



HAL
open science

Genetic structure of *Malus sylvestris* and potential link with preference/performance by the rosy apple aphid pest *Dysaphis plantaginea*

Thomas Denoirjean, Géraldine Doury, Amandine Cornille, Xilong Chen,
Thierry Hance, Arnaud Ameline

► To cite this version:

Thomas Denoirjean, Géraldine Doury, Amandine Cornille, Xilong Chen, Thierry Hance, et al.. Genetic structure of *Malus sylvestris* and potential link with preference/performance by the rosy apple aphid pest *Dysaphis plantaginea*. *Scientific Reports*, 2021, 11, pp.5732. 10.1038/s41598-021-85014-x . hal-03298990

HAL Id: hal-03298990

<https://hal.science/hal-03298990>

Submitted on 26 Jul 2021

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

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

1 **Title: Genetic structure of *Malus sylvestris* and potential link with preference/performance by the rosy apple**
2 **aphid pest *Dysaphis plantaginea***

3
4 **Denoirjean Thomas¹, Doury Géraldine¹, Cornille Amandine², Chen Xilong², Hance Thierry³, Ameline Arnaud^{1*}.**

5 1. UMR CNRS 7058 EDYSAN (Écologie et Dynamique des Systèmes Anthropisés), Université de Picardie Jules
6 Verne, 33 rue St Leu, F-80039 Amiens Cedex, France.

7 2. Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France

8 3. Earth and Life Institute, Biodiversity Research Centre, UC Louvain, ELIB – Croix du sud 4–5 bte L7.07.04, 1348
9 Louvain-la-Neuve, Belgium.

10 * Correspondence: Tel.: +33 3 22 82 75 56; Fax: +33 3 22 82 75 47

11 *E-mail address:* arnaud.ameline@u-picardie.fr

12
13
14
15
16
17 **Abstract**

18 The European crabapple *Malus sylvestris*, a crop wild relative of *Malus domestica*, is a major contributor to
19 the cultivated apple genome and represents a potential source of interesting alleles or genes, particularly pest
20 resistance traits. An original approach was used to explore the trophic interaction between *M. sylvestris* populations
21 and its pest, the rosy apple aphid (*Dysaphis plantaginea*). Using 13 microsatellite markers, population genetic
22 structure and level of crop-to-wild introgressions were inferred between *M. sylvestris* seedlings from three sites in
23 Europe (Denmark, France, Romania), and *M. domestica* seedlings. Genetically characterized plants were also used to
24 analyze aphid feeding behavior and fitness parameters. First, aphids submitted to two genetically close *M. sylvestris*
25 populations (the Danish and French) exhibited similar behavioral parameters, suggesting similar patterns of resistance
26 in these host plants. Second, the Romanian *M. sylvestris* population was most closely genetically related to *M.*
27 *domestica*. Although the two plant genetic backgrounds were significantly differentiated, they showed comparable
28 levels of sensitivity to *D. plantaginea* infestation. Third, aphid fitness parameters were not significantly impacted by
29 the host plant's genetic background. Finally, crop-to-wild introgression seemed to significantly drive resistance to *D.*
30 *plantaginea* independent of host plant population genetic structure, with hybrids being less suitable hosts.

31 **Keywords:** Crop wild relatives (CWR) - *Malus sylvestris* - Genetic structure - *Dysaphis plantaginea* - Aphid
32 preference/performance - Electropetrography - Crop-to-wild gene flow - Fruit trees - Pest.

33 |

34

35

36

37 Introduction

38 Intense farming practices can lead to harmful impacts on the environment and human health. It is therefore
39 urgent to promote eco-friendly agricultural management while feeding a growing world population^[1]. In this case, new
40 breeding strategies making use of wild untapped genetic diversity could become a promising opportunity to provide
41 farmers with crops less dependent on chemical inputs without heavy drawbacks to productivity^[2]. Crop breeding
42 programs often rely on wild species that are either phylogenetically close to a crop and/or that have played a primary
43 role in the crop domestication history. Such plants have been named “crop wild relatives” (or CWR). Although the use
44 of CWR in breeding programs has the potential to fulfill numerous agronomic needs (e.g. yield increase or abiotic
45 stress adaptation), the major focus of CWR research concerns their ability to enhance crop resistance against
46 pathogens and pests^[3].

47 In the context of breeding programs relying on CWR genetic traits, the cultivated apple tree *Malus domestica*
48 Borkh appears to be an ideal model system for reducing the environmental impacts of food production. Indeed, *M.*
49 *domestica* is one of the most important fruit tree crops in the world (<http://faostat.fao.org/>). In Europe, the apple tree is
50 severely attacked by several pests and pathogens, and therefore apple production relies heavily on the use of
51 pesticides. To reduce reliance on pesticide applications, apple breeding programs should test potential sources of
52 resistance alleles such as the three potential local wild apple species in Eurasia. Population genetics analyses revealed
53 that the cultivated *M. domestica* originated from the wild apple *Malus sieversii* Ledeb. in the Tian Shan mountains
54 located in Central Asia^[4,5]. From there, the cultivated apple continued its journey along the Silk Routes where it
55 hybridized with local wild apple tree species: First to a little extent with *Malus orientalis* Uglitz. in the Caucasus, and
56 later on (~1500 YA), massively in Europe with *Malus sylvestris* (L.) Mill.^[4]. *Malus sieversii* is therefore considered
57 the progenitor of the cultivated apple while *M. sylvestris* is considered the second main contributor of apple genetic
58 diversity through recent wild-to-crop introgressions^[5]. Through introgression and phylogenetic closeness, *M. sieversii*,
59 *M. orientalis*, and *M. sylvestris* are all considered CWR of cultivated apple. The need to assess their value as future
60 sources for future breeding programs is urgent because these species are currently threatened by local crop-to-wild
61 gene flow^[6-9]

62 The search for resistance on *M. domestica* wild relatives has been carried out mostly against pathogens,
63 whereas studies describing resistance to pests are less frequent^[10]. Among these, the *M. floribunda* Siebold clone 821
64 proved to be a major source of resistance genes to apple scab *Venturia inaequalis* (Cooke) Winter^[11], fire blight
65 *Erwinia amylovora* Burrill^[12] and the rosy apple aphid (RAA), *Dysaphis plantaginea* Passerini^[13]. However, only a
66 handful of studies have investigated resistance to pests and pathogens of CWR involved in the domestication of the
67 apple. One study^[14] compared the levels of resistance against fire blight among 51 different genotypes of *M. sieversii*
68 collected in Kazakhstan and Kyrgyzstan. Similarly the resistance of 194 *M. sieversii* accessions belonging to four
69 distinct genetic groups^[15] was evaluated concurrently with nine different *M. domestica* cultivars^[16]. Various fire blight
70 resistance levels were observed among *M. sieversii* genetic clusters with two of them exhibiting the highest resistance
71 levels among tested genotypes. Furthermore, a distinct resistance mechanism was revealed when several wild
72 accessions were compared with cultivated apple cultivars using shoot inoculation in orchards. Specifically, fire blight
73 infection rates were lower for several *M. sieversii* accessions but when successful, infection led to greater damage in
74 these trees. Concerning resistance against insect pests, another study^[17] quantified the resistance of 19 *M. domestica*
75 cultivars and two of its CWR, *M. sylvestris* and *M. kirghisorum*, to the florivorous apple blossom weevil, *Anthonomus*
76 *pomorum* L.. The authors compared weevil resistance levels between the cultivated apple to the wild species. The
77 species *M. sylvestris* and *M. kirghisorum* Al. Fed. & Fed. appeared to be more sensitive to *A. pomorum* but also
78 supported a more abundant community of the weevil’s natural enemies.

