Genetic diversity of Flavescence dorée phytoplasmas at vineyard scale

Genetic diversity of Flavescence dorée phytoplasmas at vineyard scale

Running title: FD phytoplasma diversity at vineyard scale

Marika Rossi[a], Mattia Pegoraro[a], Matteo Ripamonti[b], Simona Abbà[a], Dylan Beal[b*], Alessia Giraudo[a,b], Flavio Veratti[a], Sylvie Malembic-Maher[c], Pascal Salar[c], Domenico Bosco[a,b], Cristina Marzachì[a#]

[a] Istituto per la Protezione Sostenibile delle Piante, CNR, Strada delle Cacce 73, 10135 Torino, Italy

[b] Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco (TO). Italy

[c] INRA, Univ. Bordeaux, UMR 1332 de Biologie du Fruit et Pathologie, CS20032, 33882, Villenave d'Ornon Cedex, France

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

*Present Address: Department of Environmental Science, Policy, & Management, UC Berkeley, Berkeley, CA 94720

Keywords: *Vitis vinifera*, *Scaphoideus titanus*, *Clematis*, *dna*K, *vmp*A, *mal*G

ABSTRACT

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

IMPORTANCE

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

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

control of disease spread. A possible approach to evaluate disease dispersal patterns over spatial scales is through analyses of pathogen genetic markers (7). Based on sequence and restriction fragment length polymorphism (RFLP) analysis of the 16S rRNA and 16S-23S intergenic spacers, two FDp taxonomic groups were described: 16SV-C and 16SV-D (8). Moreover, sequencing of two non-ribosomal loci, *sec*Y and *rps*C, allowed for the identification of three genetic clusters within FDp populations sampled in France and Italy (9). Arnaud and co-workers (2007) analyzed the sequences of two other genetic loci, *map* and *deg*, and confirmed the existence of three genetic clusters of FDp characterized by different geographical distribution and genetic variability: strain Cluster FD1, characterized by low genetic variability and high incidence in Southwestern France; strain Cluster FD2, including isolates FD-92 and FD-D, with no genetic variability and present both in France and Italy; and strain cluster FD3 comprising FD-C, showing high variability and found only in Italy (10). Both FD-C and FD-D phytoplasma are found in Piedmont (9). This genetic classification was not sufficiently accurate to describe the genetic variability present at a single agroecosystem. In fact, the low variability of the genetic loci considered until now represented a limit in the study of populations from a small geographical area. In the present study, two new genetic markers for FDp characterization, *dna*K and *mal*G, were developed. These two genes, together with the *vmp*A locus, already described by Renaudin and co-workers (11), were used to genetically characterize the FDp populations at seven agroecosystems in Piedmont. To identify the components of the FD epidemiological cycle and to study the epidemic flow of the disease, *V. vinifera*, *V. vinifera* and "wild" rootstocks (hybrids of *V. riparia*, *V. rupestris* and *V. berlandieri*) from abandoned vineyards, *Clematis* spp. (12) and *S. titanus* were sampled both within and outside the vineyards. Due to the robustness of *dna*K for discriminating between FD-C and FD-D isolates as well as the high variability of *mal*G sequence, these two markers 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: *dna*K, *mnt*A, *nrd*F, *mal*G, *mal*F, *map*,
114 *rpo*C, *rps*E, *rsm*A, *htmp*1, *htmp*2, *htmp*3, *htmp*4, *htmp*5, *lol*D, *gly*A, *vmp*A. PCR amplification of the

