Large-scale geographic survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa

Olvera-Vazquez S.G.[1], Remoué C.[1], Venon A.[1], Rousselet A.[1], Grandcolas O.[1], Azrine M.[1], Momont L.[1], Galan M.[2], Benoit L.[2], David G. M.[3], Alhmedi A.[4], Beliën T.[4], Alins G.[5], Franck P.[6], Haddioui A.[7], Jacobsen S.K.[8], Andreev R.[9], Simon S.[10], Sigsgaard L. [8], Guibert E.[11], Tournant L.[12], Gazel F.[13], Mody K.[14], Khachtib Y. [7], Roman A.[15], Ursu T.M.[15], Zakharov I.A. [16], Belcram H.[1], Harry M.[17], Roth M.[18], Simon J.C.[19], Oram S.[20], Ricard J.M.[11], Agnello A.[21], Beers E. H.[22], Engelman J.[23], Balti I.[24], Salhi-Hannachi A[24], Zhang H.[25], Tu H. [25], Mottet C.[26], Barrès B.[26], Degrave A.[27], Razmjou J. [28], Giraud T.[3], Falque M.[1], Dapena E.[29], Miñarro, M.[29], Jardillier L.[3], Deschamps P.[3], Jousselin E.[2], Cornille, A.[1]

1. Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France
2. INRAE, UMR 1062, Centre de Biologie pour la Gestion des Populations CBGP (INRAE, IRD, CIRAD, Montpellier SupAgro), Montferrier-sur-Lez, 34980, France.
3. Ecologie Systematique et Evolution, CNRS, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France
4. Department of Zoology, Fruit Research Center (pcfruit *npo*), Sint-Truiden, Belgium.
5. Institut de Recerca i Tecnologia Agroalimentàries, IRTA-Fruit Production, PCiTAL, Parc de Gardeny, edifici Fruitcentre, 25003 Lleida, Spain
6. INRAE, UR1115, Plantes et Système de cultures Horticoles (PSH), Agroparc CS 40509, Avignon, F-84000, France.
7. Laboratory of Biotechnology and Valorisation of Plant Genetic Resources, Faculty of Sciences and Techniques, University of Sultan Moulay Slimane, Beni Mellal, Morocco

8. Department of Plant and Environmental Sciences, University of Copenhagen, Denmark.
9. Department of Entomology, Agricultural University, Plovdiv, Bulgaria.
10. Unité Expérimentale UERI 695, INRAE, 460 Chemin de Gotheron, F-26320 Saint-Marcel-lès-Valence, France
11. CTIFL, Centre opérationnel de Balandran
12. Fédération Régionale de Défense contre les Organismes Nuisibles Nord Pas-de-Calais, Loos-en-Gohelle, France
13. BORDX ARBORI UE - Unité d'arboriculture
14. Department of Applied Ecology, Hochschule Geisenheim University, Geisenheim, Germany
15. NIRDBS, Institute of Biological Research Cluj-Napoca, 400015, Cluj-Napoca, Romania.
16. Vavilov Institute of General Genetics RAS, Moscow, Russia
17. Laboratoire Évolution, Génomes, Comportement, Écologie, UMR9191 CNRS/IRD/Université Paris-Saclay, Gif-sur-Yvette, France
18. GAFL, INRAE, 84140, Montfavet, France
19. INRAE, UMR IGEPP, Domaine de la Motte, Le Rheu, France
20. People's Trust for Endangered Species, London, UK
21. Department of Entomology, Cornell AgriTech, Cornell University, Geneva, NY
22. Department of Entomology, Tree Fruit Research and Extension Center, Washington State University, Wenatchee, WA
23.Virginia Tech, Alson H. Smith, Jr. Agricultural Research and Extension Center, 595 Laurel Grove Road, Winchester, VA 22602, USA
24. Department of Biology, Faculty of Science of Tunis, University of Tunis El Manar, 2092, Tunis, Tunisia, Salhi-Hannachi A: Department of Biology, Faculty of Science of Tunis, University of Tunis, Tunisia.

25. Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, 450009, Zhengzhou, China.

26. Université de Lyon, Anses, INRAE, USC CASPER, 69007 Lyon, France

27. IRHS, INRAE, Agrocampus-Ouest, Université d'Angers, SFR 4207 QUASAV, Beaucouzé, France

28. Department of Plant Protection, Faculty of Agriculture and natural Resources, University of Mohaghegh Ardabili, Iran

29. Servicio Regional de Investigación y Desarrollo Agroalimentario, Ctra. AS-267, PK 19, E-33300, Villaviciosa, Asturias, Spain

Corresponding author: amandine.cornille@inrae.fr

Summary

With frequent host shifts involving the colonization of new hosts across large geographical ranges, crop pests are good models for examining the mechanisms of rapid colonization. The microbial partners of pest insects may also be involved in or affected by colonization processes, which has been little studied so far. We investigated the demographic history of the rosy apple aphid, *Dysaphis plantaginea*, a major pest of the cultivated apple (*Malus domestica*) in Europe, North Africa and North America, as well as the diversity of its microbiota. We genotyped a comprehensive sample of 714 colonies from Europe, Morocco and the US using mitochondrial (*CytB* and *CO1*), bacterial (16s rRNA and TrnpB), and 30 microsatellite markers. We detected five populations spread across the US, Morocco, Western and Eastern Europe and Spain. Populations showed weak genetic differentiation and high genetic diversity, except the Moroccan and the North American that are likely the result of recent colonization events. Coalescent-based inferences revealed high levels of gene flow among populations during the colonization but did not allow determining the sequence of colonization of Europe, America and Morroco by *D. plantaginea*, likely because of the weak genetic differentiation and the occurrence of gene flow among populations. We found that *D. plantaginea* rarely hosts other endosymbiotic bacteria than its obligate nutritional symbiont *Buchnera aphidicola*. This suggests that secondary endosymbionts did not play an important role in the rapid spread of the rosy apple aphid. These findings have fundamental importance for understanding pest colonization processes and implications for sustainable pest control programs.

Keywords: colonization, expansion, pest, apple, aphid, endosymbiotic bacteria, *Buchnera*, gene flow, ABC-RF.

Introduction

Understanding the evolutionary processes underlying the colonization of new environments and the range expansion of species is a key goal in evolutionary biology (Angert, Bontrager, & Ågren, 2020; Austerlitz, Jung-Muller, Godelle, & Gouyon, 1997; Excoffier, Foll, & Petit, 2009; Hoffmann & Courchamp, 2016; Rius & Darling, 2014). Crop pests in agro-ecosystems, with their frequent colonization of new hosts across large geographic ranges, are good models to study the mechanisms of rapid colonization and range expansion (Garnas et al., 2016; Gladieux et al., 2014; Stukenbrock & McDonald, 2008). The key questions relating to the evolutionary processes underlying the colonization, spread and success of crop pests pertain to the geographic origin of the source population, the location of the colonization routes, the extent to which genetic diversity is reduced via founder effects (Blakeslee et al., 2019) and the extent of gene flow among populations during the spread of the plant parasite (Stukenbrock, 2016). Current and future threats to biodiversity and their consequences on ecosystem health and services make these questions more relevant than ever. Understanding the routes of pest colonization contributes greatly to the efforts to protect crops against future pest emergence and therefore has direct implications for breeding and agronomic programs that develop biological methods of parasite control (Estoup & Guillemaud, 2010; Fraimout et al., 2017; Lawson Handley et al., 2011; Turcotte, Araki, Karp, Poveda, & Whitehead, 2017).

Historically, gene genealogies have been a rich source of information into a species' evolutionary history that could be applied to the study of colonization (Bloomquist, Lemey, & Suchard, 2010; Hickerson et al., 2010; Posada & Crandall, 2001). The characterization of population structure, genetic diversity and demographic history (divergence time, migration rates among populations and effective population size) are also essential to understand the evolutionary processes underlying rapid colonization and range of geographic expansion (Excoffier et al., 2009). Approximate Bayesian computation methods (referred to as "ABC"

115 hereafter) provide a robust framework for inferring a species' history by allowing the
116 comparison of alternative demographic models and the estimation of their associated
117 parameters (divergence time, migration rate, effective population size) (Bertorelle, Benazzo,
118 & Mona, 2010; Csillery, Blum, & Francois, 2012; Csilléry, Blum, Gaggiotti, & François, 2010;
119 Estoup, Raynal, Verdu, & Marin, 2018; Raynal et al., 2019; Roux & Pannell, 2015). The power
120 of the ABC methods has made it possible to retrace the evolutionary history of notorious plant
121 parasites (*e.g., Plasmopara viticola* (Berk & Curtis) Berl. & de Toni (Fontaine et al., 2020),
122 *Microbotryum lychnidis-dioicae* (DC. Ex Liro) G. Deml & Oberw. (Gladieux et al., 2015)),
123 insect crop pests (*e.g.*, *Batrocera dorsalis* Hendel (Aketarawong et al., 2014), *Drosophila*
124 *suzukii* Matsumura (Fraimout et al., 2017), *Daktulosphaira vitifoliae* Fitch (Rispe et al., 2020))
125 and invasive alien species posing a threat to the native fauna (*e.g.*, *Harmonia axyridis* Pallas
126 (Lawson Handley et al., 2011)). These studies have identified the source populations,
127 reconstructed complex colonization routes, and determined the pace of geographic range
128 expansion, often emphasizing the role of human transportation in the spread of these noxious
129 species. Recently, the ABC approach, combined with machine learning (*i.e.*, random forest,
130 referred to as "ABC-RF" hereafter, (Estoup et al., 2018; Raynal et al., 2019)), was used to
131 demonstrate that the African arid-adapted locust pest species *Schistocerca gregaria* Forsskål
132 colonized Africa through major migration events driven by the last glacial climatic episodes
133 (Chapuis et al., 2020). Yet, ABC methods are underused for estimating the extent of gene flow
134 during parasite colonization (but see (Fraimout et al., 2017)). Most studies assume punctual
135 admixture events among populations, but rarely continuous gene flow per generation among
136 populations. Only recently has the new ABC-RF approach been used to infer the invasion
137 routes, evolutionary history and extent of gene flow in the spotted-wing *D. suzukii* Matsumura
138 from microsatellite markers (Fraimout et al., 2017) (referred to as SSR for simple sequence
139 repeat hereafter). Beyond population genomics approaches, and in the special case of insect

140 pests, the investigation of colonization history could also benefit from the characterization of
141 insect endosymbiotic bacterial communities. Indeed, many insect pests host a consortium of
142 endosymbiotic bacteria that mediate their adaptation to new environmental conditions (Frago,
143 Zytynska, & Fatouros, 2020). Variations in the endosymbiotic consortium along colonization
144 routes could facilitate the rapid adaptation of insect pests to different environments (Lenhart &
145 White, 2020) and, in consequence, its spread. Alternatively, if the colonization stems from only
146 a few populations, it might be accompanied by a loss of endosymbiont diversity along the
147 colonization routes.

148 Aphids are a good study system to investigate the evolutionary processes involved in
149 range expansion and the colonization of new environments. Aphids infest a wide range of host
150 species and can be major pests of many crop plants (Blackman & Eastop, 2000). Some aphid
151 species have now become cosmopolitan following the dissemination of crops around the globe
152 (Brady et al., 2014; Kirk, Dorn, & Mazzi, 2013; Zepeda-Paulo et al., 2010; Zhang, Edwards,
153 Kang, & Fuller, 2014). The clonal reproduction of aphids during spring and summer is one of
154 the reasons put forward to account for their remarkable success worldwide. Indeed, asexual
155 reproduction allows for a rapid increase in population size after the colonization of a favorable
156 new environment (Simon, Rispe, & Sunnucks, 2002; Simon, Stoeckel, & Tagu, 2010; Figueroa,
157 Fuentes-Contreras, Molina-Montenegro, & Ramírez, 2018). So far, only a handful of studies
158 have reconstructed the colonization history of aphid crop pests by combining population
159 genetics approaches using the information from SSR, sequence or single nucleotide
160 polymorphism (SNP) markers (Fang, Chen, Jiang, Qu, & Qiao, 2018; Giordano et al., 2020;
161 Kim, Hoelmer, & Lee, 2016; Leclair, Buchard, Mahéo, Simon, & Outreman, 2021; Morales-
162 Hojas, Sun, Iraizoz, Tan, & Chen, 2020; Peccoud et al., 2008; Piffaretti et al., 2013; Zepeda-
163 Paulo et al., 2010; Zhang et al., 2014; Zhou et al., 2015). These studies demonstrated that aphid
164 species can spread very quickly across the world, probably via plants transported by humans

165 and/or wind. Colonization was found to involve several populations with high genetic diversity,
166 possibly with gene flow (Wang, Hereward, & Zhang, 2016; Wei, Zuorui, Zhihong, &
167 Lingwang, 2005) and/or a few "super-clones", *i.e.*, predominant genotypes widespread in space
168 and time (Vorburger, Lancaster, & Sunnucks, 2003), which are highly differentiated from each
169 other and have high dispersal capacities (Piffaretti et al., 2013). Assessing the genetic diversity,
170 genetic structure and the extent of gene flow among populations in aphids are therefore central
171 to determining the evolutionary processes that have occurred during aphid colonization.

