Ecological and evolutionary drivers of phenotypic and genetic variation in the European crabapple (*Malus sylvestris* (L.) Mill.), a wild relative of the cultivated apple

Chen X.1, Avia K.2, Forler A.1, Remoué C.1, Venon A.1, Rousselet A., Lucas G.3, Kwarteng A.O.1, Rover R.1, Le Guilloux M.1, Belcram H.1, Combes V.1, Corti H.1, S. Vazquez1, Falque M.1, G. Alins4, T. Kirisits5, Ursu T.M.6, Roman, A.6, Volk. G.M.7, Stéphane Bazot8, Cornille A.1

**Affiliations**

1. Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France.
2. Université de Strasbourg, INRAE, SVQV UMR-A 1131, F-68000 Colmar, France
3. Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette 91198, France
4. Institut de Recerca i Tecnologia Agroalimentàries, IRTA-Fruit Production, PCiTAL, Parc 21 de Gardeny, edifici Fruitcentre, 25003 Lleida, Spain
5. Institute of Forest Entomology, Forest Pathology and Forest Protection (IFFF), Department of Forest and Soil Sciences, University of Natural Resources and Life Sciences, Vienna (BOKU), Peter-Jordan-Straße 82 (Franz Schwackhöfer-Haus), A-1190 Vienna, Austria
6. NIRDBS, Institute of Biological Research Cluj-Napoca, 48 Republicii St., Cluj-Napoca, Romania
7. USDA-ARS National Laboratory for Genetic Resources Preservation, 1111 S. Mason St., Fort Collins, Colorado, U.S.A.
8. Ecologie Systématique et Evolution, CNRS, AgroParisTech, Ecologie Systématique Evolution, Université Paris-Saclay, Orsay, France

**corresponding author:** amandine.cornille@cnrs.fr
Abstract (250 words max with bullet points)

- Characterizing the phenotypic and genetic variation among populations of crop wild relatives help understanding the ecological and evolutionary processes involved in population divergence, and better harness their diversity to mitigate the impact of climate change on crops. We assessed genetic and phenotypic diversity of the European crabapple, *Malus sylvestris*, a main contributor to the cultivated apple genome (*Malus domestica*), and investigated for ecological divergence.

- We assessed variation in growth rate and traits related to carbon uptake between seedlings measured in a common garden, and related it to the genetic ancestry of the seedlings, assessed using 13 microsatellite loci and Bayesian clustering method. The occurrence of patterns of isolation-by-distance, -by-climate and -by-adaptation that might have caused genetic and phenotypic differentiation among *M. sylvestris* populations was also tested.

- Seedlings belonged to seven *M. sylvestris* populations in Europe, with 11.6% of seedlings introgressed by *M. domestica*. Significant trait variation among *M. sylvestris* populations was observed, which for some was of moderate to high heritability. Lack of association between trait and genetic divergence suggests that this significant phenotypic variation is not adaptive, but strong association between genetic variation and the climate during the last glacial maximum suggests local adaptation of *M. sylvestris* to past climates.

- This study provides an insight into the ecological and evolutionary drivers of phenotypic and genetic differentiation among populations of a wild apple species and relative of cultivated apples, which is a starting point for future breeding programs.

**Keywords:** population structure, isolation-by-distance, isolation-by-ecology, local adaptation, climate change, apple tree, crop wild relatives.

Societal impact Statement (113 words needs to be reduced to 100)

Apple is a major fruit crop worldwide and a model species for understanding the evolutionary processes underlying perennial crop domestication. Several wild species have contributed to the genetic make-up of the cultivated apple, yet phenotypic and genetic diversity data across their natural distribution is lacking. This study revealed phenotypic variation between populations of the European crabapple, and showed that both geography, and surprisingly, past but not current climate, shaped its genetic structure. We provide a starting point for harnessing wild apple diversity for apple breeding programs to mitigate the impact of climate change on this perennial crop.
Introduction

Knowledge of the spatial phenotypic and genetic variation among populations is essential for understanding the ecological (biotic and abiotic) and evolutionary (gene flow, selection, drift, mutation) processes involved in population divergence and adaptation (Savolainen et al., 2013; Sork, 2018). The global biodiversity crisis, and its consequences on ecosystem health and services, makes investigating these questions, and identifying the taxa most vulnerable to anthropogenic change, all the more relevant (Hoffmann et al., 2021).

Plant species distributed across climatic gradients typically experience spatial variation in selection, genetic drift and gene flow, processes that drive genetic and phenotypic divergence among populations (Svenning et al., 2015). Climate influence demography such as population expansion and contraction, the extent of gene flow among populations, and ultimately the extent of genetic divergence among populations (Edwards et al., 2022). For instance, changes in the climate since the last glacial maximum (LGM) 20,000 years ago have driven the genetic composition of the European crabapple and many other tree species (Comes & Kadereit, 1998; Kremer et al., 2002; Pyhäjärvi et al., 2008; Cornille et al., 2013a; Gugger et al., 2013; Riordan et al., 2016; Lander et al., 2021; Yamada et al., 2021; Parisod, 2021). Climate can also shape phenotypic variation among populations. Populations occurring under the same climate may share physiological tolerances to climatic conditions, including plant carbon uptake via photosynthesis. Carbon uptake traits condition plant size and growth, reproduction and survival under different climatic conditions (Nicotra et al., 2010; Hartmann et al., 2020). In some cases, local climate can impose divergent selection on carbon uptake traits and lead to long-term reduction in gene flow among populations and local adaptation (Keller et al., 2011; Franks et al., 2014; Aitken & Bemmels, 2015; Ramírez-Valiente et al., 2017; Alexandre et al., 2020). Whether the phenotypic variation observed in species distributed across large climatic ranges results from their demographic or adaptive history remains an intense topic of investigation (Li et al., 2012; Tiffin & Ross-Ibarra, 2014). Furthermore, investigating this question can help predict how plants may respond to climate change and how species adapt to their environment.

