
468 two flavescence doree-related phytoplasmas in Veneto (Italy). *Plant Dis* 83:925–930.
- 469 9. Martini M, Botti S, Marcone C, Marzachi C, Casati P, Bianco PA, Benedetti R, Bertaccini A.  
470 2002. Genetic variability among flavescence doree phytoplasmas from different origins in Italy and  
471 France. *Mol Cell Probes* 16:197–208.
- 472 10. Arnaud G, Malembic-Maher S, Salar P, Bonnet P, Maixner M, Marcone C, Boudon-Padieu E,  
473 Foissac X. 2007. Multilocus sequence typing confirms the close genetic interrelatedness of three  
474 distinct flavescence doree phytoplasma strain clusters and group 16SrV phytoplasmas infecting  
475 grapevine and alder in Europe. *Appl Environ Microbiol* 73:4001–4010.
- 476 11. Renaudin J, Beven L, Batailler B, Duret S, Desque D, Arricau-Bouvery N, Malembic-Maher S,  
477 Foissac X. 2015. Heterologous expression and processing of the flavescence doree phytoplasma  
478 variable membrane protein VmpA in *Spiroplasma citri*. *BMC Microbiol* 15:12.

- 479 12. Angelini E, Squizzato F, Lucchetta G, Borgo M 2004. Detection of a phytoplasma associated  
480 with grapevine Flavescence dorée in Clematis vitalba. *Eur J Plant Pathol* 110:193-201.
- 481 13. Firrao G, Martini M, Ermacora P, Loi N, Torelli E, Foissac X, Carle P, Kirkpatrick BC, Liefing  
482 L, Schneider B, Marzachi C, Palmano S. 2013. Genome wide sequence analysis grants unbiased  
483 definition of species boundaries in “Candidatus Phytoplasma.” *Syst Appl Microbiol* 36:539–548.
- 484 14. Carle P, Malembic-Maher S, Arricau-Bouvery N, Desque D, Eveillard S, Carrere S, Foissac X.  
485 2011. “Flavescence doree” phytoplasma genome: a metabolism oriented towards glycolysis and protein  
486 degradation. *Bull Insectology* 64:S13–S14.
- 487 15. Arricau-Bouvery N, Duret S, Dubrana M-P, Batailler B, Desque D, Beven L, Danet J-L,  
488 Monticone M, Bosco D, Malembic-Maher S, Foissac X. 2018. Variable membrane protein A of  
489 flavescence doree phytoplasma binds the midgut perimicrovillar membrane of *Euscelidius variegatus*  
490 and promotes adhesion to its epithelial cells. *Applied and environmental microbiology*.
- 491 16. Cimerman A, Pacifico D, Salar P, Marzachi C, Foissac X. 2009. Striking Diversity of vmp1, a  
492 Variable Gene Encoding a Putative Membrane Protein of the Stolbur Phytoplasma. *Appl Environ*  
493 *Microbiol* 75:2951–2957.
- 494 17. Marzachi C, Boarino A, Vischi A, Palermo S, Morone C, Loria A, Boccoardo G. 2001.  
495 Flavescenza dorata, legno nero e giallume dell’astro in vitigni del Piemonte sud orientale. *Informatore*  
496 *Fitopatologico*, 9, pp. 58-63.
- 497 18. Pavan F, Mori N, Bigot G, Zandigiacomo P. 2012. Border effect in spatial distribution of  
498 Flavescence doree affected grapevines and outside source of *Scaphoideus titanus* vectors. *Bull*  
499 *Insectology* 65:281–290.
- 500 19. Maggi F, Marzachi C, Bosco D. 2013. A Stage-Structured Model of *Scaphoideus titanus* in  
501 Vineyards. *Environ Entomol* 42:181–193.

