


# A few north Appalachian populations are the source of European black locust

Xavier Paul Bouteiller<sup>1</sup>  | Cindy Frédérique Verdu<sup>2</sup> | Emmi Aikio<sup>3</sup> | Paul Bloese<sup>4</sup> | Kasso Dainou<sup>2</sup> | Adline Delcamp<sup>1</sup> | Olivier De Thier<sup>2</sup> | Erwan Guichoux<sup>1</sup> | Coralie Mengal<sup>2</sup> | Arnaud Monty<sup>5</sup> | Marion Pucheu<sup>1</sup> | Marcela van Loo<sup>6</sup> | Annabel Josée Porté<sup>1</sup> | Ludivine Lassois<sup>2</sup> | Stéphanie Mariette<sup>1</sup>

<sup>1</sup>BIOGECO, INRA, Univ. Bordeaux, Cestas, France

<sup>2</sup>Biodiversity and Landscape Unit, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

<sup>3</sup>Department of Genetics and Physiology, University of Oulu, Oulu, Finland

<sup>4</sup>Department of Forestry, Michigan State University, East Lansing, Michigan

<sup>5</sup>Forest Management Unit, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

<sup>6</sup>Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

## Correspondence

Xavier Paul Bouteiller and Stéphanie Mariette, BIOGECO, INRA, Univ. Bordeaux, Cestas, France.  
Emails: [bouteiller.xavier@gmail.com](mailto:bouteiller.xavier@gmail.com); [stephanie.mariette@inra.fr](mailto:stephanie.mariette@inra.fr)

## Funding information

ANR-10-EQPX-16 Xyloforest; Société Française d'Ecologie; BioGeCo INRA-Univ Bordeaux Research Unit; Special Research Fund of the University of Liège; Forest and Nature Management Research Unit of Gembloux Agro-BioTech; Agence de l'Eau Adour Garonne; EU COST Action FP1403; Trees4Future project, Grant/Award Number: 284181

## Abstract

The role of evolution in biological invasion studies is often overlooked. In order to evaluate the evolutionary mechanisms behind invasiveness, it is crucial to identify the source populations of the introduction. Studies in population genetics were carried out on *Robinia pseudoacacia* L., a North American tree which is now one of the worst invasive tree species in Europe. We realized large-scale sampling in both the invasive and native ranges: 63 populations were sampled and 818 individuals were genotyped using 113 SNPs. We identified clonal genotypes in each population and analyzed between and within range population structure, and then, we compared genetic diversity between ranges, enlarging the number of SNPs to mitigate the ascertainment bias. First, we demonstrated that European black locust was introduced from just a limited number of populations located in the Appalachian Mountains, which is in agreement with the historical documents briefly reviewed in this study. Within America, population structure reflected the effects of long-term processes, whereas in Europe it was largely impacted by human activities. Second, we showed that there is a genetic bottleneck between the ranges with a decrease in allelic richness and total number of alleles in Europe. Lastly, we found more clonality within European populations. Black locust became invasive in Europe despite being introduced from a reduced part of its native distribution. Our results suggest that human activity, such as breeding programs in Europe and the seed trade throughout the introduced range, had a major role in promoting invasion; therefore, the introduction of the missing American genetic cluster to Europe should be avoided.

## KEYWORDS

biological invasion, bottleneck, introduction history, population genetics, *Robinia pseudoacacia*, single-nucleotide polymorphism

## 1 | INTRODUCTION

Since their first definition in Charles Elton's book (Elton, 1958), biological invasions have been increasingly studied over the last few decades. Compared to the ecological impacts of many invasive species and the management issues surrounding them, the role of evolution in biological invasions has long been overlooked (Colautti & Lau, 2015). In order to fill this knowledge gap, it is crucial to identify the source populations of the introduction for a better understanding of the evolutionary mechanisms behind invasiveness, such as the role of selection, local adaptation or admixture (Colautti & Lau, 2015; Dlugosch, Anderson, Braasch, Cang, & Gillette, 2015; Keller & Taylor, 2008). The practical applications of such studies are the identification of source risk and the prediction of the invasive potential of a population (Chown et al., 2015).

When a species is introduced to a new range, it is generally expected to experience a genetic bottleneck leading to a loss of genetic diversity (allelic richness or heterozygosity) (Dlugosch et al., 2015). For example, the invasive plant *Heracleum mantegazzianum* exhibited a lower diversity in the invasive range attesting a strong founder event (Henry et al., 2009). However, some studies (Dlugosch et al., 2015; Dlugosch & Parker, 2008) have emphasized that the loss of genetic diversity within the native and invasive ranges was generally weak (15%–20% on average); this can be explained by multiple introductions that have limited the loss of diversity, as in the case of *Phalaris arundinaceae* or *Prunus serotina* (Lavergne & Molofsky, 2007; Pairon et al., 2010). Genetic diversity is even likely to increase in the invasive range if population admixture is high (Dlugosch et al., 2015; Dlugosch & Parker, 2008), although a large increase is rare (Uller & Leimu, 2011); for example, the invasion of *Phalaris arundinaceae* was shown to have been promoted by an increased genetic variation (Lavergne & Molofsky, 2007). Genomic admixture was likely to have favored the success of *Silene vulgaris* in its new American range (Keller, Fields, Berardi, & Taylor, 2014; Keller & Taylor, 2010).

Additionally to propagule pressure during introduction, the mating system can have a high impact on the diversity and genetic structure of populations. Clonal or self-fertilizing species are likely to experience a greater loss of genetic diversity, whereas a bottleneck effect may be reduced for outcrossing species (Baker, 1967; Pappert, Hamrick, & Donovan, 2000). For example, the loss of genetic diversity between native and invasive ranges was greater for purely clonal populations of the invasive *Oxalis pes-caprae* than for sexual ones (Ferrero et al., 2015).

Few studies have been carried out on the numerous invasive species of trees and shrubs, despite their great impact on ecosystems (Richardson & Rejmánek, 2011). Contrary to many herbaceous species, invasive trees (comprising 357 tree species, i.e., nearly 0.5% of all trees species; Richardson & Rejmánek, 2011) have often been voluntarily introduced to their new ranges for horticultural or forestry purposes (Richardson & Rejmánek, 2011), resulting in multiple repeated introductions which may have shaped the diversity of the trees in the introduced range (Hirsch, Richardson, & Le Roux, 2017). Furthermore, invasive trees are characterized by a longer generation

time compared to invasive herbaceous species; this life-history trait may influence differentiation rate between ranges. In addition, the fact that a few centuries and tree generations have passed since the first introduction presents a challenge to the study of evolutionary processes in invasive trees (Hirsch et al., 2017). Little research has been carried out on the population genetics of invasive trees in both their invasive and native ranges. To our knowledge, a few studies have evidenced mostly multiple introductions to the invasive range from the native range; for example, *Acacia saligna* (Thompson, Bellstedt, Richardson, Wilson, & Le Roux, 2015) and *Pinus tadea* (Zenni, Bailey, & Simberloff, 2014) have been widely sampled from the native range exhibiting a high level of admixture within the invasive range, and *Prunus serotina* (Pairon et al., 2010) has been introduced several times, but from a limited part of the native range.

Black locust *Robinia pseudoacacia* L. (Fabaceae) is native to North America and is considered as invasive on a world scale (eight regions among the 14 defined by Richardson & Rejmánek, 2011). In Europe, it is now recognized as one of the 100 worst invasive species (Basnou, 2009; DAISIE, 2006).

The native range of this species consists of two disjoint areas on both sides of the Mississippi watershed (Little, 1971); the largest area corresponding to the Appalachian Mountains and partially encompassing several current States (Pennsylvania, Maryland, West Virginia, Virginia, North Carolina, South Carolina, Georgia, Alabama, Tennessee, Kentucky, and Ohio) and the smallest area being located further west in the Ozark Mountains (Missouri, Arkansas and Oklahoma). In America, black locust was intensively displaced by settlers due to the undeniable interest in its wood as stated by Michaux in "*Histoire des arbres d'Amérique septentrionales tome III*" (1813), or by Cobbett in his book "The Woodlands" (1825). Such displacement has sometimes led to the misinterpretation of its native distribution; for example, in the *Gardeners Dictionary (1756–1759)*, Miller wrongly stated that black locust was native to Massachusetts (Michener, 1988). To date, this species has spread to every state in the contiguous USA and also to British Columbia, Québec, Newfoundland and Labrador in Canada (Schütt, 1994). It was introduced to Europe during the early 17th century and it is now present in all European countries; however, when and how the first introduction occurred is not precisely known and is still shrouded in mystery. It is widely written that Jean Robin (1550–1629), botanist of King Henri the Fourth in 1601, was responsible for the first European introduction. A century later, Carl Von Linné gave the black locust its current name, *Robinia*, in recognition of the work carried out by Jean Robin and son (Vespasien) in acclimating it to Europe. Actually, 1601 seems a very unlikely date since black locust is absent (Wein 1930 cited by Cierjacks et al., 2013) from the lists edited by Jean and Vespasien Robin in 1601 ("*catalogue de son jardin*") and in 1620 ("*histoire des plantes nouvellement trouvées en l'isle de Virginie*"). To our knowledge, the first citations of the species appeared in England in the John Tradescant list in "*plantarum in horto Iohannem Tradescanti*" (1634) (cited in "Early British botanists and their gardens" by Gunther, 1922) under the name of *Locusta virginiana arbor*, and quasi-simultaneously in France in the book *Plantarum canadesium*

*historia* (1635) by Vespasien Robin's friend, Jacques Philippe Cornuti, under the name of *Acacia americana robini*. Moreover, Tradescant's son travelled to Virginia before 1640 (Hamel, 1854) and as it has been documented that he brought several plants back with him, it is likely that he was the one who first introduced black locust seeds to Europe from Virginia. It has been established that Tradescant the elder corresponded and exchanged seeds with the Robins from 1601 (Gunther, 1922); thus he probably sent some seeds to Vespasien Robin in France, who would have sown them and cultivated the trees (such as the one planted in the King's garden in Paris in 1634 ("*jardin des plantes*"; in *Biographie Universelle*, 1824)).

