

## ORIGINAL ARTICLE

# Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples

Amandine Cornille,<sup>1,2,3,\*</sup> Alice Feurtey,<sup>1,2</sup> Uriel Gélín,<sup>4</sup> Jeanne Ropars,<sup>1,2</sup> Kristine Misvanderbrugge,<sup>5</sup> Pierre Gladieux<sup>1,2</sup> and Tatiana Giraud<sup>1,2,\*</sup>

1 Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France

2 CNRS, Orsay, France

3 Department of Plant Ecology and Evolution, Uppsala University, Uppsala, Sweden

4 Département de biologie, Université de Sherbrooke, Sherbrooke, QC, Canada

5 Institute for Forestry and Game Management, Geraardsbergen, Belgium

## Keywords

admixture, crabapple, dispersal, glacial refugia, global changes, pollinators, SPIPOLL.

## Correspondence

Amandine Cornille, Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Sud, F-91405 Orsay, France.

Tel.: +4618 471 6408;

fax: +33 1 69 15 46 97;

e-mail: amandine.cornille@gmail.com

\*Both authors codirected the work.

Received: 4 September 2014

Accepted: 9 December 2014

doi:10.1111/eva.12250

## Introduction

Over the last century, researchers in evolutionary and applied biology have paid much attention to gene flow, which is considered to play a key role in species persistence, evolution, and diversification. Many studies have investigated the consequences of intra- and interspecies gene flow for evolution (e.g., Garant et al. 2007; Feder et al. 2012). However, the anthropogenic and natural determinants of gene flow remain largely unexplored (Manel et al. 2003; Manel and Holderegger 2013) because multiple data sources (genetic, environment, ecology) are required to unravel these issues. One central question concerns the extent to which climate changes and recent global changes

## Abstract

Gene flow is an essential component of population adaptation and species evolution. Understanding of the natural and anthropogenic factors affecting gene flow is also critical for the development of appropriate management, breeding, and conservation programs. Here, we explored the natural and anthropogenic factors impacting crop-to-wild and within wild gene flow in apples in Europe using an unprecedented dense sampling of 1889 wild apple (*Malus sylvestris*) from European forests and 339 apple cultivars (*Malus domestica*). We made use of genetic, environmental, and ecological data (microsatellite markers, apple production across landscapes and records of apple flower visitors, respectively). We provide the first evidence that both human activities, through apple production, and human disturbance, through modifications of apple flower visitor diversity, have had a significant impact on crop-to-wild interspecific introgression rates. Our analysis also revealed the impact of previous natural climate change on historical gene flow in the nonintrogressed wild apple *M. sylvestris*, by identifying five distinct genetic groups in Europe and a north–south gradient of genetic diversity. These findings identify human activities and climate as key drivers of gene flow in a wild temperate fruit tree and provide a practical basis for conservation, agroforestry, and breeding programs for apples in Europe.

due to human disturbances (e.g., land-use management and fragmentation) affect gene flow (Manel and Holderegger 2013). This question must be addressed, to make it possible to forecast the impact of these changes on the evolution of populations and to develop appropriate management, breeding, and conservation programs.

Within-species gene flow, which is mediated by pollen and seed dispersal, is a key factor underlying population persistence (Ellstrand 2014). Gene flow shapes the spatial genetic structure and can therefore be inferred from the amount of genetic differentiation between populations by the use of neutral marker loci (Vekemans and Hardy 2004). Over the last 20 years, an array of methods, including autocorrelation, parentage analyses, and assignment methods,

has been developed for the identification of natural factors affecting gene flow (Pritchard et al. 2000; Manel et al. 2003; Smouse and Sork 2004; Storfer et al. 2010). Studies based on these methods have identified climate as a key driver of within-species historical gene flow in many temperate species in Europe during the Pleistocene (Hewitt 2004). In particular, climate-driven range fluctuations during the Pleistocene have continuously reshaped genetic structure within species. However, the impact of climate on gene flow remains largely unexplored in scattered species, because of the difficulties of large-scale sampling. Yet, the determination of genetic diversity, differentiation, and admixture across the geographical range of such scattered species can be a practical basis for sustainable management and conservation programs.

Human activities also affect gene flow. In particular, the anthropization of landscapes and the cultivation of crops over extended areas have increased the opportunities for contact between cultivated plants and their wild relatives. Over the last 20 years, crop-to-wild gene flow has been considered a human disturbance and has received close attention, particularly in the context of the cultivation of transgenic crops and the ecological consequences of such gene flow for the fitness and genetic integrity of wild relatives (Ellstrand et al. 1999; Ellstrand 2003). Crop-to-wild gene flow has been investigated using neutral markers and shown to be widespread (Ellstrand et al. 1999; Coart et al. 2003; Delplancke et al. 2011; Aerts et al. 2013). However, only a few studies (Arnaud et al. 2003) have investigated the human disturbances and anthropogenic landscape drivers affecting crop-to-wild gene flow, despite the potentially large number of such drivers at work. Local human food crop production can affect crop-to-wild gene flow directly, by modifying the distribution and abundance of the crop. Humans can also have indirect effects on introgression rates, through their land-use practices. For instance, clear evidence has been obtained of a decline in the populations of both wild and domesticated pollinators associated with the intensification of cultivation and landscape management (Kremen et al. 2007). Pollinators are key vectors of gene flow and allele distribution in landscapes. Land-use changes can therefore have also an indirect impact on both crop-to-wild and within wild gene flow, through anthropogenic effects on the distribution and movements of the pollinator in the landscape.

Fruit trees, particularly scattered trees, are key species in ecosystems (Manning et al. 2006), and fruit tree crops are essential to human society, due to their contribution to worldwide food production (e.g., apple, grape, orange, coffee, cocoa, olive). However, despite the ecological and economic importance of these trees, much uncertainty remains concerning the impact of recent global changes on gene flow from crop-to-wild species and within wild

species, particularly in temperate fruit trees. The cultivated apple (*Malus domestica*) is the one of the most important fruit tree crop in terms of worldwide production (<http://faostat.fao.org/>). It was introduced into Europe 1500 years ago and has been subject to the introgression of genes from the local wild European apple, *Malus sylvestris*, to the extent that *M. domestica* is now more closely related to *M. sylvestris* than to its primary progenitor in the Tian Shan, *M. sieversii* (Cornille et al. 2012). Reciprocally, crop-to-wild introgressions have been reported in the European wild apple *M. sylvestris* (Coart et al. 2006; Larsen et al. 2006; Gross et al. 2012; Cornille et al. 2013b). Investigations of the demographic history of the European wild apple have revealed the existence of three main genetic groups shaped by past climate change during the Pleistocene (Cornille et al. 2013a). However, both a sampling gap in the distribution of *M. sylvestris* in France and a lack of information about the pollination ecology of this species have limited our understanding of the structure of the population and the identification of drivers of intra- and interspecies gene flow. Yet, cultivated and wild apples in Europe, with their high levels of gene flow from crop-to-wild species and within wild species, easily detectable due to relatively recent secondary contact, constitute an ideal model for investigations of the anthropogenic and natural drivers of gene flow in fruit trees at large geographic scales.