17 genes from total DNA of periwinkle-maintained FD-C (13) and FD-D (14) isolates provided amplicons of the expected sizes for 15 and 17 targets of FD-C and –D, respectively. In particular, primers designed on *htmp*1 and 4 failed to amplify their targets from FD-C (Table 2). These two genes were excluded from further analyses. In preliminary experiments on FD-infected grapevine samples collected in 2013, amplification with specific primers designed on the remaining 15 genes provided amplicons of the expected sizes for most of the genes analyzed. Amplicons from *htmp*1, 2, 3, and 4 genes were not efficiently amplified from most of the field samples and were, consequently, excluded from successive analysis (Table 2).All the obtained amplicons were sequenced in both directions. Upon sequencing, *rpo*C and *rsm*A amplicons provided a single genotype and due to their low sequence variability, they were excluded from further analysis. Among the tested target genes, only *mal*G provided more than two genotypes and, due to this high sequence variability, it was selected for FDp characterization. Sequencing of the remaining target genes always provided 2 genotypes, corresponding to FD-C (13) and FD-D (14) reference isolates. Among these, *dna*K was arbitrary selected for further studies. Gene *vmp*A was also chosen due the adhesion role of VmpA that could be essential in the colonization of the insect by FDp (15). These three genes were used for subsequent characterization of FDp from the different geographical sites.

Genetic diversity of the selected target genes

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

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

***dnaK* and *vmp*A**. Amplification with primers *dnaK*_F/R produced a specific amplicon of the expected size from 46 cultivated vines, 26 *Vitis* spp. plants from outside the vineyards, 42 *S. titanus* collected inside the vineyards, and 38 *S. titanus* from outside the vineyards, as well as the six *Clematis* spp from wild vegetation. One hundred and fifty-eight *dna*K sequences were analyzed, and three *dna*K genotypes were identified as *dna*K1, 2 and 3. A *dnaK*1 genotype was found in 116 isolates (73 %) and the reference isolate FD92 (FD-D), *dnaK*2 genotype was obtained from 36 isolates (23 %) and the reference isolate FD-C, *dnaK*3 genotype was obtained from the six *Clematis* spp. (Figure 1). Genotype *dna*K2 differed by three SNPs from *dna*K1 at positions 624, 888 and 969, and by one SNP from *dna*K3 at position 789. All the mutations were synonymous (Table 3). Mixed infections were evident from the chromatograms of seven samples (3 VV, 2 WG, and 2 ST, not shown), from AS, PA, and PC, so PCR amplicons were cloned in the plasmid vector pGEM-T (Promega, Madison, WI) for further investigations. Sequencing of five clones for each of the seven samples confirmed double infection with *dna*K1 and *dna*K2 isolates for all samples. The incidence of *dnaK*1 was higher than that of *dna*K2 for all analyzed sample categories (χ^2, p=0.025). Interestingly, the *dna*K1/*dna*K2 frequency ratio for VV (1.4) was lower than for WG (3.8), ST_IN (3.6) and ST_OUT (8.3) (not shown).

The phytoplasma variable membrane protein VmpA gene is characterized by a stretch of 234 nt repeated sequences (R) (11, 16). Insertion/deletion of one repeat sequence determines size variability of the gene. Amplification with primers VmpAF3/VmpAR yielded two possible amplicons of 1.488 bp (A) and 1.254 bp (B) respectively. Sequencing of R1 repeats from 158 samples identified two R1 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 *vmp*A gene except for a WG and
171 a ST_OUT both collected in PA. This plant showed a mixed profile also for the *vmp*A amplicon size (A
172 and B) and *dna*K genotype (*dna*K1 and *dna*K2). The ST_OUT showing both *vmp*A amplicon types had
173 a R1_2 profile associated with *dna*K mixed profile. When the length of *vmp*A gene was considered
174 together with R1 sequence and *dna*K genotypes, 7 types of *dna*K-*vmp*A profiles were detected and
175 listed in Table 4.
176 ***mal*G.** Partial sequencing of *mal*G gene detected a mix of divergent sequences in 108 of the 158
177 analyzed samples. The *mal*G 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 *mal*G genotypes were