There are multiple ways to investigate whether the genetic and phenotypic variation among populations distributed across climatic gradients results from selection, genetic drift and/or gene flow. A first step could be to use a common garden experiment to investigate the genetic basis of phenotypic variation among populations. Indeed, different populations occurring across a climatic gradient may display clinal variation, i.e., differences in a trait that may be the result of plasticity or local genetic adaptation (Savolainen et al., 2013; de Villemereuil et al., 2016). Measuring candidate traits for adaptation to climate, e.g., phenology (Brachi et al., 2013) or traits related to plant carbon uptake (Savolainen et al., 2013; de Villemereuil et al., 2016) in individuals from different populations under the same environmental conditions can help elucidate the genetic basis of phenotypic variation across populations without the confounding effects of the environment. Ideally, common garden should include the main genetic groups across the species’ distribution (de Villemereuil et al., 2020).
association of neutral genotypic variation with ecological variation can also be used as evidence of adaptive divergence among populations (Shafer & Wolf, 2013; Wang & Bradburd, 2014). The correlation between neutral genetic differentiation and environmental or phenotypic divergence among populations, independent of geographic distance, referred to as isolation-by-ecology (IBE), is an extension of the isolation-by-distance (IBD hereafter) model (Wright, 1943), and has increasingly been used as an indicator of adaptive divergence between populations. In the IBE model, natural selection, which results from several factors including climate, can indirectly increase neutral genetic and phenotypic differentiation between populations by promoting general barriers to gene flow (Nosil et al., 2009; Orsini et al., 2013, p. 201; Shafer & Wolf, 2013; Wang & Bradburd, 2014). The IBE pattern is agnostic with respect to the underlying processes that generated it (Wang & Bradburd, 2014); this pattern can be generated by different processes including natural selection against immigrants, sexual selection against immigrants, reduced hybrid fitness and biased dispersal. Although it can be difficult to map one or more processes to this pattern, testing for the IBE pattern is valuable for better understanding the ways in which natural selection shapes neutral genetic and phenotypic variation. Evidence from common garden experiments and IBE patterns can therefore contribute to understanding how genotypes, phenotypes and the environment interact to ultimately influence population divergence and potentially local adaptation.

Fruit trees are a major component of terrestrial ecosystems (Petit & Hampe, 2011) and are grown in managed plantations and orchards to provide a variety of economically important products (Boyd et al., 2013). Recent breeding efforts have involved the repeated use of a limited number of cultivars sources of genetic material leading to a reduction in genetic diversity and the loss of valuable alleles at genes that are not directly targeted by human selection (Myles et al., 2011; Warschefsky & von Wettberg, 2019; Migicovsky et al., 2021). Wild relatives of crop fruit trees (or “CWR” for crop wild relative) harbor phenotypic and genetic diversity that are potentially highly valuable for future breeding programs in the context of climate change (Zhang et al., 2017; Hoban et al., 2018; Hübner & Kantar, 2021). However, rare of the studies which thoroughly investigate the phenotypic variation of CWR of fruit trees in relation to response to climate; most studies so far have focused on forest trees (Kremer & Hipp, 2019). Key traits to study in this context are related to plant carbon uptake. Indeed, climate impacts plant carbon uptake (Aubin et al., 2016), which latter known to impact fruit quality characteristics and production (Demestithas et al., 2017). These questions are urgent as native CWRs can be threatened by crop-to-wild gene flow from nearby domesticated trees (Delplancke et al., 2011; Cornille et al., 2015; Diez et al., 2015; Feurtey et al., 2017; Flowers et al., 2019; Liu et al., 2019). Therefore, the study of the genetic and phenotypic variation among populations of CWR fruit tree species is timely to guide future breeding programs; in addition, it may contribute to our understanding of the evolutionary and ecological drivers of population divergence, including climate.

The European crabapple, Malus sylvestris (L.) Mill, is a CWR and a major contributor to the cultivated apple genome (Cornille et al., 2012, 2014, 2019; Peace et al., 2019). Substantial crop-to-wild
gene flow has been observed across *M. sylvestris* populations in Europe (up to 23.1% of natural populations are introgressed by *M. domestica* (Cornille et al., 2015)). Crop-wild hybrids sampled in a forest in France and grown in controlled conditions showed superior fitness compared to wild seedlings (Feurtey et al., 2017). Population genetics analyses also identified five pure (i.e., not introgressed by *M. domestica*) populations in Scandinavia, western France, eastern France, Eastern Europe and Italy (Cornille et al., 2015). These five populations resulted from past contractions and expansions associated with the LGM (Cornille et al., 2013a, 2015). It remains unclear, however, whether these five populations, distributed across a large area with different climatic conditions, present phenotypic variation that could be the result of local adaptation to past and/or present climates.

We investigated the spatial phenotypic and genetic variation among populations of a major wild contributor to the cultivated apple, *M. sylvestris*, to test for adaptive divergence. Variation in plant growth and traits related to carbon uptake was measured in 584 *M. sylvestris* seedlings grown under controlled conditions and genotyped for 13 microsatellite markers. We first assessed the genetic status of each seedling (pure vs. crop-wild hybrid). Then, we compared growth traits and traits related to plant carbon uptake among seedlings belonging to different European genetic groups. We also formally tested the impact of geography (IBD) and ecology (IBE tested with phenotypic traits and climate) on genetic variation observed from 13 microsatellite markers. We investigated the following questions: 1) Does growth rate and carbon uptake trait vary between populations of the European crabapple? Are those traits heritable, and thus can population history predict seedling phenotype? 2) Is there any association between phenotypic variation and genetic variation, taking into account geographic distance? and 3) Is climate associated with neutral genetic diversity of the European crabapple, which could suggest local adaptation (i.e., ecological/adaptive divergence)?

Materials and Methods

Plant material, experimental design and trait measurements

A total of 584 seeds were collected from 90 *M. sylvestris* mother trees (three to 15 seeds per mother tree, Table S1) from six different geographical regions in Europe: Austria (*N* = 89, two sites), Denmark (*N* = 91, three sites), Spain (*N* = 39, one site), France (*N* = 220, eight sites), Italy (*N* = 32, one site), Romania (*N* = 117, seven sites) (Table S1).

In mid-April 2019, the 584 seeds were washed, sterilized (in 0.5% chlorine for 20 min), and vernalized for three months at 4°C in the dark in a mix of damp sand and vermiculite. Then, seeds were sowed in jiffy pellets and each pellet was randomly placed in a 20-hole array. Seeds were grown in controlled conditions for two months (from mid-July to mid-September 2019: 22 ±1°C, 60 ±5 % relative humidity, a 16:8 (L:D) photoperiod and a light level of 40–60 µmol m⁻² s⁻¹). Each 20-hole array was rotated daily in the growth chamber to avoid any micro-environmental variation in plant response, and plants were watered weekly.
During the course of the two-month experiment, the number of leaves and the height of each seedling were recorded. Some accessions, due to the low germination rate, could not be recorded (i.e., height and number of leaves could not be recorded for 19 seedlings out of 584, \( N = 565 \), Table 1). Seedlings were measured every two or three days, starting from day 7-11 after the start of the experiment.