We report here an unprecedented dense sampling of *M. sylvestris* across Europe (1889 individuals), with a focus in France (1092 individuals), unique for a temperate fruit tree. We also sampled a large number of *M. domestica* varieties (339 varieties), and combined genetic (microsatellites), ecological, and environmental data, to investigate the natural and anthropogenic drivers of gene flow in apples. Using genetic data across Europe, we estimated the degree of crop-to-wild introgressions into the European wild apple, across its geographic distribution and investigated the impact of climate on population structure and diversity, through estimates of historical gene flow within the nonintrogressed wild apple gene pool. Furthermore, by estimating recent crop-to-wild introgressions in France, and using environmental and ecological data, we investigated the effects of apple production on these crop-to-wild introgressions. Finally, we determined whether human activities had indirectly impacted rates of gene flow through their effects on the flower visitor community.

Altogether, we highlighted the impact of previous natural climate change on historical gene flow across Europe in the wild apple and identify human activities as key drivers of gene flow in a wild temperate fruit tree. Our landscape genetics study provides useful information for *in situ* and *ex situ* conservation management for this emblematic fruit tree species in Europe, for agroforestry programs aiming at integrating wild apples in agrosystems and for future

breeding programs, the European wild apple being a major contributor to the cultivated apple genome.

## Materials and methods

### Sampling, DNA extraction, and microsatellite genotyping

In addition to the 599 samples of *M. sylvestris* from Europe previously studied (Cornille et al. 2013a,b) (38 sites), we collected a further denser sample of 1290 *M. sylvestris* trees (28 sites), mostly in France (20 sites,  $N = 1092$ ), but also in Denmark (6 sites,  $N = 162$ ), Italy (1 site,  $N = 5$ ) and Belgium (1 site,  $N = 31$ ). In total, 1889 *M. sylvestris* trees from 66 sites were analyzed for this study. In addition to the 299 *M. domestica* previously analyzed (Cornille et al. 2012), we sampled additional *M. domestica* trees in the field in Denmark ( $N = 24$ ) and in French apple orchards ( $N = 16$ ). In total, we analyzed 339 *M. domestica* trees in this study. The dataset for this study thus corresponds to 2228 individuals (details of the samples are provided in Table S1 and Fig. S1).

DNA was extracted with the NucleoSpin<sup>®</sup> plant II DNA extraction kit (Macherey & Nagel, Düren, Germany). Multiplex microsatellite PCR amplifications were performed with a multiplex PCR kit (Qiagen Inc.) as previously described (Cornille et al. 2012; Patocchi et al. 2009), with 26 microsatellite markers. We retained only multilocus genotypes for which <20% of the data were missing. The suitability for population genetic analysis of the markers used was demonstrated in previous studies (Cornille et al. 2012, 2013a,b).

### Crop-to-wild gene flow, population structure, and historical gene flow (large-scale analyses, Europe)

#### Bayesian inference of crop-to-wild gene flow, population structure, and admixture

We used the individual-based Bayesian clustering methods implemented in STRUCTURE 2.3.3 (Pritchard et al. 2000) to estimate recent introgression from *M. domestica* into *M. sylvestris*. STRUCTURE uses Markov chain Monte Carlo (MCMC) simulations to infer the proportion of ancestry of genotypes from  $K$  distinct clusters. The underlying algorithms attempt to minimize deviations from Hardy–Weinberg within clusters and linkage disequilibrium among loci. A lack of consideration of intraspecies genetic structure in STRUCTURE analyses can bias the interpretation of introgression rates (Kalinowski 2011; Cornille et al. 2013b). We ran STRUCTURE, assuming  $K = 1$  to  $K = 9$  clusters, on the whole dataset including *M. sylvestris* ( $N = 1889$ ) and *M. domestica* ( $N = 339$ ). Based on 10 repeated runs of MCMC sampling comprising 500 000 iterations after a burn-in of 50 000 steps, we determined the amount of additional information explained by increasing  $K$ , using the

$\Delta K$  statistic (Evanno et al. 2005). As a genetic structure was observed within the *M. sylvestris* dataset (see below), we assumed that the total membership coefficient of *M. sylvestris* individuals in the *M. sylvestris* gene pool was the sum of their membership coefficients in each *M. sylvestris* cluster. All *M. sylvestris* individuals with a cumulative membership coefficient for the *M. sylvestris* gene pool >0.9 were considered to be pure *M. sylvestris* trees. Conversely, all *M. sylvestris* individuals with a cumulative membership coefficient >0.1 in the *M. domestica* gene pool were considered to be admixed with *M. domestica*. *Malus sylvestris* individuals with a cumulative membership coefficient <0.1 in the *M. sylvestris* gene pool were considered to be misidentified individuals actually belonging to *M. domestica*. The proportion admixture of the *M. domestica* gene pool with the *M. sylvestris* gene pool is hereafter denoted  $P_{\text{DOM} \rightarrow \text{SYL}}$  and referred to as ‘introgression rate’. In the absence of genomewide genotyping data, we probably failed to identify advanced generation backcrosses. However, our approach aimed to identify early generations of admixture and introgression, because our focus here was the effect of recent human disturbance on gene flow.

To investigate the population structure and admixture of *M. sylvestris*, we excluded hybrids previously identified as displaying introgression from *M. domestica* (i.e., individuals assigned to a wild gene pool with a membership coefficient <0.9;  $N = 1376$  individuals, 62 sites) and used the individual-based spatially explicit Bayesian clustering method implemented in TESS 2.3.1 (Chen et al. 2007). TESS is similar to STRUCTURE, except that it incorporates a spatial component into the clustering procedure, such that genotypes from areas located close together geographically are considered more likely to belong to the same cluster. We used the conditional autoregressive (CAR) Gaussian model of admixture with linear trend surface, setting the spatial interaction parameter ( $\rho$ ) to 0.6. These parameters ( $\rho$  and trend) affect the weight given to spatial distance when clustering genotypes. Ten independent analyses were carried out for each number of clusters  $K$  ( $2 \leq K_{\text{max}} \leq 9$ ), using 500 000 MCMC iterations after a burn-in of 50 000 steps. We determined the amount of additional information explained by increasing  $K$ , using the rate of change of the deviation index criterion (DIC) with increasing  $K$ .

For both TESS and STRUCTURE analyses, outputs were processed with CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007), to identify distinct clustering solutions in the replicated runs for each  $K$  value.

#### Genetic variation within *M. sylvestris*

For the analysis of within-species genetic variation, we excluded hybrids identified as displaying introgression from *M. domestica*. Descriptive statistical analyses were performed for each site (i.e., geographical location) and

each population (i.e., clusters inferred by TESS, including hybrids with a membership coefficient of up to 0.55 for the focal cluster).

Allelic richness ( $A_R$ ) and private allelic richness ( $A_P$ ) were calculated with ADZE (Szpiech et al. 2008), using standardized sample sizes of  $N = 2$  (one individual  $\times$  two chromosomes) and  $N = 100$  (50 individuals  $\times$  two chromosomes), respectively, corresponding to the minimal number of observations across sites and populations. Heterozygosities (expected and observed), Weir and Cockerham  $F$ -statistics and deviations from Hardy–Weinberg equilibrium were calculated with GENEPOP 4.2 (Raymond and Rousset 1995; Rousset 2008). Only sampling sites for which at least four individuals were successfully genotyped were included in site-specific calculations (51 of the 62 sites). Spatial patterns of genetic variability were visualized by mapping variation ( $A_R$ ), for 43 sites (i.e., geographical locations with at least four individuals and for which at least two individuals were successfully genotyped for each marker) and genetic diversity ( $H_C$ ) for 51 sites across space, with the kriging and surface functions (*field* R package).