- 502 20. Maggi F, Bosco D, Galetto L, Palmano S, Marzachi C. 2017. Space-Time Point Pattern  
503 Analysis of Flavescence Doree Epidemic in a Grapevine Field: Disease Progression and Recovery.  
504 *Front Plant Sci* 7.
- 505 21. Bosco D, Mori N. 2013. Flavescence dorée control in Italy. *Phytopathogenic Mollicutes*, 3(1),  
506 pp 40-43.
- 507 22. Lessio F, Tota F, Alma A. 2014. Tracking the dispersion of *Scaphoideus titanus* Ball  
508 (Hemiptera: Cicadellidae) from wild to cultivated grapevine: use of a novel mark-capture technique.  
509 *Bulletin of Entomological Research* 104:432–443.
- 510 23. Malembic-Maher S, Salar P, Filippin L, Carle P, Angelini E, Foissac X. 2011. Genetic diversity  
511 of European phytoplasmas of the 16SrV taxonomic group and proposal of “*Candidatus Phytoplasma*  
512 *rubi*.” *Int J Syst Evol Microbiol* 61:2129–2134.
- 513 24. Filippin L, Jovic J, Cvrkovic T, Forte V, Clair D, Tosevski I, Boudon-Padieu E, Borgo M,  
514 Angelini E. 2009. Molecular characteristics of phytoplasmas associated with Flavescence doree in  
515 clematis and grapevine and preliminary results on the role of *Dictyophara europaea* as a vector. *Plant*  
516 *Pathol* 58:826–837.
- 517 25. Jeger M, Bragard C, Caffier D, Candresse T, Chatzivassiliou E, Dehnen-Schmutz K, Gilioli G,  
518 Miret JAJ, MacLeod A, Navarro MN, Niere B, Parnell S, Potting R, Rafoss T, Rossi V, Urek G, Van  
519 Bruggen A, Van Der Werf W, West J, Winter S, Bosco D, Foissac X, Strauss G, Hollo G, Mosbach-  
520 Schulz O, Gregoire JC, PLH EPPH. 2016. Risk to plant health of Flavescence doree for the EU  
521 territory. *Efsa Journal* 14.
- 522 26. Malembic-Maher S, Constable F, Cimerman A, Arnaud G, Carle P, Foissac X, Boudon-Padieu  
523 E. 2008. A chromosome map of the Flavescence doree phytoplasma. *Microbiology-(UK)* 154:1454–  
524 1463.

- 525 27. Pelletier C, Salar P, Gillet J, Cloquemin G, Very P, Foissac X, Malembic-Maher S. 2009.  
526 Triplex real-time PCR assay for sensitive and simultaneous detection of grapevine phytoplasmas of the  
527 16SrV and 16SrXII-A groups with an endogenous analytical control. *Vitis* 48:87–95.
- 528 28. Boudon-Padieu E, Bejat A, Clair D, Larrue J, Borgo M, Bertotto L, Angelini E. 2003.  
529 Grapevine yellows: Comparison of different procedures for DNA extraction and amplification with  
530 PCR for routine diagnosis of phytoplasmas in grapevine. *Vitis* 42:141–149.
- 531 29. Schneider B, Cousins MT, Klinkong S, Seemuller E. 1995. Taxonomic relatedness and  
532 phylogenetic positions of phytoplasmas associated with diseases of faba bean, sunnhemp, sesame,  
533 soybean, and eggplant. *J Plant Dis Prot* 102:225–232.
- 534 30. Lee IM, Gundersen DE, Hammond RW, Davis RE. 1994. Use of mycoplasma-like organism  
535 (MLO) group-specific oligonucleotide primers for nested-PCR assays to detect mixed-MLO infections  
536 in a single host-plant. *Phytopathology* 84:559–566.
- 537 31. Hall. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program  
538 for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41 : 95-98.
- 539 32. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis  
540 Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870–1874.
- 541 33. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and  
542 space complexity. *BMC Bioinformatics* 5:1–19.
- 543 34. Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific  
544 phylogenies. *Mol Biol Evol* 16:37–48.
- 545 35. Orkin S, Sambrook J, Fritsch EF, Maniatis T. 1990. *Molecular-cloning - A laboratory manual*,  
546 2nd edition. *Nature* 343:604–605.
- 547

548

549 **Tables**

550 **Table 1** Number of FDP PCR positive plants and *S. titanus* at each location, and percentage of FD-  
 551 infected samples for each category. Total number of analyzed samples in brackets. VV: *Vitis vinifera*;  
 552 WG: wild *Vitis* spp. plants; CL: *Clematis* spp; St: *S. titanus* (ST) sampled inside (IN) and outside  
 553 (OUT) each vineyard.