79 Among the pool of CRW apple species with putative benefits for cultivated apple breeding programs, *M.*
80 *sylvestris* appears to be largely underexploited. For European apple production, *M. sylvestris* breeding presents several
81 advantages including a shared local environment with *M. domestica*. Previous population genetic analyses of *M*
82 *sylvestris* populations using microsatellite markers revealed five genetic clusters spread throughout Scandinavia,
83 Western Europe (mostly in France), Eastern Europe, Central Europe, and Italy^[6,7]. These five populations may
84 possess adaptive alleles associated with specific environmental conditions or local parasites. However, responses to
85 pathogens and pests among these wild apple genetic groups have yet to be explored.

86 The rosy apple aphid *D. plantaginea* (RAA) is the major aphid pest of the cultivated apple in Europe, Maghreb,
87 and North America. This aphid species feeds on sap drawn from the phloem and develops at the apex of branches
88 and/or on the most recently developed leaves. In addition to sap extraction, RAA saliva secretion provokes leaf-rolling
89 and impairs shoot growth, greatly reducing yield^[18]. Plant defenses against aphids include various strategies disrupting
90 aphid preference, particularly through their host-plant colonization process^[19] that can be characterized by the
91 potential success throughout six behavioral phases : (1) long- and short-range of host-plant perception, (2) plant
92 contact and assessment of surface cues, (3) epidermal probing, (4) stylet pathway activity, (5) phloem penetration and

93 salivation, and (6) phloem acceptance and ingestion. This colonization is finalized by entering a reproductive phase
94 which can be characterized by fecundity and adult survival traits as aphid performance. In this study, a host plant was
95 considered resistant to RAA from the moment it negatively impacted the preference and/or performance of the pest.

96 The current work explores the RAA behavior/physiology on a crop wild relative of the cultivated apple, taking
97 into account intra-specific CWR population structure and level of crop-to-wild gene flow. This manuscript describes
98 four research objectives: 1) An original sample collection was built from 42 wild apple plants grown from field-
99 collected seeds derived from three of the five known *M. sylvestris* European populations^[7] and 14 cultivated apple
100 plants derived from breeding crosses of *M. domestica* cultivars. 2) The 56 plants sampled were genetically
101 characterized using 13 microsatellite markers^[4]. Using population genetics inference, their genetic statuses were
102 determined (i.e. belonging to the Western, Scandinavian and Eastern *M. sylvestris* populations, or to the *M. domestica*
103 gene pool). The degree of crop-to-wild introgression was also assessed among these plants. These genetically
104 characterized plants were then used for behavioral and physiological assays that tested for putative RAA resistance in
105 a CWR of *M. domestica*. 3) For each plant population identified, aphid preference was tested based on feeding
106 behavior measured with the electropenetrography (EPG) technique. 4) As a proxy for measuring fitness, adult
107 fecundity, survival, and adult weight measures were used to determine their performance on each of the identified
108 plant populations^[20].

110 **Materials and methods**

111 **Plant and insect materials**

112 A total of 56 apple plants were grown from seeds and sampled for this study. Cultivated apple plants resulting
113 from crosses between various cultivated apple varieties were used (*M. domestica*, referred to as “Dom”, $N=14$, Table
114 S1). The seeds were kindly provided by INRAE IRHS Angers that performed every year crosses for apple breeding
115 programs. A total of 42 *M. sylvestris* plants were grown from field-collected seeds. These wild apple seeds originated
116 from three out of the five known European wild apple populations (referred to as Danish: Syl_Dk, French: Syl_Fr and
117 Romanian: Syl_Ro, $N=14$ per population). Each population was represented by a single sampling site, and within each
118 site, each seed was sampled on a single mother tree, so that each seedling has a different parental origin. Though *M.*
119 *domestica* is usually grafted, new plants were grown from seed to eliminate the rootstock effect.

120 After field sampling, seeds were stored at -20°C before vernalization for the experiment. Seeds were then
121 vernalized for three months at 4°C in the dark, then grown in controlled conditions for two months before being
122 individually transferred to 3 L pots containing commercial sterilized potting soil. Potted plants were grown in a
123 growth chamber for four weeks under the following conditions: $20\pm 1^{\circ}\text{C}$, $75\pm 5\%$ Relative Humidity (RH), and a 16:8
124 light:dark (L:D) photoperiod. The 56 plants were then genotyped using 13 previously published microsatellite markers
125 (see below) to confirm their genetic status (i.e., belonging to one of the *M. sylvestris* European populations or crop-to-
126 wild/wild-to-wild hybrid).

127 A single colony of *D. plantaginea* (Hemiptera: Aphididae) was used and provided by INRAE which were
128 sampled as a population in spring 2018 from an apple tree at the Agrocampus Ouest orchard (Angers, France)
129 (Philippe Robert, personal communication). This aphid population was mass reared without differentiating individual
130 aphid clones on *M. domestica* cv. “Jonagold” plants obtained by *in vitro* multiplication^[21]. Pots containing three plants
131 ($90 \times 90 \times 70$ mm) were placed in a Plexiglas cube (50 cm). Mass rearing and experiments were performed in growth
132 chambers under $20\pm 1^{\circ}\text{C}$, $60\pm 5\%$ RH, and a 16:8 L:D cycle.

133 Synchronized first instar nymphs were obtained by placing parthenogenetic adult females on plantlets for 24
134 hours before removing them. They were then reared on *M. domestica* cv. “Jonagold” plants inside Plexiglas aerated
135 boxes ($36 \times 24 \times 14$ cm) for ten days then used as the young adult RAA for the behavioral/performance experiments.

137 **Apple population genetic diversity and structure**

138 Genomic DNA was extracted with the NucleoSpin plant DNA extraction kit II (Macherey & Nagel, Düren,
139 Germany) according to the manufacturer’s instructions. Microsatellites were amplified by multiplex PCR, with the
140 Multiplex PCR Kit (QIAGEN, Inc.). We used 13 microsatellite markers, Ch01f02, Ch01f03, Ch01h01, Ch01h10,
141 Ch02c06, Ch02c09, Ch02c11, Ch02d08, Ch03d07, Ch04c07, Ch05f06, GD12, and Hi02c07 in four multiplexes
142 (MP01, MP02, MP03, MP04)^[4]. PCR were performed in a final reaction volume of 15 ml (7.5 ml of QIAGEN

143 Multiplex Master Mix, 10–20 mM of each primer, with the forward primer labelled with a fluorescent dye and 10 ng
144 of template DNA) (See ^[4] for more details). The final volume was achieved with distilled water. A touch-down PCR
145 program (initial annealing temperature of 60°C, decreasing by 1°C per cycle down to 55°C) was used. Genotyping was
146 performed on the GENTYANE platform (INRAE Clermont-Ferrand) using an ABI PRISM X3730XL, with 2 ml of
147 GS500LIZ size standard (Applied Biosystems). Alleles were scored with GENEMAPPER 4.0 software (Applied
148 Biosystems). Only multilocus genotypes with <10% missing data were retained.