### Spatial autocorrelation and historical gene flow

We estimated spatial genetic structure within the five populations previously identified with TESS. Analyses were performed with SPAGeDI 1.3 (Hardy & Vekemans 2002) for each population, using the estimator of kinship coefficient of Nason (Loiselle et al. 1995). Kinship coefficient values ( $F_{ij}$ ) were regressed against the natural logarithm,  $\ln(d_{ij})$ , to obtain the regression slope,  $b$ . The spatial positions of individuals were permuted 9999 times to test for the existence of a spatial genetic structure. We compared the extent of spatial genetic structure between sites over the same spatial scale, by calculating  $b_{Ld}$ , the regression slope of  $F_{ij}$  against  $\ln(d_{ij})$  for  $d_{ij} = 10$  km (the maximum distance providing the best representation of pairs of individuals for each population). We then calculated the  $Sp$  statistic, defined as  $Sp = -b_{(Ld10)} / (1 - F_N)$ , where  $F_N$  is the mean  $F_{ij}$  between neighboring individuals, which was approximated by  $F_{(d)}$  for the first distance interval (Vekemans and Hardy 2004; Hardy et al. 2006).

### Factors affecting crop-to-wild gene flow and the diversity of flower visitors (fine-scale analyses, France)

All statistical analyses were performed with R version 2.15.3 (R Foundation for Statistical Computing, Vienna, Austria).

#### *Anthropogenic factors affecting crop-to-wild introgression*

We expected the density of cultivated apple trees to be the principal determinant of recent crop-to-wild introgression

rate detected by STRUCTURE (i.e.,  $P_{DOM \rightarrow SYL}$ ). We first investigated the effect of the density of cultivated apple trees (based on apple production in tons/ha/year for the French administrative region concerned) on  $P_{DOM \rightarrow SYL}$  in wild apple. We then included cultivated apple density as a fixed factor in the models 1 and 2 described below, as the source of the crop-to-wild gene flow. Because the number and area of apple orchards were highly correlated ( $r = 0.91$ ,  $P < 0.001$ ), we investigated the effects of the number and area of apple orchards on  $P_{DOM \rightarrow SYL}$  independently in two distinct models (hereafter called models 1 and 2, respectively), controlling for cultivated apple tree density. With this approach, we aimed to determine the relative contributions of orchard number and orchard area to crop-to-wild gene flow in apples. For these analyses, we used the dense sample obtained in France ( $N = 1028$  individuals from 20 sites, Fig. S1), and we ran a generalized linear model (GLM) with a quasi-Poisson distribution, as the data appeared to be overdispersed. Apple production (tonnes/year) and the number of and area covered by apple orchards were extracted, for each French administrative region, from the AGRESTE (<http://agreste.agriculture.gouv.fr/>).

#### *Influence of cultivation and management on the distribution of taxa visiting apple flowers*

Anthropogenic effects on introgression rates may also be indirectly mediated by effects on pollinators. An intensification of cultivation and management practices may affect the diversity of flower visitors and, thus, gene flow. We therefore explored the impact of the intensity of cultivation and management on the diversity of the taxa visiting flowers in France (hereafter called model 3). We investigated whether the frequency of apple taxa visiting flowers depended on the intensity of cultivation and management. We ran a GLM with a quasi-Poisson distribution, as the data appeared to be overdispersed. In total, 306 observations corresponding to different intensities of cultivation were recorded (Fig. S1 and Table S3).

The intensity of cultivation and management was obtained from the GLC2000 dataset (The Land Cover Map for Europe in the Year 2000, <http://edit.csc.es/Soil-Vegetation-LandCover.html>). We defined two classes of cultivation and management intensity, distinguishing between areas with higher and lower intensities of cultivation and management, corresponding to the 1–15 and 17–18 classes of the GLC2000 dataset, respectively. These two classes were defined qualitatively on the basis of the standard classification in the worldwide GLC database as ‘natural’ or ‘cultivated and managed’. Individual geographical coordinates were assigned to these two classes with the extract function (*raster* R package).

The diversity of the insect taxa visiting apple flowers (i.e., the number of observed taxa per insect order) was



obtained from a nationwide monitoring exercise based on observations by citizens. Using the database generated by this exercise, entomology experts used the photographs provided to check the preliminary classifications obtained with a simple online key by the citizens participating in the project (for further information on the SPIPOLL project methodology, see Deguines et al. 2012). As it is difficult to differentiate between *M. sylvestris* and *M. domestica* morphologically, we considered that it would not be possible to assign the observations of citizens to one species or the other; these observations were therefore pooled.

## Results

### Crop-to-wild gene flow, population structure, and historical gene flow (large-scale analyses, Europe)

#### Large spatial scale of crop-to-wild gene flow in apple

Failing to take the population structure of the wild species into account can lead to spurious signals of introgression from crop species (Kalinowski 2011). We therefore analyzed the structure of *M. sylvestris* and identified five distinct clusters within the wild species, and in addition a cluster corresponding to *M. domestica* (Fig. S2). The use of  $K$  values  $>6$  uncovered no further structure within *M. sylvestris*, indicating that the *M. sylvestris*  $\times$  *M. domestica* hybrids detected at  $K = 6$  were not artifacts due to genetic structure within the wild apple gene pool. We therefore used cumulative membership coefficient in the five *M. sylvestris* clusters in subsequent analyses to identify crop-to-wild hybrid genotypes.

For the total dataset for individuals identified *a priori* as *M. sylvestris* ( $N = 1889$ ), 436 genotypes (23.1% of the total dataset) showed signs of introgression from *M. domestica* (i.e., cumulative membership coefficients  $<0.9$  and  $>0.1$  in the *M. sylvestris* gene pool,  $P_{\text{DOM} \rightarrow \text{SYL}}$ ) and 77 genotypes were found to correspond to misidentified *M. domestica* trees (i.e., cumulative membership coefficients  $<0.1$  in the *M. sylvestris* gene pool; 4.1% of the total dataset). The mean introgression rates and numbers of hybrids are presented in Fig. 1, Tables S1 and S2. Our results indicated large-scale crop-to-wild gene flow in apple in Europe.

We also observed recent wild-to-crop introgressions into the *M. domestica* cultivars (i.e., cultivars with membership coefficients  $<0.9$  and  $>0.1$  in the *M. domestica* gene pool), as previously demonstrated (Cornille et al. 2012). The mean wild-to-crop introgression rate was 0.15 (Table S3).