Site	VV	WG	CL	ST_IN	ST_OUT
CI	6 (6)	7 (43)	0 (1)	26 (50)	15 (50)
AS	6 (8)	6 (19)	0 (0)	17 (50)	14 (50)
CR	26 (33)	3* (22)	1 (1)	18 (50)	0 (12)
LM	6 (15)	1 (20)	4 (13)	10 (100)	29 (50)
MO	12 (18)	7 (21)	0 (1)	24 (50)	29 (50)
PA	9 (11)	8 (28)	0 (5)	6 (100)	8 (50)
PC	9 (13)	9 (39)	2 (11)	22 (50)	17 (50)
Total (%)	71,1	21,4	21,9	27,3	35,9

554 \* The collected 19 wild or abandoned grapevines growing nearby CR vineyard were negative for FD  
 555 presence. Therefore, three cultivated grapevines from neighboring vineyards were sampled. These three  
 556 samples were PCR positive for FDP.

557

558

559

560

**Table 2** List of Flavescence dorée phytoplasma (FDp) genes selected for preliminary identification of potential genetic markers to map FDp diversity at each location, based on a set of 13 DNA isolates sampled in Piedmont in 2013 and reference isolates FD-C and FD-D.

Gene name	Gene product name	Primer name	Primer sequence	Amplified fragment length (bp)	Fragment size (bp)	Number of identified genotypes
<i>dnaK</i>	chaperone protein <i>DnaK</i>	dnaK_F dnaK_R	TTAGGCGGAGGAACCTTCGAC AAGCTCCCATCGCAACTACT	559	492	2
<i>mntA</i>	Mn/Zn ABC transporter solute binding component	mntA_F mntA_R	GGATCCTTTAATGGGAGTAGG TATTCGCTTCTGTTTGGGTT	554	462	2
<i>nrdF</i>	ribonucleoside-diphosphate reductase 2, beta subunit	nrdF_F nrdF_R	AAAATGCTGTTCACGCTAAA TAACGGACAAAAGCGTTTAC	541	459	2
<i>malG</i>	probable ABC transporter, permease component	malG_F malG_R malGtestF malGtestR2 malGtestR5	GCTTTCGAGGCCAATTCCA ATTCTGGCCAAGCATAAGCG GTCTCAGGAGAAAATAAAAGTGGT CTTCTGGATGTTCTGAAGTTA GAAACAGCTACTAAAGCGG	496	373	9
<i>malF</i>	maltose transporter (subunit)	malF_F malF_R	TGCTTAATGATCGCCTTAGCTT GCCGCTGTTGTTCTTTAGC	591	510	2
<i>map</i>	methionine aminopeptidase	map_F map_R	GTTATCAAGGCTTCGGTGGTT CGGAAGTAACAGCAGTCCAA	498	435	2
<i>rpoC</i>	RNA polymerase, beta prime subunit	rpoC_F rpoC_R	AGCTGTGCGGAGTAATAGCAGC GTCGACCTACGGCTAACGAT	614	530	1
<i>rpsE</i>	30S ribosomal subunit protein S5	rpsE_F rpsE_R	TAGTTCAAGAGACAAAATAATT TTGTTTACCTTTAAATCTTGCTATC	518	417	2
<i>rsmA</i>	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase	rsmA_F2 rsmA_R2	ATAAAAAATGTTGTTGAAAATCGGTCC CATCAACTTAGGTTGTGGGAAA	450	372	1
<i>htmp1</i>	hypothetical transmembrane protein 1	1htmp_F 1htmp_R	TGACTATTTATGAGGTTTTGG CCGATAAAGCAAATTAACCA	500	144	1 <sup>ab</sup>
<i>htmp2</i>	hypothetical transmembrane protein 2	2htmp_F 2htmp_R	TGCATCTGATGAAAAAGAAA TGTTTATTACGCCAGTCATTT	476	393	2 <sup>c</sup>
<i>htmp3</i>	hypothetical transmembrane protein 3	3htmp_F 3htmp_R	TTTTTAAGAAGTGTGTTTTTG TCAACAAAATCAACAAGAAAA	475	321/312	2 <sup>b</sup>
<i>htmp4</i>	hypothetical transmembrane protein 4	4htmp_F 4htmp_R	TCCGATAGAAAATACGGAAA GCTCTTGGCAAGGTTAATA	535	468	1 <sup>ac</sup>
<i>htmp5</i>	hypothetical transmembrane protein 5	5htmp_F 5htmp_R	AAAACAAGAAGAAACGCAAAA CCAAGATTCTTCTAAACATTTTAA	376	260	2
<i>lolD</i>	probable ABC transporter ATP-binding component	lolD_F lolD_R	AAAATATCCAAGAAAGAAACGA TTCTTAAAAATAGGGTGCCAAATT	760	630	2
<i>glyA</i>	serine hydroxymethyltransferase	glyA_F glyA_R	ATTGCTGGATTAATTGTTGC CATTGCTGGAGTTCCTATTC	501	392	2
<i>vmpA</i>	variable membrane protein A	vmpA-F3 vmpA_R vmpA-F5	GATGGA AAAACA AAAATGATAG AATAAATCAATAAAAAACTCAC CCTTATCAACTGGATATGGT	1488/1254	A/B	2