149 The genetic status of each seedling was assessed using the individual-based Bayesian clustering method
150 implemented in STRUCTURE 2.3.3^[22]. STRUCTURE makes use of Markov Chain Monte Carlo (MCMC)
151 simulations to infer the proportion of ancestry of genotypes from K distinct clusters. The underlying algorithm
152 attempts to minimize deviations from Hardy–Weinberg and linkage disequilibria. STRUCTURE was run from $K=1$ to
153 $K=8$, ten independent runs were carried out for each K and 500,000 MCMC iterations were used after a burn-in of
154 50,000 steps. CLUMPAK (Greedy algorithm)^[23] was used to look for distinct modes among the 10 replicated runs of
155 each K . STRUCTURE analyses were run for the full dataset ($N = 55$, DNA could not be extracted from one Romanian
156 seedling), and included as well 40 *M. domestica* genotypes as a reference for the cultivated apple gene pool^[6]. We
157 determined the strongest level of genetic structure using ΔK ^[24], as implemented in the online post processing software
158 Structure Harvester^[25]. However, the K identified by this criterion often does not correspond to the finest biologically
159 relevant population structure^[6,7,26,27]. A lack of consideration of intraspecies genetic structure in STRUCTURE
160 analyses can bias the interpretation of introgression rates. We therefore visualized the bar plots and chose the K value
161 for which all clusters had well assigned individuals while no further well-delimited and biogeographically relevant
162 clusters could be identified for higher K values.

163 Once the best K chosen, wild plants assigned to the cultivated gene pool with a membership coefficient >0.1
164 were classified as crop-to-wild hybrids (i.e., introgressed by *M. domestica*). Once crop-wild hybrids removed, plants
165 assigned to a given wild gene pool with a cumulated membership coefficient >0.9 were defined as “pure wild”
166 individuals. Plants assigned to the wild gene pool with a cumulated membership coefficient <0.9 to a given wild apple
167 gene pool were defined as wild-wild-hybrids. The pure, crop-to-wild and wild-wild hybrids were included as factors in
168 the statistical analyses. Pure seedlings were then assigned to a population (i.e., group of plants with a cumulated
169 membership coefficient of up to 0.90 for a given wild apple cluster). Pure populations from the same geographic
170 origin (i.e., Romania or France or Denmark) which showed 1) weak genetic differentiation with other wild
171 populations 2) low number of individuals were merged. The “population” was then used as a factor for statistical
172 analyses on physiological and behavioral assays. Population genetics statistics were estimated with Genodive^[28] for
173 each “pure” wild apple population including expected and observed heterozygosities, Weir and Cockerham F -
174 statistics, Jost’s D , and deviations from Hardy–Weinberg equilibrium.

176 *Dysaphis plantaginea* feeding behavior

177 The feeding behavior of the RAA was investigated using the electrical penetration graph (EPG) method^[29].
178 Individual aphids were connected to the Giga-8 DC-EPG amplifier, each being placed on the abaxial side of a new
179 growing leaf of an individual plant. The recordings were performed continuously for 8 h during the photophase inside
180 a Faraday cage. Acquisition and analysis of the EPG waveforms were carried out using the PROBE 3.5 software (EPG
181 Systems, www.epgsystems.eu). Parameters from the recorded waveforms were calculated with the EPG-Calc 6.1.7
182 software^[30]. They were based on different EPG waveforms^[31] corresponding to: (Pr) stylet activity within plant
183 tissues; (C) stylet pathways in plant tissues except phloem and xylem; (E1) salivation in phloem elements; (E2)
184 passive phloem sap ingestion; (G) active xylem sap ingestion; and (F) derailed stylet mechanics. A total of eight plants
185 per *M. sylvestris* population (Syl_Dk, Syl_Fr, Syl_Ro) or *M. domestica* (Dom) genetic group were used for the EPG
186 measurements. EPG records were obtained from 25 aphids for *M. domestica*, and from 27 aphids for each *M.*
187 *sylvestris* population.

189 *Dysaphis plantaginea* performance

190 Two-to-three clip-cages were installed on 12 plants per genetic group identified in this study. Each cage
191 contained an individual, synchronized aphid and was enclosed on a newly grown leaf. For each synchronized adult
192 observations were assessed every 24 hours for 10 days. Survival (i.e., the duration of adult survival over the period of
193 10 days) and daily fecundity (i.e., the number of newly larviposited nymphs) were collected for 25 adults for *M.*
194 *domestica*, and 28 to 29 adults for each of the three *M. sylvestris* populations.

195

196 To measure aphid weight, newly larviposited nymphs were enclosed for nine days in clip-cages on newly
197 grown leaves similar to the above. For each plant genetic background, up to 20 aphids (*i.e.* young adults that had not
198 larviposited yet) were then collected and stored in a freezer at -80°C . Each individual aphid was weighed using an
199 electronic precision balance (Mettler M3, class 1, Max: 3g Low: 1 μg , T = -3G [dd] = 1 μg).

200

201

202 **Statistical analyses**

203 All statistical analyses were performed using the R software version 3.6.2 (The R Foundation, [https://www.r-](https://www.r-project.org/)
204 [project.org/](https://www.r-project.org/)). Generalized linear models (GLM) with a likelihood ratio and Chi-square test were used to assess
205 whether there was an effect of the host plant on aphid feeding behavior and performance. The apple tree genotype was
206 included as the main factor. Data on daily aphid fecundity and some EPG parameters describing the number of
207 occurrences of a particular phase (e.g. “n_E2”) were not normally distributed (count data), accordingly a GLM was
208 carried out using respectively a quasi-Poisson and a Poisson distribution; a quasi-likelihood function was used to
209 correct for over-dispersion, and Log was specified as the link function in the model. EPG data on feeding phase
210 durations (e.g. duration of phloem sap ingestion “s_E2”) and aphid weight were not normally distributed, so a GLM
211 using a Gamma (link = “inverse”) distribution was carried out. Analysis of the time before the first probe (“t.1Pr”) and
212 before the first phloem sap ingestion (“t.1E2”) and adult survival has been carried out using the Cox proportional
213 hazards (CPH) regression model, which is adapted to treat time-dependent parameters. Absence of an EPG reading
214 were treated as missing values. The assumption of validity of proportional hazards was validated using the function
215 “coxph” (package R: “survival”, version 3.1.8: <https://github.com/therneau/survival>). To assess whether the crop-to-
216 wild hybrid status had a plant-mediated effect on RAA feeding behavior and performance, the same statistical tests
217 were carried out with the hybrid status (*i.e.* “wild pure” or “wild-crop hybrid”) as the fixed factor while restricting the
218 data set to the *M. sylvestris* populations only.

219 Finally, because of their close genetic relatedness, the Danish and French wild apples plants (Syl_Dk and
220 Syl_Fr, respectively) were grouped together, as well as the *M. domestica* and the Romanian wild apple plants (Dom
221 and Syl_Ro, respectively). A Monte-Carlo permutation test (999 replicates) was conducted to test for the significance
222 of the differences of median of EPG phases duration, daily fecundity and weight between aphids submitted to these
223 two host groups. Analysis of the time before the first probe (“t.1Pr”) and before the first phloem sap ingestion
224 (“t.1E2”) and adult survival has been carried out using the CPH regression model. The function “randtest” (package
225 R: “ade4”: <https://cran.r-project.org/web/packages/ade4/ade4.pdf>) was run to access the significance of the observed
226 differences.

227 The fit of all GLM was controlled by a visual evaluation of residuals and QQ plots. Concerning QQ plots, the
228 distribution of the series were considered to follow the chosen theoretical law if the points of the graph were roughly
229 aligned on a straight line. Any other structuring of the points (curvature(s), many distant points, etc.) indicated the
230 opposite. GLM post-hoc comparisons were carried out by pairwise comparisons using estimated marginal means
231 (package R: “emmeans”, <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>).