172 Associating the reconstruction of aphid colonization history with the characterization
173 of their endosymbiotic bacterial community can shed light on the processes of their dispersal.
174 Aphids harbor both obligate symbionts that supply them with the nutrients missing from their
175 diet (Buchner, 1965) and facultative symbionts that can provide various selective advantages
176 in specific environmental conditions (Haynes et al., 2003; Oliver, Degnan, Burke, & Moran,
177 2010). The obligate aphid endosymbiont bacterium *Buchnera aphidicola* has strictly
178 codiverged with most aphid species (Emmanuelle Jousselin, Desdevises, & Coeur d'acier,
179 2009). Bacterial markers (*e.g.*, *TrpB*), along with other typical markers (*e.g.*, *CO1* or *Cytb*),
180 can be used to infer the phylogenetic history of aphid species and to investigate signals of recent
181 range expansions (Popkin et al., 2017; Zhang et al., 2014). Beyond the use of bacterial genomes
182 to help reconstruct aphid phylogeography, the composition of the bacterial populations in
183 aphids might also help to assess the importance of facultative bacteria for colonizing new
184 geographic regions. Many studies have investigated the variation in bacterial communities
185 associated with the pea aphid (*Acyrthosiphon pisum* Harris), revealing geographical variation,
186 host-specific differentiation and associations with environmental factors such as temperature,
187 host plants, and natural enemies (Russell et al., 2013; Tsuchida, Koga, Shibao, Matsumoto, &
188 Fukatsu, 2002; Zepeda-Paulo, Ortiz-Martínez, Silva, & Lavandero, 2018; Leclair et al., 2021).
189 The effect of endosymbionts on aphid fitness has been confirmed experimentally in certain

190 cases (Frago et al., 2017; Leclair et al., 2016). Nevertheless, studies on a global scale in non-
191 model aphid species are still scarce (Zytynska & Weisser, 2016), and there are as of yet no
192 studies simultaneously investigating the colonization routes of an aphid species and the
193 changes in symbiotic associations along this route.

194 *Dysaphis plantaginea* Passerini, the rosy apple aphid, is one of the most harmful aphid
195 pests attacking cultivated apple trees (*Malus domestica* Borkh), causing major economic losses
196 every year, especially in Europe, North Africa and North America (Guillemaud, Blin, Simon,
197 Morel, & Franck, 2011; Warneys et al., 2018; Wilkaniec, 1993). This aphid species occurs
198 across temperate regions (Central and Southwest Asia, North Africa, America, and Europe)
199 (Blackman & Eastop, 2000) in cultivated apple orchards. The rosy apple aphid completes its
200 life cycle on two successive host plants: the cultivated apple trees as its sole primary host plant,
201 from early autumn to late spring, and the plantain herb *Plantago* spp. as a secondary host plant
202 during summer (Bonnemaison, 1959). The rosy apple aphid reproduces through cyclical
203 parthenogenesis whereby clonal reproduction alternate with a sexual reproduction, the latter
204 taking place in autumn when females lay fertilized overwintering eggs on apple trees. Eggs
205 hatch in early spring (Blommers, Helsen, & Vaal, 2004). While its phylogenetic relationships
206 with other aphid species is quite well resolved, the evolutionary history of *D. plantaginea* has
207 been little explored to date (but see (Guillemaud et al., 2011)). The native geographical range
208 of *D. plantaginea* and its ancestral host range are not known. It might have been associated
209 with *Malus sieversii,* the primary ancestor of the cultivated apple (*M. domestica*) (Harris,
210 Robinson, & Juniper, 2002), and then colonized Europe during the journey of the cultivated
211 apple along the Silk Routes from Asia to Europe. Alternatively, the rosy apple aphid may have
212 colonized its cultivated apple host in Europe rapidly and recently, about 1,500 years ago when
213 the Greeks brought the cultivated apple to Europe from Central Asia (Cornille et al., 2019;
214 Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux, 2014). These scenarios are derived from

215 our knowledge of the domestication history of apples; there are no data on the rosy apple aphid
216 that would support any of these scenarios. The colonization routes of the rosy apple aphid are
217 unknown, except that historical records document the introduction of the rosy apple aphid in
218 America was recent (*ca.* 1890s) (Foottit, Halbert, Miller, Maw, & Russell, 2006). More
219 generally, the population structure and the extent of gene flow among populations throughout
220 the geographic distribution of *D. plantaginea* are largely unknown, including the regions where
221 it causes the most damage in apple orchards, *i.e.*, North America, North Africa and Europe.
222 There are also no data so far regarding the diversity of endosymbionts across a large
223 geographical range in aphids.

224 Here, we investigate the colonization history of this major fruit tree pest using multiple
225 approaches and genetic datasets, drawn from comprehensive samples taken from the primary
226 host, the cultivated apple in Europe, North America and Morocco. Note that, despite repeated
227 attempts, we failed to collect *D. plantaginea* in its putative source region where apple trees
228 originated in Central Asia, preventing us from fully addressing its earliest colonization history.
229 Therefore, we aimed to answer the following questions focusing on regions most negatively
230 affected by the rosy apple aphid, *i.e.*, Europe, North America and North Africa: i) What is the
231 spatial genetic diversity and population structure of *D. plantaginea* across Europe, the US and
232 Morocco? Can we detect genetically differentiated populations, and/or recent bottlenecks in
233 colonizing populations? ii) Did gene flow occur among populations during colonization? iii)
234 Did *D. plantaginea* populations lose or gain symbionts during their colonization?

235

236 **Materials and Methods**

237 **Samples and DNA extraction**

238 Each sample described hereafter consisted of a single aphid colony of 10-15 females collected
239 on a host plant during the spring of 2017 and 2018. Sampling only one colony per tree ensured

240 that the sample did not contain different clones, which can occur on the same tree. Each colony
241 was kept in ethanol (96%) at -20°C until DNA extraction. For the three methods described
242 below (*i.e.*, SSR genotyping, Sanger sequencing of the aphid mitochondrial *CO1*, *Cytb*
243 markers, and *TrpB* bacterial marker, and metabarcoding of the 16S rRNA bacterial marker),
244 DNA was isolated from a single individual per colony using a new standardized protocol
245 (Supplementary material Text S1). We used different individuals to obtain DNA for the
246 amplification of the different markers. For the sake of simplicity, colonies are referred to as
247 'individuals' or 'samples' hereafter. Note that samples used for 16S rRNA sequencing
248 underwent two extra chemical washes before DNA extraction to remove the external bacteria
249 that could be present on the aphid's cuticle. The extra chemical washes consisted of a first wash
250 with dithiothreitol/DTT (50 mM) for 4 minutes, followed by a second wash with potassium
251 hydroxide/KOH (200 mM) for 4 minutes. The KOH wash was performed twice. Five negative
252 controls were also included ((Jousselin et al., 2016), Table S1).

253 Different sample sizes were used for each of the three methods (*i.e.*, SSR genotyping,
254 metabarcoding of the 16S rRNA bacterial marker, Sanger sequencing of the aphid *CO1*, *Cytb*
255 and *TrpB* markers). For each sample, the locality, sample collector identity, host plant species,
256 latitude, longitude and use in this study (genotyping, Sanger sequencing and/or metabarcoding)
257 are given in Table S1.

258 The largest sample was collected for SSR genotyping, comprising 667 *D. plantaginea*
259 samples (colonies) from Europe, Morocco and the US, from three hosts: *M. domestica* (50 sites,
260 *i.e.*, orchards, $N = 654$), but also *M. sylvestris* (one site, Alta Ribagorça in Spain, $N = 7$), the
261 European wild apple, and *P. lanceolata*, the secondary host (one site, Loos-en-Gohelle in
262 France, $N=6$). The 667 samples originated from 52 different geographic sites (*i.e.*, 52 orchards)
263 spread over 13 countries; seven to 15 individuals were collected at each site (Table S1, Figure
264 S1). We tried to obtain samples from Eastern Asia and Central Asia during fieldwork in 2017

265 and 2018, and through our collaborative network. However, despite our attempts and although
266 *D. plantaginea* is referenced in the literature on various hosts in several Central Asian countries
267 (Aslan & Karaca, 2005; CAB International, 2020; Holman, 2009), it was not observed in these
268 areas.

269 For the investigation of the bacterial 16S rRNA region, we used 178 *D. plantaginea*
270 individuals out of the same 667 colonies used for SSR genotyping (Table S1). We selected two
271 to three samples (colonies) per site to cover a wide and even spatial distribution across Europe,
272 North America and Morocco. The selected 178 individuals were collected across 12 countries
273 on *M. domestica,* except eight on *M. sylvestris* (Table S1).

274 For Sanger sequencing of *CO1*, *TrpB* and *CytB,* we used a total of 84 samples belonging
275 to eight aphid species (*D. plantaginea*, *Dysaphis* sp., *Aphis citricola* van der Goot, *Aphis pomi*
276 de Geer, *Aphis spiraecola* Patch, *Melanaphis pyraria* Passerini*, Myzus persicae* Sulzer*,*
277 *Rhopalosiphum insertum* Walker) sampled on five host plant species (*M. domestica*, *M.*
278 *sylvestris, Sorbus aucuparia, Prunus persica, Pyrus communis*). The species were chosen to
279 represent the aphid genera known to be closely related to the rosy apple aphid (Choi, Shin,
280 Jung, Clarke, & Lee, 2018). One to three individuals from each of the geographic sites listed
281 above and in Table S1 were sampled, for a total of 67 samples for *D. plantaginea*. The 17
282 additional samples from seven other aphid species (one to two samples per aphid species) were
283 collected on the cultivated apple and other fruit tree species (Table S1).

284

285 **PCR and Sanger sequencing**

286 We amplified the coding regions from the aphid mitochondrial cytochrome c oxidase subunit
287 I (*CO1*) gene and the cytochrome B (*CytB*) gene, as well as the Tryptophan synthase subunit
288 B (*TrpB*). Fragments were amplified following the protocol reported in Popkin et *al.* (2017)
289 with some modifications (Table S2). The final PCR volume was 30 µL, containing Buffer (1X),

290 $MgCl_2$ (1.5 mM), dNTP (0.1 mM), Forward and Reverse primers (0.7 µM), 5 µL of Taq
291 polymerase (1 U), and 5 µL of DNA (1/10 dilution). Amplification products were visualized
292 on an agarose gel 1.5% stained with ethidium bromide under ultraviolet light. We prepared
293 four 96-well plates with 15 µL of PCR products and a negative control. Plates were sent to
294 Eurofins Genomics France SAS for sequencing.

295 Chromatograms were inspected and corrected manually with CodonCode Aligner
296 version 8.0.1 (www.codoncode.com), assigned as an 'N' when two peaks overlapped.
297 Alignment, evaluation of all coding genes for frameshifts and elimination of pseudogenes were
298 performed with MEGA version 7.0.26 (Kumar, Stecher, & Tamura, 2016). The neutral
299 evolution of each gene, and thus its suitability for phylogenetic and population genetic
300 analyses, was assessed with the McDonald and Kreitman test (Egea, Casillas, & Barbadilla,
301 2008; McDonald & Kreitman, 1991) (Table S3). Two samples of *Brachycaudus helichrysi*
302 Kaltenbach, a pest of *Prunus*, for which sequences of *CO1* (NCBI sequence identifiers:
303 KX381827.1, KX381828.1), *CytB* (KX381989.1, KX381990.1), and *TrpB* (KX382153.1,
304 KX38215.1) were available (Popkin et al., 2017), were added in the phylogenetic analyses.
305 Sequences were concatenated, resulting in a data matrix of 86 concatenated sequences.
306

307 **Phylogenetic tree and taxonomic assignation**

308 We checked the taxonomic assignation of the samples used in this study by running a
309 phylogenetic analyses including *D. plantaginea* and other aphid species found on fruit trees
310 with a Bayesian approach implemented in MrBayes v3.2.7 (Huelsenbeck & Ronquist, 2001)
311 and Randomized Axelerated Maximum Likelihood-RAxML (Stamatakis, 2014), using the
312 GTR "Generalized time-reversible" mutational model. The two *M. pyraria* individuals were
313 used as an external group. We chose the default parameters (unlink statefreq = (all) revmat =
314 (all) shape = (all); prset applyto = (all) ratepr = variable; mcmcp ngen = 1000000 nruns = 2

315 nchains = 4 samplefreq = 1000 printfreq = 1000). Inferred trees were visualized with FigTree
316 v1.4 (http://tree.bio.ed.ac.uk/software/figtree/).
317
318 **SSR genotyping**
319 We used 30 SSR markers, including one that was previously used for *D. plantaginea*
320 (Guillemaud et al., 2011), and 29 that were newly developed from the sequencing of a low
321 coverage genome (see details of the protocol in supplementary material Texts S2 and S3). We
322 tested the neutrality, and the absence of linkage disequilibrium of the 29 SSR markers using
323 the Ewens-Watterson neutrality test ((Watterson, 1978); Text S3 and Table S4). Each SSR was
324 amplified separately by PCR. PCR was performed in a final volume of 20 µL (0.2 µM of each
325 forward and reverse primer, with the forward primer labeled with a fluorescent dye, 0.2 µM of
326 dNTPs, between 1 and 1.5 mM of $MgCl_2$, 1X Buffer (5X), 2 µL of a homemade Taq, 5 µL of
327 DNA (1/30 dilution) and sterile H_2O to reach the final volume). We used the following PCR
328 program: 94°C for 5 minutes, followed by 35 cycles of 30s at 94°C, 30s at 60 to 65°C and 45s
329 at 72°C, then 5 minutes at 72°C and finally 10 minutes at 4°C. In the PCR program, the
330 annealing temperature varied between 55°C and 65°C. Annealing temperatures for each SSR
331 marker are detailed in Supplementary Material Table S5. PCR products were then pooled
332 according to the four multiplexes described in Table S5.
333 SSR genotyping was performed at the Gentyane platform (INRAE, Clermont-Ferrand,
334 France). Alleles of each SSR marker were identified, and their size scored with Genemapper
335 v.4.0 (Applied Biosystems TM, Foster City, USA) by two people independently. In case of
336 discrepancy, the electropherogram was triple checked for a final decision. Allele scoring
337 resulting from Genemapper was then processed with the Autobin Excel Macro
338 (https://www6.bordeaux-aquitaine.inra.fr/biogeco_eng/Scientific-Production/Computer-
339 software/Autobin). We retained only multilocus genotypes with less than 30% missing data