In Europe, the first introduction appears to have been followed by a period of interest in its ornamental aspect; however, it subsequently fell into disuse in the early 18th century, as explained in a dictionary from 1722 about the black locust, which was quoted by Nicolas François de Neufchateau in his book "*lettre sur le robinier*" (1807). In the middle of the 18th century, American explorers returned to Europe and promoted the use of black locust in forestry: For instance, Michaux (1813) described the abundance of this tree in the Allegheny mountains throughout Pennsylvania and West Virginia and indicated that after the end of the 18th century, the tree was appreciated more for the excellent qualities of its wood than for the beauty of its foliage and flowers. At the same time, the English politician William Cobbett, who emigrated to America in the late 18th century, emphasized all the qualities of this tree and promoted its plantation in Europe: "I sold the plants; and, since that time, I have sold altogether more than a million of them," adding that "My seed has always come from the neighborhood of Harrisburgh in Pennsylvania" (Cobbett, 1825).

From this information, we can conclude that the European dissemination of the black locust seems to have experienced a lag phase between the tree species' first introduction to Europe—possibly from Virginia during the early 17th century—and its rediscovery in the middle of the 18th century, leading to a new wave of introductions of the species, which probably came from Pennsylvania and the Virginias. More recently, black locust breeding programs have been carried out in central Europe since the beginning of the 20th century (Keresztesi, 1983; Liesebach, Yang, & Schneck, 2004; Straker, Quinn, Voigt, Lee, & Kling, 2015). Currently, Hungary is the European leader in the production of black locust seedlings, and their selected provenances for wood production are now widely distributed in Europe for new forest plantations (Keresztesi, 1983; Liesebach et al., 2004; Straker et al., 2015). In Europe, the black locust is now recognized as one of the 100 worst invasive species (Basnou, 2009; DAISIE, 2006) and it is considered as an invasive tree on a world scale (eight regions out of the fourteen defined by Richardson & Rejmánek, 2011).

Although knowledge about the black locust's genetic diversity is key to developing further ecological or evolutionary studies (Lawson Handley et al., 2011), little information exists about its genetic diversity and structure in introduced ranges, nor regarding its origin and differentiation from the population sources in North America. The only studies we know of in Europe compared four American populations with sixteen German and Hungarian

populations (Liesebach & Schneck, 2012; Liesebach et al., 2004), but although the results suggested a high genetic differentiation among American populations, they were mostly inconclusive. Modern molecular and statistical tools used in population genetics have proved to be useful for finding the geographic origins of invasive species, complementing or providing a solution to the lack of available historical knowledge (Besnard et al., 2014; Chown et al., 2015; Cristescu, 2015; Hoos, Whitman Miller, Ruiz, Vrijenhoek, & Geller, 2010). Using SNP markers developed for the black locust (Verdu et al., 2016), we investigated its introduction history and genetic diversity in its native range and European invasive range, in particular by answering the following questions: (a) Can we identify the native population sources of European black locust? (b) What is the genetic differentiation within and between ranges? (c) Can we detect a founder event associated with a loss of genetic diversity?

## 2 | MATERIAL AND METHODS

### 2.1 | Sampling

Sixty-three populations of black locust were sampled in both the native range (29 populations) and the European invasive range (34 populations). Sampling was conducted between spring 2014 and fall 2016 (Table 1 and Supporting Information Appendix S1) by different collaborators using the same protocol: Between 10 and 30 trees were sampled in each population. Samples were collected either in common gardens or in natural populations. A total of 818 individuals were sampled: 402 from Europe and 416 from North America.

Black locust propagates through sexual and asexual reproduction. In common gardens, since trees were grown from seeds of known origin, there was no risk of collecting clones. However, in natural populations, a minimal distance of 25 m was kept between two sampled trees in order to minimize the risk of collecting clones.

Either leaves, cambium, buds, or seeds were harvested depending on the season. For leaf sampling, a few leaflets on a green healthy leaf were collected using a manual tree pruner. For cambium sampling, external bark was removed from the trunk with a knife, then five rings of wood were collected using a 1-cm-diameter punch. In the field, samples were put into referenced tea bags and then placed into plastic boxes containing silica gel in order to dry the samples. The silica gel was renewed after 24 hr and 48 hr and then until it no longer changed color. The plastic boxes were then stored at ambient temperature in closed cupboards.

In natural populations, GPS coordinates of either the population or each sampled tree were recorded using a portable GPS (GPSMAP62, Garmin, Olathe, KS, USA). On the campus of Michigan State University, the geographic origins of each mother tree were known and were used for the coordinates of the sampled trees, and the populations were defined by gathering trees from a close geographic location (see Supporting Information Appendix S1).

**TABLE 1** General genetic information regarding the sampled populations

Number	Range	Country/State	Pop	X Long	Y Lat	N	G	R	F <sub>IS</sub> Mean	F <sub>IS</sub> LC195	F <sub>IS</sub> HC195	Ho	Hs
1	EU	France	Barthelasse Avignon	4.818	43.965	19	18	0.944	0.112	0.048	0.179	0.232	0.262
2	EU	Czech Republic	Brno	16.518	49.042	11	5	0.400	-0.026	-0.131	0.084	0.240	0.234
3	EU	Hungary	Budapest	19.107	47.663	20	13	0.632	0.115	0.047	0.184	0.244	0.276
4	EU	Romania	Carei	22.449	47.661	11	11	1.000	0.096	0.026	0.165	0.259	0.286
5	EU	Belgium	Corphalie	5.259	50.539	10	10	1.000	0.120	0.042	0.195	0.251	0.285
6	EU	Poland	Drewnica	21.251	52.253	10	10	1.000	0.060	-0.015	0.138	0.231	0.246
7	EU	Spain	Gafos Galicia	-8.617	42.383	12	5	0.364	0.056	-0.047	0.166	0.269	0.285
8	EU	Bulgaria	Gorna Oryahovitsa	25.694	43.119	12	12	1.000	0.100	0.026	0.176	0.246	0.273
9	EU	Netherlands	Kelpen-Oler	5.825	51.205	12	12	1.000	0.119	0.039	0.197	0.245	0.278
10	EU	Germany	Klein	9.089	49.991	11	11	1.000	0.010	-0.099	0.124	0.240	0.243
11	EU	France	La Flotte	-0.305	44.385	6	3	0.400	0.062	-0.087	0.205	0.240	0.259
12	EU	France	La Gouaneyre	-0.273	44.376	6	5	0.800	0.152	0.020	0.281	0.213	0.251
13	EU	England	London Street Ham	0.143	51.433	10	7	0.667	0.144	0.023	0.265	0.239	0.279
14	EU	England	London Wandworth	0.163	51.446	10	8	0.778	0.085	0.002	0.166	0.240	0.262
15	EU	Macedonia	Macedonia	21.571	41.507	12	12	1.000	0.086	0.017	0.157	0.242	0.265
16	EU	Germany	Meppen	7.377	52.704	12	8	0.636	-0.097	-0.211	0.026	0.257	0.235
17	EU	Spain	Montseny	2.512	41.831	12	12	1.000	0.068	-0.009	0.147	0.230	0.246
18	EU	Germany	Munchenberg	14.046	52.559	12	4	0.273	-0.180	-0.319	-0.030	0.223	0.190
19	EU	Bulgaria	Novi Pazar-Kulevcha	27.195	43.345	12	12	1.000	0.091	0.024	0.161	0.242	0.267
20	EU	Hungary	Nyirsegi	19.041	47.581	12	12	1.000	0.087	0.004	0.173	0.242	0.265
21	EU	Germany	Obermeidenrich	6.816	51.476	12	12	1.000	0.066	0.000	0.134	0.254	0.272
22	EU	Poland	Pinczow	20.702	50.265	10	10	1.000	0.154	0.067	0.249	0.212	0.251
23	EU	Poland	Poznan	16.808	52.311	12	12	1.000	0.119	0.037	0.203	0.230	0.261
24	EU	Germany	Priesterweg Naturpark	13.358	52.461	12	12	1.000	0.105	0.035	0.177	0.249	0.278
25	EU	Hungary	Pusztavacs	19.506	47.172	10	10	1.000	0.059	-0.021	0.147	0.256	0.273
26	EU	France	Remy	2.675	49.460	12	8	0.636	0.104	0.014	0.193	0.227	0.254
27	EU	Greece	Rhodos	27.944	36.287	12	9	0.727	0.104	0.014	0.197	0.227	0.254
28	EU	Slovakia	Slovakia	19.867	48.720	12	12	1.000	0.161	0.075	0.251	0.218	0.260
29	EU	Poland	Szczecin	14.548	53.337	12	12	1.000	0.076	0.003	0.151	0.242	0.262
30	EU	Turkey	Turkey	32.904	40.159	11	11	1.000	0.081	-0.005	0.172	0.224	0.245
31	EU	Netherlands	Uden	5.618	51.685	11	5	0.400	0.133	0.015	0.247	0.242	0.280
32	EU	Spain	Valencia	-0.784	39.397	19	6	0.278	0.066	-0.048	0.179	0.247	0.264
33	EU	Spain	Vitoria	-1.942	43.216	13	13	1.000	0.132	0.054	0.210	0.237	0.273

(Continues)

TABLE 1 (Continued)