232

233 **Results**

234 **Population structure and detection of crop-wild and wild-wild hybrids**

235 **Figure 1. Assessment of the genetic status of the wild and cultivated apple seedlings (*Malus sylvestris* and *Malus***
236 ***domestica*, respectively) used in this study.** Upper barplot: Population structure inferred with STRUCTURE for
237 *K*=14 for the Romanian, French, Danish *M. sylvestris* seedlings, and the *M. domestica* seedlings. STRUCTURE
238 detected eight cultivated gene pools, including the 40 *M. domestica* reference cultivars (DOM_REF, 40 reference *M.*
239 *domestica*) and the cultivated apple seedling used in this study (DOM_NA, *N* = 14). The Romanian seedlings
240 (SYL_RO, *N*=13) clustered into three clusters (orange, red and yellow colors), the French seedlings (SYL_FR, *N*=14)
241 into two clusters (light and dark blue color), and the Danish seedlings (SYL_DK, *N*=14) into one cluster (green color).
242 Lower barplot : For the sake of visualization the eight cultivated gene pools were coloured in dark blue (lower
243 barplot), the three Romanian clusters in red, the two French clusters in light blue, and the Danish stayed light green.

244 STRUCTURE analyses revealed a clear split between *M. domestica* and *M. sylvestris* seedlings for $K=3$
245 (Supplementary Fig. S1). However, failing to take the population structure of the wild species into account can lead to
246 spurious signals of introgression from crop species^[26]. We therefore analyzed the structure of *M. sylvestris* and
247 identified that the Romanian, French, and Danish *M. sylvestris* seedlings formed distinct genetic clusters from each
248 other for $K=14$. For $K=14$ the *M. domestica* seedlings used in the experiment grouped with the 40 reference *M.*
249 *domestica* (Fig. 1 and Supplementary Fig. S1). The use of K values >14 uncovered no further structure within *M.*
250 *sylvestris*, indicating $K=14$ captured $>99\%$ of the genetic variance. STRUCTURE analysis detected eight clusters for
251 *M. domestica* and six clusters for *M. sylvestris*. For $K=14$, STRUCTURE revealed a clear partition between four
252 discrete groups (Fig. 1 and Supplementary Fig. S1): 1) *M. domestica*, including *M. domestica* seedlings and the 40
253 reference samples, divided in eight admixed genetic groups, 2) the Romanian seedlings, divided into three genetic
254 groups (orange, red, yellow), 3) the French seedlings divided into two genetic groups (dark and light blue,
255 respectively), and 4) the Danish seedlings formed a single distinct genetic group (green). Note that the French and
256 Danish samples only split from $K=11$. This weak genetic structure was further validated by the relatively low F_{ST} and
257 Jost's D values among those two groups ($F_{ST}=0.08$, $P<0.01$, Table S2). We therefore used cumulative membership
258 coefficient of each seedling in the six *M. sylvestris* or the eight *M. domestica* genetic groups in subsequent analyses to
259 identify crop-to-wild and wild-wild hybrid genotypes.

260 For the 41 seedlings identified *a priori* as *M. sylvestris*, four genotypes (three French seedlings and one
261 Romanian seedling, 9.7% of the *M. sylvestris* seedlings) showed signs of introgression from *M. domestica* (i.e.,
262 cumulative membership coefficients >0.1 into the cumulated *M. domestica* gene pool, Supplementary Table S1). A
263 total of four wild-wild hybrids were detected (i.e., individuals with cumulative membership coefficients into the three
264 French genepools <0.9 , Supplementary Table S1), including three French-Danish, and one Romanian-French hybrids.
265 Genetic diversity estimates for each population, and genetic differentiation estimates among populations (excluding
266 crop-wild and wild-wild hybrids) are provided in Tables S1 and S2. Note that, once the hybrids were removed, the
267 Romanian wild apple population (FR_RO) was the genetically closest wild apple population to the cultivated apple
268 (i.e. DOM_REF and DOM_NA, $F_{ST(REF_DOM-SYL_RO)}=0.09$ and $F_{ST(REF_NA-SYL_RO)}=0.11$, respectively, $P<0.001$), and the
269 Danish and French wild apple populations were still the most genetically closely related, followed by the Romanian
270 and the French wild apple populations (Supplementary Table S2).

Table 1 Feeding phases (mean \pm standard error of the mean) of *Dysaphis plantaginea* feeding on plants belonging to the three *Malus sylvestris* populations (i.e., Danish, French and Romanian, hereafter referred to as “Syl_Dk”, “Syl_Fr”, “Syl_Ro”, respectively) and to the *Malus domestica* genepool (“Dom”).

EPG classes	GLM/Cox models	<i>Dom</i>	<i>Syl_Dk</i>	<i>Syl_Fr</i>	<i>Syl_Ro</i>
	P value				
General probing behavior and pathway phase		(<i>n</i> =25)	(<i>n</i> =27)	(<i>n</i> =27)	(<i>n</i> =27)
1. Time to first probe (min)	0.68 (NS)	20.67 \pm 6.52	22.77 \pm 7.00	14.67 \pm 2.56	25.88 \pm 6.12
2. Total duration of probing (Pr) (min)	0.60 (NS)	374.66 \pm 12.55	361.50 \pm 10.83	377.63 \pm 11.31	359.90 \pm 11.13
3. Number of probes (Pr)	0.14 (NS)	11.44 \pm 1.26	12.96 \pm 1.98	12.78 \pm 1.50	13.70 \pm 1.37
4. Total duration of pathway phase (C) (min)	0.37 (NS)	218.08 \pm 17.05	250.30 \pm 11.53	228.86 \pm 12.47	217.33 \pm 16.81
5. Number of pathway phases (C)	0.14 (NS)	16.32 \pm 1.42	18.96 \pm 1.98	17.56 \pm 1.40	18.19 \pm 1.32
Phloem phase		(<i>n</i> =21)	(<i>n</i> =22)	(<i>n</i> =20)	(<i>n</i> =20)
6. Total duration of phloem salivation (E1) (min)	<0.001(***)	9.09 \pm 1.62 a	3.04 \pm 0.59 b	2.58 \pm 0.46 b	8.28 \pm 1.62 a
7. Number of phloem salivation (E1)	<0.001(***)	4.80 \pm 0.76 a	3.22 \pm 0.53 bc	2.52 \pm 0.51 c	4.07 \pm 0.68 ab
		(<i>n</i> =19)	(<i>n</i> =20)	(<i>n</i> =18)	(<i>n</i> =19)
8. Time to first phloem ingestion (E2) (min)	0.48 (NS)	174.38 \pm 25.13	230.56 \pm 28.55	253.90 \pm 30.16	183.95 \pm 20.93
9. Total duration of phloem sap ingestion (E2) (min)	0.008 (**)	91.39 \pm 18.68 a	32.25 \pm 7.44 b	71.15 \pm 19.19 ab	86.75 \pm 18.04 a
10. Number of phloem ingestion (E2)	0.006 (**)	3.24 \pm 0.57 a	2.37 \pm 0.42 ab	1.96 \pm 0.38 b	3.19 \pm 0.57 a
Other phases		(<i>n</i> =5)	(<i>n</i> =0)	(<i>n</i> =2)	(<i>n</i> =2)
11. Total duration of xylem ingestion (G) (min)	-	134.85 \pm 52.83	-	50.04 \pm 24.70	39.07 \pm 12.62
		(<i>n</i> =13)	(<i>n</i> =14)	(<i>n</i> =10)	(<i>n</i> =12)
12. Total duration of stylet derailment (F) (min)	0.59 (NS)	82.76 \pm 13.23	93.28 \pm 10.06	108.66 \pm 19.94	109.42 \pm 23.26

The letters within a row indicate significant differences associated with pairwise comparisons using estimated marginal means.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ associated with GLM models (using respectively Poisson and Gamma distribution for the number and total duration of feeding phases) or Cox models (for “time to first phase”) (degree of freedom = 3 for each test).