340 and containing less than 5% null alleles. Null alleles were detected with GENEPOP v4.7
341 (Rousset, 2008).
342
343 **Clonal population structure**
344 Each individual was classified according to its multilocus genotype (*MLG*) with GenoDive
345 2.0b23 ((Meirmans & Van Tienderen, 2004), Table S1). We used the stepwise mutation model
346 with a threshold of 0 and the corrected Nei's diversity index as a statistic to test clonal
347 population structure. To reduce the influence of clonal copies produced by asexual
348 reproduction on Hardy–Weinberg equilibrium, allele frequency and genetic differentiation
349 estimates, the dataset was pruned to include only one copy of each *MLG* for further analyses.
350
351 **Population genetics descriptive statistics**
352 Observed and expected heterozygosity (H_O and H_E) and inbreeding coefficient (F_{IS}) were
353 calculated with GENEPOP v4.7 (Rousset, 2008) from the SSR dataset for each SSR marker,
354 each site (*i.e.*, geographic location/orchard) and each population (*i.e.*, clusters inferred with the
355 STRUCTURE software, to comprise individuals with a membership coefficient of at least
356 62.5% to the given cluster, see results). The 62.5% cut-off was chosen based on the distribution
357 of individual membership coefficients across the clusters detected for the most likely *K* value
358 (see results). Pairwise genetic differentiation (F_{ST}) between sites and between populations was
359 also calculated with GENEPOP (Rousset, 2008). Only sites with at least five successfully
360 genotyped samples were included for the site-specific computations. Allelic richness and
361 private allelic richness for each site and each population were calculated with ADZE (Szpiech,
362 Jakobsson, & Rosenberg, 2008) using a sample size corresponding to the smallest number of
363 observations per site or population, multiplied by two chromosomes (*e.g.*, a sample size of 20
364 represents ten individuals x two chromosomes).

365 We also estimated Nei's nucleotide diversity index π (Nei, 1987), Watterson's index θ
366 (Watterson, 1975), haplotype diversity, Tajima's *D* (Tajima, 1989) and Fu's *Fs* (Fu, 1997) with
367 DNAsp (Rozas, Sànchez-DelBarrio, Messeguer, & Rozas, 2003) using the concatenated three-
368 marker dataset (*i.e.*, *CO1*, *TrpB* and *CytB*) for each population (*i.e.*, cluster inferred with
369 STRUCTURE including individuals with membership coefficient > 62.5% to this cluster, see
370 results).
371
372 **Detecting recent bottlenecks during population range expansion**
373 We tested whether a bottleneck occurred during the range expansion of each population with
374 the method implemented in BOTTLENECK (Cornuet & Luikart, 1996; Piry, Luikart, &
375 Cornuet, 1999). Inferences regarding historical changes in population size are based on the
376 principle that the expected heterozygosity estimated from allele frequencies decreases faster
377 than the expected heterozygosity estimated under a given mutation model at mutation-drift
378 equilibrium in populations that have experienced a recent reduction in size. The tests were
379 performed under the stepwise-mutation model (SMM) and a two-phase model (TPM) allowing
380 for 30% multi-step changes.
381
382 **Spatial distribution of allelic richness and observed heterozygosity**
383 Spatial patterns of allelic richness and observed heterozygosity were visualized by mapping
384 the variation in allelic richness and observed heterozygosity at 48 sites in total (*i.e.*, sites with
385 at least five individuals) with the geometry-based inverse distance weighted interpolation in
386 QGIS (Quantum GIS, GRASS, SAGA GIS). The correlation between genetic variability (H_O)
387 and latitude, and between H_O and longitude, was tested using a linear model.
388

389 **Population subdivision**

390 We inferred the finest population structure by comparing the results obtained with three
391 population genetic tools: STRUCTURE 2.3.2 (Pritchard, Stephens, & Donnelly, 2000), TESS
392 2.3.1 (Chen, Durand, Forbes, & François, 2007), and a discriminant analysis of principal
393 components (DAPC) (Jombart, Devillard, & Balloux, 2010). STRUCTURE is based on the use
394 of Markov chain Monte Carlo (MCMC) simulations to infer the assignment of genotypes to K
395 distinct clusters. In addition to this, TESS also considers a spatial component, so that genotypes
396 from sites that are geographically closer to each other are considered more likely to be in the
397 same cluster. For both STRUCTURE and TESS, ten independent analyses were carried out for
398 each value of K ($1 \leq K \leq 10$ and $2 \leq K \leq 10$, respectively) with 500,000 MCMC iterations after
399 a burn-in of 50,000 steps. STRUCTURE and TESS outputs were processed with CLUMPP
400 1.1.2 (Jakobsson & Rosenberg, 2007) to identify potential distinct modes (*i.e.*, clustering
401 solutions) in replicated runs (10) for each K. We also assessed the population subdivision with
402 DAPC with the R package 'adegenet' (Jombart & Collins, 2015), which does not rely on any
403 assumption about the underlying population genetics model, in particular concerning Hardy-
404 Weinberg equilibrium or linkage equilibrium. The number of genetic clusters was investigated
405 with the *find.cluster* function (Jombart & Collins, 2015; Ripley & Ripley, 2001), which runs
406 successive K-means for clustering. The automatic cluster selection procedure '*diffNgroup*' was
407 used with *n.iter* set to 10^6 and *n.start* set to 10^3. The ordination analysis (DAPC) was performed
408 using the *dapc* function. The statistically optimal number of principal components was assessed
409 using the *optim.a.score* function. Assessment of the samples assigned to a genetic cluster was
410 performed using the *compoplot* function.

411 We used Pophelper (Francis, 2016) to run the Evanno method on the STRUCTURE
412 outputs. The Evanno method detects the strongest level of population subdivision (Evanno,
413 Regnaut, & Goudet, 2005). For TESS, we used the rate of change of the deviation index

414 criterion (*DIC*) to determine the amount of additional information explained by increasing *K*.
415 For DAPC, we looked at the Bayesian information criterion (*BIC)* obtained with the *adegenet*
416 package to estimate the optimal *K* value. However, the *K* identified with the *DIC, BIC* and *ΔK*
417 statistics does often not correspond to the finest biologically relevant population structure
418 (Cornille et al., 2015; Kalinowski, 2011; Puechmaille, 2016). We therefore visualized the bar
419 plots with Pophelper (Francis, 2016) and chose the *K* value for which all clusters had well
420 assigned individuals while no further well-delimited and biogeographically relevant clusters
421 could be identified for higher *K* values. For further analyses, we considered an individual to be
422 assigned to a cluster when its membership coefficient was ≥62.5% to this cluster (see results
423 below).

424 The spatial pattern of genetic structure was visualized by mapping the mean
425 membership coefficients for each site, as inferred from each of the three population genetics
426 structure analyses, with QGIS 3.12 'Las Palmas' (https://qgis.org). We further explored
427 relationships among populations with a principal component analysis performed on a table of
428 standardized alleles frequencies (PCA, dudi.pca, ade4 R package (Dray & Dufour, 2007)).

429

430 **Isolation-by-distance**

431 We tested whether there was a significant isolation-by-distance (IBD) pattern. A Mantel test
432 with 10,000 random permutations was performed between the individual coefficient of
433 relatedness F_{ij} (Loiselle, Sork, Nason, & Graham, 1995) and the matrix of the natural logarithm
434 of geographic distance. We also performed a correlation between $F_{ST}/(1-F_{ST})$ and the natural
435 logarithm of geographic distance. These analyses were performed using SPAGeDI 1.3 (Hardy
436 & Vekemans, 2002) separately for each *D. plantaginea* population (*i.e.*, cluster containing
437 individuals with a membership coefficient > 0.625) identified with TESS, STRUCTURE, and
438 DAPC.

439

440 **Demographic and divergence history using ABC-RF**

441 We used ABC to investigate whether the spatial patterns of genetic clustering, diversity and
442 differentiation observed in *D. plantaginea* resulted from the occurrence of gene flow among
443 populations during colonization. We also attempted to infer the sequence of colonization events
444 in each population. We used the recently developed ABC method based on a machine learning
445 tool named "random forest" (ABC-RF) to perform model selection and parameter estimations
446 (Estoup et al., 2018; Pudlo et al., 2016; Raynal et al., 2019). In brief, this method creates a
447 "forest" of bootstrapped decision trees to classify scenarios based on the summary statistics of
448 the datasets. Some simulations are not used to build the trees and can thus be used to cross-
449 validate the analysis by computing a "prior error rate". This approach allows the comparison
450 of complex demographic models (Pudlo et al., 2016) by comparing groups of scenarios with a
451 specific type of evolutionary event with other groups with different types of evolutionary
452 events (instead of considering all scenarios separately) (Estoup et al., 2018).

453 We used a nested ABC approach with two key steps. First, we inferred the divergence and
454 demographic history of the rosy apple aphid in Europe (step 1). Then, we tested the divergence
455 and demographic history of the rosy apple aphid outside of Europe (step 2). Each ABC step
456 compared different sequences of colonization events, with and without bidirectional gene flow
457 among populations (Figure S2). This two-step nested approach avoids the need to compare
458 models that are too complex, which would require the simulation of too many populations and
459 parameters, and is more powerful than testing all scenarios individually to determine the main
460 evolutionary events that characterize demographic history and divergence (Estoup et al., 2018).
461 Populations were defined as the clusters detected with STRUCTURE (see results), removing
462 putative hybrid individuals (*i.e.*, individuals with a membership coefficient < 62.5 to any given
463 cluster). The model parameters used were the divergence time between *X* and *Y* populations

464 ($T_{X\text{-}Y}$), the effective population size of population X ($N_{E\text{-}X}$), the migration rate per generation
465 between X and Y populations ($m_{X\text{-}Y}$). Prior values for divergence time were drawn for the log-
466 uniform distribution bounded between the distributions used in the ABC and are given in Table
467 S6.

468 For all models, identical SSR datasets were simulated for 29 out of the 30 markers that
469 had perfect repeats (we excluded the L4 marker because it did not have perfect repeats, Tables
470 S4 and S5), increasing confidence in the simulated model. We preliminarily checked that the
471 population structure inferred with 29 SSR markers did not differ significantly from the
472 inferences obtained with 30 SSR markers (data not shown). We assumed a generalized stepwise
473 model of SSR evolution. Mutation rates were allowed to vary across loci, with locus-specific
474 mutation rates drawn from a gamma distribution (α, α/μ) where μ is the mutation rate per
475 generation and α is a shape parameter. We assumed a log-uniform prior distribution for μ (1e-
476 4, 1e-3) and a uniform distribution for α (1.30).

477 We used ABCtoolbox (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010)
478 with fastsimcoal 2.5 (Excoffier & Foll, 2011) to simulate datasets, using model parameters
479 drawn from prior distributions (Table S6). We performed 10,000 simulations per scenario how
480 has been suggested (Pudlo et al., 2016b). For each simulation, we calculated six summary
481 statistics per population with arlsumstats v 3.5 (Excoffier and Lischer 2010): *H*, the mean
482 heterozygosity across loci, *sd(H)*, the standard deviation of the heterozygosity across loci, *GW*,
483 the mean Garza-Williamson statistic across loci (Garza and Williamson, 2001), *sd(GW),* the
484 standard deviation of the mean Garza-Williamson statistic over populations, *NGW*, the mean
485 modified Garza-Williamson statistic over loci, *sd(NGW),* the standard deviation of the mean
486 modified Garza-Williamson statistic over populations. We also computed pairwise F_{ST} (Weir
487 and Cockerham, 1984) and genetic distances $(\delta\mu)^2$ (Goldstein et al., 1995) between pairs of
488 populations.