Number	Range	Country/State	Pop	X Long	Y Lat	N	G	R	$F_{IS}$ Mean	$F_{IS}$ LC195	$F_{IS}$ HC195	Ho	Hs
34	EU	Austria	Wien	16.473	48.252	12	12	1.000	0.069	-0.013	0.155	0.249	0.268
35	US	PA—Pennsylvania	ALTOONA	-78.383	40.489	11	10	0.900	<b>0.115</b>	0.036	0.192	0.249	0.281
36	US	VA—Virginia	Barbours Creek	-80.110	37.580	21	20	0.950	<b>0.126</b>	0.064	0.192	0.215	0.246
37	US	NC—North Carolina	Blue Ridge	-82.672	35.457	22	21	0.952	<b>0.118</b>	0.060	0.176	0.210	0.238
38	US	WV—West Virginia	CAMP CREEK	-81.103	37.488	12	12	1.000	0.073	-0.008	0.154	0.237	0.255
39	US	TN—Tennessee	CHATTANOOGA	-85.783	35.120	12	11	0.909	0.065	-0.020	0.151	0.210	0.224
40	US	KY—Kentucky	DANIEL BOONE NF	-83.645	37.751	12	12	1.000	<b>0.147</b>	0.067	0.230	0.222	0.260
41	US	KY—Kentucky	Erline	-83.540	37.040	22	22	1.000	<b>0.082</b>	0.027	0.138	0.222	0.241
42	US	AR—Arkansas	FAYETTEVILLE	-94.205	36.071	10	9	0.889	<b>0.186</b>	0.095	0.277	0.202	0.248
43	US	WV—West Virginia	FORT MILL RIDGE	-78.797	39.327	12	12	1.000	0.056	-0.012	0.124	0.236	0.250
44	US	AR—Arkansas	FORT SMITH	-94.290	35.343	12	10	0.818	<b>0.117</b>	0.027	0.215	0.188	0.213
45	US	OH—Ohio	Ironton	-82.460	38.800	22	17	0.762	<b>0.105</b>	0.033	0.181	0.231	0.258
46	US	WV—West Virginia	LEWISBURG	-80.381	37.783	7	7	1.000	0.037	-0.061	0.137	0.239	0.249
47	US	NC—North Carolina	Locust Cove	-83.710	35.360	22	18	0.810	<b>0.115</b>	0.056	0.179	0.213	0.241
48	US	KY—Kentucky	Morehead	-83.466	38.091	22	19	0.857	<b>0.135</b>	0.064	0.209	0.196	0.226
49	US	AR—Arkansas	OJACHITA	-93.837	34.449	12	9	0.727	<b>0.097</b>	0.015	0.178	0.230	0.254
50	US	WV—West Virginia	Perry	-78.660	39.000	22	22	1.000	<b>0.097</b>	0.026	0.168	0.229	0.253
51	US	AR—Arkansas	Pleasant Hill	-93.460	35.590	4	3	0.667	-0.058	-0.200	0.099	0.245	0.234
52	US	NC—North Carolina	SHOOTING CREEK	-83.628	35.055	11	11	1.000	<b>0.111</b>	0.039	0.186	0.209	0.235
53	US	WV—West Virginia	Slatyfork	-80.000	38.180	17	17	1.000	<b>0.108</b>	0.040	0.175	0.215	0.242
54	US	VA—Virginia	Stokesville	-79.300	38.280	22	22	1.000	<b>0.131</b>	0.060	0.206	0.204	0.235
55	US	MD—Maryland	US GRP 1	-78.750	39.650	15	15	1.000	<b>0.095</b>	0.030	0.159	0.224	0.247
56	US	WV—West Virginia	US GRP 2	-81.100	39.067	11	11	1.000	<b>0.133</b>	0.052	0.216	0.214	0.247
57	US	VA—Virginia	US GRP 3	-79.933	37.267	6	6	1.000	<b>0.123</b>	0.018	0.228	0.247	0.281
58	US	KY—Kentucky	US GRP 4	-84.500	38.033	6	6	1.000	0.095	-0.001	0.196	0.206	0.228
59	US	KY—Kentucky	US GRP 5	-84.533	38.650	8	8	1.000	<b>0.175</b>	0.084	0.268	0.200	0.242
60	US	KY—Kentucky	US GRP 6	-83.683	36.750	7	7	1.000	<b>0.157</b>	0.047	0.260	0.215	0.255
61	US	AR—Arkansas	Victor	-93.050	35.650	22	17	0.762	<b>0.151</b>	0.085	0.224	0.193	0.227
62	US	OH—Ohio	WAYNE NF	-82.594	38.658	12	12	1.000	<b>0.132</b>	0.049	0.216	0.222	0.256
63	US	VA—Virginia	Whiteop	-81.656	36.769	22	20	0.905	<b>0.114</b>	0.049	0.183	0.217	0.245

Note. The range corresponds either to Europe (EU) or the USA (US) and either the country or the state is indicated. X Long and Y Lat (longitude and latitude, respectively) corresponded to the GPS coordinates of the sampled population provided in the WGS84 geographic projection. N is the number of individuals genotyped per population. G is the number of unique genotypes in each population. R is the index of clonal diversity, as defined in the material and methods section.  $F_{IS}$  mean,  $F_{IS}$  LC95 and  $F_{IS}$  HC95 indicate, respectively, mean  $F_{IS}$  value and the 95% confidence interval computed using the hierstat R package for each population. The  $F_{IS}$  values in bold indicate that the 95% confidence interval, calculated using 1,000 bootstrap replicates, does not include zero. Ho is the observed heterozygosity and Hs the expected heterozygosity. Genetic diversity values were calculated using the initial dataset after clone removal (i.e., 113 SNPs and 720 individuals).



## 2.2 | DNA extraction and genotyping

For each individual, either a 1 cm<sup>2</sup> leaf sample was collected on a leaflet, cambium was manually extracted from one ring of wood or five buds were collected. The plant material was then crushed using an automated grinder (2010 Geno/Grinder, SPEX SamplePrep, Metuchen, NJ, USA). For four populations (Corphalie, Drewnica, Pinczow and Lewisburg), a few seeds from ten sampled mother trees were scarified and grown in the laboratory (Bouteiller, Porté, Mariette, & Monty, 2017). The first fresh leaf on each specimen was then used for genotyping. DNA was extracted and isolated from all populations using DNeasy 96 Plant Kit (Qiagen, Venlo, Netherlands) following the manufacturer's protocol. One negative control was set on each plate. DNA concentration was measured using an UV spectrophotometer NanoDrop 8000 (Thermo Fisher Scientific Inc., Wilmington, Delaware, USA) and confirmed using Quant-iT™ dsDNA Assay Kit (Thermo Fisher Scientific Inc., Wilmington, Delaware, USA). Besides DNA concentration, 260/280 and 260/230 absorbance ratio provided information about DNA purity. DNA concentrations were standardized to 10 ng/μl before SNP genotyping.

SNPs have recently been developed on black locust (Verdu et al., 2016) using the double-digest RAD approach. Nine samples (from six North American trees, two European trees, and one Iranian tree) were digested with *EcoRI/MseI* and subsequent libraries were sequenced using Illumina technology. The resulting sequences were submitted to a bioinformatics pipeline and more than 300 SNPs were validated by carrying out individual genotyping using the Sequenom MassARRAY System (Agena Bioscience, San Diego, USA). It is assumed that the SNPs used in the present study were located within neutral regions since we did not use restriction enzymes targeting nonneutral restriction sites.

Two genotyping experiments were performed using these SNPs: (a) All collected samples (initial dataset) were genotyped using 113 SNPs, which were selected according to the procedure presented in Bouteiller et al. (2018); (b) after clone removal (see Section 3), 163 individuals were subsampled randomly within the populations in each range to maintain the sampling design between ranges (additional dataset: 69 individuals from the USA and 96 from Europe) and genotyped using a total of 251 SNPs. These additional SNPs were among the SNPs developed by Verdu et al. (2016), which had first been discarded as we had initially chosen to prioritize the more polymorphic SNPs. Four additional multiplexes of SNPs (138 SNPs in total) were thus designed using the MassArray Assay Editor 4.0.1.4 software (see Verdu et al., 2016 for more details regarding the procedure). SNP genotyping was performed using the Sequenom MassARRAY System (Agena Bioscience, San Diego, USA) at the Bordeaux Genome Transcriptome Facility (<https://pgtb.cgfb.u-bordeaux.fr/en>), and using the iPLEX Gold chemistry genotyping kit according to the manufacturer's instructions. SNP data were visualized and validated using ViClust, a R program that we implemented for Galaxy (<https://usegalaxy.org/>); the program was also made available as a

standalone R script for Linux or Windows (Bouteiller et al., 2018). The whole dataset will be made available on the Open Science Framework repository after acceptance.

## 2.3 | Data analysis

### 2.3.1 | Clone removal

For the analysis of genetic diversity and structure within and between ranges, we chose to identify and remove clones from the analysis using R version 3.3.1 (R Development Core Team, 2016). Within populations, only markers without missing values were kept, and a pairwise comparison of each genotyped individual was carried out in order to detect putative clones.

For each population, the index of clonal diversity  $R$  (Arnaud-Haond, Duarte, Alberto, & Serrão, 2007) was calculated as:

$$R = \frac{(G-1)}{(N-1)} \quad (1)$$

where  $G$  is the number of unique genotypes in the considered population and  $N$  the sample size of the population. This index varies from 0 in purely clonal populations to 1 when all the individuals corresponded to different genotypes.

As some populations were sampled from trees in common gardens or from laboratory-grown seedlings which originated from seeds (Table 1), they were unlikely to contain clones; however, we checked that no clone was present in these populations and removed them before carrying out the subsequent analysis. The difference in clonality between ranges was determined using a Pearson  $\chi$  squared test with Yate's continuity correction using R version 3.3.1 (R Development Core Team, 2016).

### 2.3.2 | Molecular genetic structure

After removing clones from the dataset, molecular genetic differentiation was explored both between ranges and among populations within ranges using two approaches.

First, the typology of all sampled individuals from both ranges was assessed using a principal component analysis (PCA), developed in the R *ade4* library (Jombart, 2008; Jombart, Devillard, & Balloux, 2010).

Second, individual membership was analyzed using the Bayesian clustering approach, *Markov Chain Monte Carlo* (MCMC), developed in the *STRUCTURE* v2.3.4 software (Porrás-Hurtado et al., 2013; Pritchard, Stephens, & Donnelly, 2000), using an admixture model. Each model run assumed that the overall diversity was structured into  $K$  clusters and, according to the SNP data, each individual was assigned proportionally to each cluster; thus, for each individual membership we obtained coefficients with a 90% confidence interval for the associated  $K$  clusters by setting the *ANCESTRY* parameter from 0 to 1.