The last week of the experiment, the superficial flavonol and chlorophyll content and the nitrogen balance index (NBI) were measured in three leaves per seedling. Superficial chlorophyll content is the concentration of chlorophyll in the leaf epidermis (µg/cm²), and superficial flavonol content is an index of the flavonoid concentration (µg/cm²) in this upper layer and is related to phenol accumulation and UV protection. Leaf chlorophyll and flavonol content and NBI are parameters correlated with plant carbon uptake via photosynthesis. Flavonol is a phenolic compound that is also known to contribute to plant resistance, acclimation and adaptation to environmental constraints through various mechanisms, including its antioxidant activity. These traits were measured using a portable Dualex device (Force-A, Orsay, France), which uses a combination of fluorescence signals at various excitation bands to quantify pigments and chemical compounds. As carbon uptake related traits must be measured in the same day, a subsample of 257 seedlings out of the 565 seedlings (Table 1, numbers in brackets) was measured because of time limitation in a day. Seedlings measured for carbon uptake related traits were selected based on two criteria: having at least one seedling per mother tree and three seedlings per geographic site.

DNA extraction, microsatellite genotyping and genetic ancestry of the seedlings

At the end of the experiment, leaves of each seedling were sampled for microsatellite genotyping. Genomic DNA was extracted with the NucleoSpin plant DNA extraction kit II (Macherey & Nagel, Düren, Germany) according to the manufacturer’s instructions. Microsatellites were amplified by multiplex PCR with the Multiplex PCR Kit (QIAGEN, Inc.). We used 13 microsatellite markers, Ch01f02, Ch01f03, Ch01h01, Ch01h10, Ch02c06, Ch02c09, Ch02c11, Ch02d08, Ch03d07, Ch04c07, Ch05f06, GD12 and Hi02c07 in four multiplexes (MP01 to MP04; (Cornille et al., 2012)).

PCR was performed in a final reaction volume of 15 ml (7.5 ml of QIAGEN Multiplex Master Mix, 10–20 mM of each primer with the forward primer labelled with a fluorescent dye, and 10 ng of template DNA). We used a touch-down PCR program (initial annealing temperature of 60°C, decreasing by 1°C per cycle down to 55°C). Genotyping was performed at the GENTYANE platform (INRAE Clermont-Ferrand) on an ABI PRISM X3730XL, with 2 ml of GS500LIZ size standard (Applied Biosystems). Alleles were scored with the GENEMAPPER 4.0 software (Applied Biosystems). We retained only multilocus genotypes presenting less than 10% missing data.

Clones or closely related individuals can bias inferences of population structure. We estimated the kinship coefficient between pairs of individuals (\( F_{ij} \)) with SPAGeDI 1.5d (Loiselle et al., 1995; Hardy & Vekemans, 2002), and removed highly genetically related individuals with \( F_{ij} > 0.5 \).
The individual-based Bayesian clustering method implemented in STRUCTURE 2.3.3 (Pritchard et al., 2000) was used to estimate the admixture between M. domestica and M. sylvestris, and the population genetic structure of M. sylvestris. STRUCTURE uses Markov Chain Monte Carlo (MCMC) simulations to infer the proportion of ancestry of genotypes from K distinct clusters. The underlying algorithm attempts to minimize deviations from Hardy–Weinberg and linkage disequilibria. K ranged from 1 to 10. Ten independent runs were carried out for each K and 500,000 MCMC iterations after a burn-in of 50,000 steps were used. CLUMPAK (Greedy algorithm) (Kopelman et al., 2015) was used to identify distinct modes in the 10 replicated runs for each K. STRUCTURE analyses were run for the full dataset (N = 584), plus 40 M. domestica genotypes included as a reference for the cultivated apple gene pool (Cornille et al., 2013b). The R package pophelper v2.3.0 was used (Francis, 2016) to visualize bar plots. The amount of additional information explained by increasing K was determined using the ΔK statistic (Evanno et al., 2005), as implemented in Structure Harvester (Earl & vonHoldt, 2012). However, ΔK provides statistical support for the strongest but not the finest population structure (Puechmaille, 2016). Natural populations can display a hierarchical genetic structure with fine-scale population structure. Visual inspection of the barplots was used to identify the K value for which all clusters have well assigned individuals, and where additional clusters at higher K values do not have well assigned individuals (indicating that we have reached the highest K value for which no new genuine clusters could be delimited). The K value we therefore considered corresponded to the finest one, which can be higher than the K value of the strongest population structure identified by ΔK.

Using the best K value inferred with STRUCTURE, we defined \( P_{\text{dom}} \), the membership proportion of a seedling to the M. domestica gene pool; membership coefficients were used to define the genetic ancestry of each seedling: 1) seedlings with \( P_{\text{dom}} > 0.9 \), whose mother tree was likely misidentified in the field (referred to as "dom" hereafter); 2) seedlings with \( P_{\text{dom}} > 0.1 \), i.e., crop-wild hybrids (referred to as "cw" hereafter); then 3) seedlings with a membership coefficient > 0.9 to a given wild apple cluster were considered to be «pure» wild seedlings (referred to as "pure" hereafter); and 4) seedlings with a membership coefficient < 0.9 to a given wild apple gene pool were considered to be wild-wild hybrids ("ww", hereafter). In addition, “pure” seedlings were assigned to different populations (i.e., groups of seedlings with membership coefficient > 0.9 to a given wild gene pool). Two effects were then tested using statistical models below: the genetic status effect (i.e., dom, cw, ww, pure), and the wild apple population effect (i.e., corresponding to the “pure” populations detected with STRUCTURE).

Fitness proxy estimates

The fitness of each seedling was therefore estimated from growth and carbon related trait proxies.

The absolute growth rate (\( AGR \) (Radford, 1967)), relative growth rate (\( RGR \) (Briggs et al., 1920)), and whole \( AGR \) were estimated as follows (the traits considered were the height and the number of leaves of the seedling):
\[ AGR(cm/day) = \frac{\text{trait}_{i+1} - \text{trait}_i}{\text{date}_{i+1} - \text{date}_i} \] (1)

\[ RGR(cm/day/day) = \frac{AGR}{\text{date}_i} \] (2)

\[ \text{WholeAGR}(cm/day) = \frac{\text{trait}_{end} - \text{trait}_{beginning}}{\text{date}_{end} - \text{date}_{beginning}} \] (3)

Note that for the whole AGR, the beginning of the experiment corresponded to days 7 and 11 for leaf and height measurements, respectively, while the last measurement was done at day 60. The internode ratio, which represents the ratio between the number of leaves and the height of the seedling at day 60, was also considered a fitness trait, as this value plays an important role in apple tree architecture (Ripetti et al., 2008).