184 left in the list, *mal*G1 identical to *mal*G of the FD-D isolate (*dna*K1 profile), *mal*G2 identical to *mal*G
185 of the FD-C isolate (*dna*K2 profile) and *mal*G3 (*dna*K1 profile) being the predominant ones (28,2 %,
186 30,4 %, 9,4 % of the 447 analyzed sequences, respectively). *mal*G38 and *mal*G39 were identified only
187 in *Clematis* spp. (*dna*K3, sec-*map*: M50/M51/ KJ911219; Figure 2); *mal*G38 was always associated
188 with *vmp*A_R1_3A, and *mal*G39 with *vmp*A_R1_4A (Figure 2). Most of the samples (78), showed
189 multiple profiles, predominantly *mal*G1 and *mal*G3, and among these, 43 samples had more than two
190 *mal*G types. Profiles *mal*G1 and *mal*G3 were found 7 and 11 times as pure profiles, respectively. Type
191 *malG2* was found with other *mal*G genotypes in four samples, of which three (1 VV, 1 WG, and 1
192 St_OUT) had FD-C/FD-D mixed infections and one (WG) had just the *dnaK*1 (FD-D) profile.
193 Overall, the frequencies of *mal*G1 and *mal*G3, both associated to *dna*K1 profile, were similar (around
194 30 %) suggesting a possible gene duplication of this gene locus. Southern blot confirmed the
195 duplication of *mal*G gene at least in the FD92 chromosome (Figure 4, Panel A). Copy-specific PCR
196 amplification confirmed the duplication (Figure 4, Panel B), as primer pair malGtestF / malGtestR2 and
197 malGtestF / malGtestR5 were able to amplify both *mal*G copies from total DNA of FD-D reference
198 isolate. These primers were designed on the sequences of the contigs 002 and 005 of FD92 draft
199 genome, which included two identical *mal*G1 sequences. Primer malGtestF was designed on the *mal*G
200 coding sequence whereas malGtestR2 and malGtestR5 were designed to align downstream of the
201 identical region, so that they could amplify specifically contig 002 and 005 sequences, respectively. A
202 single copy of *mal*G (in the contest of contig 002) was detected in the chromosome of the FD-C isolate
203 (Figure 4).
204 Overall, 32 *mal*G types were found in insects, 25 of which were present with less than 2 % frequency
205 (Figure 2). Thirteen *mal*G genotypes were detected in cultivated grapevines and wild growing *Vitis* spp.

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

The AS and PC sites, with 15 *mal*G types each, showed the highest genetic variability, which was mainly determined by insect and wild grapevine isolates (Figure 2). Six genotypes were identified at PA, which was the site with the lowest variability. This vineyard was also characterized by the prevalence of FD-C (*dna*K2, *mal*G2, *vmp*A2A) both inside and outside the vineyard.

The DNA sequence identity among the *mal*G genotypes ranged between 99.7 % and 97.3 %. Among the three most frequent genotypes, only the SNP at position 380 (G to A) determined a non-synonymous substitution (V127I) (Figure S6). This was the most frequent mutation and determined a clear distinction between *mal*G1 and *mal*G3 clusters (Figure S6). A second mutation at position 332, determining a valine to leucine substitution, characterized *mal*G38 and explained its relationship with *mal*G103 and *mal*G126 genotypes (Figure 3 and Figure S6) found only in WG category at PC. The silent mutation T/C at position 629 was found only in isolates from CI.

The Median-joining network analysis of the 50 *mal*G genotypes (Figure 4) identified seven main nodes (based on sequence and frequency): *mal*G1, *mal*G2, *mal*G3, *mal*G6, *mal*G16, *mal*G18, *mal*G34. In particular, most of the genetic diversity was linked to *mal*G1 and 3, especially *mal*G1. Eighteen minor genotypes were linked to *mal*G1, 15 of which were found only in insects. Eleven minor genotypes were linked to *mal*G3, and seven of those were found only in insects. Two genotypes were directly linked to *mal*G2 (141 and 146). Genotypes *mal*G103 and *mal*G126 were found only in the WG category. These two genotypes were linked to *mal*G38 of CL, with no direct link to *mal*G genotypes of other sample categories.