489 We used the abcrf v.1.7.0 R statistical package (Pudlo et al., 2016) to carry out the
490 ABC-RF analysis. This analysis provides a classification vote that represents the number of
491 times a scenario is selected as the best one among *n* trees in the constructed random forest. For
492 each ABC step, we selected the scenario, or the group of scenarios, with the highest number of
493 classification votes as the best scenario, or best group of scenarios, among a total of 500
494 classification trees (Breiman, 2001). We then computed the posterior probabilities and prior
495 error rates (*i.e.*, the probability of choosing a wrong group of scenarios when drawing model
496 index and parameter values from the priors of the best scenario) over 10 replicate analyses
497 (Estoup et al., 2018). We also checked visually that the simulated models were compatible with
498 the observed dataset by projecting the simulated and the observed datasets onto the two first
499 linear discriminant analysis (LDA) axes (Pudlo et al., 2016), and checking that the observed
500 dataset fell within the clouds of simulated datasets.
501 We calculated parameter inferences using the final selected model following the two-
502 step ABC procedure. Note that the ABC-RF approach includes the model checking step that
503 was performed *a posteriori* in previous ABC methods.
504
505 **Characterization of the bacterial community associated with the rosy apple aphid using**
506 **the 16S rRNA bacterial gene**
507
508 In order to investigate the bacterial diversity in *D. plantaginea* populations, we amplified a 251
509 bp portion of the V4 region of the 16S rRNA gene (Mizrahi-Man, Davenport, & Gilad, 2013)
510 and used targeted sequencing of indexed bacterial fragments on a MiSeq (Illumina) platform
511 (Kozich, Westcott, Baxter, Highlander, & Schloss, 2013) following the protocol described in
512 (Jousselin et al., 2016). We used 178 aphid DNA extracts (Table S1), comprising 175 *D.*
513 *plantaginea* individuals and three *M. pyraria* individuals (Table S1). We wanted to represent a
514 range as large as possible from our samples. We also added eight randomly chosen samples

515 that did not undergo the two extra chemical washes (see Materials and methods, and Table S1).
516 Each sample was amplified twice along with negative controls (DNA extraction and PCR
517 controls). PCR replicates were conducted on distinct 96-well microplates. We obtained a total
518 of 390 PCR products (186 DNA extracts, by 2 for PCR duplicates, plus PCR controls), which
519 were pooled and then separated by gel electrophoresis. Bands based on the expected size of the
520 PCR products were excised from the gel, purified with a PCR clean-up and gel extraction kit
521 (Macherey-Nagel), and quantified with the Kapa Library Quantification Kit (Kapa
522 Biosystems). Paired-end sequencing of the DNA pool was carried out on a MISEQ (Illumina)
523 FLOWCELL with a 500-cycle Reagent Kit v2 (Illumina).

524 We first applied sequence filtering criteria following Illumina's quality control
525 procedure. We then used a pre-processing script from (Sow et al., 2019) to merge paired
526 sequences into contigs with FLASH V.1.2.11 (Magoč & Salzberg, 2011) and trim primers with
527 CUTADAPT v.1.9.1 (Martin, 2011). We then used the FROGS pipeline (Escudié et al., 2018)
528 to generate an abundance table of bacterial lineages across samples. In brief, we first filtered
529 out sequences > 261 bp and < 241 bp with FROGS, then we clustered variants into operational
530 taxonomic units (OTUs) with SWARM (Mahé, Rognes, Quince, Vargas, & Dunthorn, 2014)
531 using a maximum aggregation distance of three. We identified and removed chimeric variants
532 with VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016). We only kept OTUs that
533 were present in both PCR replicates of the same sample and then merged the number of reads
534 for each OTU for each aphid sample.

535 Taxonomic assignment of OTUs was carried out using RDPtools and Blastn (Altschul,
536 Gish, Miller, Myers, & Lipman, 1990) against the Silva138-16s database (https://www.arb-
537 silva.de) as implemented in FROGS. From the abundance table of OTUs across samples, we
538 transformed read numbers per aphid sample into frequencies (percentages); sequences
539 accounting for < 0.5 % of all the reads for a given sample were excluded following Jousselin

540 e*t al.* (2016). All filters resulted in excluding reads found in low abundance that could represent
541 sequencing errors and which were also often found in the negative controls.
542
543 **Results**
544 **Taxonomic status of aphid samples**
545 The Bayesian phylogenetic tree built with the 86 sequences representing nine aphid species
546 resulted in a polytomy for the 67 *D. plantaginea* samples from Europe, Morocco and the USA
547 (Figure S3), showing very little sequence variation at the intraspecific level (low bootstrap
548 values < 0.6, Figure S3a). We therefore kept only two representatives out of the 67 *D.*
549 *plantaginea* individuals in subsequent phylogenetic analyses. Bayesian analysis of this pruned
550 dataset comprising 19 individuals, representing nine aphid species, confirmed known
551 phylogenetic relationships in Aphidinae ((Choi et al., 2018), Figure S3b). The species *A. pomi,*
552 *A. citricola* and *A. spiraecola* were grouped in the same clade, the three species appearing
553 polyphyletic (Figure S3b). *Dysaphis plantaginea* appeared closely related to other
554 representatives of the Macrosiphini tribe (*B. helichrysi* and *M. persicae*). Interestingly, the
555 samples from Iran clustered apart from the *D. plantaginea* clade, the former belonging to the
556 *Dysaphis* sp. species. This result suggests that the two *D. sp.* Iranian samples belong to a yet
557 unidentified species or population exhibiting strong differentiation from European populations.
558 Those two samples were not included in the population genetics analyses using SSR, as we
559 only had two representatives from this Caucasian region (Table S1).
560
561 **Clone detection**
562 Overall, the proportion of unique genotypes was variable among sites. We found 582 unique
563 genotypes. Clones therefore represented 85 individuals grouped in 29 unique genotypes (12.7%
564 of the total dataset, Table S7), mainly coming from Belgium, (86% of the clones, eight sites;

565 mean proportion of unique genotypes (mean G/N) = 0.36 ± 0.2), Bulgaria (6%, one site, G/N
566 = 0.67), France (5%, three sites, mean *G/N* = 0.87 ± 0.05), the USA (1 %, one site; *G/N* = 0.93),
567 Spain, (1%, one site; *G/N* = 0.91) and from the lab-reared aphids (1%, *G/N* = 0.09). We kept
568 only the 582 unique genotypes for the analyses presented below.
569
570 **Spatial distribution of allelic variation**
571 The map of allelic richness (Figure 1a) showed that genetic diversity decreased along a north-
572 east to south-west gradient, with the highest allelic richness found in northeastern Europe and
573 the lowest in Morocco and the USA, except for Belgium that showed a lower level of genetic
574 diversity. We found a significant correlation between allelic richness and longitude (r = 0.2294,
575 *P-value* = 0.000573), as well as between allelic richness and latitude (r = 0.2684, *P-value* =
576 0.000162). The map of observed heterozygosity (Figure 1b) confirmed that genetic diversity
577 decreased along a north-east to south-west gradient, with the highest allelic richness in
578 Denmark and the lowest in Morocco. We found a significant correlation between observed
579 heterozygosity and longitude (r = 0.1731, *P-value* = 0.00327), and between observed
580 heterozygosity and latitude (r = 0.4998, *P-value* = 1.95e-08).
581
582 **Population structure and subdivision**
583 The spatial genetic structures inferred for *D. plantaginea* with TESS, DAPC and
584 STRUCTURE, and the respective *DIC*, *BIC* and *ΔK,* are shown in Figure 2 and supplementary
585 material Figures S4 to S9. For each K value, CLUMPP analyses produced highly similar
586 clustering patterns among the 10 repetitions (average G' > 95%). We therefore only presented
587 here the major modes.
588 With TESS, increasing K above 3 did not reveal any additional cluster (Figure S4). For
589 K = 3, TESS analyses revealed a clear partition between the Moroccan samples (blue), and

590 other samples (European and North American, orange). An additional cluster was identified
591 comprising only one Italian individual (yellow) (Figure 2a). With DAPC, increasing *K* above
592 4 did not reveal well-delimited new clusters, *i.e.*, only individuals with multiple admixtures
593 were assigned to the new clusters (Figure S5). For $K = 4$, DAPC identified three well-delimited
594 clusters (Figure 2b): one in Eastern Europe (*i.e.*, Bulgaria, Italy and Romania, in yellow), one
595 in Morocco (blue), one in the USA (orange), plus another cluster that comprised the rest of the
596 Western European (Spain, France, Belgium, the UK) and Danish samples (pink). With
597 STRUCTURE, increasing *K* above 5 did not reveal additional clusters. For $K = 5$ (Figure 2c),
598 samples were partitioned as follows: Morocco (blue), the USA (orange), Spain (pink), Eastern
599 Europe and Italy (green), and Western Europe and Denmark (yellow).

600 We used the inferences from STRUCTURE in subsequent analyses because it revealed
601 the finest population genetic structure. Genotypes were then assigned to a given population if
602 their membership coefficient for that population exceeded 0.625 (Tables 1 and S1). We chose
603 this threshold based on the bimodal distribution of cumulative coefficients inferred with
604 STRUCTURE for K=5 (Figure S10). A total of 175 individuals (30% of the dataset) could not
605 be assigned to any population and were not included in subsequent analyses.

606 We further visualized and quantified the genetic differentiation among populations with
607 a principal component analysis (PCA, Figure 2d) and F_{ST} estimates (Figure 2e), respectively.
608 Among all populations, F_{ST} values were low but all significant (Figure 2e). Pairwise genetic
609 differentiation was the lowest between the European populations ($F_{ST} = 0.02$, *P-value* < 0.0001,
610 Figure 2e); the Moroccan and US populations were the most differentiated (Figure 2e). The
611 PCA showed similar genetic relationships among the five populations (Figure 2d).

612

613 **Genetic diversity, bottleneck and range expansion**

614 The mean expected heterozygosity was relatively high (average = 0.74 with values ranging
615 from 0.55 to 0.75, Table S8). The mean F_{ST} across all loci was low (mean F_{ST} = 0.08, range: -
616 0.006 - 0.5) but significant for all pairs of sites (*P-value* < 0.001, Figure S11).

617 Allelic richness and private allelic richness were significantly different among the five
618 populations, except between the North American and Moroccan populations for allelic
619 richness, and among the Spanish, Moroccan and Eastern European populations for private
620 allelic richness (Table S9). The North American and Moroccan populations showed
621 significantly lower levels of allelic richness and private allelic diversity than the European
622 populations (*i.e.,* Spanish, Eastern and Western European populations; Tables 1 and S9). In
623 Europe, the Eastern European/Italian population had the highest level of allelic richness and
624 private allelic diversity, followed by the Western European/Danish population, and lastly the
625 Spanish population.

626 We also tested whether a strong and recent bottleneck occurred for each population.
627 BOTTLENECK analyses showed no significant deviation from the mutation-drift equilibrium
628 in any of the populations (Table S10). Tajima's *D* and Fu's F_S statistics did not reveal any
629 signature of demographic range expansion either (Table 1).

630

631 **Isolation-by-Distance (IBD)**

632 A significant but weak IBD pattern was observed (r=0.057, *P-value* ≤ 0.05) for the rosy apple
633 aphid (Figure S11). The *Sp* statistic can be used to quantify spatial structure and is useful for
634 comparing populations and/or species. Low *Sp* values are associated with greater dispersal
635 capacities and/or effective population sizes. Here, *Sp* values were extremely low (close to 0)
636 and were only significant for the Moroccan and the Western European/Danish populations
637 (Table 1). These results suggest that *D. plantaginea* has high dispersal capacities and/or large
638 effective population sizes.

639

640 **Inference of the divergence and demographic history of the rosy apple aphid**

641 First, we reconstructed the divergence and demographic history of the rosy apple aphid in
642 Europe (*i.e.*, including only the Spanish, Eastern European/Italian and Western
643 European/Danish populations). We defined 12 scenarios assuming different divergence
644 histories of the Eastern European/Italian, Western European/Danish and Spanish populations
645 (Figure S2). The 12 scenarios were tested with and without gene flow among populations. We
646 therefore ended up comparing 24 scenarios. Classification votes from the first round were the
647 highest ten times out of ten for the group of scenarios that assumed gene flow among the three
648 populations (295 votes out of the 500 RF trees, posterior probability P = 0.61, prior error rate
649 = 3.04%, Table S12, Figure S12). Projection of the reference table datasets and the observed
650 dataset on a single axis showed that the observed data fell within the distribution of the
651 simulated summary statistics of the group of scenarios that assumed gene flow among the three
652 populations, suggesting this analysis had the power to discriminate between the two groups of
653 scenarios and to identify the most likely scenario (Figure S12). The second round of ABC
654 inferences testing the sequence of colonization of the European populations requires caution in
655 interpretation as prior error rates were high and posterior probabilities low (Figure S12 and
656 Table S13).

657 We then investigated the colonization history of the rosy apple aphid outside of Europe,
658 *i.e.,* of the Moroccan and American populations. Given the lack of power to discriminate
659 between different scenarios of colonization sequence of the rosy apple aphid in Europe, and
660 the weak genetic differentation among the three European populations (mean F_{ST}= 0.02, Figure
661 2e), we merged the three European populations into a single European population (N = 316)
662 for this analysis. We then defined six scenarios of sequence of colonization starting either from
663 Europe or Morocco (Figure S2). We excluded the hypothesis that the rosy apple aphid

664 originated in North America. Indeed, historical records show that the introduction of the rosy
665 apple aphid was very recent in America (*ca.* 1890s) (Foottit et al., 2006). Furthermore, the
666 American population had the lowest levels of private allelic diversity and allelic richness, and
667 the American samples clustered with the European samples in the DAPC and TESS analyses.
668 For each of the six scenarios, five scenarios of gene flow among populations were tested: no
669 gene flow, gene flow among all populations and gene flow between each population pair (*i.e.*,
670 Europe/Morocco, Europe/the USA, and Morocco/the USA). We simulated these specific
671 models of gene flow among specific pairs of populations for the ABC-RF as we observed
672 variable admixture levels among populations (Figure 2c). In total, 30 scenarios were compared
673 (six colonization sequences x five gene flow modes, Figure S2). ABC-RF analyses showed
674 relatively high support for scenarios assuming gene flow (ABC-RF round 1, 10 out of the 10
675 replicates for the groups of scenarios assuming gene flow, 337 votes out of the 500 RF trees,
676 posterior probability P = 0.65, prior error rate = 6.55%, Table S14, Figure S13). Projection of
677 the reference table datasets and the observed dataset on a single axis showed that the observed
678 data fell within the distribution of the simulated summary statistics of the group assuming gene
679 flow, suggesting the occurrence of gene flow (Figure S13). However, although the observed
680 dataset fell within the distribution of simulated summary statistics (Figure S13), we lacked the
681 power to infer the sequence of colonization of the rosy apple aphid outside of Europe (*i.e.*,
682 posterior probability P = 0.65, and high prior error rate = 73.82%, Table S15).
683 Altogether, ABC-RF inferences supported the occurrence of gene flow during the
684 colonization history of *D. plantaginea*. However, ABC-RF did not allow to determine the
685 sequence of colonization of Europe, America and Morroco by the rosy apple aphid.
686

687 **16S rRNA amplicon sequencing**

688 After sequence filtering using the FROGS pipeline, high-throughput sequencing of 16S rRNA
689 bacterial genes from 186 aphids resulted in 5.7 M sequencing reads with an average of 30,800
690 reads per aphid sample. We found an extremely low bacterial diversity in *D. plantaginea*. The
691 5.7 M sequencing reads were clustered into 18 OTUs (Figure S14, Table S1). Seven OTUs
692 were assigned to *B. aphidicola* and made up 97.8% of the sequencing reads, 92 % of the reads
693 were assigned to a single *B. aphidicola* OTU, which was found associated with all *D.*
694 *plantaginea* individuals, the remaining *B. aphidicola* OTUs were found associated with
695 "outgroups". The remaining reads were mainly assigned to two known aphid endosymbionts,
696 *Serratia symbiotica* and *Regiella insecticola*, which were found in eight and three aphid
697 specimens, respectively. The three aphids hosting *Regiella* belonged to a population of *M.*
698 *pyraria* collected in Switzerland, while aphids hosting *Serratia* were found in distantly related
699 populations including *D. plantaginea* collected on *M. domestica* from various apple orchards
700 in France, and *P. communis* in Iran (Table S1). These results highlight the extremely limited
701 diversity of symbionts across a large geographical scale.