272 Feeding phases and associated analytical results are described in Table 1 for aphids reared on *M. domestica*
273 and trees of each *M. sylvestris* population. Concerning general probing (Pr, parameters 1-3), the total duration of stylet
274 activity in the plants (Pr, parameter 2) lasted on average about 6 hours (out of 8 recorded hours) and was not
275 significantly different among aphids, whatever the genetic background of their host plant (GLM using Gamma
276 distribution: $\chi^2=0.60$, Df=3, P=0.60). This global activity was composed of an average of 13 probing events (Pr,
277 parameter 3), again without any significant difference between aphids on the host plants with different genetic
278 background (GLM using Poisson distribution: $\chi^2=0.14$, Df=3, P=0.14). This is despite the fact that the mean number
279 of probes tended to be smaller on *M. domestica* host plants. Similarly, there was no significant difference between the
280 time to first probe (Pr, parameter 1) (CPH: $\chi^2=1.49$, Df=3, P=0.68). The pathway phase (C, parameters 4 and 5)
281 comprised on average of 60 % of the total activity. There was no significant difference for its total duration (C,
282 parameter 4) (GLM using Gamma distribution: $\chi^2=3.13$, Df=3, P=0.37) or for the number of occurrence (C, parameter
283 5) (GLM using Poisson distribution: $\chi^2= 5.49$, Df=3, P=0.14).

284 Concerning the phloemian phase (parameters 6-10), most parameters revealed differences among aphids fed
285 on trees belonging to the three wild apple populations and the cultivated gene pool. The mean duration of salivation
286 within phloem was significantly shorter for aphids submitted to French and Danish *M. sylvestris* plants compared with
287 aphids submitted to *M. domestica* and Romanian *M. sylvestris* plants (E1, parameter 6) (GLM using Gamma
288 distribution: $\chi^2=35.02$, Df=3, P < 0.001). Aphids fed on Danish and French *M. sylvestris* plants salivated three times
289 less compared to aphids fed on *M. domestica* and Romanian *M. sylvestris* plants. Aphids submitted to French and
290 Danish *M. sylvestris* plants displayed a smaller number of salivations within phloem compared to aphids on *M.*
291 *domestica* whereas aphids submitted to French *M. sylvestris* plants had a significantly shorter number of salivations
292 within phloem compared to those on Romanian *M. sylvestris* plants (E1, parameter 7) (GLM using Poisson
293 distribution: $\chi^2=21.55$, Df=3, P < 0.001). The mean duration of phloem ingestion (E2, parameter 9) was significantly
294 shorter for aphids on the Danish *M. sylvestris* compared to aphids on *M. domestica* and the Romanian *M. sylvestris*
295 plants (GLM using Gamma distribution: $\chi^2=12.50$, Df=3, P=0.08). The mean proportion of the time dedicated to
296 phloem ingestion (E2) was variable depending on the host plant genetic background: from 9 % (Syl_Dk) to 24 %
297 (Syl_Ro/Dom) within the general probing activity. The number of phloem ingestion (E2, parameter 10) was
298 significantly smaller for aphids submitted to the French *M. sylvestris* compared to those submitted to *M. domestica*
299 and the Romanian *M. sylvestris* plants (GLM using Poisson distribution: $\chi^2=11.89$, Df=3, P=0.06).

300 Finally, the duration of the time needed by an aphid to reach the phloem (E2, parameter 8) tended to be shorter
301 on *M. domestica* or Romanian *M. sylvestris* plants (around 3 hours) than on Danish and French *M. sylvestris* plants
302 (around 4 hours), though no significant difference was observed (CPH: $\chi^2= 2.48$, Df=3, P=0.48). Considering xylem
303 ingestion (G, parameter 11) aphids submitted to Danish *M. sylvestris* plants did not ingest raw sap, whereas a few
304 aphids ingested xylem on the French and Romanian *M. sylvestris* plants. Altogether, not enough aphids displayed this
305 behavior to conduct statistical analysis. Almost half of the aphids presented stylet derailment (F, parameter 12) for an
306 average total duration of roughly 1.5 hours, that was not statistically different between the different host plant genetic
307 backgrounds (GLM using Gamma distribution: $\chi^2=0.59$, Df=3, P=0.59).

308 Considering the genetic proximity of the Danish and French *M. sylvestris* populations (Table S2), as well as
309 the genetic proximity of the Romanian *M. sylvestris* and *M. domestica*, the two-by-two pairing of datasets (i.e.,
310 Syl_Dk/Syl_Fr vs Dom/Syl_Ro) revealed that the total duration of phloem salivation (Monte-Carlo permutation test,
311 P=0.001) and the total duration of phloem sap ingestion (Monte-Carlo permutation test, P=0.016) were significantly
312 longer for the Dom/Syl_Ro pair. Regardless *M. sylvestris* populations, the duration of phloem sap ingestion phase was
313 significantly affected by the host plant hybrid status and was shorter for hybrids compared to pure *M. sylvestris* (GLM
314 using Gamma distribution: $\chi^2=4.23$, Df=1, P=0.04).

315

316

317 **Effect of the host plant population on *Dysaphis plantaginea* fitness parameters**

318 **Table 2** Fitness parameters (mean \pm standard error of the mean) for *Dysaphis plantaginea* reared on plants belonging
 319 to three *Malus sylvestris* populations (Danish, French and Romanian, i.e. hereafter referred as to “Syl_Dk”, “Syl_Fr”,
 320 “Syl_Ro”, respectively) and to the *Malus domestica* genepool (“Dom”).

Parameters	GLM/Cox models	<i>Dom</i>	<i>Syl_Dk</i>	<i>Syl_Fr</i>	<i>Syl_Ro</i>
	P value	<i>n</i> =25	<i>n</i> =28	<i>n</i> =29	<i>n</i> =28
Daily fecundity	0.69 (NS)	2.33 \pm 0.26	2.46 \pm 0.26	2.38 \pm 0.22	2.08 \pm 0.22
Survival (days)	0.34 (NS)	9.20 \pm 0.34	9.75 \pm 0.14	9.28 \pm 0.39	9.50 \pm 0.31
		<i>n</i> =56	<i>n</i> =73	<i>n</i> =80	<i>n</i> =80
Aphid weight (μ g)	0.043 (*)	557.88 \pm 39.05 a	427.92 \pm 28.54 b	476.30 \pm 22.84 ab	490.26 \pm 32.20 ab

321 The letters within a row indicate significant differences associated with pairwise comparisons using estimated marginal means.
 322 *P<0.05; **P<0.01; ***P<0.001 associated with GLM models (using respectively a quasi-Poisson for daily fecundity and Gamma
 323 distribution aphid weight) or Cox models (for survival) (degree of freedom = 3 for each test).