Genetic variability of FDp at different sampling sites

229 The distribution and frequency of the most represented *mal*G 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, *mal*G2 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 *dna*K2 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, *dna*K1 and *dna*K2
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 *dna*K3 profiles. *Clematis*,
243 instead, differed in terms of *Vmp*A 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, *dna*K, *vmp*A, and *mal*G 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 *mal*G 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 *mal*G2 type of FD-C showed no variability at all, our results cannot exclude the presence of a *mal*G
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 *mal*G copy in this context. Poor quality
272 of FD-C draft genome (13) in that region does not support any of the hypotheses. Yet, *mal*G 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 *mal*G1 and *mal*G3 genotypes

showed the highest variability; interestingly, 15 and 7 of the 18 *mal*G1 and 11 *mal*G3 were found only in insects, indicating that most of the variability was detected in this host. All *mal*G1 and 3-derived types were associated to a *dna*K1 profile, and again, this was more abundant in insects, compared to vines, especially cultivated ones. In contrast, the unique *mal*G2 profile was associated only with the *dna*K2 type, which is more abundant in plants, in particular in cultivated grapevines, than in insects. Infections with different *dna*K profiles were rare, especially in vectors, indicating a possible antagonism between the two *dna*K genotypes, which is an issue currently under investigation. Such large genetic variability of FDp described on a very restricted geographical scale, supports the hypothesis of a European origin of the phytoplasma (10). Interestingly, similar levels of FDp genetic variation were recorded in the cultivated and wild areas for either grapevines or leafhoppers. This finding, together with the co-occurrence of many FDp genotypes inside and outside the vineyards, confirms the importance of wild or abandoned *Vitis* spp. plants and associated *S. titanus* in the epidemiology of the disease. The large overlap of FDp genotypes in the two environments is confirmed at the single site level, with the partial exception of the PA site. In this site, very few *S. titanus* from the wild area were found infected, so epidemiology of the disease at this site, should be mainly explained by within-vineyard spread ("secondary infection"). At the other sites, FDp genetic diversity was consistent with the hypothesis of "primary infection" by incoming vectors from outside the vineyard. This hypothesis is further supported by the "edge" effect recorded for FDp-infected grapevines (18–20). Primary infections are likely to occur late in the season, when cultivated grapevines are no longer protected by insecticides, due to the need of respecting a safety period before grape harvest (21). Data on *S. titanus* dispersal capability (22) indicate that 80% of adults do not fly beyond 30 m, although few can move up to 300 m. In our study, wild or abandoned grapevines and associated *S. titanus* adults 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, *mal*G 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*

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

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

Total DNA extraction and FDp diagnosis

Total nucleic acids were extracted from 1 g of leaf midribs and petioles and from single leafhoppers according to the method of Pelletier (27, 28). Total DNA extracts from plants and insects were then suspended in 100 µL or 75 µL or of Tris-HCl 10mM pH 8, respectively. DNA concentration was measured with NanoDrop 2000^{TM} Spectrophotometer (Thermo Scientific, Waltham, MA), and all samples were then diluted to 20 ng/µL.

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

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

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

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

Applied and Environmental Microbiology

388 purified using the DNA Clean and Concentrator kit™ (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* DH5α 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*, *vmp*A (R1
400 repeat), and *mal*G 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 *mal*G, 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 *mal*G 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 *Eco*RI (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 synthetized using the PCR DIG Probe Synthesis Kit
425 following the manufacturer's instructions (Roche, Basel, Switzerland). pGEM-T-*mal*G1 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**