702

703 **Discussion**

704 We investigated the demographic history of a major fruit tree pest, the rosy apple aphid, in the
705 regions where it impacts the most cultivated apple orchards (*i.e.*, Europe, North Africa and
706 North America). Using multiple lines of evidence, we showed that the colonization of Europe
707 by the rosy apple aphid is likely recent, was not accompanied by strong bottlenecks and
708 involved gene flow between and within populations. The high level of gene flow among
709 populations and within populations was supported by the weak spatial genetic structure
710 observed across Europe, and coalescent-based simulations combined with ABC-RF. We also
711 found that *D. plantaginea* rarely hosts endosymbiotic bacteria other than their primary

712 symbiont, *B. aphidicola*. Our results provide further understanding of the evolutionary
713 processes at play during pest range expansion. We discussed the impacts of our results for pest
714 control programs.
715
716 **Colonization of apple trees by the rosy apple aphid is recent, and did not involve drastic**
717 **bottlenecks**
718 After removing clones from our analyses, we detected five main panmictic populations of the
719 rosy apple aphid: three in Europe, one in Morocco and one in the USA. The genetic diversity
720 for each population was within the same range as other aphid species such as *B. helichrysi*
721 (Popkin et al., 2017), *Eriosoma lanigerum* (Zhou et al., 2015) and *M. persicae nicotianae*
722 (Zepeda-Paulo et al., 2010). However, genetic diversity (private alleles and allelic richness) in
723 *D. plantaginea* was higher in Europe and lower in the USA and Morocco. More generally,
724 patterns of genetic structure and diversity suggest different demographic histories for the
725 European, Moroccan and American populations of *D. plantaginea*.
726 Individuals from the European populations had partial memberships to multiple
727 clusters, with similar membership coefficients for most individuals. This pattern of high
728 admixture along a spatial transect might reflect a continuous gradation in allele frequencies
729 (*i.e.*, a cline) across regions that cannot be detected by the methods used in this study. Indeed,
730 a major limitation of all clustering approaches is the risk of inferring artefactual discrete groups
731 in populations where genetic diversity is distributed continuously. DAPC and STRUCTURE
732 are not immune to this bias and may erroneously identify clusters within a cline (Jombart et al.,
733 2010). TESS is more sensitive to allelic gradient; this program includes a decay of the
734 correlation between membership coefficients and distance within clusters (Chen et al., 2007).
735 In the presence of clines and with evenly distributed sampling, TESS may detect fewer clusters
736 than STRUCTURE (Durand, Chen, & François, 2009). The larger admixed clusters found with

737 DAPC and STRUCTURE may therefore reflect a cline of allele frequency across Europe for
738 the rosy apple aphid. Allele frequency clines can result from admixture between genetically
739 distinct populations (Currat & Excoffier, 2005; Menozzi, Piazza, & Cavalli-Sforza, 1978)
740 and/or from subsequent founder events during range expansion (Barbujani, Sokal, and Oden
741 1995; Fix 1997; Currat and Excoffier 2005). Genetic diversity is also expected to decrease
742 along the expanding range (François, Blum, Jakobsson, & Rosenberg, 2008; Prugnolle,
743 Manica, & Balloux, 2005). Founder events associated with a recent range expansion may have
744 resulted in the decreasing east-west gradient of genetic diversity and the large number of
745 admixed individuals observed in Europe. It is therefore possible that the European populations
746 of the rosy apple aphid underwent a recent expansion. Note that we did not detect any signature
747 of range expansion with the three-markers dataset (*CO1*, *cytB* and *trpB*). This lack of signature
748 of range expansion may be due to limited number of samples used in our test (at least for the
749 Moroccan and American populations), but also that the range expansion is so recent that we
750 cannot catch its footprint with our sequence markers. Several preliminary ABC-RF tests of
751 range expansion of the rosy apple aphid in Europe that failed (*i.e.*, very low posterior
752 probabilities and prior error rates, data not shown) suggest that the second hypothesis is
753 possible.

754 The Moroccan and North American populations displayed a different pattern of
755 population differentiation and diversity compared to the European populations, suggesting an
756 even more recent colonization history. The two populations were well circumscribed and had
757 the highest level of genetic differentiation from the European populations, the lowest genetic
758 diversity and the lowest number of private alleles. In the TESS analysis, the American samples
759 did not cluster separately from the European samples, and in the DAPC analysis, samples from
760 America and Western Europe clustered together, suggesting that the American population
761 originated recently from Europe. This is in agreement with the earliest record of *D. plantaginea*

762 in the Eastern US dating back to 1890 (Foottit et al., 2006). Since then, there have probably
763 been multiple introductions into the USA that prevented genetic differentiation from Europe.
764 Similar scenarios have been described for the tobacco aphid, which was introduced into
765 America from different European pools (Zepeda-Paulo et al., 2010), and the leaf-curl plum
766 aphid *B. helichrysi* (Piffaretti et al., 2013), for which population genetic tools showed very little
767 differentiation between European and North American populations. As for the Moroccan
768 population, we lacked the power to infer its origin with the ABC-RF method. However, the
769 significantly lower genetic diversity and number of private alleles in this population compared
770 with that of Europe, the high level of admixture and the close genetic relationship with the
771 Spanish population, suggest that the Moroccan population resulted from a recent colonization
772 event from Southern Europe. Nevertheless, the history of the rosy apple aphid in North Africa
773 deserves further investigation requiring additional sampling in this region. Altogether, the
774 significantly lower genetic diversity observed in the Moroccan and American populations
775 suggest that these originated recently through founder events. Indeed, we did not find any
776 evidence that these two populations underwent a recent strong bottleneck despite having
777 significantly lower diversity in both SSR and sequence markers. Thus, the founder effect
778 underlying the colonization of North America and Morocco may involve genetic drift in small
779 populations rather than severe bottlenecks at the introduction event. Genetic diversity was also
780 found to be higher in native populations within their native range in the tobacco aphid (Zepeda-
781 Paulo et al., 2010) and the Russian wheat aphid (Zhou et al., 2015).

782 Altogether, a recent range expansion of the rosy apple aphid on its cultivated apple host
783 is a plausible explanation of the observed spatial genetic structure and diversity. Rapid range
784 expansion has also been described in the Russian wheat aphid (Zhang et al., 2014).
785 Unfortunately, the ABC-RF method was not powerful enough to disentangle the different
786 scenarios of colonization of the rosy apple aphid. However, the ABC-RF method was powerful

787 enough to identify the occurrence of gene flow within and outside Europe. Another recent study
788 on the main invasion routes of *D. suzukii* also reported a moderate level of confidence in model
789 choice (*i.e.*, low posterior probabilities between 0.50 to 0.63) and high prior error rates (ranging
790 from 0.30 to 0.40, using ABC-RF (Fraimout et al., 2017)). The absence of samples from key
791 places during the species colonization and of individuals of the species ancestral group have
792 been discussed to impact in the resolution of the colonization history analysis using ABC
793 approaches (Lombaert et al., 2014). This might be a source of the lack of support in our
794 proposed scenarios using ABC-RF. Additional information on the mutation rates of the newly
795 developed SSR in the present study improve support to some of our colonization scenarios.
796 The mutation rate of transcribed and untranscribed SSR was indeed successfully used to
797 reconstruct the main migration routes of *S. gregaria* to Africa using ABC-RF (Chapuis et al.,
798 2020).
799
800 **Colonization with gene flow, likely driven by humans**
801 We found that the expansion of the rosy apple aphid involved several populations with high
802 genetic diversity each, and inter- and intra- population gene flow across Europe, North America
803 and Morocco. ABC-RF analyses strongly support scenarios with bidirectional gene flow
804 among populations. The *Sp* parameter estimated within each population revealed large extent
805 of historical gene flow within population. We used the *Sp* parameter estimates to compare the
806 dispersal capacities of *D. plantaginea* with existing estimates in plants (Vekemans & Hardy,
807 2004). The rosy apple aphid showed dispersal capacities equivalent to or even higher than that
808 of wind-dispersed trees. The rosy apple aphid can therefore spread over kilometers, as
809 suggested previously (Guillemaud et al., 2011), and as frequently found in aphids (Loxdale et
810 al., 1993). However, despite its high dispersal capacities and the large amount of gene flow
811 among and within populations, we can still observe a significant (but weak) spatial genetic

812 structure across its distribution. The observed subtle spatial genetic structure may be associated
813 with human agricultural practices. The interchange of apple materials (*i.e.*, cultivars in the form
814 of scions and/or trees) is nowadays, and probably has been historically, frequent and a potential
815 way of moving the rosy apple aphid in the form of overwintering eggs among regions where
816 apple is cultivated. This hypothesis agrees with the differentiated population of *D. plantaginea*
817 observed in Asturias (Northwestern Spain). This region is known for the main utilization of the
818 apple for cider production and is based on local cider cultivars (Tardío, Arnal, & Lázaro, 2020).
819 Most of our samples from Spain are from Asturias, except the Catalan samples that were
820 clustering with the European ones. The exchange of apple material within Asturias, and within
821 other regions in Europe, may have occurred, implying higher risk of aphid movement within
822 regions than between regions. Of course, this is not the only explanation of the spatial genetic
823 structure observed, physical barriers (*i.e*., Atlantic Ocean) could also explain the higher genetic
824 differentiation of the Moroccan and American populations.
825
826 **Putative center(s) of origin of the rosy apple aphid**
827 The geographical origin of *D. plantaginea* remains unresolved. Our phylogenetic analyses
828 confirmed previous relationships between the rosy apple aphid and other aphids (Bašilova &
829 Rakauskas, 2012; Rebijith et al., 2017). However, estimates of the divergence time of *D.*
830 *plantaginea* from closely related species is now required, with denser sampling of the *Dysaphis*
831 genus across several regions in Eurasia and molecular dating analyses.
832 Genetic diversity estimates from SSR markers suggest that the source population came
833 from Eastern Europe, but it may also have originated even further east in Central Asia where
834 its fruit tree host was originally domesticated (Cornille et al., 2014; Harris et al., 2002). Our
835 failure to collect *D. plantaginea* samples in this area, despite our attempts in China and
836 Kazakhstan, prevents us from testing this scenario. However note that, while *D. plantaginea*

837 has been recorded in Central Asia according to several faunistic surveys (Kadyrbekov, 2002),
838 it is hard to find in these regions. Furthermore, a lack of records of this species in the Global
839 Biodiversity Information Facility (GBIF, 2020) suggests *D. plantaginea* is uncommon in this
840 area.