26

---

vmpA-R3	CTGATGCGTTAGCCACTTC
vmpA-F8	TTATAGAAATTATTCTCACA
vmpA-R9	TAAAA(C/A)AGT(C/A)GATAATTCAAC

---

(<sup>a</sup>) no amplification of FD-C reference isolate; (<sup>b</sup>) no amplification of field isolates; (<sup>c</sup>) no amplification for the majority of field isolates

561

562

563

564 **Table 3** List of single nucleotide polymorphisms (SNPs) of each *dnaK* profile, and their location on the  
565 coding sequence, starting from the ATG codon of FD-D *dnaK* gene (FD92 draft genome) (14).  
566

Genotype	SNPs			
	624	789	888	969
<i>dnaK1</i>	T	C	T	C
<i>dnaK2</i>	C	C	C	T
<i>dnaK3</i>	C	A	C	T

567

568

569 **Table 4** Percentage of the different *dnak-vmpA* genotype combinations. The most frequent  
 570 combinations, *dnak1-vmpA\_R1\_1B* and *dnak2-vmpA\_R1\_2A*, correspond to those of the two FD-D  
 571 and FD-C reference strains respectively (14; 13). %: percentage of analysed amplicons showing each  
 572 *dnak/vmpA* profile.

	<i>vmpA_R1</i>						<b>Total</b>
	1A	1B	2A	2B	3A	4A	
<i>dnak1</i>	0,7	71,1	-	1,3	-	-	73,1
<i>dnak2</i>	0,7	-	22,4	-	-	-	23,1
<i>dnak3</i>	-	-	-	-	2,5	1,3	3,8
	1,4	71,1	22,4	1,3	2,5	1,3	100

573

574

575

576

577

578 **Captions to Figures**

579

580 **Figure 1** Distribution of *dnaK* and *vmpA* R1 profiles found among the different sample categories  
581 (VV: cultivated *V. vinifera*, WG: wild or abandoned *Vitis* spp. plants; CL: *Clematis* spp.; ST:  
582 *Scaphoideus titanus*) sampled outside (OUT) and inside (IN) the seven analyzed vineyards. Black  
583 rectangles indicate “presence” and white ones indicate “absence” of a particular *dnaK* or *vmpA*  
584 genotype. The grey line separates samples collected inside the vineyards (upper part) from those  
585 collected outside (lower part) for each location. Cisterna d’Asti, CI; CREA-Asti (AS); Castel Rocchero  
586 (CR); La Morra (LM); Montà (MO); Paderna (PA); Portacomaro (PC). \*: *dnaK* and *vmpA* profiles of  
587 FD-92 strain; §: *dnaK* and *vmpA* profile of FD-C.