Sequence data were submitted to GenBank and the following accession numbers were provided: MH547710 - 12 (*dna*K1 - 3); MH547713 - 47 (*mal*G1 - 35); MH547748 - 93 (*mal*G38 - 183); MH547894 - 96 (*vmp*A_R1_1 - 3); MK091396-97 (*mnt*A1-2); MK091398-99 (*nrd*F1-2); MK091400-01 (*mal*F1-2); MK091402-3 (*map*1-2); MK091404 (*rpo*C1); MK091405-6 (*rps*E1-2); MK091407 (*rsm*A1); MK091408 (*htmp*1_1); MK091409-10 (*htmp*2_1-2*htmp*); MK091411-12 (*htmp*3_1-2) MK091413 (*htmp*4_1); MK091414-15 (*htmp*5_1-2); MK091416-17 (*lol*D1-2); MK091418-19 (*gly*A1-2) .

AKNOWLEDGEMENTS

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

REFERENCES

1. Schvester D, Carle P, Moutous G. 1967. Testing susceptibility of vine varieties to (Flavescence doree) by means of inoculation by Scaphoideus littoralis Ball. Annales Des Epiphyties 18:143.

2. Chuche J, Thiery D. 2014. Biology and ecology of the Flavescence doree vector Scaphoideus 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 flaveseence 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.

12. Angelini E, Squizzato F, Lucchetta G, Borgo M 2004. Detection of a phytoplasma associated with grapevine Flavescence dorée in Clematis vitalba. Eur J Plant Pathol 110:193-201.

13. Firrao G, Martini M, Ermacora P, Loi N, Torelli E, Foissac X, Carle P, Kirkpatrick BC, Liefting L, Schneider B, Marzachi C, Palmano S. 2013. Genome wide sequence analysis grants unbiased definition of species boundaries in "Candidatus Phytoplasma." Syst Appl Microbiol 36:539–548.

14. Carle P, Malembic-Maher S, Arricau-Bouvery N, Desque D, Eveillard S, Carrere S, Foissac X. 2011. "Flavescence doree" phytoplasma genome: a metabolism oriented towards glycolysis and protein degradation. Bull Insectology 64:S13–S14.

15. Arricau-Bouvery N, Duret S, Dubrana M-P, Batailler B, Desque D, Beven L, Danet J-L, Monticone M, Bosco D, Malembic-Maher S, Foissac X. 2018. Variable membrane protein A of flavescence doree phytoplasma binds the midgut perimicrovillar membrane of Euscelidius variegatus and promotes adhesion to its epithelial cells. Applied and environmental microbiology.

16. Cimerman A, Pacifico D, Salar P, Marzachi C, Foissac X. 2009. Striking Diversity of vmp1, a Variable Gene Encoding a Putative Membrane Protein of the Stolbur Phytoplasma. Appl Environ Microbiol 75:2951–2957.

17. Marzachì C, Boarino A, Vischi A, Palermo S, Morone C, Loria A, Boccardo G. 2001. Flavescenza dorata, legno nero e giallume dell'astro in vitigni del Piemonte sud orientale. Informatore Fitopatologico, 9, pp. 58-63.

18. Pavan F, Mori N, Bigot G, Zandigiacomo P. 2012. Border effect in spatial distribution of Flavescence doree affected grapevines and outside source of Scaphoideus titanus vectors. Bull Insectology 65:281–290.

19. Maggi F, Marzachì C, Bosco D. 2013. A Stage-Structured Model of Scaphoideus titanus in Vineyards. Environ Entomol 42:181–193.

20. Maggi F, Bosco D, Galetto L, Palmano S, Marzachì C. 2017. Space-Time Point Pattern Analysis of Flavescence Doree Epidemic in a Grapevine Field: Disease Progression and Recovery. Front Plant Sci 7.

21. Bosco D, Mori N. 2013. Flavescence dorée control in Italy. Phytopathogenic Mollicutes, 3(1), pp 40-43.

22. Lessio F, Tota F, Alma A. 2014. Tracking the dispersion of Scaphoideus titanus Ball (Hemiptera: Cicadellidae) from wild to cultivated grapevine: use of a novel mark-capture technique. Bulletin of Entomological Research 104:432–443.