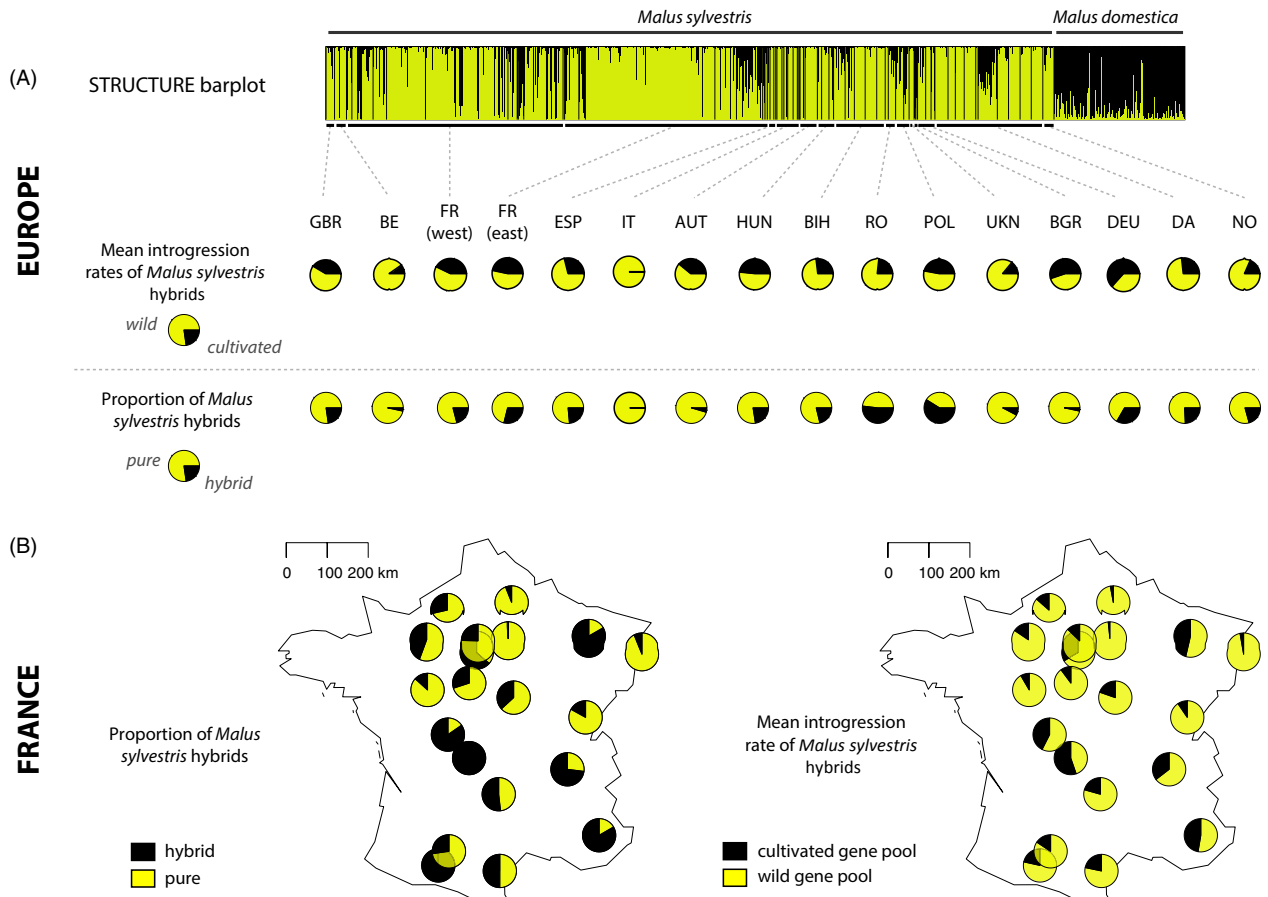
#### Spatial genetic variation and population structure in *Malus sylvestris*

For analyzing the population structure of *M. sylvestris*, introgressed individuals (i.e., genotypes assigned to a wild gene pool with a membership coefficient  $<0.9$ ) were removed. The final dataset included 1376 individuals (62

sites). For each site, the summary statistics for genetic diversity are shown in Table S4. On average, gene diversity was high across sites and across markers [ $H_E = 0.79 \pm 0.07$  (min–max: 0.57–0.89)] and  $F_{\text{IS}}$  values were low [ $F_{\text{IS}} = 0.10 \pm 0.06$  (0–0.21)], although there was a highly significant species-wide heterozygote deficit compared to Hardy–Weinberg expectations ( $P < 0.001$ ). Mean  $F_{\text{ST}}$  across loci were small [ $F_{\text{ST}} = 0.025 \pm 0.04$  (0–0.56)], but significant for 1349/1831 pairs of sites ( $P < 0.05$ ). Overall, these results indicate that spatial genetic structure is weak in *M. sylvestris*.

The map of interpolated  $A_R$  (Fig. 2C) showed that genetic variability decreased with increasing latitude, with the highest values found in Southern and Eastern Europe. This latitudinal trend was confirmed by the highly significant negative correlation observed between latitude and  $A_R$  ( $r = -0.75$ ,  $P < 0.001$ ). A significant positive correlation between longitude and  $A_R$  was also found ( $r = 0.33$ ,  $P = 0.02$ ). The map of interpolated  $H_O$  (Fig. 2D) showed a significant negative correlation between longitude and  $H_O$  ( $r = -0.30$ ,  $P = 0.03$ ); no correlation between latitude and  $H_O$  was also found ( $r = 0.19$ ,  $P = 0.16$ ). This latter map highlights the main admixture zones among *M. sylvestris* populations as regions of high diversity.

Spatially explicit clustering analyses carried out with TESS confirmed the existence of five well-defined clusters within *M. sylvestris* (Figs 2A,B and S3): two of these clusters spanned Western (green) and Central Europe (yellow); one spread over Eastern Europe and the Balkan Peninsula (red); one covered Northern Europe (blue); the last one was located in Italy (purple). DIC values decreased monotonically from  $K_{\text{max}} = 2$  to  $K_{\text{max}} = 9$ , indicating that increases in the number of clusters continually improved the fit of the model to the data, with a sharper decrease in DIC was observed at  $K_{\text{max}} = 8$  (Fig. S4) indicating that  $K_{\text{max}} = 8$  was the best-fit model. However, at  $K_{\text{max}} = 8$ , only five clusters still appeared (Fig. S3), as in three of the eight clusters, the membership coefficients of all individuals were close to 0. We therefore concluded that the biologically most relevant number of clusters was five, as observed in STRUCTURE analyses; we nevertheless used the TESS membership coefficient inferred at  $K_{\text{max}} = 8$  to define the five populations used in subsequent analyses, as it had the highest DIC value and therefore the best fit to the data. Individuals were assigned to populations for which their membership coefficients exceeded 0.55. The five populations are hereafter referred to as the Western (green,  $N = 421$ ), Central (yellow,  $N = 362$ ), Southern (purple,  $N = 83$ ), Eastern (red,  $N = 177$ ), and Northern (blue,  $N = 245$ ) populations. In total, 88 genotypes could not be assigned to any population;  $N = 1288$  individuals were thus retained for population-specific analyses.



**Figure 1** Geographical quantification of introgression rates and numbers of hybrids among European wild apple sites in Europe and, more specifically, in France. (A) STRUCTURE barplot for  $K = 6$ , for the European wild apple gene pool (yellow) and the cultivated apple gene pool (black) at the European scale ( $N = 2228$ , 66 sites); below, associated country-averaged introgression rates ( $P_{\text{DOM} \rightarrow \text{SYL}}$ ) and number of hybrids. GB: Great Britain, BE: Belgium, FR (west): Western France, FR (east): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway. (B) Site-averaged introgression rates ( $P_{\text{DOM} \rightarrow \text{SYL}}$ ) and number of hybrids in France ( $N = 1092$ , 20 sites).