Each run corresponded to a MCMC model with a burn-in period of 500,000 iterations followed by 500,000 iterations, which was repeated 10 times (Gilbert et al., 2012). The analysis was first performed using the initial dataset after clone removal to determine the structure

of populations from both ranges ( $K$  varying from 1 to 20), then it was performed for each range separately ( $K$  varying from 1 to 15). The most probable number of clusters was determined according to Evanno, Regnaut, & Goudet (2005) using the peak in the  $\Delta K$  parameter calculated with the STRUCTURE HARVESTER software (Earl & vonHoldt, 2012). All runs were computed on the GenoToul bioinformatics cluster (<http://bioinfo.genotoul.fr/>) using the StrAuto script (Chhatre & Emerson, 2017) to produce STRUCTURE mainparams and extra-params and to automatize and parallelize the STRUCTURE analysis. Finally, the CLUMPAK software (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) was used to synthesize STRUCTURE outputs and compute graphs of membership into the most likely  $K$  cluster number, in the form of line charts for individuals or pie charts for populations.

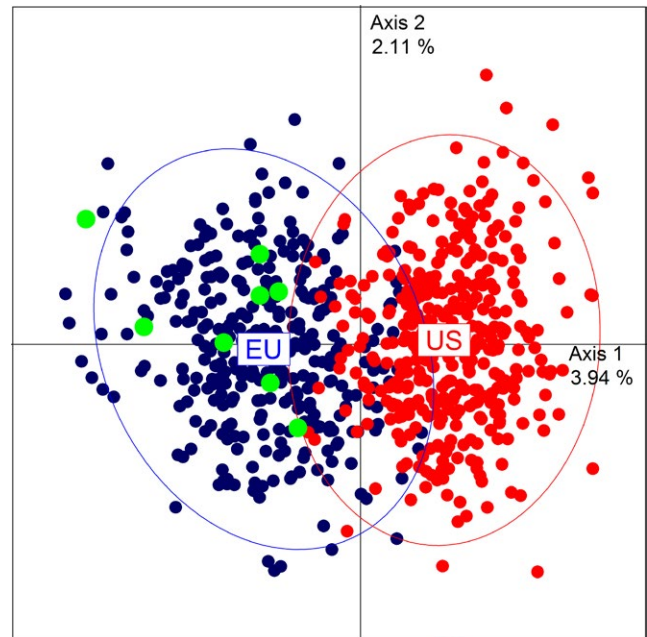
We determined if each individual was significantly assigned to one of the  $K$  clusters or if it was admixed by using individual membership coefficient confidence intervals for the  $K$  clusters. When the confidence intervals were not overlapping, an individual was considered as significantly assigned to the  $K$ th cluster for which it had the highest membership coefficient. Thus, for each population we were able to calculate the ratio of individuals assigned to each  $K$ th cluster or which were admixed. Finally, we assigned the population to the cluster for which the proportion of assigned individuals was the highest. In the case of equality, or if 100% of individuals were admixed, the population was declared as admixed.

To visualize within range genetic structure, we computed spatial interpolation using individual membership coefficients for the most likely  $K$  within each range. An inverse distance weighting (IDW) interpolation with a power value of 1 was carried out using the ArcGIS v10.2.2 geostatistical analyst tool (ESRI, 2011). The neighborhood was searched using a four sectors circle with a maximal value of 25 neighbors and a minimum of 0.

### 2.3.3 | Analysis of genetic differentiation and diversity

The genetic differentiation between populations (with values between 0 and 1, none-full differentiation) was analyzed using  $F_{ST}$  indexes (Wright, 1931). Within and between ranges,  $F_{ST}$  were calculated with the hierfstat v 0.04-28 R package (Goudet, 2005) according to the Weir and Cockerham method (1984). 95% confidence intervals (CI) were estimated by performing 1,000 bootstraps over loci.

Two datasets were analyzed in order to compare genetic diversity between ranges: the initial dataset (818 individuals minus 98 clones = 720 genotyped individuals using 113 SNPs) and the additional dataset (163 genotyped individuals using 251 SNPs, see Section 2). The second dataset specifically aimed to test for a potential bias in allelic frequency due to SNP selection. First, allelic frequencies were evaluated by plotting the MAF (Minor Allelic Frequency) distributions per locus and per range. Using the R package Hierfstat v 0.04-28 (Goudet, 2005) and Fstat software v2.9.3 (Goudet, 2013), diversity indices were calculated (a) between ranges, (b) among populations within ranges,



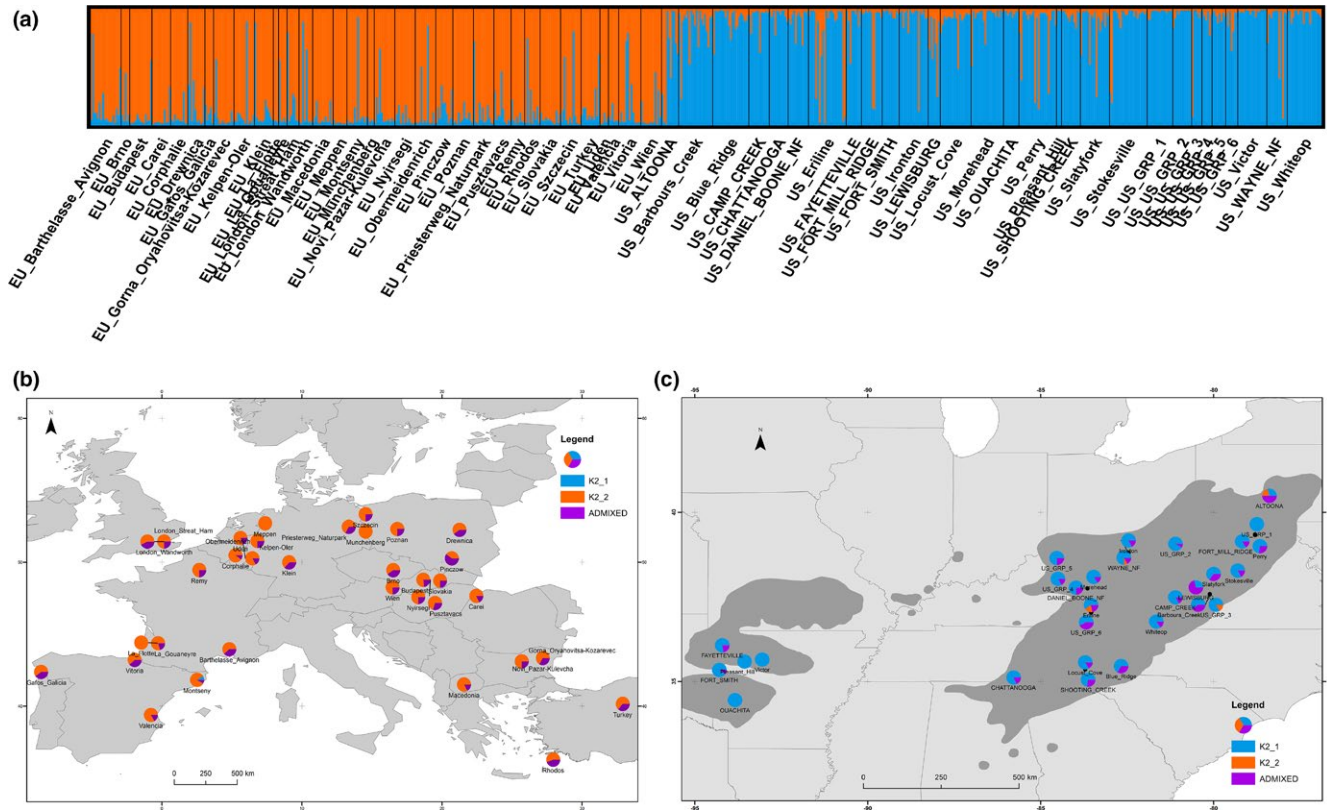
**FIGURE 1** Principal component analysis performed at the individual level. European individuals were plotted using blue dots, whereas American individuals were plotted with red and green dots. Green dots represented American individuals located within the European dot cloud beyond the limit of the ellipse related to American individuals. Ellipses were plotted to illustrate the identified genetic clusters. They encompassed roughly 95% of the individuals of each range

and (c) within populations and were as follows: observed heterozygosity ( $H_o$ ), which quantifies the proportion of heterozygous individuals; expected heterozygosity ( $H_e$ ), also known as genetic diversity, that measures the expected proportion of heterozygous individuals under Hardy–Weinberg Equilibrium; allelic richness (AR), which corresponds to the number of alleles weighted by the number of individuals in the smallest population; inbreeding coefficient  $F_{IS}$  within each population that measures the proportional deviation of observed from expected heterozygosity within subpopulation; and the total number of alleles (TNA) per range, calculated by summing within each range the number of allele over all loci.

Differences in AR were determined by performing a nonparametric Wilcoxon paired test among loci between ranges. In order to evaluate differences in total number of alleles between ranges, a bootstrap over all loci and individuals was computed using 1,000 simulations and the differences were determined using a nonparametric Mann–Whitney test using R version 3.3.1 (R Development Core Team, 2016).