324 The impact of host plant genetic background on RAA fitness and associated statistical analyses are presented
 325 in Table 2. There was no significant difference for daily fecundities (GLM using quasi-Poisson distribution: $\chi^2=1.49$,
 326 Df=3, P > 0.05) between aphids raised on the four plant genetic backgrounds. Similarly, there was no significant
 327 difference for survival (CPH: $\chi^2=3.39$, Df=3, P > 0.05). Aphid weight was significantly impacted by the plant genetic
 328 background: the weight of aphids raised on *M. domestica* was significantly greater (GLM using Gamma distribution:
 329 $\chi^2=8.10$, Df=3, P < 0.05) than that of aphids raised on the Danish *M. sylvestris* plants. Again taking into consideration
 330 the genetic proximity between the Danish and French *M. sylvestris* as well as the Romanian *M. sylvestris* and *M.*
 331 *domestica*, the two-by-two pairing of datasets (Syl_Dk/Syl_Fr vs Dom/Syl_Ro) revealed no significant difference for
 332 all fitness parameters. Regardless *M. sylvestris* population, aphid weight was significantly affected by the host-plant
 333 hybrid status: aphids raised on hybrids displayed smaller weights compared to those on pure *M. sylvestris* (GLM using
 334 Gamma distribution: $\chi^2=5.16$, Df=1, P = 0.02), whereas neither fecundity nor survival were impacted.

335
 336 **Discussion**

337 This is the first study reporting differences in phytophagous pest preference that are congruent with the
 338 genetic relationship between a wild relative and a cultivated plant with which it has introgressed. Population genetic
 339 analyses also revealed weak genetic differentiation between the Danish and French *M. sylvestris* wild apple
 340 populations. Accordingly, behavioral assays of aphids submitted to plants from these two populations showed similar
 341 patterns suggesting antixenosis resistance^[32]. Likewise, the Romanian *M. sylvestris* host population was strongly
 342 differentiated from *M. domestica* but was also the most closely genetically related to the wild apple population.
 343 Consistent with this observation, the Romanian wild apple population and *M. domestica* showed comparable levels of
 344 sensitivity to RAA. Crop-to-wild introgression appeared to drive resistance to RAA independent of population genetic
 345 structure.

346 This study revealed a putative link between aphid preference and the genetic structure among wild and
 347 cultivated apple populations. The population structure inferred here stingingly matched the one previously observed
 348 for the European wild apple^[7,33], with five main populations in Europe of which, an Eastern, a French and a
 349 Scandinavian. We showed that the French and Danish populations were the genetically closest and sharing the highest
 350 number of wild-wild hybrids. Accordingly, we can note that aphids showed similar patterns of feeding behavior and
 351 performance when submitted to the two most closely related wild apple populations (the Danish and French
 352 populations). When aphids were subjected to the Romanian wild apple population and the cultivated apple, similar
 353 patterns of aphid behavior and performance were recorded and were associated with higher preference for the
 354 Romanian host-plant than aphids submitted to the Danish and French populations. Aphid behavior was actually
 355 congruent with the level of genetic differentiation between the Romanian wild apple and the cultivated apple. Once
 356 the recent crop-to-wild hybrids were removed, the Romanian population appeared to be the closest wild apple relative

357 to cultivated *M. domestica*. Genetic proximity of populations is known to drive patterns of resistance against
358 pathogens in the wild apple *M. sieversii*^[15,16]. Previous studies revealed variable resistance against pests and pathogens
359 among *M. domestica* CWR^[14,16]. In this study, the genetic differentiation between the two paired groups
360 (Syl_Dk/Syl_Fr vs Dom/Syl_Ro) might be associated with a phenotypic differentiation associated with the differences
361 observed in terms of RAA feeding behavior. The genetic proximity of the French/Danish and Romanian/cultivated
362 apple may reflect common evolutionary history, however further investigations are required concerning the
363 evolutionary history of the cultivated apple in Europe. In particular, the relative contributions of each wild apple
364 population, especially the Romanian, to the cultivated apple gene pool remains unknown. Addressing this issue would
365 require much larger sampling among European apple seeds.

366 Not considering genetic proximity among populations, but only population structure, behavioral analyses with
367 EPG demonstrated a generalized activity for the rosy apple aphid, which did not depend on the genetic background of
368 the host plant. Our results showed that whatever the plant genetic background, the time to first probe was not delayed,
369 meaning that the possible influence of epidermal barriers and/or putative VOC repulsive effects could be excluded;
370 plant volatile organic compound (VOC) on leaf surface could indeed impact aphids behavior^[19]. A delayed aphid
371 stylet activity is considered to be due to epidermic factors, as the second phase of host selection involves the
372 assessment of plant surface cues by the aphid. Features such as a thick cuticle or the presence of trichomes are
373 physical parameters that may play a role in aphid resistance^[34]. Stylet derailment was displayed on every plant genetic
374 background in the same range of mean duration and the pathway phase was not influenced by the plant genetic
375 background. This means that putative mild physicochemical resistance is present of mesophyll tissues in both wild and
376 cultivated apple. In contrast with the above, significant differences were observed between the four plant genetic
377 backgrounds in phloem-related behavior. The phloemian activity was significantly reduced in terms of the duration of
378 both salivation and ingestion for aphids submitted to the Danish and French *M. sylvestris* populations. Since the
379 average time to reach the phloem was not significantly different between the four plant genetic backgrounds, these
380 differences did not appear to be linked to physical characteristics but due to the phloem chemical composition.
381 Comparison of ascorbic acid glycoside (AAG) content in *M. domestica*, *M. sylvestris* and *M. sieversii* apple fruits
382 revealed that accessions of *M. sylvestris* were distinguished by higher concentrations of AAG^[35]. A difference in terms
383 of phenolic compounds among *M. sylvestris* populations could be a possible factor explaining the contrasted
384 phloemian activities observed. In fact, phenolic profile of various *M. domestica* cultivars apple fruits can be linked to
385 field RAA resistance^[36]. Further studies involving choice assays towards the four genetic backgrounds should provide
386 a better understanding of RAA preference, and especially of long and short range host-plant perception. Despite the
387 differences recorded in the feeding behavior, no differences were observed in two of the three RAA fitness parameters
388 (survival and fecundity) regardless of host-plant genetic background. As high proportions of aphids could initiate
389 reproduction before accessing the phloem^[37], the results concerning fecundity may be consistent with the absence of
390 significant differences in pathway phase parameters. Aphid biomass assays revealed that only aphids submitted to the
391 cultivated apple tree had greater weights than those submitted to the Danish *M. sylvestris*. This is consistent with the
392 fact that the shortest sap ingestion was observed for aphids submitted to the Danish *M. sylvestris*, whereas aphids
393 reared on *M. domestica* exhibited the longest sap ingestion. Sap ingestion is known to be positively correlated with
394 growth^[38], thus the contrasted preference of RAA was not reflected in RAA performance, except for adult weight.