588

589 **Figure 2** Distribution of *malG* profiles found among the different sample categories (VV: cultivated  
590 *V. vinifera*, WG: wild or abandoned *Vitis* spp. plants; CL: *Clematis* spp.; ST: *Scaphoideus titanus*)  
591 sampled outside (OUT) and inside (IN) the seven analyzed vineyards. Colored rectangles indicate  
592 “presence” and white ones indicate “absence” of a particular *malG-vmpA* genotype. The grey line  
593 separates samples collected inside the vineyards (upper part) from those collected outside (lower part)  
594 for each location. CI: Cisterna, AS: Asti, CR: Castel Rocchero, LM: La Morra, MO: Montà, PA:  
595 Paderna, PC: Portacomaro \*: *malG* profile of FD-92 strain; #: *malG* profile of FD-C.

596

597 **Figure 3** Median-joining network inferred from *malG* genotypes. Genotypes are represented by circles  
598 and the circle size shows the genotype frequency. Two genotypes are connected by a line, each SNP  
599 mutation is represented by a hatch mark. VV\_IN: cultivated grapevines inside the vineyard (purple),  
600 VV\_OUT: cultivated grapevines in neighboring vineyards (pink); WG: wild *Vitis* spp. plants (green);  
601 CL: *Clematis* spp. (yellow); ST\_IN: *Scaphoideus titanus* inside the vineyard (light grey); ST\_OUT:  
602 *Scaphoideus titanus* outside the vineyard (dark grey).

603

604 **Figure 4** A) Southern blot of EcoRI digested total DNA from FD-C and FD-D infected and healthy  
605 periwinkles (H) probed with DIG labelled *malG* gene amplicon obtained through PCR driven by  
606 *malG\_F/malG\_R* primer pair (C+: probe positive control represented by pGEM-T-*malG1* plasmid). B)  
607 Electrophoresis separation of amplicons obtained following PCR of total DNA from FD-C and FD-D  
608 infected periwinkles with copy-specific primer pairs (002 and 005), according to the draft genome of  
609 FD92, and from healthy periwinkle. (\* : nonspecific PCR product).

610

611 **Figure 5** Distribution and frequency of the most frequent *malG* genotypes found among the different  
612 categories (VV: cultivated grapevines, WG: wild or abandoned *Vitis* spp. plants; CL: *Clematis* spp.;  
613 ST: *Scaphoideus titanus*) sampled outside (OUT) and inside (IN) the seven analyzed vineyards, and  
614 graphic overview of the recorded overall *malG* diversity. CI: Cisterna, AS: Asti, CR: Castel Rocchero,  
615 LM: La Morra, MO: Montà, PA: Paderna, PC: Portacomaro.

Location	Sample category	<i>dna k</i> genotype			<i>vmp A R1</i> genotype						
		1	2	3	1A	1B	2A	2B	3A	4A	
CI	VV	■				■					
	ST_IN	■				■					
	ST_OUT	■				■		■			
	WG	■				■					
AS	VV	■	■			■	■				
	ST_IN	■	■			■	■				
	ST_OUT	■	■		■	■					
	WG	■	■		■	■					
CR	VV	■				■					
	ST_IN	■				■					
	ST_OUT	■				■					
	VV_OUT	■				■					■
	CL			■							■
LM	VV	■	■			■	■				
	ST_IN	■	■			■	■				
	ST_OUT	■				■		■			
	WG	■				■					
	CL			■						■	
MO	VV	■	■		■	■	■				
	ST_IN	■	■		■	■	■				
	ST_OUT	■				■		■			
	WG	■				■					
PA	VV		■	■		■	■				
	ST_IN		■	■		■	■				
	ST_OUT		■	■		■	■				
	WG		■	■		■	■				
PC	VV	■	■			■	■				
	ST_IN	■	■			■	■				
	ST_OUT	■	■			■	■				
	WG	■	■			■					
	CL			■							■