### 2.3.4 | Isolation by distance analysis

In natural populations, genetic similarity is expected to be high between spatially close populations and then to decrease among populations with geographic distance; this pattern is known as isolation by distance (IBD). IBD was tested within each range, using two approaches: (a) The genetic distances between populations



**FIGURE 2** (a) Individual assignment for the most likely number of clusters (where  $K = 2$ ) as a result of the between range STRUCTURE analysis. Each colored vertical line represents one individual ancestry membership between the two clusters (orange, cluster K2\_1, and blue cluster K2\_2). Black vertical lines separate different populations. Both analyses were computed on the initial dataset after clone removal (720 individuals from 63 populations genotyped using 113 SNPs). (b and c) Pie charts of the population assignment in Europe and the USA for the most likely number of clusters (where  $K = 2$ ) as a result of the STRUCTURE analysis between ranges. In blue, proportion of individuals significantly assigned to cluster K2\_1; in orange, proportion of individuals significantly assigned to cluster K2\_2; and in Purple, proportion of individuals admixed in each population. The native distribution of black locust within America (Little, 1971) is plotted in gray shading and in Europe it is present almost everywhere from Southern to Northern Europe

were calculated as the ratio  $F_{ST}/(1-F_{ST})$  using pairwise  $F_{ST}$  plotted against the logarithm of the pairwise geographic distances among populations (Rousset, 1997) and the correlation was tested using a Pearson coefficient test; and (b) A Mantel test between the matrices of pairwise geographic distances and pairwise genetic distances was performed using the R ade4 library, with 9,999 permutations of matrices. For both methods, pairwise geographic distances among populations were calculated with the GPS coordinates of each population using the R CalcDists function provided by Scott Chamberlain on GitHub (<https://gist.github.com/sckott/931445>). The matrix of pairwise genetic distances was estimated using the Cavalli-Sforza and Edwards Chord distance with hierfstat v 0.04-28 (Takezaki & Nei, 1996).

### 3 | RESULTS

#### 3.1 | More asexual reproduction in European populations

Overall, a higher clonality was detected in the European populations compared to the American ones, with a significant range effect

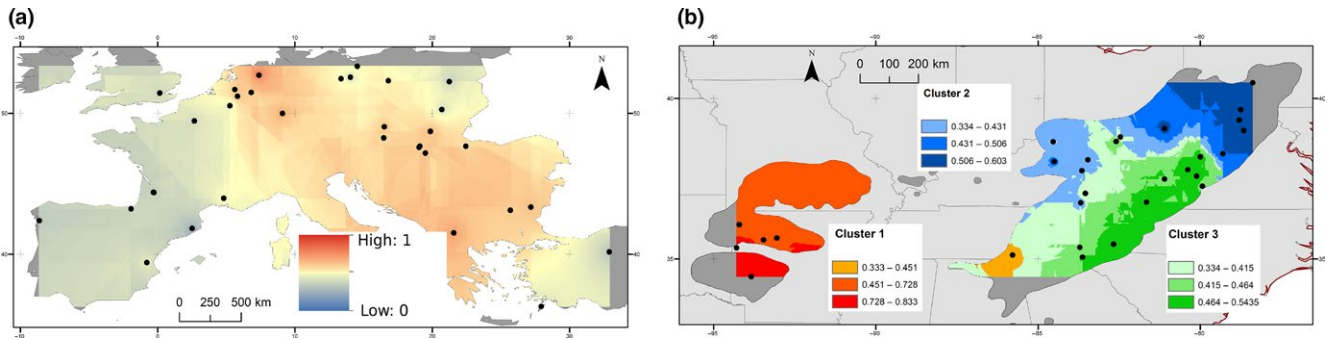
**TABLE 2** Isolation by distance correlation tests

Range	Pearson test		Mantel test	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
USA	<b>0.53</b>	<b><math>2.9 \times 10^{-30}</math></b>	<b>0.479</b>	<b><math>3 \times 10^{-4}</math></b>
Appalachians	<b>0.193</b>	<b><math>1.28 \times 10^{-3}</math></b>		
Ozarks	0.386	0.27		
Europe	<i>r</i>	0.75	-0.028	0.562
K2_1:	0.105	0.223		
K2_2:	0.0692	0.483		

Note. Both regression of pairwise  $F_{ST}/(1 - F_{ST})$  on logarithm of pairwise geographic distance (Rousset, 1997) and a Mantel test were performed within each range or within a subselection of the population in each range. Significant results are in bold.

( $\chi^2 = 29.04$ ,  $df = 1$ ,  $p = 7.10 \times 10^{-8}$ ). As expected, no clone was found within the common garden populations, nor in the populations obtained from seedlings germinated in the laboratory. When removing these populations from the analysis (thus leaving 280 European and 356 American individuals), 98 genotypes were found with a least one





**FIGURE 3** (a and b) The graphical IDW interpolation computed on the STRUCTURE for individual ancestry membership for each within range analysis. Results are shown for the most likely  $K$  in Europe and in the USA,  $K = 2$  and  $K = 3$ , respectively. IDW within America is plotted over native distribution of black locust. The two European clusters are represented by a continuous color scale from blue (K2\_1\_EU) to red (K2\_2\_Eu). The three clusters in the USA are represented by a continuous red color scale (K3\_1\_US) through to a blue color scale (K3\_2\_US) and green color scale (K3\_3\_US)

duplicated version (i.e., clones): 68 clones out of the 280 European samples and 30 clones out of the 356 American samples. Keeping only one sample per genotype resulted in a dataset that contained 720 genotypes out of the 818 sampled individuals after clone removal, which was distributed as 334 genotypes in 34 European populations and 386 genotypes in 29 American populations.

Among European populations (Table 1), the index of clonal diversity  $R$  ranged from 0.27 to 1 (mean = 0.82,  $SD = 0.26$ ), with significant differences in clonality between populations ( $\chi^2 = 88.33$ ,  $df = 60$ ,  $p = 0.01$ ). The most clonal populations ( $R < 0.5$ ) were Munchenberg Germany ( $R = 0.27$ ), Valencia Spain ( $R = 0.28$ ), Gafos Galicia Spain ( $R = 0.36$ ), Brno Czech republic ( $R = 0.4$ ), La Flotte France ( $R = 0.4$ ), and Uden the Netherlands ( $R = 0.4$ ). Among American populations, the index of clonal diversity  $R$  ranged from 0.67 to 1 (mean = 0.93,  $SD = 0.10$ ) and there was no significant effect of population on the clonality ( $\chi^2 = 76.77$ ,  $df = 60$ ,  $p = 0.07$ ). The lowest  $R$  value was observed for Pleasant Hill, in Arkansas ( $R = 0.67$ ).

The overall  $F_{IS}$  that was calculated among European populations of the "clonal dataset" (0.019, 95%CI: -0.018-0.062, estimated by bootstrapping over loci) was significantly lower than the overall  $F_{IS}$  calculated among American populations of the "clonal dataset" (0.11, 95%CI: 0.077-0.14, estimated by bootstrapping over loci). At the population level, negative  $F_{IS}$  values (Table 1) were observed in three European populations (Brno: -0.0256, Meppen: -0.0974, Munchenberg: -0.180) and in one American population (Pleasant Hill: -0.0582) but only Munchenberg was significant. All these populations exhibited high clonal reproduction ( $R$  ranging from 0.28 to 0.67). Significantly positive  $F_{IS}$  values were estimated among 22 of the 34 European populations (ranging from 0.066 to 0.161) and among 23 of the 29 American populations (ranging from 0.082 to 0.186).

### 3.2 | Introduced populations are genetically close to a few northeastern native populations

Introduced European individuals of black locust are genetically close to only a few native individuals located in four populations from

the northeastern part of its native range. Among the 29 sampled American populations, only four (Altoona, Eriline, US Grp 3, and Wayne National Forest) showed individuals with a high level of genetic relatedness to the European individuals. The structure and multivariate analyses gave congruent results.

First we used a PCA (Figure 1) to analyze the position of individuals on the factorial plan. Axis 1 roughly separated European (in blue) and American (in red and green) individuals in two partially overlapping clusters. A few American individuals (in green) were present inside the European dot cloud beyond the limit of the ellipse, whereas the contrary is not true. These American individuals belonged to four American populations (Altoona PA, Eriline KY, US Grp 3 VA, Wayne National Forest OH) located in the Northern Appalachian part of the native distribution.

Second, we used STRUCTURE to cluster the individuals according to molecular genetic similarity. An optimal number of  $K = 2$  clusters (Supporting Information Appendix S2A,B) was identified in STRUCTURE output using the initial dataset after clone removal (720 individuals from the 63 populations from both ranges). In the first cluster (Figure 2, in blue; K2\_1; Supporting Information Appendix S3), the majority of individuals came from America (76.8%), and only one European population (Montseny: 8.3%) contained samples that could be significantly attributed to this cluster using the individuals' ancestry confidence intervals (0.245% of the total number of European individuals).

The second cluster (Figure 2c in orange; K2\_2; Supporting Information Appendix S3) contained most of the European individuals (73.6% of the total number of European samples) and a few samples from North America (2.18%). The American samples that could be significantly assigned to cluster K2\_2 belonged to only four American populations (Supporting Information Appendix S3), and only a small proportion of individuals were significantly attributed to cluster K2\_2 (Altoona PA: 20%, Eriline KY: 18.2%, US Grp 3 VA: 16.7%, Wayne National Forest OH: 8.33%). These four populations were all located in the northern part of the Appalachian Mountains (Figure 2c). Finally, admixed individuals (i.e., those not attributed significantly to one cluster) were found in both ranges with relatively

similar proportion (Europe: 26.1%, America: 21.0%, Global: 23.8%). In America, the proportion of admixed individuals per population ranged from 0% (Ouachita, Pleasant Hill, US Grp 1, US Grp 3, Victor) to 71% (Lewisburg), whereas in Europe it ranged from 0% (Meppen, Munchenberg) to 60% (Pinczow).

### 3.3 | Significant genetic differentiation among populations in the native range, contrary to the introduced range

There was a significant genetic differentiation among all populations: Estimated  $F_{ST}$  among all populations was 5.23% (95% CI: 4.77%–5.70%). Overall, within the native range, black locust populations were clearly genetically differentiated, matching with geographic structure, whereas in the introduced European range, the differentiation between populations was low and no structure was detected across the continent.