395 For the first time, our study also shed light on the impact of domestic introgression in *M. sylvestris* on RAA
396 preference and performance. Previous studies already detected substantial crop-to-wild gene flow in apple trees in
397 Europe^[7,8]. Here, we confirmed the occurrence of ongoing crop-to-wild gene flow for the European wild apple. We
398 detected 11% of crop-to-wild hybrids in our dataset, which is half less than previous estimates (23%)^[7]. Several
399 reasons can explain this discrepancy, including the narrow spatial geographic area investigated here (four locations
400 versus 62 locations previously) and the lower number of samples used (here 42 vs 1889 *M. sylvestris* trees previously).
401 Note however that the aim of our study was not to investigate the large-scale crop-wild gene flow in *M. sylvestris* but
402 assessing the genetic status of the seedlings used for aphid physiological and behavioral assays to control this effect in
403 the statistical models. Yet, in comparison to previous studies, the detection of crop-wild hybrids in seeds collected in
404 2016-2017 are proof that recent ongoing crop-to-wild gene flows are still at work in apples, especially in the French
405 populations. Indeed, previous studies rather investigated historical crop-to-wild gene flow as they did not sampled
406 seeds but much older mother trees. It is also interesting to note that we detected a clear effect of the hybrid status on
407 RAA preference - in terms of phloem sap ingestion - and on RAA performance - in terms of aphid weight. Our
408 bioassays on aphids revealed that pure *M. sylvestris* were more suitable hosts to RAA than crop-to-wild hybrids. A
409 previous study revealed that crop-to-wild hybrids showed higher plant growth and pollination rates compared to pure
410 wild apples^[8]. Here our results would suggest that higher fitness of crop-to-wild hybrids is not only expressed for early
411 developmental traits but is also associated with higher resistance abilities to RAA attacks. However, this question
412 would require further investigation.

413 Finally, it is worth questioning to what extent *Malus sylvestris* could represent a putative genetic source of
414 resistance for *Malus domestica* breeding programs. The *M. domestica* genetic group studied here could be considered

415 as susceptible when compared with the resistance/susceptibility of *M. domestica* to RAA demonstrated by a previous
416 study also using EPG^[39]. Regarding general probing activity and phloemian phases, the mean values of
417 electropenetrography parameters obtained here for aphids submitted to the *M. domestica* genetic group appeared to be
418 close to values obtained on a susceptible cultivar (*M. domestica* cv. Golden Delicious) in comparison to a resistant one
419 (*M. domestica* cv. Florina)^[39]. The latter has been identified as strongly resistant to the RAA in numerous studies<sup>[13,40-
420 42]</sup>. In our study, we observed a gradient of resistance to the RAA for the European wild apple, but less marked than
421 for *M. domestica*. Among *M. domestica* cultivars, resistant cultivars impacted RAA preference through both epidermic
422 and phloemian factors^[39]. Indeed, compared to that of aphids on susceptible controls, the feeding behavior of aphids
423 on the resistant Florina cultivar revealed shorter durations for general probing (Pr), phloem salivation phase (E1),
424 xylem sap ingestion (G). Also, none of the aphids ingested phloem sap (i.e the E2 phase was null) and a significantly
425 longer total duration of stylets derailment was observed when submitted to this resistant cultivar. In our case, *M.*
426 *sylvestris* did only impact RAA preference through the phloemian phase, with most individuals able to ingest sap,
427 although lasting for short duration. Therefore, RAA preference in the European wild apple studied here only involved
428 one factor, the phloemian phase, in contrast to what was previously observed among *M. domestica* cultivars. Besides,
429 the two-factor aphid response on the *M. domestica* host is also associated with lower RAA performance^[13,40-42].
430 Strikingly, our study showed that only the Danish *M. sylvestris* population negatively impacted both RAA preference
431 and performance compared to *M. domestica*. Thus, the Danish *M. sylvestris* population may be more resistant against
432 RAA than *M. domestica* and would appear to represent a potential source of resistance for *M. domestica* breeding
433 programs, although this CRW candidate did not impact RAA fecundity or survival. However, our results are
434 consistent with previous investigations of pests and pathogens resistance in *Malus domestica* CWR which were
435 involved in its domestication, in which some wild accessions were as sensitive as the cultivated ones^[16,17]. Resistance
436 against the pest of CWR would be mainly indirect, as they support greater communities of natural enemies^[17]. Thus,
437 as we did not proceed to field validation of our results, we may have overlooked some components of CWR
438 resistance. Investigating RAA performance for several aphid generations would also be worth carrying out on the
439 Danish *M. sylvestris* population, as this host negatively impacted aphid behavior and fitness in terms of weight and
440 these negative effects may have a greater impact over generations. A greater diversity of new resistance genes or
441 alleles against RAA may be present in CWR gene pool involved in the apple domestication. CWR are however largely
442 neglected when it comes to studying their resistance to RAA. To better understand resistance differences at an
443 interspecific level, more CWR species have to be included in experiments, such as *M. orientalis* and *M. sieversii*.
444 However, this would involve exposing RAA to apple CWRs that are absent in its natural environment. Future studies
445 may also investigate cross infestations of aphid populations from different parts of Europe onto CWR populations to
446 truly test for RAA local adaptation. To test for the influence of maternal priming, it would also be interesting to
447 compare relative preferences among the aphids reared on *M. domestica* to the relative preference among aphids reared
448 on *M. sylvestris*.

449 To conclude, this work tested for the first time preferences and survival of a main pest of apple trees, among
450 genetically distinct groups including wild and cultivated host plants. Identification of resistance adaptations among
451 wild genotypes may help design strategies to improve *M. domestica* plant productivity. But above all, the present
452 study reveals that the search for resistant CWR must not only be based on a genetic structure of wild populations but
453 also on the crop-to-wild gene flow that appears to substantially drive resistance to RAA. In that sense French wild
454 apples, which showed high level of crop-to-wild gene flow, may be good candidates for future breeding programs.

457 References

- 458 1. Lechenet, M., Dessaint, F., Py, G., Makowski, D. & Munier-Jolain, N. Reducing pesticide use while preserving crop
459 productivity and profitability on arable farms. *Nat. Plants* **3**, 1–6 (2017).
- 460 2. Zhang, H., Mittal, N., Leamy, L. J., Barazani, O. & Song, B.-H. Back into the wild—Apply untapped genetic diversity of
461 wild relatives for crop improvement. *Evol. Appl.* **10**, 5–24 (2017).
- 462 3. Hodgkin, T. & Hajjar, R. Using crop wild relatives for crop improvement: trends and perspectives. in *Crop wild relative
463 conservation and use* (eds. Maxted, N. *et al.*) 535–548 (CABI, 2007).
- 464 4. Cornille, A. *et al.* New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple
465 to the Genome of Cultivated Varieties. *PLOS Genet.* **8**, e1002703 (2012).