Seven fitness proxies were therefore calculated for the full dataset (565 seedlings, Table 1): height_AGR, height_RGR, whole_height_AGR, leaf_AGR, leaf_RGR, whole_leaf_AGR and internode.

In addition, chlorophyll (Chl) and flavonol (Flav) content, and NBI, were measured in the subsample (257 seedlings, Table 1). A preliminary exploration of correlation and variation among phenotypic traits was carried out using a principal component analysis (PCA) with the FactoMineR R package (Lê et al., 2008).

Statistical analyses of fitness variation

A previous study demonstrated that crop-to-wild gene flow has an impact on early-stage growth rate (Feurtey et al., 2017). The effect of the genetic status of seedlings (i.e., \textit{dom}, \textit{cw}, \textit{ww}, \textit{pure}, Table 1) on fitness variation among seedlings was therefore tested. A linear mixed model was fitted to the data as follows:

\[ \text{Fitness}_{ijkl} \sim \mu + \text{wild population of origin}_i + \text{genetic ancestry status}_j + \text{wild population of origin}_i + \text{genetic ancestry status}_j + \text{mother}_k + e_{ijkl} \] (4),

where \( \mu \) is the overall mean, “wild population of origin” is the fixed effect of the population of origin of the seedling inferred with STRUCTURE, “status” is the fixed effect of the genetic status of the seedling (i.e., “\textit{pure}”, “\textit{dom}”, “\textit{ww}”, “\textit{cw}”), the interaction between the two fixed effects, and “mother” is a normally distributed random effect with its own mean and variance parameters, and \( e \) is the residual. The mother tree of each seedling was used as a random factor to avoid pseudo-replication due to the presence of multiple half-siblings (i.e., from the same mother tree). We ran the model (4), but replaced the “status” effect by the “\textit{Pdom}” fixed effect. We gradually removed interactions and effects depending on their significance. In addition, we evaluated the differences in the effect on trait variation using a contrast analysis. We fitted the data to the model using the \textit{lm4} R package (Bates et al., 2015).
For fitness proxies defined from the number of leaves (leaf\_AGR, leaf\_RGR and whole\_leaf\_AGR), a log link function was used and the residual distribution was fitted to a negative binomial distribution (function glm.nb in R package lme4). For fitness proxies defined from the height of the seedling (height\_AGR, height\_RGR, whole\_height\_AGR), and for chlorophyll and flavonol content, and NBI, a similar linear mixed model was run, but with a residual term that was assumed to be normally distributed.

Heritability estimates

Heritability estimates were computed using only pure and wild-wild M. sylvestris seedlings detected as above. We fitted each fitness proxy with a linear mixed model as follows: \[ Y_{ijk} = \mu + F_i + C_j + e_{ijk} \] (5), where \( Y_{ijk} \) is the fitness proxy (growth rate or carbon uptake related trait) of the \( k \)th seedling belonging to family \( i \), member of the \( j \)th genetic cluster, \( \mu \) the overall fixed mean of the population, \( F_i \) the random effect of the \( i \)th family, \( C_j \) the fixed effect of the \( j \)th genetic cluster and \( e_{ijk} \) the random error term. The model was fitted using REML (restricted maximum likelihood). Calculations were performed by the lme-function of the R-library nlme (Pinheiro et al., 2022). The output of lme provides estimates for the variance components, the corresponding standard deviations (sd), and the best unbiased linear predictors (BLUP) for random effects. Genetic parameters were then calculated as follows:

- the additive genetic variance: \( VA = 4\sigma_F^2 \) with \( \sigma_F^2 \) representing the between-family variance
- the corresponding coefficient of variation: \( CVA = \frac{\sqrt{VA}}{\mu} \)
- the phenotypic variation: \( VP = \sigma_Y^2 + \sigma_E^2 \), with \( \sigma_E^2 \) representing the residual variance
- the corresponding coefficient of variation: \( CVP = \frac{\sqrt{VP}}{\mu} \)
- Narrow-sense heritability: \( h^2 = \frac{VA}{VP} \)
- Dickerson’s approximation for its standard deviation: \( sd(h^2) \approx \frac{4sd(\sigma_F^2)}{VP} \)

Test for isolation-by-ecology

Only pure and wild-wild hybrid M. sylvestris seedlings were selected for IBE analysis \( (N = 449, 21 \) sites, Table 1). The IBE pattern, i.e., the contribution of climate and phenotypic distances to the genetic structure taking into account geographical distance, was evaluated using a distance-based redundancy analysis (db-RDA). db-RDA can be used when the response variable is a distance matrix, here a genetic distance matrix \( (F_{ST}) \) across 21 sampled sites, and the explanatory variables are in vector form. Explanatory variables were as follows: (i) the geographical distance between sampled sites which underlies an IBD process, represented by vectors with positive eigenvalues of a principal coordinate of neighbor matrix \( (PCNM) \) (Borcard & Legendre, 2002), which was applied to the geographical pairwise distance matrix between sampled sites computed with SPAGeDI 1.5d (Hardy & Vekemans, 2002); 19 bioclimatic variables downloaded from the Worldclim2 database (30s resolution,
https://www.worldclim.org/data/worldclim21.html) representing annual and seasonal trends and extremes averaged (ii) over the years 1970-2000 and averaged (iii) for the Pleistocene period (20K years ago) (Gamisch, 2019), which were used to test for an isolation-by-climate pattern (IBC, hereafter); and (iv) growth rate ($N = 551$, Table 1) and carbon uptake related traits ($N = 239$, Table 1) averaged per site, as well as chlorophyll and flavonoid content, which were used to test for an isolation-by-adaptation pattern, referred as to IBA hereafter.

To identify the variables that explained the genetic structure of $M. sylvestris$, a db-RDA using the “capscale” function (Oksanen et al., 2014) was run on a full model which included all investigated variables (i.e., PCNM components, growth rates, carbon uptake related traits, 19 bioclimatic variables). The best variables were selected for an optimum model with the function “step” based on the Akaike Information Criterion (AIC). Because db-RDA does not provide information on the relative contribution of each variable of the model, a variance partitioning analysis was run using the “varpart” function from the R-package “vegan” (Peres-Neto et al., 2006).