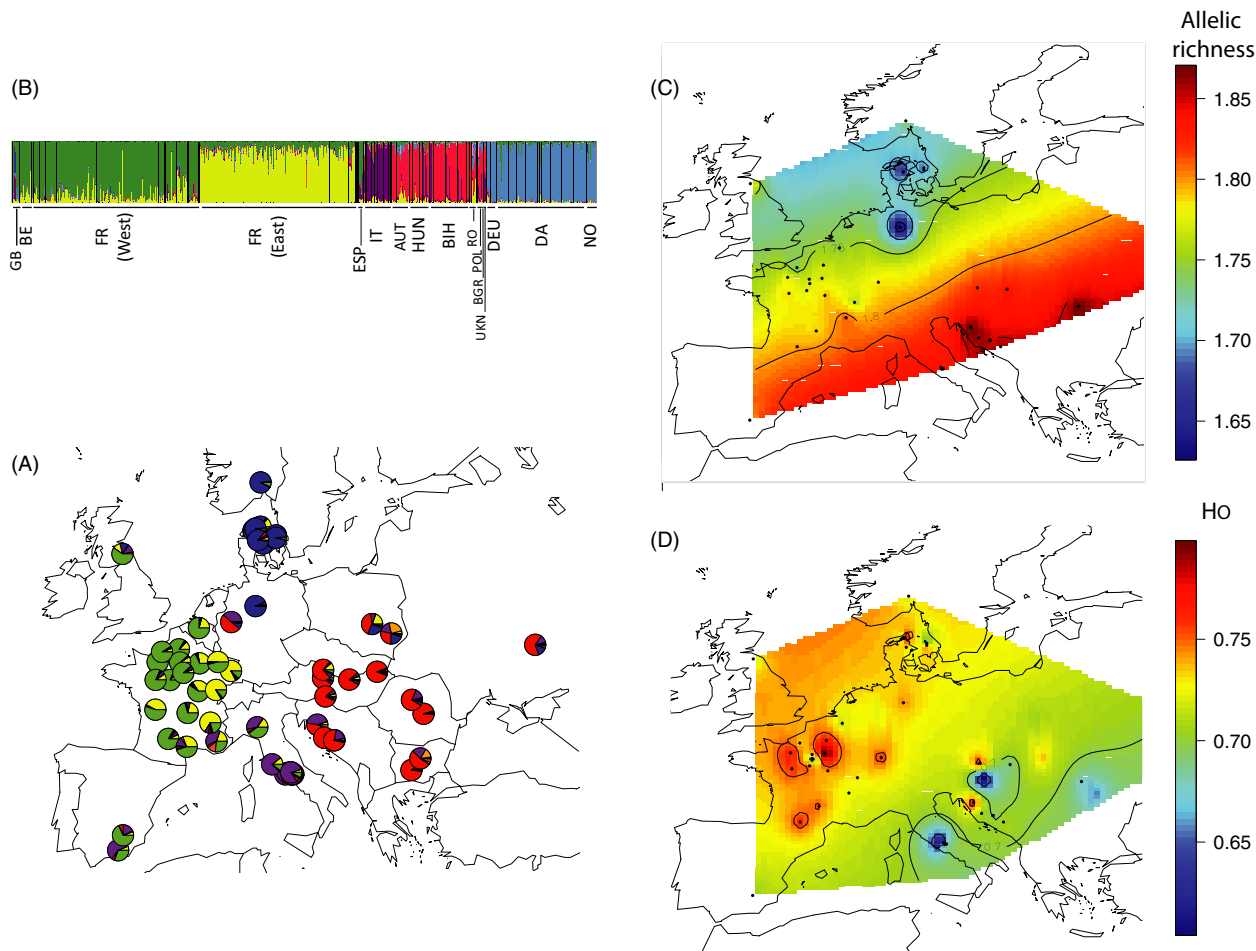
For each population, summary statistics for genetic diversity are shown in Table 1. Genetic variability ( $H_E$  and  $H_O$ ,  $A_R$  and  $A_P$ ) differed significantly between all populations (Wilcoxon signed rank (WSR) test,  $P < 0.001$ ) other than the Western and Central populations (WSR,  $P > 0.60$ ), and the Eastern and Southern populations (WSR,  $P > 0.44$ ). Overall, the Eastern and Southern populations displayed significantly higher levels of genetic variability, in terms of allelic richness ( $A_R$  and  $A_P$ ), than the Northern, Western, and Central populations. Higher heterozygosities ( $H_E$  and  $H_O$ ) were considered to be footprints of recent zones of admixture between different genetic clusters (i.e., the Western/Central populations and Central/Eastern populations).

Proportions of admixture differed between the five populations (Fig. 2, Tables 1 and 2). Substantial admixture was detected between the Western, Southern, and Northern populations, with hybrid numbers highest for the Western and Central populations ( $N = 321$  admixed individuals)

and then for the Northern and Western populations ( $N = 41$ ) (Table 2). The Eastern cluster mostly displayed admixture with the Southern ( $N = 31$ ), Northern ( $N = 21$ ), and Central ( $N = 21$ ) populations. The Southern population displayed admixture mostly with the Eastern ( $N = 31$ ) population and the Western population ( $N = 23$ ). The admixture detected is indicative of gene flow between these populations. Genetic differentiation ( $F_{ST}$ ) was greatest between the Central/Western/Northern and Southern/Eastern populations (Table 2). These results indicate a strong west–east differentiation of allelic frequencies, that was detected at  $K = 2$  by TESS and STRUCTURE (Figs S2 and S3).

#### *Spatial autocorrelation, isolation by distance (IBD), and historical gene flow*

Within each of the five populations, genetic differentiation and geographical distance were significantly correlated,



**Figure 2** Geographical genetic diversity, genetic structure, and admixture of *Malus sylvestris* in Europe (A) Population structure for the nonintrogressed *Malus sylvestris* dataset ( $N = 1376$ , 62 sites), inferred with TESS, for  $K_{\max} = 8$ , showing five distinct genetic clusters. Each vertical line represents an individual. Individuals were grouped by country. Colors represent the inferred ancestry from  $K$  ancestral populations. For clarity, the 62 sites are grouped by country: GB: Great Britain, BE: Belgium, FR (West): Western France, FR (East): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway; (B) Map representing the mean membership proportions inferred by TESS for the five detected clusters, for samples of *Malus sylvestris* collected from the same site (62 sites across Europe). (C) and (D) Maps of interpolated allelic richness  $A_R$  and observed heterozygosity  $H_O$ , respectively, across Europe.

consistent with an IBD model (Table 1). The  $S_p$  statistic was low suggesting weak spatial genetic structure and high levels of historical gene flow and/or effective population size. Overall, the mean  $S_p$  value was 0.01.

#### Factors affecting crop-to-wild gene flow and flower visitors (finer scale analyses in France)

##### Anthropogenic factors affecting crop-to-wild introgression

For model 1, the proportion of admixture of the *M. domestica* gene pool into the *M. sylvestris* gene pool ( $P_{\text{DOM} \rightarrow \text{SYL}}$ ) was found to be dependent on the interaction between cultivated apple density and the number of apple orchards (Table 3). Indeed, for small numbers of apple orchards,  $P_{\text{DOM} \rightarrow \text{SYL}}$  were unaffected by cultivated apple density, whereas cultivated apple density had a positive effect on

$P_{\text{DOM} \rightarrow \text{SYL}}$  for large numbers of orchard apple farms. For model 2,  $P_{\text{DOM} \rightarrow \text{SYL}}$  appeared to depend on the interaction between cultivated apple density and the area of apple orchards (Table 3) with the same trend found in model 1.

##### Factors affecting the diversity of flower visitors

Apple species were visited by 86 taxa from six orders of insects (Fig. 3 and Table S5). In model 3, the frequency of taxa visiting apple flowers was significantly lower in intensively cultivated and managed areas (26 taxa, 59 observations) than in less intensively cultivated and managed areas (83 taxa, 247 observations) (Fig. 3 and Table 3), after controlling for the order to which each taxon belonged. This relationship was also observed for all the insect orders considered (Fig. 3 and Table 3,  $P < 0.0001$ ).

**Table 1.** Genetic polymorphism, population structure, and the spatial pattern of genetic differentiation in *Malus sylvestris* ( $N = 1288$  individuals).