23. Malembic-Maher S, Salar P, Filippin L, Carle P, Angelini E, Foissac X. 2011. Genetic diversity of European phytoplasmas of the 16SrV taxonomic group and proposal of "Candidatus Phytoplasma rubi." Int J Syst Evol Microbiol 61:2129–2134.

24. Filippin L, Jovic J, Cvrkovic T, Forte V, Clair D, Tosevski I, Boudon-Padieu E, Borgo M, Angelini E. 2009. Molecular characteristics of phytoplasmas associated with Flavescence doree in clematis and grapevine and preliminary results on the role of Dictyophara europaea as a vector. Plant Pathol 58:826–837.

25. Jeger M, Bragard C, Caffier D, Candresse T, Chatzivassiliou E, Dehnen-Schmutz K, Gilioli G, Miret JAJ, MacLeod A, Navarro MN, Niere B, Parnell S, Potting R, Rafoss T, Rossi V, Urek G, Van Bruggen A, Van Der Werf W, West J, Winter S, Bosco D, Foissac X, Strauss G, Hollo G, Mosbach-Schulz O, Gregoire JC, PLH EPPH. 2016. Risk to plant health of Flavescence doree for the EU territory. Efsa Journal 14.

26. Malembic-Maher S, Constable F, Cimerman A, Arnaud G, Carle P, Foissac X, Boudon-Padieu E. 2008. A chromosome map of the Flavescence doree phytoplasma. Microbiology-(UK) 154:1454–1463.

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 mycoplasmalike 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

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
dnaK	chaperone protein *DnaK*	dnaK_F dnaK_R	TTAGGCGGAGGAACTTTCGAC AAGCTCCCATCGCAACTACT	559	492	2
mntA	Mn/Zn ABC transporter solute binding component	mntA_F mntA_R	GGATCCTTTAATGGGAGTAGG TATTCGCTTCTGTTTGGGTT	554	462	2
nrdF	ribonucleoside-diphosphate reductase 2, beta subunit	nrdF_F nrdF_R	AAAATGCTGTTCACGCTAAA TAACGGACAAAAGCGTTTAC	541	459	2
***mal*G**	probable ABC transporter, permease component	malG_F malG_R malGtestF malGtestR2 malGtestR5	GCTTTCCGAGGCCAATTCCA ATTCTGGCCAAGCATAAGCG GTCTCAGGAGAAAATAAAAGTGGT CTTTCTGGATGTTCTGAAGTTA GAAACAGCTACTAAAGCGG	496	373	9
malF	maltose transporter (subunit)	malF_F malF_R	TGCTTTAATGATCGCCTTAGCTT GCCGCTGTTGTTCCTTTAGC	591	510	2
map	methionine aminopeptidase	map_F map_R	GTTATCAAGGCTTCGGTGGTT CGGAAGTAACAGCAGTCCAA	498	435	2
rpoC	RNA polymerase, beta prime subunit	rpoC_F rpoC_R	AGCTGTCGGAGTAATAGCAGC GTCGACCTACGGCTAACGAT	614	530	1
rpsE	30S ribosomal subunit protein S5	rpsE_F rpsE_R	TAGTTCAAGAGACAAAACTAATT TTGTTTACCTTTAAATCTTGCTATC	518	417	2
rsmA	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase	rsmA_F2 rsmA_R2	ATAAAAATGTTGTTGAAATCGGTCC CATCAACTTTAGGTTGTGGGAAA	450	372	1
htmp1	hypothetical transmembrane protein 1	1htmp_F 1htmp_R	TGACTATTTATGAGGTTTTGG CCGATAAAGCAAATTAAACCA	500	144	1[a,b]
htmp2	hypothetical transmembrane protein 2	2htmp_F 2htmp_R	TGCATCTGATGAAAAAGAAA TGTTTATTACGCCAGTCATTT	476	393	2[c]
htmp3	hypothetical transmembrane protein 3	3htmp_F 3htmp_R	TTTTTAAGAAGTGTCGTTTTTG TCAACAAAATCAACAAGAAAA	475	321/312	2[b]
htmp4	hypothetical transmembrane protein 4	4htmp_F 4htmp_R	TCCGATAGAAAATACGGAAA GCTCTTGGCAAGGTTTAATA	535	468	1[a,c]
htmp5	hypothetical transmembrane protein 5	5htmp_F 5htmp _R	AAAACAAGAAGAAACGCAAAA CCAAGATTCTTCTAAACATTTTAA	376	260	2
lolD	probable ABC transporter ATP-binding component	lolD_F lolD_R	AAAATTATCCAAGAAAGAAACGA TTCTTAAAATAGGGTGCCAAATT	760	630	2
glyA	serine hydroxymethyltransferase	glyA_F glyA_R	ATTGCTGGATTAATTGTTGC CATTGCTGGAGTTCCTATTC	501	392	2
vmpA	variable membrane protein A	vmpA-F3 vmpA_R vmpA-F5	GATGGAAAACAAAATGATAG AATAAATCAATAAAAAACTCAC CCTTATCAACTGGATATGGT	1488/1254	A/B	2