**Results**

**Genetic ancestry of seedlings**

No clones or closely related individuals were detected (Figure S1) and therefore all 584 seedlings were included in the STRUCTURE analyses.

STRUCTURE revealed a clear spatial population genetic structure of $M. sylvestris$ in Europe as well as crop-wild admixture (Figures 1 and S2). For $K = 2$, the analysis recovered a group that included $M. domestica$ and the western $M. sylvestris$ samples (green) and another group that consisted of the Eastern European samples (red). For $K = 3$, the western group was split into two groups, one comprising $M. domestica$ and Spanish and Italian $M. sylvestris$ seedlings (black), and another comprising the remaining western samples (green); the Central European group was also recovered. For $K = 4$, the Eastern European samples were split into two groups: an Austrian group and a Romanian group. For $K = 5$, there was a clear east/west substructure in France. For $K = 6$, a sixth cluster comprising the Danish individuals was found. For $K = 7$, the Italian population split into two groups. For $K = 8$, the population from one site in eastern France (Lor) was identified as a new cluster. When $K > 9$, STRUCTURE did not reveal any further substructures, with only additional cluster with highly admixed individuals (Figure S2). Therefore, although the $\Delta K$ indicated that the most likely $K$ value was five (Figure S3), $K = 8$ was the finest population structure and was therefore retained in subsequent analyses.

For $K = 8$, we found that the $M. domestica$ reference varieties were admixed with the Italian and Western European $M. sylvestris$. Conversely, we detected 68 $M. sylvestris$ seedlings with $P_{dom} > 0.1$ (considered as crop-wild hybrids), corresponding to 11.6% of the seedlings ($N = 584$, Figures 1, S4 and S5, Table 1). We also found 21 seedlings with a membership coefficient to the $M. domestica$ gene pool > 0.9, corresponding to 4% of the seedlings. Nearly all Spanish seedlings were assigned to the $M. domestica$ gene pool with membership coefficients > 0.1 (i.e., 26 crop-wild hybrids and 13 individuals
assigned to the *M. domestica* gene pool) and showed admixture only with the wild Italian purple gene pool (Figure S4). 33 individuals could not be assigned to any cluster (i.e., individuals with a membership coefficient < 0.5 to any cluster).

We therefore identified 68 “cw”, 21 “dom”, 167 “ww” and 282 “pure” seedlings (Table 1, *N* = 551). After removing *cw* hybrids (*N* = 68), seedlings sampled from misidentified mother trees (*N* = 21), the *M. domestica* reference samples (*N* = 40) and individuals with a membership coefficient < 0.5 to any cluster, seven wild apple populations (i.e., groups of seedlings with a membership coefficient > 0.5 to a wild apple cluster) were defined: French Western (“FR-W”, *N* = 77), French Eastern (“FR-E”, *N* = 50), French Lorraine (“FR-Lor”, *N* = 28), Danish (DA, *N* = 78), Italian (“IT”, *N* = 27), Austrian (“AUT”, *N* = 81) and Romanian (“RO”, *N* = 108) (Figure S6). Each *M. sylvestris* population showed a high level of genetic variation (Table S2). The Romanian population was the most genetically differentiated population and was close to the Austrian population; the Danish and French Western populations were genetically similar (Figure S7).

**No effect of seedling status on phenotypic variation**

Variation and correlations among traits are presented in Figures S8 to S11. Heritability estimates were moderate to high for all traits except growth rate based on leaf number (Table S3). However, these estimates need to be taken with caution given the limited sample size, as reflected by the large standard deviations.

We did not find any significant effect of seedling status (i.e., *pure*, *ww*, *cw*, *dom*) (Figure 5, Table S4, Figure S12) or *P*<sub>dom</sub> (Table S5, Figure S13) on phenotypic traits (i.e., leaf and height AGR, leaf and height RGR, whole leaf and height AGR). We therefore removed the seedling status and *P*<sub>dom</sub> effects from the model 4, as well as *dom* and *cw* individuals. We therefore only considered wild apple seedlings (i.e., *pure* and *ww*, *N* = 449), and focused on the ‘wild population of origin’ effect (Table S6).

**Significant variation in growth rates and chlorophyll content among populations**

Mean height variation along the course of the experiment among seedlings from different populations is shown in Figure 2. There was significant variation among seedlings from different populations in certain growth-related traits (Table 2). On average, seedlings belonging to the Austrian population were taller (+11 cm, *P* = 0.047) whereas Romanian (-14.9 cm, *P* = 0.008) and Italian (-18.7 cm, *P* = 0.044) seedlings were shorter (Figure 2) than seedlings from other populations. Seedlings from other populations did not show a significant difference in height. In addition, the number of leaves and height traits were negatively correlated, *r* = -0.3, *P* < 0.001. The Austrian population presented the lowest number of leaves (average = 5, sd = 4) whereas seedlings belonging to the Romanian population had the highest number of leaves (average = 8, sd = 7, Figure S15). The Romanian population also had the largest internode (+ 0.02 leaf/cm, *P* = 0.024).
Chlorophyll content differed among populations with seedlings from the Italian population producing on average more chlorophyll (+4.14 μg/cm², \( P = 0.039 \), Figure S16) than seedlings from other populations. Flavonol content and NBI did not differ significantly among populations.

**Significant IBD and IBC**

Correlation plots between bioclimatic variables are provided in Figures S17 and S18; however, all variables were included in the analysis as db-RDA can cope with correlated variables. The optimal model was chosen according to its best AIC value. The optimal model explained up to 25.9% of the genetic structure (\( Adj-R^2 = 69.9\% \), \( P = 0.001 \)) and contained seven variables (four geographic and three bioclimatic variables) (Table 3): the geographical distance is represented by the 1st, 2nd, 3rd and 6th axis of the PCNM analysis and three past climatic variables (Bio3: isothermality; Bio6: minimum temperature of coldest month; Bio 9: mean temperature of driest quarter). In total, IBD explained 47% of the variance of the wild apple tree population genetic structure, whereas IBC explained 22% (Figure 3). Taking geographical distance into account, we did not find a pattern of IBA i.e., covariation between phenotype and genetic divergences.