	Northern	Eastern	Southern	Western	Central
$N$	245	177	83	421	362
$H_O$	0.66	0.79	0.84	0.72	0.74
$H_E$	0.72	0.89	0.89	0.78	0.79
$F_{IS}$	0.08***	0.11***	0.06***	0.07***	0.07***
$A_R$	1.72	1.89	1.89	1.78	1.79
$A_P$	0.67	1.17	1.19	0.73	0.79
<b>Population structure (colors as in Fig. 2)</b>					
$Q_1$ (blue)	0.92	0.02	0.02	0.01	0.04
$Q_2$ (red)	0.04	0.85	0.13	0.04	0.02
$Q_3$ (purple)	0.02	0.11	0.84	0.04	0.05
$Q_4$ (green)	0.00	0.01	0.01	0.84	0.04
$Q_5$ (yellow)	0.01	0.01	0.01	0.07	0.86
<b>Spatial pattern</b>					
Mean(Ln(dist))	0.53	0.34	1.17	0.77	1.55
$b$	-0.014***	-0.021***	-0.007***	-0.011**	-0.007**
$F_{(1)}$	0.04	0.06	0.06	0.06	0.008
$r^2$	0.08	0.15	0.03	0.05	0.005
Sp	0.014	0.022	0.007	0.011	0.007

$N$ , sample size of each population;  $H_O$  and  $H_E$ , observed and expected heterozygosity, respectively;  $F_{IS}$ , inbreeding coefficient;  $A_R$ , mean allelic richness, corrected by the rarefaction method, estimated for a sample size of 2;  $A_P$ , private allelic richness, corrected by the rarefaction method, estimated for a sample size of 100;  $Q_{1-5}$ , mean membership coefficient for each of the five clusters inferred by TESS analyses, including hybrids with membership coefficients of up to 0.55 for the cluster concerned; Sp, Sp parameter; mean(Ln(dist)), mean of the logarithm of the geographic distance between genotypes;  $b$ , slope of the regression between  $F_{ij}$  and the logarithm of geographic distance;  $F_{(1)}$ , mean  $F_{ij}$  between individuals from the first distance class;  $r^2$ , squared correlation coefficient for the relationship between the logarithm of geographic distance and  $F_{ij}$ ; \*\*,  $0.01 < P \leq 0.001$ ; \*\*\*,  $P < 0.001$ .

**Table 2.** Genetic differentiation (upper triangle,  $F_{ST}$ ) and total number of admixed genotypes (lower triangle, i.e. individuals assigned to the population with membership coefficients of 0.55–0.9) among the five populations of *Malus sylvestris*.

	Northern	Western	Central	Southern	Eastern
Northern	–	0.03	0.06	0.13	0.11
Western	41	–	0.04	0.09	0.08
Central	14	321	–	0.09	0.08
Southern	2	23	11	–	0.04
Eastern	21	13	21	31	–

All  $F_{ST}$  values were highly significant ( $P < 0.001$ ).

## Discussion

Previous studies investigating crop-to-wild gene flow in fruit trees have provided estimates of introgression rates, but did not identify the factors influencing these rates (Delpiancke et al. 2011; Hoban et al. 2012; Aerts et al. 2013). This study provides new insight into the impact of human activities on the extent of recent crop-to-wild gene flow in fruit trees. We detected (i) large-scale recent introgressions from the cultivated apple to *M. sylvestris*, the extent of which was influenced by the intensity of apple production, and (ii) a decrease in the diversity of apple flower visitors with increasing intensity of cultivation and management

practices, suggesting an indirect impact of land-use on pollen flow.

Our study also unraveled the genetic structure of a scattered temperate European fruit tree, through comprehensive sampling. It provides new insight into the history of the wild apple, by revealing the influence of past climatic changes and barriers to dispersal in shaping the genetic structure of this forest fruit tree, resulting in the formation of differentiated genetic clusters in Europe. We also estimated levels of inter- and intracluster gene flow, and we discuss the fundamental and applied impact and significance of these findings below.

## Human impact on crop-to-wild gene flow in apples

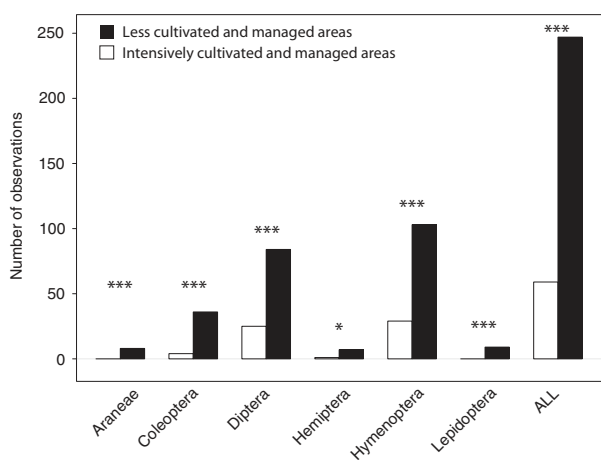
Since its introduction into Europe by humans 1500 years ago, the cultivated apple has acted as a source of genes for introgression into the European wild apple (Cornille et al. 2013b). We found that the rates of crop-to-wild introgression varied across Europe and could be accounted for by direct and indirect anthropogenic factors. Interestingly, the density of cultivated apple trees, as estimated from apple production, was found to be positively related to recent introgression rates in wild apple populations, with the number of apple orchards surrounding wild populations and with the total orchard area. These results suggest that



**Table 3.** Summary of minimal models 1, 2, and 3, which were tested for the identification of factors affecting crop-to-wild gene flow and the diversity of visitors to apple flowers.

Model	Response variable	Explanatory variable	F-value (df)	P-value
Model 1	$P_{\text{DOM} \rightarrow \text{SYL}}$	Cultivated apple density	0.67 (1,563)	0.387
		Number of apple orchards	46.64 (1,563)	<0.001
		Number of apple orchards * Cultivated apple density	31.53 (1,563)	<0.001
Model 2	$P_{\text{DOM} \rightarrow \text{SYL}}$	Cultivated apple density	10.01 (1,563)	<0.001
		Area of apple orchards	27.20 (1,563)	<0.001
		Area of apple orchards * Cultivated apple density	25.92 (1,563)	<0.001
Model 3	Diversity of taxa	Intensity of cultivation and management	125.12 (1,306)	<0.001
		Order	41.97 (1,306)	<0.001

$P_{\text{DOM} \rightarrow \text{SYL}}$ , proportion of admixture of the *Malus domestica* gene pool into the *Malus sylvestris* gene pool; df, degrees of freedom;  $x*y$ , interaction between explanatory variables  $x$  and  $y$ .



**Figure 3** Number of observations for the six orders of insect visitors to apple flowers in France (N = 306), in areas with more intensive (white) and less intensive (black) cultivation and management. ALL: number of observations for all orders considered together. \*\*\*:  $P < 0.0001$ ; \*:  $P < 0.01$ .

crop-to-wild introgression in apple is both dependent on the distribution of orchards within the landscape and on their area. It seems intuitive that several large apple orchards scattered across the landscape would act as a more efficient source of crop-to-wild gene flow than a smaller number of small orchards. Our results indicate that human management practices have also had an impact on the insect community visiting apple flowers in France. The numbers of insect orders visiting apple trees were smaller in managed and cultivated areas, suggesting that human disturbance has an indirect impact on gene flow, through negative effects on potential pollinator populations. Further investigations are required to determine the extent to which the changes in apple visitor communities according to land-use affect gene flow from crop-to-wild species and between populations in wild species.