- 466 5. Cornille, A. *et al.* A Multifaceted Overview of Apple Tree Domestication. *Trends Plant Sci.* **24**, 770–782 (2019).
- 467 6. Cornille, A., Gladieux, P. & Giraud, T. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of
468 the cultivated apple. *Evol. Appl.* **6**, 737–748 (2013).
- 469 7. Cornille, A. *et al.* Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A basis for conservation and
470 breeding programs in apples. *Evol. Appl.* **8**, 373–384 (2015).
- 471 8. Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A. & Giraud, T. Crop-to-wild gene flow and its fitness consequences for a
472 wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. *Evol. Appl.* **10**, 180–188 (2017).
- 473 9. Ruhsam, M., Jessop, W., Cornille, A., Renny, J. & Worrell, R. Crop-to-wild introgression in the European wild apple *Malus*
474 *sylvestris* in Northern Britain. *For. Int. J. For. Res.* **92**, 85–96 (2018).
- 475 10. Keller-Przybyłkiewicz S. & Korbin M. U. The history of mapping the apple genome. *Folia Hort.* **25/2**, 161-168 (2013).
- 476 11. Dunemann, F., Kahnau, R. & Schmidt, H. Genetic Relationships in *Malus* Evaluated by RAPD ‘Fingerprinting’ of Cultivars
477 and Wild Species. *Plant Breed.* **113**, 150–159 (1994).
- 478 12. Durel, C.-E., Denancé, C. & Brisset, M.-N. Two distinct major QTL for resistance to fire blight co-localize on linkage group
479 12 in apple genotypes ‘Evereste’ and *Malus floribunda* clone 821. *Genome* **52**, 139–147 (2009).
- 480 13. Dall’Agata, M. *et al.* Identification of candidate genes at the Dp-fl locus conferring resistance against the rosy apple aphid
481 *Dysaphis plantaginea*. *Tree Genet. Genomes* **14**, 12 (2018).
- 482 14. Luby, J. J., Alspach, P. A., Bus, V. G. M. & Oraguzie, N. C. Field Resistance to Fire Blight in a Diverse Apple (*Malus sp.*)
483 Germplasm Collection. *J. Am. Soc. Hortic. Sci.* **127**, 245–253 (2002).
- 484 15. Richards, C. M. *et al.* Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated
485 apple. *Tree Genet. Genomes* **5**, 339–347 (2009).
- 486 16. Harshman, J. M. *et al.* Fire Blight Resistance in Wild Accessions of *Malus sieversii*. *Plant Dis.* **101**, 1738–1745 (2017).
- 487 17. Knuff, A. K., Obermaier, E. & Mody, K. Differential susceptibility and suitability of domestic and wild apple species for a
488 florivorous weevil and its parasitoids. *J. Appl. Entomol.* **141**, 285–299 (2017).
- 489 18. Qubbaj, T., Reineke, A. & Zebitz, C. P. W. Molecular interactions between rosy apple aphids, *Dysaphis plantaginea*, and
490 resistant and susceptible cultivars of its primary host *Malus domestica*. *Entomol. Exp. Appl.* **115**, 145–152 (2005).
- 491 19. Powell, G., Tosh, C. R. & Hardie, J. Host plant selection by aphids: Behavioral, Evolutionary, and Applied Perspectives.
492 *Annu. Rev. Entomol.* **51**, 309–330 (2006).
- 493 20. Beukeboom, L. W. Size matters in insects – an introduction. *Entomol. Exp. Appl.* **166**, 2–3 (2018).
- 494 21. Druart, P. Optimization of culture media for in vitro rooting of *Malus domestica* Borkh. cv. Compact Spartan. *Biol. Plant.* **39**,
495 67–77 (1997).
- 496 22. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*
497 **155**, 945–959 (2000).
- 498 23. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering

- 499 modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
- 500 24. Evanno, G., S. Regnaut, & J. Goudet. Detecting the number of clusters of individuals using the software STRUCTURE: a
501 simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
- 502 25. Earl D. A., vonHoldt B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and
503 implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
- 504 26. Kalinowski, S. T. The computer program STRUCTURE does not reliably identify the main genetic clusters within species:
505 simulations and implications for human population structure. *Heredity* **106**, 625–632 (2011).
- 506 27. Puechmaille, S. J. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling
507 and new estimators alleviate the problem. *Mol. Ecol. Resour.* **16**, 608–627 (2016).
- 508 28. Meirmans, P. G. & Tienderen, P. H. V. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of
509 asexual organisms. *Mol. Ecol. Notes* **4**, 792–794 (2004).
- 510 29. Tjallingii, W. F. Electrical nature of recorded signals during stylet penetration by aphids. *Entomol. Exp. Appl.* **38**, 177–186
511 (1985).
- 512 30. Giordanengo, P. EPG-Calc: a PHP-based script to calculate electrical penetration graph (EPG) parameters. *Arthropod-Plant*
513 *Interact.* **8**, 163–169 (2014).
- 514 31. Tjallingii, W. F. & Esch, T. H. Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiol.*
515 *Entomol.* **18**, 317–328 (1993).
- 516 32. Smith, C. M. *Plant Resistance to Arthropods: Molecular and Conventional Approaches*. (Springer Science & Business
517 Media, 2005).
- 518 33. Cornille, A. *et al.* Postglacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to
519 the domesticated apple. *Mol. Ecol.* **22**, 2249–2263 (2013).
- 520 34. Nalam, V., Louis, J. & Shah, J. Plant defense against aphids, the pest extraordinaire. *Plant Sci.* **279**, 96–107 (2019).
- 521 35. Richardson, A. T. *et al.* Discovery of a stable vitamin C glycoside in crab apples (*Malus sylvestris*). *Phytochemistry* **173**,
522 112297 (2020).
- 523 36. Berrueta, L. A. *et al.* Relationship between hydroxycinnamic acids and the resistance of apple cultivars to rosy apple aphid.
524 *Talanta* **187**, 330–336 (2018).
- 525 37. Tosh, C. R., Powell, G. & Hardie, J. Maternal reproductive decisions are independent of feeding in the black bean aphid,
526 *Aphis fabae*. *J. Insect Physiol.* **48**, 619–629 (2002).
- 527 38. Kang, Z.-W., Liu, F.-H., Zhang, Z.-F., Tian, H.-G. & Liu, T.-X. Volatile β -Ocimene Can Regulate Developmental
528 Performance of Peach Aphid *Myzus persicae* Through Activation of Defense Responses in Chinese Cabbage *Brassica*
529 *pekinensis*. *Front. Plant Sci.* **9**, 708 (2018).
- 530 39. Marchetti, E. *et al.* Tissue location of resistance in apple to the rosy apple aphid established by electrical penetration graphs.
531 *Bull. Insectology* **62** (2), 203–208 (2009).
- 532 40. Arnaudov, V. & Kutinkova, H. Infestation by the rosy apple aphid (*Dysaphis plantaginea* PASS., *Homoptera: Aphididae*).

- 533 *J. Fruit Ornament. Plant Res.* **14**, 137–142 (2006).
- 534 41. Dapena, E., Miñarro, M. & Blázquez, M. D. Evaluation of the resistance to the rosy apple aphid using a genetic marker. *Acta*
535 *Hortic.* **814**, 787–790 (2009).
- 536 42. Pagliarani, G. *et al.* Fine mapping of the rosy apple aphid resistance locus Dp-fl on linkage group 8 of the apple cultivar
537 ‘Florina’. *Tree Genet. Genomes* **12**, 56 (2016).
- 538

539 **Author Contribution Statement**

540 Conceived and designed the experiments: AC XC AA GD TD. Performed the experiments: AC XC TD. Analyzed the data: AC
541 XC TD. Wrote the paper: AC XC AA GD TD TH

542

543 **Acknowledgments :** This work was supported by project PROVERBIO (Protection of orchards by biological control: an adapted
544 selection of auxiliaries) financed by the European Union, under the scope of the FEDER program and INTERREG initiative :
545 Programme 2014 - 2020 INTERREG V-A Belgium - France (France - Wallonie - Vlaanderen), and the ATIP-Avenir CNRS-
546 Inserm program. We thank Philippe Robert (INRAE Agrocampus Ouest, Angers, France) for providing the population of
547 *Dysaphis plantaginea*. Yvelise Fourdrain is thanked for her help in plant cultures and aphid mass rearing. Carine Remoué, Agnès
548 Rousselet, and Matthieu Falque are thanked for their contribution in DNA plant extraction and microsatellite plant genotyping.
549 We thank the INRAE GENTYANE platform for SSR genotyping (Clermont Ferrand), and its group leader Charles Poncet.
550 Quentin Chesnais and Pedro Poli are thanked for their assistance in the statistical analysis of the data.

551

552 **Compliance with ethical standards:**

553 **Conflict of interest** The authors declare no conflict of interest.

554 **Ethical approval** The article does not contain any studies with human participants or vertebrate animals.