**Discussion**

This study is the first to take into account the population genetic structure as well as the phenotypic variation of a contributor to the cultivated apple genome (Cornille et al., 2012), to investigate ecological divergence. Bayesian clustering revealed seven *M. sylvestris* populations across Europe with a substantial number of seedlings (11.6%, mainly from Western Europe) introgressed by *M. domestica*, although this figure is less substantial than previously reported (Cornille et al., 2013b, 2015; Feurtey et al., 2017). Although the crop-wild hybrid status of seedlings did not impact phenotypic variation, we observed phenotypic variation among crabapple populations when grown in controlled conditions. Phenotypic variation was found for growth and chlorophyll content among populations of the European crabapple from different climates in Europe. Based on the IBA pattern, this phenotypic variation was not adaptive. However, the IBC pattern revealed that climate was a driver of genetic differentiation between populations. Given that the IBC pattern was still found after accounting for IBD, this implies that there are sufficient levels of local adaptation to LGM climate to reduce gene flow among populations. The European crabapple may therefore be locally adapted to the past climate conditions of the LGM. The lack of signal of adaptive phenotypic divergence suggests that traits other than the ones we investigated in this study may be under divergent selection. The results of this study pinpoints adaptive divergence related to climate in a wild contributor to a fruit tree crop genome, which is a starting point for future breeding programs and mitigating the impact of climate change of CWR of an emblematic temperate fruit tree.

**Ongoing crop-to-wild gene flow in the European crabapple**
We revealed substantial gene flow from *M. domestica* to the European crabapple, with 11.8% of seedlings, mostly from Western Europe, introgressed by *M. domestica*. Introgression rates were lower compared to previous studies (*i.e.*, 37% in Cornille *et al.* (2013b) and 23.1% in Cornille *et al.* (2015)). However, these studies genotyped more mother trees (*i.e.*, *N* = 756 and *N* = 1,889, respectively), which could explain the difference in estimates of crop-to-wild gene flow. The results described here highlight the fact that crop-to-wild gene flow is still ongoing in the European crabapple. The Spanish seedlings sampled here were the progeny of trees growing in a location that is known to have high levels of *M. domestica* introgression (*pers. comment. G. Alins*). It is even possible that the mother trees of these seedlings were *M. domestica* and not *M. sylvestris*. The inclusion of reference cultivated apple samples mainly from Western Europe may decrease the probability of detecting crop-to-wild introgression events in wild populations from Eastern and Northern Europe. The lower crop-to-wild introgression rates in wild seedlings from Eastern Europe can also be explained by their physical distance from cultivated apple orchards. Indeed, distance can be a natural barrier to hybridization between *M. domestica* and its wild relative (Larsen *et al.*., 2016), which we confirmed in this study. The position of a *M. sylvestris* individual in a forest may also impact its level of introgression. Indeed, *M. sylvestris* trees are often found in forest gaps and at the forest edge, corresponding to their ecophysiological preferences (*i.e.*, preference for light and low competition). The effect of the location of the trees in the forest on the level of crop-to-wild introgression needs to be studied further.

The consequences of crop-wild introgression on phenotypic variation between crop and wild individuals have been studied more in annual crops (*e.g.*, maize, wheat, lettuce, rice) (Ellstrand *et al.*, 2013) than in perennial fruit trees. One study has shown that crop-wild hybrid seedlings of the European crabapple have higher growth rates and showed earlier germination than wild apple seedlings (Feurtey *et al.*, 2017). We did not detect any effect of the status of a seedling (pure, wild, crop-wild, dom) or the level of introgression (*P*dom) on growth and carbon uptake related fitness proxies. This could be due to the low number of samples from the *cw* and *dom* categories. Note that we did not test the variation in germination rate among seedlings as germination can be strongly impacted by stratification conditions.

**No adaptive phenotypic variation among populations, but signs of local adaptation to past climate in the European crabapple**

Under controlled conditions, seedlings from the different populations were found to have significantly different growth and morphology, but IBA analyses indicated that this variation was likely not adaptive. Seedlings belonging to the Austrian population were the tallest, had the highest absolute growth rate and the lowest number of leaves; by contrast, Romanian seedlings were the shortest, had the lowest absolute growth rate and the highest number of leaves. Italian seedlings had the highest chlorophyll content. The seedlings belonging to the Austrian population may be fitter in the climate conditions simulated in this experiment. We tested whether this phenotypic variation was adaptive. However, taking geographic distance into account, we did not find any significant covariation between genetic
and phenotypic variation. This suggests a lack of divergent selection on traits related to carbon uptake or growth that are often associated with plant responses to climate (Bussotti et al., 2015). Therefore, the phenotypic variation we observed among populations under controlled conditions may be the result of genetic drift alone; alternatively, the traits we selected (thought to be related to responses to climate) are not good candidates for investigating divergent selection. Leaf mass per area (LMA) and foliar nitrogen content demonstrated can be future targeted parameter to assess the responses of apple seedling to environmental stress (Bussotti et al., 2015). Another explanation could be that we did not phenotype enough seedlings from each genetic group. Indeed, we observed a high variation in each phenotypic trait and in their heritability estimates, suggesting that the traits we studied may be relevant but that a larger number of seedlings should be phenotyped and analyzed (e.g., (Klein et al., 1973)). However, some studies have found that even with large sample sizes, the standard error of heritability estimates can still be large and vary greatly between experimental designs (Visscher & Goddard, 2015). The fairly high heritability estimates for most of the traits considered here could be seen as consistent with rather weak within population selection, enabling the maintenance of ample additive genetic variation (Wheelwright et al., 2014). Furthermore, high variation in seedling traits combined with high heritability estimates could suggest that there is large room of genetic material for adaptation to work on. In addition, the population of origin of the seedlings did not explain all phenotypic variation. Even though we found an effect of the population of origin on phenotypic traits, its contribution was relatively low (e.g., model 4, $R^2$ for height = 0.119, and $R^2$ for height AGR= 0.047). Environment (e.g., climate) and interaction between genotype and environment could also impact fitness.