### New insights into the demographic history of the European wild apple

Past climate changes and spatial barriers to dispersal across Europe can explain the population structure of wild apples revealed here. Over the last 2.6 million years, trees species have experienced dynamic alternations between the contraction and expansion of their ranges in Europe, in response to climate changes (Hewitt 1996, 1999; Petit et al. 2003). Our investigations of the population structure and diversity of *M. sylvestris*, based on the use of unprecedented dense sampling in France, provide further insight into the distribution ranges of apples in Europe in the past, over and above that provided by previous studies (Coart et al. 2006; Larsen et al. 2006; Cornille et al. 2013a; Schnitzler et al. 2014). We identified five well-differentiated gene pools with specific spatial distributions, whereas our previous sampling campaign led to the identification of only three clusters (Cornille et al. 2013a). Furthermore, the denser sampling used in this study made it possible to identify hotspots of allelic richness located in the South of France, Italy, the Balkans, and the Carpathian Mountains. The South of France, the Balkans, and the Carpathian Mountains have already been identified as glacial refugia (Cornille et al. 2013a), but our results suggest that Italy and the south-eastern Alps may also have served as glacial refugia for wild apple trees, as already demonstrated for other tree species (Tzedakis et al. 2013). We also found a longitudinal trend for allelic richness that could be potentially accounted for by an east–west gradient of climate severity in the Mediterranean region during the last glacial maximum (Conord et al. 2012). Spain would thus have been a minor glacial refugium for *M. sylvestris*, due to its harsher climate. Three main contact zones between clusters were identified, in the South-eastern Alps (between the green and purple clusters), in France (between the green and yellow clusters), and in the Carpathian Mountains (between the red and purple clusters). These results suggest that previous contractions of the distributional range of wild

apple during the last glacial maximum in southern Europe were followed by a recolonization of Northern Europe via two main routes: a western front (South of France—Italy, up to Scandinavia) and an eastern front (Balkans—Carpathian Mountains, up to Scandinavia). In particular, the additional Italian cluster detected appears to have acted as a bridge for the recolonization of the western and eastern parts of Europe after glaciation. Barriers to dispersal during recolonization and bottlenecks may also have affected the existing population structure. For instance, populations from northern Germany, Denmark, and Norway have a low allelic richness, and a bottleneck might account for these populations belonging to a distinct gene pool.

#### Weak isolation by distance reveals high historical levels of gene flow during the recolonization of Europe

Our findings suggest that, despite the observed population structure, the level of natural intraspecific gene flow during the course of history was high for a wild fruit tree. Indeed, we detected a weak spatial genetic structure within each cluster, spanning large geographic distributions (with low  $S_p$  values), suggesting relatively high dispersal capacities and/or effective population sizes (Vekemans and Hardy 2004). Long-distance dispersal events have been reported for another temperate fruit tree, *Sorbus domestica* (Kamm et al. 2009). Nevertheless, we might have expected to find a stronger genetic structure in this zoochorous fruit tree, because fruits carry half-sib or full-sib seeds as dispersal units, rendering dispersal nonrandom. This would be expected to lead to a stronger genetic structure than is observed in wind-pollinated trees (Hardy et al. 2006). However, the spatial genetic structure detected in this study was as weak as that detected in wind-dispersed trees. This may reflect the small number of generations since European recolonization after the last glacial maximum (<10 000 years ago, corresponding to about 1000 generations). Too little time has thus elapsed for genetic differentiation to occur, particularly given the long generation time, overlapping generations and high effective population sizes of trees (Austerlitz et al. 2000). We used genetic distances between individuals to estimate gene flow. This method principally reflects historical patterns of gene flow (Manel et al. 2003). Our estimates may thus not be indicative of local-scale pollen and seed gene flows in contemporary landscapes. Only genetic analyses of parentage (Smouse and Sork 2004) would provide an estimate of the current dispersal abilities of the wild European apple.

#### Conclusion and future prospects: a basis for future conservation and breeding programs

The consequences of crop-to-wild gene flow for the wild apple remain unclear, particularly as concerns the possibility

of genetic swamping (Lenormand 2002). This phenomenon occurred in the wild ancestor of coconut (*Cocos nucifera*), in which introgression from domesticated populations led to the disappearance of all genetically pure wild individuals (Ellstrand 2014). Crop-to-wild gene flow in apple has been shown to be responsible for a decrease in fitness in populations of the European wild apple in Germany (Bleeker et al. 2007). The European wild apple may also be vulnerable to changes in the diversity of insect visitors to its flowers, resulting in less efficient pollination services, for instance. Given the emblematic nature of the wild apple in Europe and its usefulness as a resource for future breeding programs (Cornille et al. 2014), conservation measures should be urgently considered.

The implementation of conservation programs is challenging, in terms of the application of the recommended practices in the field. Landscape genetics can provide useful information for conservation management, in the form of fundamental knowledge for the identification of high-priority populations for the conservation of genetic variation. Our landscape genetic studies of within-species gene flow and crop-to-wild gene flow in apples in Europe should facilitate the allocation of priority zones for the conservation of the European wild apple, across its natural distribution. We detected five main genetic groups, within which it would be optimal to conserve populations. Southern Europe, housing the main glacial refugia with the highest levels of genetic diversity, should also be considered a priority area for conservation programs, together with France and the Carpathian Mountains, the main suture zones, with their original allelic combinations. The identification of highly introgressed European wild apple populations and of the anthropogenic drivers of crop-to-wild introgressions will also help to dynamically manage these priority conservation zones. This *in situ* conservation could also be combined with the *ex situ* conservation of genetic resources through the creation of gene banks for a core collection of nonintrogressed *M. sylvestris* populations from diverse European regions.

Our landscape genetic investigations would also help to identify the optimal provenance for seeds or germplasm for the reintroduction of wild apples into agroforestry programs (Krauss et al. 2013). Wild apple trees are, indeed, currently being planted in France in the context of agroforestry programs, but the provenance and the degree of introgression of the wild apple germplasm used have never been considered.

In conclusion, the relatively recent introduction of the cultivated apple into Europe, only about 1500 years ago, has particularly facilitated the detection of wild-to-crop and crop-to-wild gene flow and the identification of its determinants. The cultivated-wild apple complex in Europe thus appears to be a good model in which to address both

fundamental and applied issues relating to the impact of human landscape changes on gene flow.

### Statement of authorship

AC, PG, and TG conceived and designed the experiments; AC, PG, and TG obtained funding; AC and AF performed the experiments; AC and TG supervised the Master's thesis of AF; AC, AF, and UG analyzed the data. KV and JR provided samples. The manuscript was written by AC with critical input from UG and TG.

### Acknowledgements

We thank the *Région Ile de France* (PICRI and DIM R2DS), IDEEV (*Institut Diversité Ecologie et Evolution du Vivant*), SBF (*Société Botanique de France*), and laboratoire d'Excellence BASC (*Biodiversité, Agrosystèmes, Société et Climat*) for funding, and the *Plateforme de Génotypage GENTYANE* INRA UMR 1095 for assistance with genotyping. We thank all those who helped with sampling (Table S1). We thank Sébastien Rioux Paquette and Guillaume Larocq for fruitful discussions, Colin Fontaine for sharing SPIPOLL data on apples and anonymous referees for very useful suggestions. The authors declare no conflict of interest.

### Data accessibility

Data for this study (genetic, visitor flower diversity and relative to apple production) are available at the Dryad Digital Repository: <http://doi.org/10.5061/dryad.bm46p>.