841 Alternatively, *D. plantaginea* may have originated in the Caucasus or Asia Minor,
842 maybe through a host jump from *Pyrus* to *Malus*. *Dysaphis plantaginea, Dysaphis radicola,*
843 *Dysaphis devecta, Dysaphis brancoi*, *Dysaphis anthrisci, Dysaphis chaerophylli* are *Dysaphis*
844 species reported to feed on the cultivated apple, *M. domestica,* as their primary host (Blommers
845 et al., 2004; Stekolshchikov, 2006), but many aphid species also feed on pears including,
846 *Dysaphis reaumuri* Mordvilko and *Dysaphis pyri* Boyer de Fonscolombe (Barbagallo,
847 Cocuzza, Cravedi, & Komazaki, 2007). The *Pyrus* genus is known to have diverged a long
848 time ago from the genus *Malus* probably in the Caucasus (Celton et al., 2009; Xiang et al.,
849 2017). Therefore, the ancestral group from which *D. plantaginea* diverged might have had a
850 *Pyrus* species as host. Analysis of samples from these regions is required to test this hypothesis.
851

852 **Low endosymbiont bacterial diversity associated with *D. plantaginea***

853 Our results showed that bacterial diversity was strikingly low in *D. plantaginea* across Europe,
854 North America and Morocco. As expected, *B. aphidicola* was the predominant bacterial species
855 (97% of our reads). We also distinguished different *B. aphidicola* OTUs in the Iranian *D.*
856 *plantaginea* samples which fits with the genetic differentiation observed in the corresponding
857 aphid hosts. At least nine secondary endosymbionts have been reported among aphid species
858 (reviewed in Zytynska & Weisser, 2016), including *Arsenophonus* Gherma, *Hamiltonella*
859 *defensa* Moran, *Regiella insecticola*, *Rickettsia* Da Rocha-Lima, *Rickettsiela* Drobne, *Serratia*
860 *symbiotica*, *Spiroplasma* Saglio and *Wolbachia* Hertig & Wolbach, and Fukatsuia *symbiotica*.
861 These bacteria have been reported to be aphid facultative symbiont, potential pathogens, or

862 plant-associated microbiota (Gauthier, Outreman, Mieuzet, & Simon, 2015). Here, we detected
863 the presence of secondary symbionts only in a few samples. The *Serratia* bacteria was the only
864 secondary symbiont identified and it was found in only eight samples of *D. plantaginea,*
865 collected on *Pyrus communis* in Iran, and in France and Spain on *M. domestica*. A *Regiella*
866 was found in three samples from another aphid species, *M. pyraria* collected on *M. domestica*.
867 At least seven endosymbionts (*H. defensa*, *R. insecticola*, *Rickettsia* sp., *Rickettsiela* sp., *S.*
868 *symbiotica*, *Spiroplasma* sp. and *Wolbachia* sp.) have been reported across a narrower spatial
869 distribution in the model species *Acyrthosiphon pisum* and several other well-studied aphid
870 species (Zytynska & Weisser, 2016). By contrast, another study (Henry, Maiden, Ferrari, &
871 Godfray, 2015) found that neither *D. plantaginea* nor other *Dysaphis* species hosted any
872 secondary symbionts, although it relied on a small number of individuals. This low
873 endosymbiont bacterial diversity in *D. plantaginea* shows that its likely fast expansion is not
874 the result of an association with different mutualistic endosymbionts.
875
876 **Concluding remarks**
877 This study has demonstrated that the colonization of a major fruit tree aphid pest occurred
878 without a strong bottleneck, maintaining high genetic diversity, and generated differentiated
879 populations exchanging gene flow, with an isolation-by-distance pattern. The lack of
880 demographic changes in the populations of *D. plantaginea* in Europe, except for the Belgian
881 populations, indicates that seasonal selective pressures, such as insecticide application, have
882 little impact on the genetic diversity of the species. These results may have implications in
883 control and management of *D. plantaginea*, but further studies are needed to fully understand
884 how selective pressures have impacted *D. plantaginea* adaptation. In addition, the use of other
885 genetic markers, such as SNPs, promises to make great strides to elucidate the demographic
886 history of the rosy apple aphid. Finally, the origin of the rosy apple aphid is still unknown. Our

887 results suggest it may have originated in Eastern Europe, the Caucasus or Asia Minor.
888 However, the domestication of the rosy apple aphid primary host (the cultivated apple) in this
889 region remains unknown (Cornille et al., 2019, 2014; Spengler, 2019). Further investigations
890 on the history of apple domestication, additional sampling of *D. plantaginea* in the Caucasus
891 or Asia Minor, and sampling related aphid species, are required to better understand the origin
892 of this major fruit tree pest.
893

894 **Acknowledgments**

895 This research was funded by the laboratoire d'Excellence Biodiversity Agrosystem Society and
896 Climate BASC (grant « Emergence POMPUCEDOM ») and Systematic Research funding,
897 ATIP-Avenir program and the *Institut Diversité Ecologie et Evolution du Vivant* (IDEEV). AR
898 and TU were supported by a grant of the Romanian Ministry of Education and Research,
899 CCCDI - UEFISCDI, project number 384 PED-PN-III-P2-2.1-PED-2019-4924. We thank the
900 Plateforme de Génotypage Gentyane INRAE UMR 1095 for assistance in genotyping, and
901 especially the platform leader Charles Poncet. We thank the informatics team at the GQE-Le
902 Moulon, in particular Adrien Falce, Benoit Johannet and Olivier Langella. We are grateful to
903 the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics
904 Facility, doi: 10.15454/1.5572390655343293E12) for providing help and/or computing and/or
905 storage resources.
906

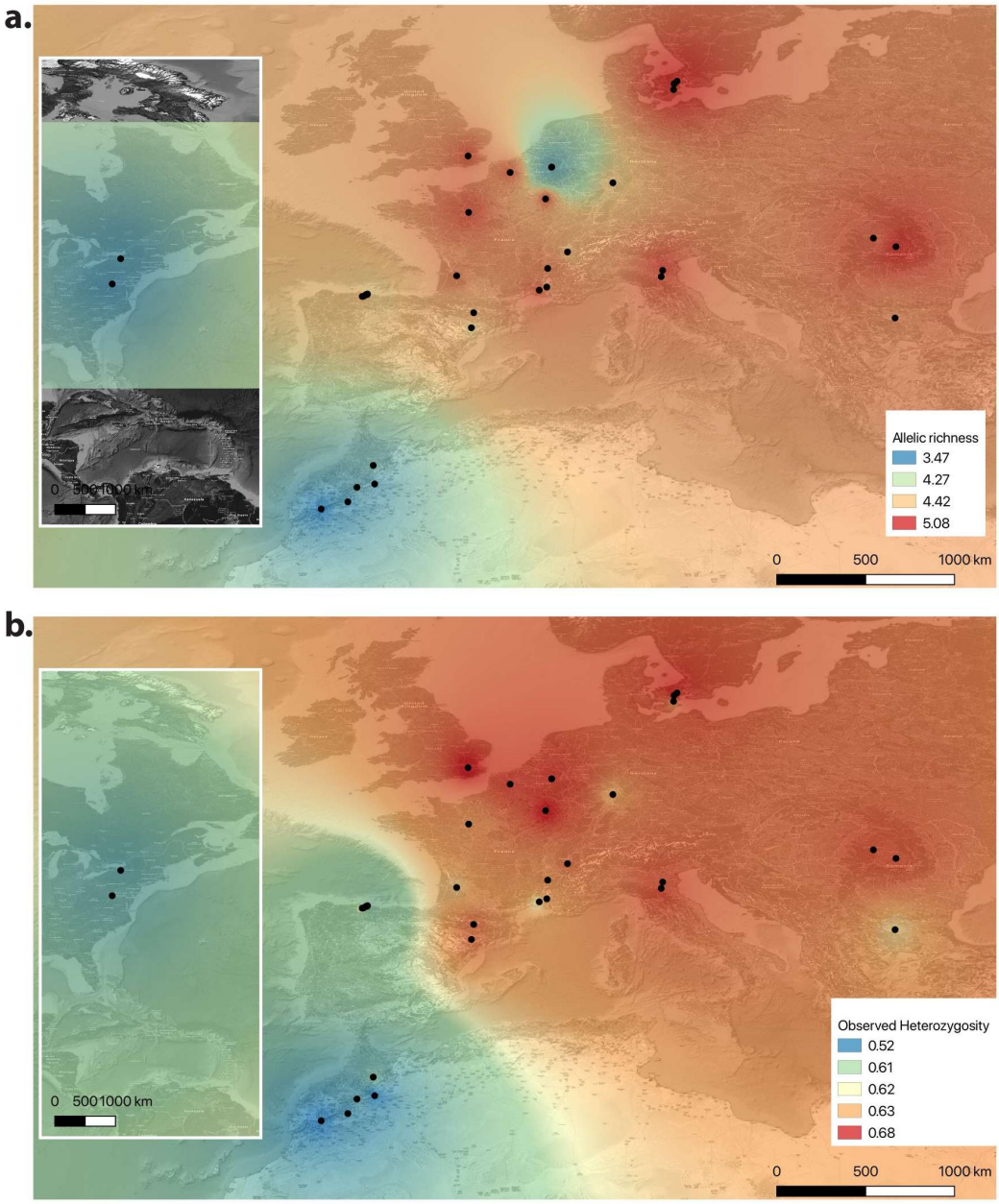
907 Script and data availability: The data and script are available at
908 https://zenodo.org/record/4537710#.YCaeomMo95c.
909

910 Authors' contributions. AC, EJ, PD, LJ conceived and designed the experiments; AC, EJ, PD,
911 LJ, MH, TG obtained funding; AC, AR, AA, TB, GA, FP, AH, SKJ, RA, SS, LS, EG, LT, FG,

KM, YK, ARom, TD, IZ, OS, RJM, AA, BEH, EJ, HZ, IB, MR, HT, CM, BB, AD sampled the material; AR, CR, AV, OG, AC, MA, LM, ML, LB, SGOV performed the molecular work; SGOV, AC, OG, MG, LB, EJ, GD analyzed the data. All co-authors discussed the results. The manuscript was written by SGOV, AC, EJ, OG with critical inputs from other co-authors.

921 **Figures and Tables**

922



923

924 **Figure 1. Spatial genetic diversity of *Dysaphis plantaginea* in Europe, Morocco and North**
925 **America (*N* = 582, clones were excluded, 52 sites, 30 SSR markers). a.** Map of allelic
926 richness per site. **b.** Map of observed heterozygosity. Sites with a sample size below five
927 individuals are not represented on this map. Each dot is a site (*i.e.*, orchard). Red means high
928 allelic richness, blue means low.

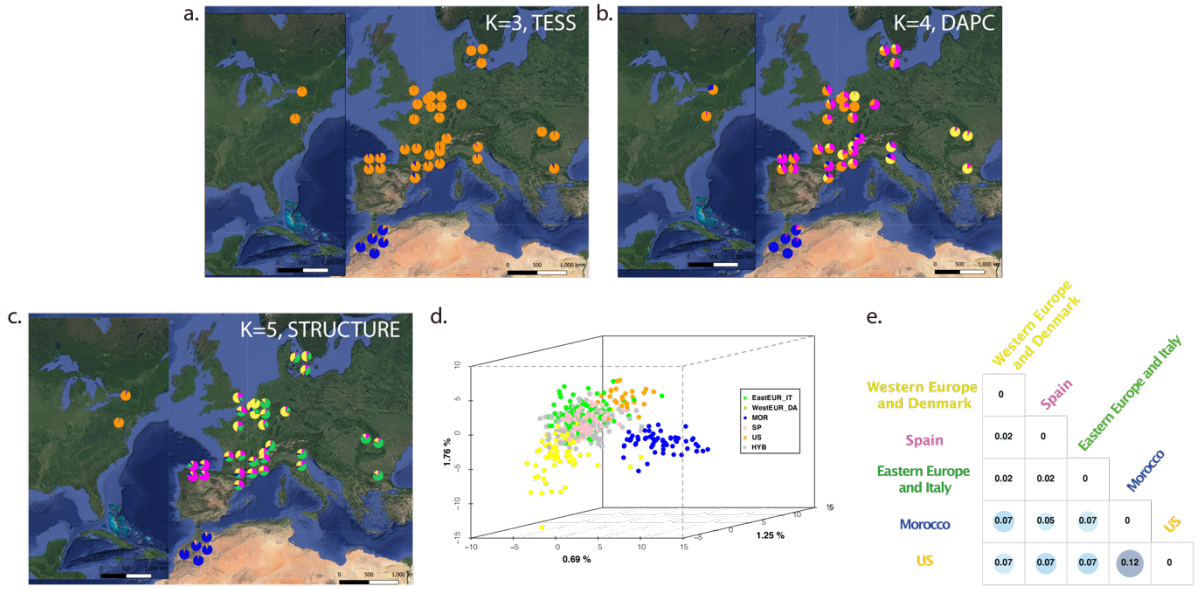


Figure 2. Spatial population structure and differentiation of the rosy apple aphid *Dysaphis plantaginea* in Europe, Morocco and North America (*N* = 582 individuals, 52 sites, 30 SSR markers). Spatial population structure inferred with **a.** TESS (*K* = 3), **b.** Discriminant Analysis of Principal Components (*K* = 4), **c.** with STRUCTURE (*K* = 5). Each figure includes a map with each pie chart representing the mean membership coefficients for each site. **d.** Principal component analysis including the 582 individuals used for running STRUCTURE analyses; each individual is colored according to its membership > 0.625 to one of the five *D. plantaginea* populations detected with STRUCTURE (individuals with a membership coefficient < 62.5% to a given cluster is referred as "admixed"). **e.** Genetic differentiation (F_{ST}) estimates among the five populations detected with STRUCTURE. The higher F_{ST} values are highlighted by circles (Morocco and the US has the highest value). US: United States of America; MOR: Morocco; SP: Spain; EasternEU_IT: Eastern Europe and Italian population (Bulgaria, Italy, Romania); WestEUR_DA: Western Europe and Danish population (France, Germany, Belgium, United Kingdom, and Denmark).

Table 1. Summary statistics obtained for the five *Dysaphis plantaginea* populations (*i.e.*, cluster including individuals with a membership coefficient > 62.5% to a given cluster) inferred with STRUCTURE (*K* = 5), based on 30 SSR markers and three concatenated genes (*CO1*, *Cytb* and *TrpB*, 67 *D. plantaginea* individuals), respectively (left and the right sections of the table).