As *M. sylvestris* is distributed across gradient, we further investigated the role of climate in shaping the genetic variation among populations of the European crabapple, without considering phenotypic trait variation. We tested for an IBE pattern, where the pattern of neutral genetic variation covaries with ecological variables (here climate). There was no combined effect of geographic and climatic distance (IBD ∩ IBC), which allowed us to assess the contribution of these processes separately (Wang & Bradburd, 2014). We showed that IBD and IBC played a significant role ($R^2_{adj} = 47\%$ and $R^2_{adj} = 22\%$, respectively) on the genetic differentiation of European crabapple populations. Weak but significant IBD has been previously identified in wild apple relatives of the cultivated apple (*i.e.*, *M. sylvestris*, *M. orientalis* and *M. sieversii*) suggesting they have high dispersal capacities (Cornille et al., 2013b,a, 2015). Weak IBD is explained by self-incompatibility systems that prevent self-fertilization (Brown, 1992), pollen dispersal by bees and flies (*Syrphidae*) and endozoochorous seed dispersal by large mammals such as ungulates, wild pigs, brown bears or humans (Larsen et al., 2006). We show that in addition to IBD, IBC persisted after taking geographic distance into account. Climate can impose divergent selection pressures on different locations and thus reduce gene flow between populations, so that IBC contributes to genetic differentiation. The main variables explaining genetic differentiation in the European crabapple were related to temperature during the LGM. This suggests that the European crabapple may be locally adapted to its past temperature but not to its current climate. Local adaptation
to current climate is well studied in wind-dispersed trees (Savolainen et al., 2013; Kremer & Hipp, 2019; Pyhäjärvi et al. 2020). However, to our knowledge no study has shown local adaptation to past climate conditions in a tree species.

Additional factors other than climate can also shape adaptive divergence between populations. *Malus* sylvestris is a pioneer species that needs high levels of light and is not very competitive. Local adaptation to biotic factors such as the presence of other species is possible. Competition for light with other species such as the European beech (*Fagus sylvatica*) could be a source of divergence between populations. Local adaptation of fruit trees to biotic factors, including parasites (Olvera-Vazquez et al., 2021), deserves further investigations. Besides selection, the potential role of phenotypic plasticity in enabling growth and optimal fitness in changing environments also needs to be carefully evaluated (Benito Garzón et al., 2011).

**Further investigations needed on local adaptation and plasticity in response to climate in the European crabapple**

Our study raises concerns regarding the future of wild apple populations and their current vulnerability to current climate change. However, the adaptation of tree species to climate remains complex (Bussotti et al., 2015). For instance, in *Eucalyptus camaldulensis*, variation in leaf traits and performance proxies was unrelated to the climate of genotype provenance (Asao et al., 2020), whereas variation in several photosynthetic traits was clearly related to the climate of genotype provenance across Australia (Dillon et al., 2018). In contrast, collective differences in leaf morphology and photosynthetic physiology, in several *Populus* species may be adaptive for differences in growth season length, temperature and insolation (Keller et al., 2011; Kaluthota et al., 2015). Further investigations on local adaptation and plasticity to climate or biotic factors in the European crabapple are therefore needed. Genomic data will be particularly useful for determining the relative influence of adaptive and neutral processes on climate- or biotic- driven divergence by screening genomes from different populations in Europe. Comparing the fitness of seedlings from different populations in reciprocal transplants will also be important to further test for local adaptation. Our study therefore raises questions regarding the processes of local adaptation of fruit trees, and is a starting point for apple breeding programs.

**Acknowledgements**

We thank the Laboratoire d’Excellence BASC (Biodiversité, Agroécosystèmes, Société, Climat), the Institut Diversité Écologie et Évolution du Vivant (IDEEV), the European LEADER program and ATIP-Avenir CNRS Inserm for funding. We are grateful to the INRAE MIGALE bioinformatics platform (http://migale.jouy.inra.fr) for providing help and support, in particular Véronique Martin, Eric Montaubon and Valentin Loux. We also thank Adrien Falce, Olivier Langella and Benoit Johannet for help and support on the INRAE Génétique Quantitative et Evolution- Le Moulon lab cluster and the Plateforme de Génotypage GENTYANE INRAE UMR 1095. We thank Einar Flensted-Jensen, Michel
Bonessee, Alberto Dominici, Jean-Luc Ricaud, Eric Wolff, Alain Guinet, Line Caminade, for providing seeds, Richard Le Menn, Bernard Gambier. USDA is an equal opportunity provider, employer, and lender. Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. AR and TU were partly funded through project PN2019-2022/19270201 – BIODIVERS 3.

Authors contributions
AC, GV, AR, SB, TU conceived and designed the experiments; AC, GV, AT, TU, TK obtained the funding; AC, GV, AT, TU, KAO, SV, RR, XC, TK, CR sampled the material; XC, CR, AV, AR, GL, KAO, RR, MLG, HB, VC, HC, SV, MF performed the molecular biology analyses; AC, AF, KomAvi and XC analyzed the data. The manuscript was written by AC, KomAvi and AF, with essential input from other co-authors.
Figures and Tables

Figure 1. Bayesian clustering of the Malus sylvestris seedlings sampled in this study (N = 584) and the reference samples of Malus domestica (N = 40) inferred with STRUCTURE for K = 8, and its associated map of mean membership per sampled site. Each individual is represented by a vertical bar partitioned into clusters. Visualization was improved by sorting genotypes by country; countries are separated by a white line. The reference M. domestica reference samples are shown on the far left of the map in the Atlantic. Circle size is proportional to the number of individuals within the cluster (scale shown on the top right-hand corner).

Figure 2. Mean height of apple seedlings measured over the time of the experiment in controlled conditions (including pure and wild-wild hybrid seedlings, N = 449, and seedlings assigned to the M. domestica gene pool, N = 21, as detected with STRUCTURE for K=8). The 40 reference M. domestica individuals were not measured under controlled conditions thus are not shown here. Vertical lines represent the standard deviation. Populations: AUT (N = 81), DA (N = 78), DOM (N = 21, includes 13 Spanish genotypes and seedlings from other countries), FR-E (N = 50), FR-Lor (N = 28), FR-W (N = 77), IT (N = 27), RO (N = 108).

Figure 3. Variance partitioning analysis of the db-RDA results obtained for Malus sylvestris (N_{sites} = 21, 13 microsatellite markers). Variation of the site pairwise genetic differentiation (F_{ST}) is explained by the variables generating isolation-by-distance (geographical distance) and isolation-by-climate (with three bioclimatic variables during the last glacial maximum: Bio3, isothermality; Bio6, minimum temperature of coldest month; Bio 9: mean temperature of driest quarter).
Table 1. Number of *M. sylvestris* seedlings used in this study for population genetic analyses inferred with STRUCTURE for \( K = 8 \) with 13 microsatellite markers and phenotyping (growth and carbon-uptake related traits).