Genetic differentiation among American populations was significant with an estimated  $F_{ST}$  of 4.46% (95% CI: 3.94%–5.06%). Clear signals of isolation by distance (IBD) were observed between populations in the native range, as both correlation tests were significant (Table 2; Supporting Information Appendix S4A). IBD remained significant among the Appalachian populations, but not among the Ozark populations. In the native range, the STRUCTURE analysis indicated that populations formed an optimal number of  $K = 3$  clusters (Supporting Information Appendix S2B). The first cluster (K3\_1\_US, Figure 3b, Supporting Information Appendix S5) corresponded to the Ozark western populations with high membership coefficients at the population level; the individuals from Ozark populations were mostly assigned to this cluster, with a minimum number of individuals for Fayetteville (22.2% of individuals assigned to K3\_1\_US and 77.8% of admixed individuals) and a maximum for Fort Smith (80.0% of individuals assigned to K3\_1\_US and 20% of admixed individuals). The mean admixture among populations assigned to this cluster was 51.2%.

The second cluster (Figure 3b, K3\_2\_US, Supporting Information Appendix S5) was located in the north of the Appalachian Mountains over Kentucky, Maryland, Ohio, Pennsylvania, Virginia, and West Virginia. The individuals assigned to this cluster were mainly found in North Appalachian populations (Altoona, Perry, US Grp 3 and US Grp 4) and ancestry membership gradually decreased toward the south and the east of the Appalachian Mountains. A proportion of individuals within the populations assigned to this cluster were significantly assigned, ranging from 4.5% (Stokesville) to 50% (Us Grp 4). Four populations also partially comprised individuals assigned to other clusters: 6.7% of US Grp 1 individuals and 16.7% of US Grp 4 individuals were assigned to cluster K3\_1\_US, and 5.6% of Locust Cove individuals were assigned to the cluster K3\_3\_US. Mean admixture among populations assigned to this second cluster was 77.5%.

The third cluster (Figure 1c, K3\_3\_US) was mainly formed in the eastern part of the Appalachians Mountains in North Carolina (Blue Ridge), Virginia (Barbours Creek, Whiteop), and in the southern part of West Virginia (Slatyfork). The other populations all comprised a

fraction of individuals significantly assigned to this cluster with values ranging from 4.5% (Eriline) to 30% (Barbours Creek). The mean admixture among populations assigned to this cluster—with the exception of the 2 fully admixed populations—was 83.9%.

Genetic differentiation among European populations was significant with an estimated  $F_{ST}$  of 3.08% (95% CI: 2.60%–3.60%), which was significantly lower than the  $F_{ST}$  observed in the native range. No significant signal of IBD was found either by testing the correlation between pairwise  $F_{ST}/(1-F_{ST})$  ratio and the logarithm of pairwise geographic distances (Supporting Information Appendix S4B; Pearson correlation test,  $r = -0.013$ ,  $p = 0.75$ ) or by realizing a Mantel test (Table 2;  $r = -0.028$ ,  $p = 0.562$ ). The correlation was not significant within each cluster either (Table 2; K2\_1: Pearson correlation test,  $r = 0.105$ ,  $p = 0.223$ , K2\_2: Pearson correlation test,  $r = 0.0692$ ,  $p = 0.483$ ).

In the introduced range, the STRUCTURE analysis indicated that populations formed an optimal number of  $K = 2$  clusters (Supporting Information Appendix S2C).

The first cluster (Figure 3a in red; K2\_1\_EU Supporting Information Appendix S5) was composed of 14 populations mainly from Central Europe. All populations in this cluster contained a fraction of significantly assigned individuals with values ranging from 8.33% (Gorna) to 62.5% (Meppen). Moreover, two populations (Macedonia, Nyirsegi) had 8.33% of their individuals significantly attributed to the second European cluster (K2\_2\_EU). Overall, the mean individual admixture in populations assigned to this cluster, with the exception of the fully admixed population, was 75.6%, to which 19% of individuals were significantly assigned, while 5.3% were assigned to the other cluster.

The second cluster (Figure 3a in blue; K2\_2\_EU Supporting Information Appendix S5) was mainly represented by 15 populations from Western and Eastern Europe. All populations contained a fraction of individuals which could be significantly assigned to this cluster, with values ranging from 9.1% (Turkey) to 100% (Munchenberg). Moreover, four populations had some individuals significantly attributed to the first European cluster (K2\_1\_EU): Vitoria (7.7%), Wien and Szczecin (8.33%), and Valencia (16.7%). Overall, mean individual admixture among populations assigned to this cluster was 70.9%.

### 3.4 | MAF distribution and detection of a bottleneck in the introduced populations

The MAF analysis performed on the initial and additional datasets highlighted a deficit in low frequency alleles when using 113 SNPs (MAF mode: 0.05–0.15) in both the native and invasive ranges, confirming that with this initial set we had oversampled high MAF SNPs, biasing the evaluation of diversity indexes (Supporting Information Appendix S6A). Consequently, the number of SNPs was increased to 251 SNPs by genotyping low MAF SNPs from our initial set of SNPs (Verdu et al., 2016) on a reduced number of samples.

When analyzing the 251 SNPs in 163 samples, no difference in heterozygosity was observed between either range. However, a

lower AR was observed in Europe compared to North America (1.91 vs. 1.96, Wilcoxon,  $v = 1,375.5$ ,  $p = 0.00453$ ), as well as a lower total number of alleles in Europe compared to North America (485 vs. 494; Mann–Whitney,  $W = 39,802$ ,  $p < 0.0001$ ) (Supporting Information Appendix S6B).

## 4 | DISCUSSION

Due to an extensive population genetics analysis in both native and European invasive ranges, we were able to show that black locust was likely to have been introduced to Europe from a limited part of its northeastern native distribution in the Appalachian Mountains. This founding effect brought about a bottleneck, detected only when we increased the number of SNPs with low MAF markers. A strong genetic structure was observed in the USA, whereas a much weaker one was detected in Europe. Moreover, asexual propagation was probably more prevalent in the invasive range than in the native one.

### 4.1 | Populations genetics and introduction history: European black locust populations close to Northern Appalachian populations

The genetic results suggest that the black locust was introduced to Europe from a restricted number of native populations located in the northeastern Appalachian Mountains. The robustness of this result is due to an intensive sampling effort on three levels: number of populations (29 in Europe and 34 in USA), number of individuals (334 in Europe and 386 in USA without clones, with a mean of 11.4 samples per population), and number of markers (113 SNPs). Furthermore, historical reviews and two types of analyses (PCA, Structure) led to the same conclusion concerning the limited origin of European populations.

The results obtained using a molecular approach are congruent with historical records pointing to the original sources of black locust in the northeastern part of its native range in the Appalachian Mountains. By reviewing historical studies, we were able to conclude that the first black locusts introduced to Europe during the early 17th century were likely to have come from Virginia and further black locusts introduced during the 17th and 18th centuries from Pennsylvania and West Virginia (Cobbett, 1825; Gunther, 1922; Michaux, 1813). Consequently, taking into account both the historical indications and the genetic proximity of all European black locust populations to a few native ones, it can be hypothesized that no subsequent introductions followed, and that the expansion of the species in Europe through asexual reproduction or seeds resulted from the original black locusts grown in Europe.

### 4.2 | Evidencing the bottleneck depends on the set of genotyped SNPs

Given that there are a few American populations close to European ones, a bottleneck is expected in European populations. The

decrease in genetic diversity was only observed when using a larger number of SNPs, due to some particular properties of SNPs.

SSRs and SNPs are two widely used markers for genotyping non-model species (Morin, Luikart, & Wayne 2004; Coates et al., 2009; Helyar et al., 2011). SNPs have many advantages: They can be easily developed using NGS; genotyping is easily replicable among laboratories; and SNPs are widely distributed throughout the genome (Coates et al., 2009; Morin et al., 2004; Helyar et al., 2011). However, more SNPs are needed than SSRs in order to reach the same level of precision, essentially because SSRs are multiallelic, whereas SNPs are mainly biallelic (Morin et al., 2004). One major problem in using SNPs is the ascertainment bias (Coates et al., 2009; Morin et al., 2004; Helyar et al., 2011). In particular, SNPs with a high minor allele frequency are more susceptible to being sampled for genotyping populations and consequently this can alter diversity estimates (Helyar et al., 2011). As a consequence, a genetic diversity analysis conducted with SNP data may lead to false negative or false positive conclusions (Helyar et al., 2011; Morin et al., 2004). Some empirical studies showed that SNPs performed better than SSRs for studying population structure, whereas SSRs were more efficient for estimating diversity (Singh et al., 2013). Nevertheless, other empirical studies reached the same conclusion regardless of type of marker used (van Inghelandt, Melchinger, Lebreton, & Stich 2010; Filippi et al., 2015).

Our study emphasizes the importance of taking the SNP ascertainment bias into account when comparing genetic diversity among several groups. Using the initial dataset, we observed more frequent minor alleles (modal class 0.05–0.15, Online Resource 3), which confirmed the sampling bias. We partially corrected this bias by using the additional dataset (251 SNPs), where actual MAF distribution is closer to the expected MAF distribution (modal class 0–0.05, Supporting Information Appendix S6A). By carrying out the analysis with this additional dataset, we were able to detect a bottleneck (decrease in allelic richness and total number of alleles in the introduced range, Supporting Information Appendix S6B), which would not have been the case if we had only used the initial dataset for studying genetic diversity between ranges (no difference in heterozygosity, allelic richness, and total number of alleles between ranges, Supporting Information Appendix S6B). Both a loss in allelic richness and in the total number of alleles point to the occurrence of a bottleneck, whereas heterozygosity is not expected to respond as well as allelic richness to a founding event (Dlugosch et al., 2015) as observed in our study.