### Literature cited

- Aerts, R., G. Berecha, P. Gijbels, K. Hundera, S. Van Glabeke, K. Vandepitte, B. Muys et al. 2013. Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications* **6**:243–252.
- Arnaud, J. F., F. Viard, M. Delescluse, and J. Cuguen 2003. Evidence for gene flow via seed dispersal from crop to wild relatives in *Beta vulgaris* (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**:1565–1571.
- Austerlitz, F., S. Mariette, N. Machon, P. H. Gouyon, and B. Godelle 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* **154**:1309–1321.
- Bleeker, W., U. Schmitz, and M. Ristow 2007. Interspecific hybridisation between alien and native plant species in Germany and its consequences for native biodiversity. *Biological Conservation* **137**:248–253.
- Chen, C., E. Durand, F. Forbes, and O. François 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**:747–756.
- Coart, E., X. Vekemans, M. J. M. Smulders, I. Wagner, J. Van Huylbroeck, E. Van Bockstaele, and I. Roldán-Ruiz 2003. Genetic variation in the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. *Molecular Ecology* **12**:845–857.
- Coart, E., S. Van Glabeke, M. De Loose, A. S. Larsen, and I. Roldán-Ruiz 2006. Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology* **15**:2171–2182.
- Conord, C., J. Gurevitch, and B. Fady 2012. Large-scale longitudinal gradients of genetic diversity: a meta-analysis across six phyla in the Mediterranean basin. *Ecology and Evolution* **2**:2600–2614.
- Cornille, A., P. Gladieux, M. J. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersisyan et al. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* **8**:e1002703.
- Cornille, A., T. Giraud, C. Bellard, A. Tellier, B. Le Cam, M. J. M. Smulders, J. Kleinschmit et al. 2013a. Post-glacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**:2249–2263.
- Cornille, A., P. Gladieux, and T. Giraud 2013b. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications* **6**:737–748.
- Cornille, A., T. Giraud, M. J. M. Smulders, I. Roldán-Ruiz, and P. Gladieux 2014. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**:57–65.
- Deguines, N., R. Julliard, M. de Flores, and C. Fontaine 2012. The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. *PLoS One* **7**:e45822.
- Delplancke, M., N. Alvarez, A. Espíndola, H. Joly, L. Benoit, E. Brouck, and N. Arrigo 2011. Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications* **5**:317–329.
- Ellstrand, N. C. 2003. Current knowledge of gene flow in plants: implications for transgene flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* **358**:1163–1170.
- Ellstrand, N. C. 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* **101**:737–753.
- Ellstrand, N. C., H. C. Prentice, and J. F. Hancock 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* **30**:539–563.
- Evanno, G., S. Regnaut, and J. Goudet 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**:2611–2620.
- Feder, J. L., S. P. Egan, and P. Nosil 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**:342–350.
- Garant, D., S. E. Forde, and A. P. Hendry 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology* **21**:434–443.
- Gross, B., A. Henk, P. Forsline, C. Richards, and G. Volk 2012. Identification of interspecific hybrids among domesticated apple and its wild relatives. *Tree Genetics and Genomes* **8**:1223–1235.
- Hardy, O. J., and X. Vekemans 2002. SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**:618–620.
- Hardy, O. J., L. Maggia, E. Bandou, P. Breyne, H. Caron, M.-H. Chevallier, A. Doligez et al. 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology* **15**:559–571.

- Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**:247–276.
- Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**:87–112.
- Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* **359**:183–195.
- Hoban, S. M., T. S. McCleary, S. E. Schlarbaum, S. L. Anagnostakis, and J. Romero-Severson 2012. Human-impacted landscapes facilitate hybridization between a native and an introduced tree. *Evolutionary Applications* **5**:720–731.
- Jakobsson, M., and N. A. Rosenberg 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**:1801–1806.
- 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**:625–632.
- Kamm, U., P. Rotach, F. Gugerli, M. Siroky, P. Edwards, and R. Holderegger 2009. Frequent long-distance gene flow in a rare temperate forest tree (*Sorbus domestica*) at the landscape scale. *Heredity* **103**:476–482.
- Krauss, S. L., E. A. Sinclair, J. D. Bussell, and R. J. Hobbs 2013. An ecological genetic delineation of local seed-source provenance for ecological restoration. *Ecology and Evolution* **3**:2138–2149.
- Kremen, C., N. M. Williams, M. A. Aizen, B. Gemmill-Herren, G. LeBuhn, R. Minckley, L. Packer et al. 2007. Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters* **10**:299–314.
- Larsen, A., C. Asmussen, E. Coart, D. Olrik, and E. Kjær 2006. Hybridization and genetic variation in Danish populations of European crab apple (*Malus sylvestris*). *Tree Genetics and Genomes* **2**:86–97.
- Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**:183–189.
- Loiselle, B. A., V. L. Sork, J. Nason, and C. Graham 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**:1420–1425.
- Manel, S., and R. Holderegger 2013. Ten years of landscape genetics. *Trends in Ecology & Evolution* **28**:614–621.
- Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**:189–197.
- Manning, A. D., J. Fischer, and D. B. Lindenmayer 2006. Scattered trees are keystone structures – Implications for conservation. *Biological Conservation* **132**:311–321.
- Patocchi, A., F. Fernández-Fernández, K. Evans, D. Gobbin, F. Rezzonico, A. Boudichevskaia, F. Dunemann, et al. 2009. Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. *Tree Genetics & Genomes* **5**:211–223.
- Petit, R. J., I. Aguinalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**:1563–1565.
- Pritchard, J. K., M. Stephens, and P. Donnelly 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- Raymond, M., and F. Rousset 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**:248–249.
- Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**:103–106.
- Schnitzler, A., C. Arnold, A. Cornille, O. Bachmann, and C. Schnitzler 2014. Wild European apple (*Malus sylvestris* (L.) Mill.) population dynamics: insight from genetics and ecology in the rhine valley. Priorities for a future conservation programme. *PLoS One* **9**:e96596.
- Smouse, P. E., and V. L. Sork 2004. Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management* **197**:21–38.
- Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits 2010. Landscape genetics: where are we now? *Molecular Ecology* **19**:3496–3514.
- Szpiech, Z. A., M. Jakobsson, and N. A. Rosenberg 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**:2498–2504.
- Tzedakis, P. C., B. C. Emerson, and G. M. Hewitt 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution* **28**:696–704.
- Vekemans, X., and O. J. Hardy 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**:921–935.

[Correction added on 25 November, 2016 after initial online publication: Supporting information was added to the article.]

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Geographical locations of the 66 sites across Europe ( $N = 1889$  *Malus sylvestris* trees) used for genetic analyses (left).

**Figure S2.** Bayesian clustering results obtained with STRUCTURE, for  $K = 2$  to  $K = 9$ , based on the dataset including *Malus domestica* ( $N = 339$ ) and *Malus sylvestris* ( $N = 1889$ ).

**Figure S3.** Bayesian clustering results for *Malus sylvestris* in Europe ( $N = 1376$ ), obtained with TESS, for  $K_{max} = 2$  to  $K_{max} = 9$ .

**Figure S4.** Estimated number of populations ( $K$ ) in *Malus sylvestris*, from TESS analyses based on the DIC.

**Table S1.** Description of the *Malus* accessions analyzed, with their geographical origins (locality, latitude and longitude) and providers.

**Table S2.** Mean rates of introgression into *Malus sylvestris* from *Malus domestica* and number of pure *Malus sylvestris* trees and hybrids in the European sample (66 sites).

**Table S3.** Mean wild-to-crop introgression rates by country.

**Table S4.** Genetic variation at each site, for *Malus sylvestris* ( $N = 1376$  individuals, 62 sites).

**Table S5.** Taxa diversity of insects visiting flowers of the apples in France.