<table>
<thead>
<tr>
<th>Clusters</th>
<th>( N_{pure} )</th>
<th>( N_{iw} )</th>
<th>( N_{ew} )</th>
<th>( N_{dom} )</th>
<th>( N_{no\ cluster} )</th>
<th>Total measured for phenotypic traits</th>
<th>Wild population name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (light green)</td>
<td>32</td>
<td>46</td>
<td>7</td>
<td>0</td>
<td>10</td>
<td>92</td>
<td>FR-W</td>
</tr>
<tr>
<td>Q2 (yellow)</td>
<td>0</td>
<td>52</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>57</td>
<td>FR-E</td>
</tr>
<tr>
<td>Q3 (lor)</td>
<td>28</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>28</td>
<td>FR-Lor</td>
</tr>
<tr>
<td>Q4 (blue)</td>
<td>61</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>85</td>
<td>DA</td>
</tr>
<tr>
<td>Q5 (purple)</td>
<td>23</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>34</td>
<td>IT</td>
</tr>
<tr>
<td>Q6 (dark green)</td>
<td>66</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>83</td>
<td>AUT</td>
</tr>
<tr>
<td>Q7 (red)</td>
<td>77</td>
<td>34</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>113</td>
<td>RO</td>
</tr>
<tr>
<td>Q8 (black – M. domestica)</td>
<td>40</td>
<td>0</td>
<td>46</td>
<td>21</td>
<td>8</td>
<td>73</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>287</td>
<td>175</td>
<td>68</td>
<td>21</td>
<td>33</td>
<td>551 (584)</td>
<td></td>
</tr>
<tr>
<td>Total measured for height and number of leaves</td>
<td>282</td>
<td>167</td>
<td>63</td>
<td>21</td>
<td>32</td>
<td>533 (565)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>129</td>
<td>82</td>
<td>22</td>
<td>6</td>
<td>18</td>
<td>239 (257)</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
<td></td>
</tr>
</tbody>
</table>

$N_{\text{pure}}$: number of seedlings assigned to a wild gene pool with a membership coefficient > 0.9; $N_{\text{ww}}$, number of wild-wild hybrids (i.e., seedlings with a membership coefficient > 0.1 to a wild gene pool other than its own wild gene pool and a membership coefficient < 0.1 to the M. domestica gene pool); $N_{\text{cw}}$: number of crop-wild hybrids (i.e., seedling assigned to the M. domestica gene pool with a membership coefficient > 0.1). $N_{\text{no cluster}}$: seedlings that could not be assigned to any defined gene pool; Total measured for phenotypic traits: number of individuals measured for each phenotypic trait and included in the statistical analyses, the number in brackets represents the initial sample size before data were filtered for statistical analyses. Wild population name: populations defined with STRUCTURE at $K$=8 excluding crop-wild hybrids and seedlings from misidentified mother trees (i.e., including only wild pure and wild-wild hybrids).
Table 2. Final model depicting effects of the *Malus sylvestris* population to which each seedling belonged (i.e., cluster inferred with STRUCTURE at *K*=8) on phenotypic traits (i.e., height, number of leaves, internode, chlorophyll and flavonol content, NBI) measured in 533 individuals. Variables in green are significant (*P* < 0.05).

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Cluster</th>
<th>Mother</th>
<th>Model</th>
<th>Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>χ</em>²</td>
<td><em>P</em>-value</td>
<td>df</td>
<td>REML</td>
</tr>
<tr>
<td>Fitness</td>
<td>17.863</td>
<td>0.007***</td>
<td>6</td>
<td>1.229</td>
</tr>
<tr>
<td>Height_AGR</td>
<td>12.846</td>
<td>0.045*</td>
<td>6</td>
<td>-2.264</td>
</tr>
<tr>
<td>Height_RGR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leaf_AGR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leaf_RGR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Height whole AGR</td>
<td>22.243</td>
<td>1.00e-03***</td>
<td>6</td>
<td>630</td>
</tr>
<tr>
<td>Leaf whole AGR</td>
<td>36.326</td>
<td>2.38e-06***</td>
<td>6</td>
<td>-1,277</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>31.623</td>
<td>1.93e-05***</td>
<td>6</td>
</tr>
<tr>
<td>Number of leaves</td>
<td>22.285</td>
<td>0.001***</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Chlorophyll</td>
<td>14.418</td>
<td>0.025*</td>
<td>6</td>
<td>1.181</td>
</tr>
<tr>
<td>Flavonol</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>NBI</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Internode (nbleaf/height)</td>
<td>17.768</td>
<td>0.007***</td>
<td>6</td>
<td>-1.328</td>
</tr>
</tbody>
</table>

***: P-value <0.001; **: 0.01<P-value<0.001; *: 0.05<P-value<0.01; AIC: Akaike Indice Criterion.
Table 3. Contribution of geography and climate to the genetic variation observed among *M. sylvestris* seedlings. Distance-based redundancy analyses tested the effects of geography, climate and phenotype on genetic differentiation among 21 sites from 13 microsatellites in the European crabapple. Only significant variables are presented.

<table>
<thead>
<tr>
<th>db-RDA</th>
<th>% of variance explained</th>
<th>d.f.</th>
<th>p-value</th>
<th>Adj-R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global analysis</td>
<td>25.9</td>
<td>7</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>11.3</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geography (IBD)</td>
<td>14.9</td>
<td>4</td>
<td>&lt;0.015</td>
<td></td>
</tr>
<tr>
<td>Environment (IBC_LGM)</td>
<td>11.04</td>
<td>3</td>
<td>&lt;0.015</td>
<td></td>
</tr>
</tbody>
</table>

Total: 69.9
| Residuals | 11.3% | - | - |

BIO3_LGM: isothermality (BIO2/BIO7) (×100); BIO6_LGM: minimum temperature of coldest month; BIO9_LGM: mean temperature of driest quarter, IBD: isolation-by-distance; IBC_LGM: isolation-by-climate during the last glacial maximum.
References


Puechmaille SJ. 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**: 608–627.

Pyhäjärvi T, Kujala ST, Savolainen O. 2027 years of forestry meets genomics in Pinus sylvestris. *Evolutionary Applications* n/a.


Malus domestica

Western France

Eastern France

Denmark

Italy

Austria

Romania

Latitud 60°N

55°N

50°N

45°N

40°N

10°W 0° 10°E 20°E 30°E

Longitude

Malus domestica

Spain

Western France

Eastern France

Denmark

Italy

Austria

Romania

Latitude

Longitude
Evolution of height over time

Population
- AUT
- DA
- DOM
- FR-E
- FR-Lor
- FR-W
- IT
- RO

Mean Height (cm)

Date
- Aug 01
- Aug 15
- Sep 01
- Sep 15
Variance partitioning of the db–RDA results

IBD

0.47

IBC_LGM

0.22

Residuals = 0.35

Values <0 not shown