Uller and Leimu (2011) demonstrated that genetic variation between native and invasive ranges was influenced by taxonomy: Invasive animals often suffered a loss of genetic diversity between the ranges, whereas invasive plants often exhibited higher genetic diversity in the invasive range (Uller & Leimu, 2011). According to these authors, one factor contributing to this pattern is that invasive animal populations are often founded by single introduction events, whereas multiple introductions associated with admixture are more common for plants (Uller & Leimu, 2011). Therefore, our results show that, similar to plants, black locust follows the original pattern of an introduction from a few

populations and a loss of genetic diversity. Many studies investigating the origins of herbaceous invasive plants have documented multiple introductions from wide areas of their native range. For the European weed *Ambrosia artemisiifolia*, multiple introductions from two distinct genetic clusters of the native American range have been evidenced (van Boheemen et al., 2017). Similarly, two clusters in Western and Eastern Europe have been identified as the sources of the American weed *Centaurea solstitialis* (Barker, Andonian, Swope, Luster, & Dlugosch, 2017). Thirdly, a wide scale study of the invasive weed *Mikania micrantha* in South-East Asia demonstrated the existence of two distinct genetics clusters that resulted from separate introductions originating from the native American range (Yang et al., 2017). Studies on invasive trees are less numerous, but they have generally concluded that multiple introductions occurred (Besnard et al., 2014; Merceron et al., 2017; Pairon et al., 2010; Thompson et al., 2015) without being able to clearly identify the population sources. A recent study on the genetic structure of invasive populations of *Acacia saligna* within several invasive ranges suggested that multiple introductions occurred from populations distributed throughout the Australian native range (Thompson et al., 2015).

### 4.3 | Structure in the native ranges was shaped by long-term evolutionary processes, whereas structure in the invasive range reflects anthropic action

Natural evolutionary processes seemed to have shaped the genetic diversity and structure of black locust populations in the native range.

Three genetic clusters were identified within the native range, with the greatest differentiation between the first cluster in the Ozark Mountains and the two clusters in the Northern and Southern Appalachian Mountains. Together with the pattern of isolation by distance, this suggests the action of natural and long-term evolutionary processes. On the contrary, in Europe, the weak structure in two clusters with a few outlying populations and no isolation by distance would suggest a recent evolutionary history marked by human actions.

The genetic structure detected in the American black locust is congruent with observations in other North American tree species. In North America, glacial refugia have been identified on both sides of the Mississippi River (Hewitt, 2000; Swenson & Howard, 2005) throughout geographic areas closely related to the genetic clusters identified in this study. A large differentiation on each side of the Mississippi River has been recorded for at least one other tree species, the loblolly pine (*Pinus taeda* L.), which exhibited two distinct genetic clusters (Lu et al., 2016). It is likely that the Mississippi River acted as a physical barrier during postglaciation recolonization after the last glacial maximum during Wisconsinan 21,000 years ago (Pessino, Chabot, Giordano, & DeWalt, 2014).

Moreover, similar to our findings, distinct genetic clusters have been identified along a north-south axis of the Appalachian mountains in several woody species, such as *Scirpus ancistrochaetus*

(Cipollini, Lavretsky, Cipollini, & Peters, 2017), *Tsuga caroliniana* (Potter, Campbell, Josserand, Nelson, & Jetton, 2017) and *Pinus strobus* (Nadeau et al., 2015). As described by Swenson and Howard (2005), a historical suture zone has been identified between the Northern Appalachian Mountains and Southern Appalachian Mountains. Consistent with this suture zone, two glacial refugia were detected for *P. strobus* in the Northern and Southern Appalachian Mountains. (Nadeau et al., 2015). It can therefore be assumed that Appalachian black locust genetic structure was driven by the same processes as for other North American trees and reflects postglacial colonization routes originating from glacial refugia on each side of the Appalachian Mountains.

On the contrary, in Europe the weak structure in two clusters with a few outlying populations and no isolation by distance would suggest a recent evolutionary history marked by human actions.

Within the European range, two distinct genetic clusters were detected throughout Central Europe (Austria, Bulgaria, Czech Republic, Central Germany, Hungary, Macedonia, the Netherlands, Slovakia, and Romania) and in Western and Eastern Europe (England, France, Eastern Germany, Poland, Spain and Turkey), but with a very weak signal (within Europe  $F_{ST} = 3.03\%$ ). Most of the individuals were admixed between these two clusters (approx. 75%) and within each cluster one outlying population drove the signal of differentiation: Meppen (62.5% of individuals assigned to K2\_1) and Munchenberg (100% of individuals assigned to K2\_2). Although weak, three signals confirmed that the observed genetic structure was significant: (a) Within each cluster, some individuals were significantly assigned to the cluster, (b) the STRUCTURE admixture parameter ( $\alpha$ ) was checked for convergence ( $\alpha$  should be relatively constant with a range of 0.2 or less -) as recommended in the STRUCTURE manual (Pritchard, 2010), and (c) when the two outlying populations were removed, similar genetic structure results were observed within Europe (data not shown).

As previously discussed, the black locust was most probably introduced to Europe from only a few populations located in a limited area of its native range in the northern part of the Appalachian Mountains. No support was found for the hypothesis that an introduction from two different American genetic clusters could have founded the two European clusters. Historical writings indicate that, in the 17th century, Vespasien Robin disseminated seeds collected from black locust trees grown in Paris throughout Europe, and that as from 1,634, the black locust was planted all over Europe for ornamental purposes. Seed orchards were created in Europe by trading European raised seeds (Cierjacks et al., 2013; Cobbett, 1825; François de Neufchateau, 1807) thus propagating the same genetic material throughout Europe. This can lead us to conclude that the origins of the central European cluster are related to human selection. As from the second half of the 18th century, extensive afforestation programs have been conducted in central Europe (in particular Germany, Hungary, and the Czech republic) promoting black locust for forestry purposes (Cierjacks et al., 2013; Vítková, Müllerová, Sádlo, Pergl, & Pyšek, 2017). Moreover, a genetic breeding program has been conducted since the beginning of the 20th



century in Hungary (Keresztesi, 1983). The genetic clustering within Europe may result from the evolution caused by artificial selection due to human-oriented selection and tree breeding, which was initiated in Central Europe in the 18th century. Thus, we can say that the European black locust is partially domesticated, and we can ask which traits influenced invasiveness. Further investigations involving common garden surveys would be necessary in order to assess whether genetic differences resulted in phenotypic differences.

#### 4.4 | The role of clonality in shaping genetic diversity in Europe

In general, we found that European populations of black locust were more clonal than American populations. This was demonstrated by comparing the number of clones detected within each range, as well as the higher index of clonal diversity  $R$ , and the analysis of the inbreeding coefficient  $F_{IS}$ . The same sampling protocol was followed by all field workers, who respected a minimum distance of 25 m between sampled individuals; it can therefore be concluded that the black locust is able to spread more than 25 m by clonal propagation. This result was confirmed when visualizing the mapping of individuals with GPS coordinates (data not shown, available upon request). A lower  $F_{IS}$  in European populations than in American populations indicated an excess of heterozygosity within the former. Clonality usually produces this pattern (Arnaud-Haond et al., 2007; Halkett, Simon, & Balloux, 2005; Stoeckel & Masson, 2014), as clonal reproduction can maintain heterozygosity over generations (Stoeckel et al., 2006). However, no significant relationship between the clonal diversity index (i.e.,  $R$ ) and  $F_{IS}$  was evidenced (data not shown).

Moreover, within the European range, the two outlying populations, Meppen and Munchenberg, were clearly differentiated no matter which of the analyses was applied. The results in a previous study on the Munchenberg population (code N° 7 – Hasenholz in Liesebach et al., 2004) showed that this population could clearly be differentiated from all the others. In addition, two major clones were detected which covered 80% of the stand (Liesebach et al., 2004). This is consistent with our finding since the Munchenberg population exhibited the lowest negative  $F_{IS}$  (−0.180), which indicated an excess of heterozygosity potentially due to a high level of clonality.

In Japan, both clonal and sexual reproduction have been found to promote the spread and invasion of black locust (Kurokochi & Hogetsu, 2014). Sexual regime is likely to influence invasiveness and a shift in the mating system has already been observed between ranges for several invasive species (Barrett, Colautti, & Eckert, 2008; Petanidou et al., 2012; Rambuda & Johnson, 2004). Clonal populations can maintain a high level of genetic diversity; however, they can be sensitive to founder events (Barrett, 2010). Clonality is likely to strongly decrease  $F_{ST}$  and to slowly decrease genotypic diversity in purely clonal populations (Balloux, Lehmann, & De Meeûs, 2003), but partially clonal populations are hard to differentiate from strictly sexual populations (Balloux et al., 2003). A theoretical study showed an advantage of clonal reproduction for species invasiveness. However, the relationship is not linear and species combining a high clonal rate with a small rate of sexual

reproduction would have a higher invasiveness (Bazin, Mathé-Hubert, Facon, Carlier, & Ravigné, 2013). Clonal reproduction provides invasive plant species with reproductive assurance (Barrett et al., 2008). Shifts in the mating system from outcrossing to clonality, due to a strong founder event, have been observed for the invasive *Eichhornia crassipes* (Barrett et al., 2008) and *Fallopia japonica* (Hollingsworth & Bailey, 2000). However, this is not systematic and some pure outcrossing species are successful invaders, such as *A. artemisiifolia* (Friedman & Barrett, 2008). It is also possible that a shift toward more clonal reproduction occurred in the mating system of the black locust between the native and the invasive range. This could have been produced by the founder event or by artificial selection, as one traditional way of managing black locust plantations in Europe was to stimulate clonal reproduction by damaging tree roots (François de Neufchateau, 1807; Saint-Jean de Crève & Coeur, 1786).

## 5 | CONCLUSION

We found a remarkable congruence between our genetic analysis and historical records regarding the geographic origins of the European black locust, with both approaches pointing to European populations originating from the Northern Appalachian Mountains. The history of black locust introduction is thus a unique pattern among invasive trees, which are commonly characterized by multiple introduction events. As a consequence, only a small part of black locust genetic diversity was introduced to Europe from its native American range.

Furthermore, in spite of the fact that black locust suffered a genetic bottleneck and a loss of diversity following introduction, this did not prevent its successful colonization throughout its European range. Moreover, we found some evidence for a shift in mating systems between ranges with an increase in clonality in Europe, resulting from either natural or artificial selection. However, our sampling method was not specifically designed to investigate changes in clonality during and after introduction. Therefore, further studies involving the sampling of extensive populations and plots would be needed to better understand the role of clonality in the success of this species, conjointly with studies on the role of sexual reproduction in the spread of the black locust.