	Statistics inferred from SSR							**Statistics inferred from concatenated genes**							
Population	***N***	***Ho***	***H_E***	***F_{IS}***	***A_R (N=28)***	***A_P (N=28)***	***Sp***	***Nseq***	***NH***	***HD***	***S***	***π***	***θ***	***Tajima's D***	***Fu's Fs***
Eastern European and Italian	114	0.68	0.77	0.12 ***	7.77	1.65	0.0006 n.s	34	6	0.59	6	0.0005	0.0008	-0.960 n.s	0.550 n.s
Western European and Danish	89	0.65	0.74	0.12 ***	7.26	1.17	0.0019	10	4	0.67	3	0.0006	0.0005	0.120 n.s	0.880 n.s
Spanish	113	0.61	0.71	0.14 ***	6.80	0.92	0.0003	19	5	0.71	4	0.0006	0.0006	-0.001 n.s	0.842 n.s
Moroccan	63	0.53	0.61	0.13 ***	5.42	0.60	-0.0038	2	2	0.67	1	0.0004	0.0003	1.630 n.s	1.280 n.s
American	28	0.57	0.64	0.10 ***	4.70	0.42	-	2	2	0	0	0	0	0	0

N: number of individuals, *Ho* and *He*: observed and expected heterozygosity, F_{IS}: inbreeding coefficient, *** *P-value* < 0.0001, A_R and A_P: allelic richness and private allelic richness, respectively, corrected by the rarefaction method, estimated for a sample size of 28; *Sp* = spatial parameter; the number of individuals from different geographic locations is too low in the American population to compute the *Sp* parameter. Distribution of haplotypes and molecular diversity based on the sequence markers *CO1*, *CytB* and *TrpB* from 67 individuals from different populations based on the genetic structure inferred with STRUCTURE. *Nseq* = Number of sequences; *NH*: number of haplotypes; *HD*: haplotype diversity; *S*= number of polymorphic sites; π = Nei's nucleotide diversity index; θ = Watterson's diversity index θ; n.s. = not significant.

Conflict of interest

The authors of this preprint declare that they have no financial conflict of interest with the content of this article. Amandine Cornille, Tatiana Giraud, Emmanuelle Jousselin, and Jean-Christophe Simon are PCI Evolutionary Ecology Biology recommenders. Tatiana Giraud is a member of the managing board of PCI Evolutionary Biology.

Script and data availability

https://zenodo.org/record/4537710#.YCaeomMo95c

References

Aketarawong, N., Guglielmino, C. R., Karam, N., Falchetto, M., Manni, M., Scolari, F., … Malacrida, A. R. (2014). The oriental fruitfly Bactrocera dorsalis ss in East Asia: Disentangling the different forces promoting the invasion and shaping the genetic make-up of populations. *Genetica*, *142*(3), 201–213.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Angert, A. L., Bontrager, M. G., & Ågren, J. (2020). What Do We Really Know About Adaptation at Range Edges? *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), 341–361. doi: 10.1146/annurev-ecolsys-012120-091002

Aslan, B., & Karaca, İ. (2005). Fruit tree Aphids and their natural enemies in Isparta region,Turkey. *Journal of Pest Science*, *78*(4), 227–229. doi: 10.1007/s10340-005-0097-2

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P.-H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, *51*(2), 148–164.

Barbagallo, S., Cocuzza, G., Cravedi, P., & Komazaki, S. (2007). 29 IPM Case Studies: Deciduous Fruit Trees. *Aphids as Crop Pests*, 651.

Barbujani, G., Sokal, R. R., & Oden, N. L. (1995). Indo-European origins: A computer-simulation test of five hypotheses. *American Journal of Physical Anthropology*, *96*(2), 109–132.

Bašilova, J., & Rakauskas, R. (2012). Phylogenetic relationships of Dysaphis pyri (BOYER DE FONSCOLOMBE) and Dysaphis reaumuri (MORDVILKO)(Hemiptera, Sternorrhyncha: Aphididae): COI and EF-1α evidence. *Organisms Diversity & Evolution*, *12*(2), 197–204.

Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, *19*(13), 2609–2625.

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: An identification and information guide.* John Wiley & Sons Ltd.

Blakeslee, A. M. H., Haram, L. E., Altman, I., Kennedy, K., Ruiz, G. M., & Miller, A. W. (2019). Founder effects and species introductions: A host versus parasite perspective. *Evolutionary Applications*, *13*(3), 559–574. PubMed (32431736). doi: 10.1111/eva.12868

Blommers, L. H. M., Helsen, H. H. M., & Vaal, F. W. N. M. (2004). Life history data of the rosy apple aphid Dysaphis plantaginea (Pass.) (Homopt., Aphididae) on plantain and as migrant to apple. *Journal of Pest Science*, *77*(3), 155–163. doi: 10.1007/s10340-004-0046-5

Bloomquist, E. W., Lemey, P., & Suchard, M. A. (2010). Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology & Evolution*, *25*(11), 626–632. doi: https://doi.org/10.1016/j.tree.2010.08.010

Bonnemaison, L. (1959). Le puceron cendré du pommier (Dysaphis plantaginea Pass.). Morphologie et biologie. Méthodes de lutte. *Annales Des Épiphyties*, (3), 257–320.

Brady, C. M., Asplen, M. K., Desneux, N., Heimpel, G. E., Hopper, K. R., Linnen, C. R., … White, J. A. (2014). Worldwide populations of the aphid Aphis craccivora are infected with diverse facultative bacterial symbionts. *Microbial Ecology*, *67*(1), 195–204.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Buchner, P. (1965). Endosymbiosis of Animals with Plant Microorganisms. In H. Malke (Ed.), *Zeitschrift für allgemeine Mikrobiologie* (Vol. 7, pp. 168–168). ohn Wiley & Sons, Inc.: Interscience Publ. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/jobm.19670070219

CAB International. (2020). *Dysaphis plantaginea. [Distribution map]*. Nosworthy Way, Wallingford, Oxfordshire, OX10 8DE, UK.: CAB International Wallingford UK. Retrieved from https://www.cabi.org/isc/abstract/20056600429

Celton, J. M., Chagne, D., Tustin, S. D., Terakami, S., Nishitani, C., Yamamoto, T., & Gardiner, S. E. (2009). Update on comparative genome mapping between Malus and Pyrus. *BMC Res Notes*, *2*. doi: 10.1186/1756-0500-2-182

Chapuis, M.-P., Raynal, L., Plantamp, C., Meynard, C. N., Blondin, L., Marin, J.-M., & Estoup, A. (2020). A young age of subspecific divergence in the desert locust inferred by ABC random forest. *Molecular Ecology*, 1–17. doi: 10.1111/mec.15663

Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes*, *7*(5), 747–756.

Choi, H., Shin, S., Jung, S., Clarke, D. J., & Lee, S. (2018). Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): An evolutionary hypothesis for the Pterocomma-group habitat adaptation. *Molecular Phylogenetics and Evolution*, *121*, 12–22. doi: 10.1016/j.ympev.2017.12.021

Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., … Roldán-Ruiz, I. (2019). A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*, *24*(8), 770–782. doi: https://doi.org/10.1016/j.tplants.2019.05.007

Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A basis for conservation and breeding programs in apples. *Evolutionary Applications*, n/a-n/a. doi: 10.1111/eva.12250

Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. doi: 10.1016/j.tig.2013.10.002

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, *144*(4), 2001–2014.

Csillery, K., Blum, M., & Francois, O. (2012). *Tools for Approximate Bayesian Computation (ABC) http://cran.r-project.org/web/packages/abc/index.html*.

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410–418.

Currat, M., & Excoffier, L. (2005). The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1564), 679–688. doi: 10.1098/rspb.2004.2999

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software; Vol 1, Issue 4 (2007)*. doi: 10.18637/jss.v022.i04

Durand, E., Chen, C., & François, O. (2009). *Tess version 2.3—Reference Manual August 2009*. 30.

Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: A website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, *36*(Web Server), W157–W162. doi: 10.1093/nar/gkn337

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., … Pascal, G. (2018). FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics*, *34*(8), 1287–1294. doi: 10.1093/bioinformatics/btx791

Estoup, A. A., Raynal, L., Verdu, P., & Marin, J.-M. (2018). Model choice using Approximate Bayesian Computation and Random Forests: Analyses based on model grouping to make inferences about the genetic history of Pygmy human populations. *Journal de La Societe Française de Statistique*, *159*(3), 167–190.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, *19*(19), 4113–4130. doi: 10.1111/j.1365-294X.2010.04773.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. doi: 10.1093/bioinformatics/btr124

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Fang, F., Chen, J., Jiang, L., Qu, Y., & Qiao, G. (2018). Genetic origin and dispersal of the invasive soybean aphid inferred from population genetic analysis and approximate Bayesian computation. *Integrative Zoology*, *13*(5), 536–552. doi: 10.1111/1749-4877.12307

Figueroa, C. C., Fuentes-Contreras, E., Molina-Montenegro, M. A., & Ramírez, C. C. (2018). Biological and genetic features of introduced aphid populations in agroecosystems. *Current Opinion in Insect Science*, *26*, 63–68. doi: 10.1016/j.cois.2018.01.004

Fix, A. G. (1997). Gene Frequency Clines Produced by Kin-Structured Founder Effects. *Human Biology*, *69*(5), 663–673.

Fontaine, M., Labbé, F., Dussert, Y., Delière, L., Richart-Cervera, S., Giraud, T. E., & Delmotte, F. (2020). *Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, Plasmopara viticola*. doi: 10.1101/2020.09.22.307678

Foottit, R. G., Halbert, S. E., Miller, G. L., Maw, E., & Russell, L. M. (2006). *Adventive aphids (Hemiptera: Aphididae) of America north of Mexico*. Retrieved from https://pubag.nal.usda.gov/catalog/8716

Frago, E., Mala, M., Weldegergis, B. T., Yang, C., McLean, A., Godfray, H. C. J., … Dicke, M. (2017). Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. *Nature Communications*, *8*(1), 1860. doi: 10.1038/s41467-017-01935-0

Frago, E., Zytynska, S. E., & Fatouros, N. E. (2020). Microbial symbionts of herbivorous species across the insect tree. *Mechanisms Underlying Microbial Symbiosis*, 111–159. doi: 10.1016/bs.aiip.2020.04.002

Fraimout, A., Debat, V., Fellous, S., Hufbauer, R. A., Foucaud, J., Pudlo, P., … Estoup, A. (2017). Deciphering the Routes of invasion of Drosophila suzukii by Means of ABC Random Forest. *Molecular Biology and Evolution*, *34*(4), 980–996. PMC (PMC5400373). doi: 10.1093/molbev/msx050

Francis, R. (2016). POPHELPER: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*, *17*, n/a-n/a. doi: 10.1111/1755-0998.12509

François, O., Blum, M. G. B., Jakobsson, M., & Rosenberg, N. A. (2008). Demographic History of European Populations of Arabidopsis thaliana. *PLOS Genetics*, *4*(5), e1000075. doi: 10.1371/journal.pgen.1000075

Fu, Y.-X. (1997). Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*, *147*(2), 915.

Garnas, J., R., Auger-Rozenberg, M.-A., Roques, A., Bertelsmeier, C., Wingfield, M. J., Saccaggi, D. L., … Slippers, B. (2016). Complex patterns of global spread in invasive insects: Eco-evolutionary and management consequences. *Biological Invasions*, *18*(4), 935–952. doi: 10.1007/s10530-016-1082-9

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, *10*(2), 305–318. doi: 10.1046/j.1365-294X.2001.01190.x

Gauthier, J.-P., Outreman, Y., Mieuzet, L., & Simon, J.-C. (2015). Bacterial Communities Associated with Host-Adapted Populations of Pea Aphids Revealed by Deep

Sequencing of 16S Ribosomal DNA. *PLOS ONE*, *10*(3), e0120664. doi: 10.1371/journal.pone.0120664

GBIF. (2020). Dysaphis plantaginea (Passerini, 1860) in GBIF Secretariat. Retrieved October 26, 2020, from https://www.gbif.org/species/2075245

Giordano, R., Donthu, R. K., Zimin, A. V., Julca Chavez, I. C., Gabaldon, T., van Munster, M., … Zhan, S. (2020). Soybean aphid biotype 1 genome: Insights into the invasive biology and adaptive evolution of a major agricultural pest. *Insect Biochemistry and Molecular Biology*, *120*, 103334. doi: 10.1016/j.ibmb.2020.103334

Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., … Giraud, T. (2015). The population biology of fungal invasions. *Molecular Ecology*, *24*(9), 1969–1986. doi: 10.1111/mec.13028

Gladieux, Pierre, Ropars, J., Badouin, H., Branca, A., Aguileta, G., de Vienne, D. M., … Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology*, *23*(4), 753–773. doi: 10.1111/mec.12631

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, *139*(1), 463–471. doi: 10.1093/genetics/139.1.463

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak Spatial and Temporal Population Genetic Structure in the Rosy Apple Aphid, Dysaphis plantaginea, in French Apple Orchards. *PLoS ONE*, *6*(6), e21263. doi: 10.1371/journal.pone.0021263

Hardy, O. J., & Vekemans, X. (2002). SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*(4), 618–620.

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends Genet*, *18*(8), 426–430.

Haynes, S., Darby, A. C., Daniell, T. J., Webster, G., van Veen, F. J. F., Godfray, H. C. J., … Douglas, A. E. (2003). Diversity of Bacteria Associated with Natural AphidPopulations. *Applied and Environmental Microbiology*, *69*(12), 7216–7223. doi: 10.1128/AEM.69.12.7216-7223.2003

Henry, L. M., Maiden, M. C. J., Ferrari, J., & Godfray, H. C. J. (2015). Insect life history and the evolution of bacterial mutualism. *Ecology Letters*, *18*(6), 516–525. doi: 10.1111/ele.12425

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., … Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, *54*(1), 291–301. doi: 10.1016/j.ympev.2009.09.016

Hoffmann, B. D., & Courchamp, F. (2016). Biological invasions and natural colonisations: Are they that different? *NeoBiota*, *29*, 1–14. doi: 10.3897/neobiota.29.6959

Holman, J. (2009). *Host Plant Catalog of Aphids: Palaearctic Region* (Springer Nature eReference).