vmpA-R3	CTGATGCGTTTAGCCACTTC
vmpA-F8	TTATAGAAATTATTCTCACAA
vmpA-R9	TAAAA(C/A)AGT(C/A)GATAATTCAAC

([a]) no amplification of FD-C reference isolate; ([b]) no amplification of field isolates; ([c]) no amplification for the majority of field isolates

561

562

563

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

Genotype	SNPs			
	624	*789*	*888*	*969*
***dnaK*1**	T	C	T	C
***dnaK*2**	C	C	C	T
***dnaK*3**	C	A	C	T

567

568

569 **Table 4** Percentage of the different *dna*k-*vmp*A genotype combinations. The most frequent
570 combinations, *dna*k1-*vmp*A_R1_1B and *dna*k2-*vmp*A_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 *dna*K/*vmp*A profile.

	*vmp*A_R1						**Total**
	1A	1B	2A	2B	3A	4A	
*dna*K1	0,7	71,1	-	1,3	-	-	73,1
*dna*K2	0,7	-	22,4	-	-	-	23,1
*dna*K3	-	-	-	-	2,5	1,3	3,8
	1,4	71,1	22,4	1,3	2,5	1,3	100

573

574

575

576

577

Captions to Figures

Figure 1 Distribution of *dna*K and *vmp*A R1 profiles found among the different sample categories (VV: cultivated *V. vinifera*, WG: wild or abandoned *Vitis* spp. plants; CL: *Clematis* spp.; ST: *Scaphoideus titanus*) sampled outside (OUT) and inside (IN) the seven analyzed vineyards. Black rectangles indicate "presence" and white ones indicate "absence" of a particular *dna*K or *vmp*AR1 genotype. The grey line separates samples collected inside the vineyards (upper part) from those collected outside (lower part) for each location. Cisterna d'Asti, CI; CREA-Asti (AS); Castel Rocchero (CR); La Morra (LM); Montà (MO); Paderna (PA); Portacomaro (PC). *: *dna*K and *vmp*A profiles of FD-92 strain; §: *dna*K and *vmp*A profile of FD-C.

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

Figure 3 Median-joining network inferred from *mal*G genotypes. Genotypes are represented by circles and the circle size shows the genotype frequency. Two genotypes are connected by a line, each SNP mutation is represented by a hatch mark. VV_IN: cultivated grapevines inside the vineyard (purple), VV_OUT: cultivated grapevines in neighboring vineyards (pink); WG: wild *Vitis* spp. plants (green); CL: *Clematis* spp. (yellow); ST_IN: *Scaphoideus titanus* inside the vineyard (light grey); ST_OUT: *Scaphoideus titanus* outside the vineyard (dark grey).

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

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