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. doi: 10.1093/bioinformatics/17.8.754

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. doi: 10.1093/bioinformatics/btm233

Jombart, T., & Collins, C. (2015). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0*. 43.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. doi: 10.1186/1471-2156-11-94

Jousselin, E., Clamens, A.-L., Galan, M., Bernard, M., Maman, S., Gschloessl, B., … Coeur d'acier, A. (2016). Assessment of a 16S rRNA amplicon Illumina sequencing procedure for studying the microbiome of a symbiont-rich aphid genus. *Molecular Ecology Resources*, *16*(3), 628–640. doi: 10.1111/1755-0998.12478

Jousselin, Emmanuelle, Desdevises, Y., & Coeur d'acier, A. (2009). Fine-scale cospeciation between Brachycaudus and Buchnera aphidicola: Bacterial genome helps define species and evolutionary relationships in aphids. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1654), 187–196. doi: 10.1098/rspb.2008.0679

Kadyrbekov, Rk. (2002). About the aphid fauna (Homoptera Aphididae) of West Tien-Shan (Kazakhstan part). *Tethys Entomological Research*, (4), 65–76.

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, *106*(4), 625–632.

Kim, H., Hoelmer, K., & Lee, S. (2016). Population genetics of the soybean aphid in North America and East Asia: Test for introduction between native and introduced populations. *Biological Invasions*, *19*. doi: 10.1007/s10530-016-1299-7

Kirk, H., Dorn, S., & Mazzi, D. (2013). Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications*, *6*(5), 842–856. doi: 10.1111/eva.12071

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.

Applied and Environmental Microbiology, *79*(17), 5112–5120. doi: 10.1128/AEM.01043-13

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. doi: 10.1093/molbev/msw054

Lawson Handley, L.-J., Estoup, A., Evans, D. M., Thomas, C. E., Lombaert, E., Facon, B., … Roy, H. E. (2011). Ecological genetics of invasive alien species. *BioControl*, *56*(4), 409–428. doi: 10.1007/s10526-011-9386-2

Leclair, M., Buchard, C., Mahéo, F., Simon, J.-C., & Outreman, Y. (2021). A Link Between Communities of Protective Endosymbionts and Parasitoids of the Pea Aphid Revealed in Unmanipulated Agricultural Systems. *Frontiers in Ecology and Evolution*, *9*. doi: 10.3389/fevo.2021.618331

Leclair, M., Pons, I., Mahéo, F., Morlière, S., Simon, J.-C., & Outreman, Y. (2016). Diversity in symbiont consortia in the pea aphid complex is associated with large phenotypic variation in the insect host. *Evolutionary Ecology*, *30*(5), 925–941. doi: 10.1007/s10682-016-9856-1

Lenhart, P. A., & White, J. A. (2020). Endosymbionts facilitate rapid evolution in a polyphagous herbivore. *Journal of Evolutionary Biology*, *33*(10), 1507–1511. doi: 10.1111/jeb.13697

Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). *American Journal of Botany*, *82*(11), 1420–1425.

Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., … Rhule, E. (2014). Complementarity of statistical treatments to reconstruct worldwide routes of

invasion: The case of the Asian ladybird Harmonia axyridis. *Molecular Ecology*, *23*(24), 5979–5997.

Loxdale, H. D., Hardie, J., Halbert, S., Foottit, R., Kidd, N. A. C., & Carter, C. I. (1993). The relative importance of short- and long-range movement of flying aphids. *Biological Reviews*, *68*(2), 291–311. doi: 10.1111/j.1469-185X.1993.tb00998.x

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. doi: 10.1093/bioinformatics/btr507

Mahé, F., Rognes, T., Quince, C., Vargas, C. de, & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593. doi: 10.7717/peerj.593

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. doi: 10.14806/ej.17.1.200

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), 652–654. doi: 10.1038/351652a0

Meirmans, P. G., & Van Tienderen, P. H. (2004). Genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes*, *4*(4), 792–794.

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science*, *201*(4358), 786–792. doi: 10.1126/science.356262

Mizrahi-Man, O., Davenport, E. R., & Gilad, Y. (2013). Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLOS ONE*, *8*(1), e53608. doi: 10.1371/journal.pone.0053608

Morales-Hojas, R., Sun, J., Iraizoz, F. A., Tan, X., & Chen, J. (2020). Contrasting population structure and demographic history of cereal aphids in different environmental and

agricultural landscapes. *Ecology and Evolution*, *10*(18), 9647–9662. doi: 10.1002/ece3.6565

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. Retrieved from https://books.google.es/books?id=UhRSsLkjxDgC

Oliver, K. M., Degnan, P. H., Burke, G. R., & Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol*, *55*. doi: 10.1146/annurev-ento-112408-085305

Peccoud, J., Figueroa, C. C., Silva, A. X., Ramirez, C. C., Mieuzet, L., Bonhomme, J., … Simon, J.-C. (2008). Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Molecular Ecology*, *17*(21), 4608–4618. doi: 10.1111/j.1365-294X.2008.03949.x

Piffaretti, J., Clamens, A.-L., Vanlerberghe-masutti, F., Gupta, R. K., Call, E., Halbert, S., & Jousselin, E. (2013). Regular or covert sex defines two lineages and worldwide superclones within the leaf-curl plum aphid (Brachycaudus helichrysi, Kaltenbach). *Molecular Ecology*, *22*(15), 3916–3932. doi: 10.1111/mec.12371

Piry, S., Luikart, G., & Cornuet, J. M. (1999). Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity*, *90*(4), 502–503. doi: 10.1093/jhered/90.4.502

Popkin, M., Piffaretti, J., Clamens, A.-L., Qiao, G.-X., Chen, J., Vitalis, R., … Jousselin, E. (2017). Large-scale phylogeographic study of the cosmopolitan aphid pest Brachycaudus helichrysi reveals host plant associated lineages that evolved in allopatry. *Biological Journal of the Linnean Society*, *120*(1), 102–114. doi: 10.1111/bij.12869

Posada, D., & Crandall, K. A. (2001). Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology & Evolution*, *16*(1), 37–45. doi: 10.1016/S0169-5347(00)02026-7

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Current Biology*, *15*(5), R159–R160. doi: 10.1016/j.cub.2005.02.038

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016a). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. doi: 10.1093/bioinformatics/btv684

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. doi: 10.1111/1755-0998.12512

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, *35*(10), 1720–1728. doi: 10.1093/bioinformatics/bty867

Rebijith, K. B., Hande, R., Joshi, S., Surveswaran, S., Ramamurthy, V., & Kumar, N. (2017). Reconstructing the macroevolutionary patterns of aphids (Hemiptera: Aphididae) using nuclear and mitochondrial DNA sequences. *Biological Journal of the Linnean Society*. doi: 10.1093/biolinnean/blx020

Ripley, B. D., & Ripley, B. D. (2001). The R project in statistical computing. *MSOR Connections*.

Rispe, C., Legeai, F., Nabity, P. D., Fernández, R., Arora, A. K., Baa-Puyoulet, P., … Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biology*, *18*(1), 90. doi: 10.1186/s12915-020-00820-5

Rius, M., & Darling, J. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, *29*. doi: 10.1016/j.tree.2014.02.003

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. doi: 10.7717/peerj.2584

Rousset, F. (2008). Genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.

Roux, C., & Pannell, J. R. (2015). Inferring the mode of origin of polyploid species from next-generation sequence data. *Molecular Ecology*, (24), 1047–1059. doi: 10.1111/mec.13078

Rozas, J., Sànchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*(18), 2496–2497. doi: 10.1093/bioinformatics/btg359

Russell, J. A., Weldon, S., Smith, A. H., Kim, K. L., Hu, Y., Łukasik, P., … Oliver, K. M. (2013). Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, *22*(7), 2045–2059. doi: 10.1111/mec.12211

Simon, J.-C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, *17*(1), 34–39. doi: 10.1016/S0169-5347(01)02331-X

Simon, J.-C., Stoeckel, S., & Tagu, D. (2010). Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, *333*(6), 488–496. doi: 10.1016/j.crvi.2010.03.003

Sow, A., Brévault, T., Benoit, L., Chapuis, M.-P., Galan, M., Coeur D'acier, A., … Haran, J. (2019). Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports*, *9*(1), 3646. doi: 10.1038/s41598-019-40243-z

Spengler, R. N. (2019). Origins of the Apple: The Role of Megafaunal Mutualism in the Domestication of Malus and Rosaceous Trees. *Frontiers in Plant Science*, *10*. doi: 10.3389/fpls.2019.00617

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Stekolshchikov, A. V. (2006). Aphids of the genus Dysaphis Börner (Homoptera, Aphididae) living on plants of the family polygonaceae. *Entomological Review*, *86*(7), 787–805. doi: 10.1134/S0013873806070049

Stukenbrock, E. H. (2016). The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. *Phytopathology*, *106*(2), 104–112. doi: 10.1094/PHYTO-08-15-0184-RVW

Stukenbrock, E. H., & McDonald, B. A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, *46*(1), 75–100. doi: 10.1146/annurev.phyto.010708.154114

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504. doi: 10.1093/bioinformatics/btn478

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Tardío, J., Arnal, A., & Lázaro, A. (2020). Ethnobotany of the crab apple tree (Malus sylvestris (L.) Mill., Rosaceae) in Spain. *Genetic Resources and Crop Evolution*. doi: 10.1007/s10722-020-01026-y

Tsuchida, T., Koga, R., Shibao, H., Matsumoto, T., & Fukatsu, T. (2002). Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, Acyrthosiphon pisum. *Molecular Ecology*, *11*(10), 2123–2135. doi: 10.1046/j.1365-294X.2002.01606.x

Turcotte, M. M., Araki, H., Karp, D. S., Poveda, K., & Whitehead, S. R. (2017). The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *372*(1712). doi: 10.1098/rstb.2016.0033

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935.

Vorburger, C., Lancaster, M., & Sunnucks, P. (2003). Environmentally related patterns of reproductive modes in the aphid Myzus persicae and the predominance of two 'superclones' in Victoria, Australia. *Molecular Ecology*, *12*(12), 3493–3504. doi: 10.1046/j.1365-294X.2003.01998.x

Wang, Y., Hereward, J. P., & Zhang, G. (2016). High Spatial Genetic Structure and Genetic Diversity in Chinese Populations of Sitobion miscanthi (Hemiptera: Aphididae). *Journal of Economic Entomology*, *109*(1), 375–384. doi: 10.1093/jee/tov294

Warneys, R., Gaucher, M., Robert, P., Aligon, S., Anton, S., Aubourg, S., … Degrave, A. (2018). Acibenzolar-S-Methyl Reprograms Apple Transcriptome Toward Resistance to Rosy Apple Aphid. *Frontiers in Plant Science*, *9*, 1795. doi: 10.3389/fpls.2018.01795

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256–276.

Watterson, G. A. (1978). The homozygosity test of neutrality. *Genetics*, *88*(2), 405.

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 116. doi: 10.1186/1471-2105-11-116

Wei, G., Zuorui, S., Zhihong, L., & Lingwang, G. (2005). Migration and population genetics of the grain aphid Macrosiphum miscanti (Takahashi) in relation to the geographic distance and gene flow. *Progress in Natural Science*, *15*(11), 1000–1004. doi: 10.1080/10020070512331343176

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. doi: 10.2307/2408641

Wilkaniec, B. (1993). The influence of feeding of the rosy apple aphid, Dysaphis plantaginea (Pass.) (Homoptera: Aphididae) on the growth of apple fruits. *Roczniki Nauk Rolniczych. Seria E, Ochrona Roślin*, *23*(1/2), 75–78.

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., … Ma, H. (2017). Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Molecular Biology and Evolution*, *34*(2), 262–281. doi: 10.1093/molbev/msw242

Zepeda-Paulo, F. A., Simon, J. -C., Ramírez, C. C., Fuentes-Contreras, E., Margaritopoulos, J. T., Wilson, A. C. C., … Figueroa, C. C. (2010). The invasion route for an insect pest species: The tobacco aphid in the New World. *Molecular Ecology*, *19*(21), 4738–4752. doi: 10.1111/j.1365-294X.2010.04857.x

Zepeda-Paulo, F., Ortiz-Martínez, S., Silva, A. X., & Lavandero, B. (2018). Low bacterial community diversity in two introduced aphid pests revealed with 16S rRNA amplicon sequencing. *PeerJ*, *6*, e4725. doi: 10.7717/peerj.4725

Zhang, B., Edwards, O., Kang, L., & Fuller, S. (2014). A multi-genome analysis approach enables tracking of the invasion of a single Russian wheat aphid (Diuraphis noxia) clone throughout the New World. *Mol Ecol*, *23*. doi: 10.1111/mec.12714

Zhou, H.-X., Zhang, R.-M., Tan, X.-M., Tao, Y.-L., Wan, F., Wu, Q., & Chu, D. (2015). Invasion Genetics of Woolly Apple Aphid (Hemiptera: Aphididae) in China. *Journal of Economic Entomology*, *108*. doi: 10.1093/jee/tov074

Zytynska, S. E., & Weisser, W. W. (2016). The natural occurrence of secondary bacterial symbionts in aphids. *Ecological Entomology*, *41*(1), 13–26. doi: 10.1111/een.12281