## ACKNOWLEDGMENT

We would like to thank all those who helped us sample black locust populations: Fred Huber from George Washington & Jefferson National Forests and all Rangers from the USDA Forest Service for their valuable help sampling black locust leaves in the USA (CV field work); Dr Gideon Bradburd from the Michigan State University for his support in America (XPB field work); Dr Santiago Gonzalez-Martinez (BIOGECO, France) who helped us organize sampling throughout Spain and all the scientists who contributed to sampling wild black locust populations throughout Europe, that is, Dennis Born (Technische Universität, Germany), Marianne Correard (INRA – Avignon, France), Christopher Galanos (Hellenic Open University, Greece), Dr Pablo



Goikoetxea (Neiker, Spain), Dr Emilio Laguna (University of Valencia, Spain), Dr Roman Longaner and Antonin Martinik (Mendel University, Czech Republic), Dr Juan Pino (Universitat Autònoma de Barcelona, Spain), Alexandra Quenu and Sébastien Irola (SMABVC, France), Dr Volker Schneck (Thünen Institute of Forest Genetics, Germany), Dr Tomasz Wojda & Dr Martin Klisz (Forest Research Institute, Poland), Dr Rafael Zas (Misión Biológica de Galicia, Spain) and Dr Kevin Walker and the Botanical Society of Britain and Ireland who kindly helped us locate English black locust populations. Special thanks go to Régis Burtlett and Raphaël Segura for their help in sampling UK populations. We also thank Christophe Orazio (EFI, France), Nicolas Cheval, and Dr Patrick Patuszka (UE INRA, France) for helping us sample European provenances within the REINFFORCE project, and we are very grateful to Amy Zanne (George Washington University) for providing logistical help.

This study was financially supported by the Forest and Nature Management Research Unit of Gembloux Agro-BioTech, the BioGeCo INRA-Univ Bordeaux Research Unit, the Special Research Fund of the University of Liège, the *Société Française d'Ecologie*, the French *Agence de l'Eau*, the ANR-10-EQPX-16 Xyloforest, the EU COST Action FP1403 "Non-Native Tree Species for European Forests - Experiences, Risks and Opportunities," and the Transnational Access to Research Infrastructures activity in the 7th Framework Program of the EC under the Trees4Future project (no. 284181). Special thanks go to Philippe Lejeune for supporting the collaborative project.

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTION

AM, AJP, LL and SM conceived the experiment. XB, CFV, PB, CM, AM, MVL and SM sampled the populations. XB, CFV, AD, EG, MP and SM made the lab work. XB analyzed the data, with the contribution of CFV, EA, KD, ODT and SM. XB wrote the paper, with the contribution of AJP and SM. All authors read and approved the submitted version. XB and SM are co-corresponding authors.

## DATA ACCESSIBILITY

Data are accessible on Open Science Framework repository <https://osf.io/k97ax/> and they are publicly available.

## ORCID

Xavier Paul Bouteiller  <https://orcid.org/0000-0001-8621-383X>

## REFERENCES

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in

- population studies. *Molecular Ecology*, 16(24), 5115–5139. <https://doi.org/10.1111/j.1365-294X.2007.03535.x>
- Baker, H. G. (1967). Support for Baker's Law as a rule. *Evolution*, 21(4), 853–856. <https://doi.org/10.1111/j.1558-5646.1967.tb03440.x>
- Balloux, F., Lehmann, L., & De Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, 164(4), 1635–1644. <https://doi.org/10.1016/j.tree.2005.01.001>
- Barker, B. S., Andonian, K., Swope, S. M., Luster, D. G., & Dlugosch, K. M. (2017). Population genomic analyses reveal a history of range expansion and trait evolution across the native and invaded range of yellow starthistle (*Centaurea solstitialis*). *Molecular Ecology*, 26(4), 1131–1147. <https://doi.org/10.1111/mec.13998>
- Barrett, S. C. H. (2010). Why reproductive systems matter for the invasion biology of plants. In D. M. Richardson (Ed.), *Fifty years of invasion ecology: The legacy of Charles Elton* (pp. 195–210). Hoboken, NJ: Wiley-Blackwell.
- Barrett, S. C. H., Colautti, R. I., & Eckert, C. G. (2008). Plant reproductive systems and evolution during biological invasion. *Molecular Ecology*, 17(1), 373–383. <https://doi.org/10.1111/j.1365-294X.2007.03503.x>
- Basnou, C. (2009). *Robinia pseudoacacia* L., black locust (Fabaceae, Magnoliophyta), Species Accounts of 100 of the Most Invasive Alien Species in Europe (Chapter 13). In P. E. Hulme, W. Nentwig, P. Pyšek & M. Vilà (Eds), *Handbook of Alien Species in Europe* (pp. 357). Springer, Dordrecht: Springer Science.
- Bazin, É., Mathé-Hubert, H., Facon, B., Carlier, J., & Ravigné, V. (2013). The effect of mating system on invasiveness: Some genetic load may be advantageous when invading new environments. *Biological Invasions*, 16(4), 875–886. <https://doi.org/10.1007/s10530-013-0544-6>
- Besnard, G., Dupuy, J., Larter, M., Cuneo, P., Cooke, D., & Chikhi, L. (2014). History of the invasive African olive tree in Australia and Hawaii: Evidence for sequential bottlenecks and hybridization with the Mediterranean olive. *Evolutionary Applications*, 7(2), 195–211. <https://doi.org/10.1111/eva.12110>
- Bouteiller, X. P., Barraquand, F., Garnier-géré, P., Harmand, N., Laizet, Y., Raimbault, A., & Porté, A. J. (2018). No evidence for genetic differentiation in juvenile traits between Belgian and French populations of the invasive tree *Robinia pseudoacacia*. *Plant, Ecology and Evolution*, 151(1), 5–17. <https://doi.org/10.5091/pleveo.2018.1403>
- Bouteiller, X. P., Porté, A. J., Mariette, S., & Monty, A. (2017). Using automated sanding to homogeneously break seed dormancy in black locust (*Robinia pseudoacacia* L. Fabaceae). *Seed Science Research*, 27, 243–250. <https://doi.org/10.1017/S0960258517000150>
- Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18(1), 192. <https://doi.org/10.1186/s12859-017-1593-0>
- Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *Evolutionary Applications*, 8(1), 23–46. <https://doi.org/10.1111/eva.12234>
- Cierjacks, A., Kowarik, I., Joshi, J., Hempel, S., Ristow, M., Von Der Lippe, M., & Weber, E. (2013). Biological Flora of the British Isles: *Robinia pseudoacacia*. *Journal of Ecology*, 101(6), 1623–1640. <https://doi.org/10.1111/1365-2745.12162>
- Cipollini, K., Lavretsky, P., Cipollini, D., & Peters, J. L. (2017). Genetic population structure across the range of endangered Northeastern Bulrush, *Scirpus ancistrochaetus*. *International Journal of Plant Sciences*, 178(1), 67–78. <https://doi.org/10.1086/688960>
- Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., & Lewis, L. C. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, 100(5), 556–564. <http://doi.org/10.1093/jhered/esp028>
- Cobbett, W. (1825). *The woodlands*. London, UK: William Cobbett.

- Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24, 1999–2017. <https://doi.org/10.1111/mec.13162>
- Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular Ecology*, 24(9), 2212–2225. <https://doi.org/10.1111/mec.13117>
- DAISIE (2006). Delivering alien invasive species in Europe. Retrieved from February 26, 2015, from <http://www.europe-aliens.org>
- Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, 24, 2095–2111. <https://doi.org/10.1111/mec.13183>
- Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, 17(1), 431–449. <https://doi.org/10.1111/j.1365-294X.2007.03538.x>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Elton, C. S. (1958). *The ecology of invasions by animals and plants*, University ed. Chicago, IL: University of Chicago Press.
- ESRI (2011). *ArcGIS desktop: Release 10.2.2*. Redlands, CA: Environmental Systems Research Institute.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <http://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Ferrero, V., Barrett, S. C. H., Castro, S., Caldeirinha, P., Navarro, L., Loureiro, J., & Rodríguez-Echeverría, S. (2015). Invasion genetics of the Bermuda buttercup (*Oxalis pes-caprae*): Complex intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native and introduced populations. *Molecular Ecology*, 24(9), 2143–2155. <https://doi.org/10.1111/mec.13056>
- Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Filippi, C. V., Aguirre, N., Rivas, J. G., Zubrzycki, J., Puebla, A., Cordes, D., ... Lia, V. V. (2015). Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology*, 15, 52. <http://doi.org/10.1186/s12870-014-0360-x>
- François de Neufchateau, N. (1807). *Lettre sur le robinier*. Paris, France: Galland.
- Friedman, J., & Barrett, S. C. H. (2008). High outcrossing in the annual colonizing species *Ambrosia artemisiifolia* (Asteraceae). *Annals of Botany*, 101(9), 1303–1309. <https://doi.org/10.1093/aob/mcn039>
- Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., ... Vines, T. H. (2012). Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program structure. *Molecular Ecology*, 21(20), 4925–4930. <https://doi.org/10.1111/j.1365-294X.2012.05754.x>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. <https://doi.org/10.1111/j.1471-8278.2004.00828.x>
- Goudet, J. (2013). FSTAT: A computer program to calculate F-Statistics. *Journal of Heredity*, 104, 586–590. <https://doi.org/10.1093/jhered/est020>
- Gunther, R. T. (1922). *Early British botanists and their gardens*. Oxford, UK: Oxford University Press.
- Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, 20(4), 194–201. <https://doi.org/10.1016/j.tree.2005.01.001>
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., ... Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, 11(SUPPL. 1), 123–136. <http://doi.org/10.1111/j.1755-0998.2010.02943.x>
- Henry, P., Le Lay, G., Goudet, J., Guisan, A., Jahodová, S., & Besnard, G. (2009). Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Molecular Ecology*, 18(13), 2819–2831. <https://doi.org/10.1111/j.1365-294X.2009.04237.x>
- Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, 405(6789), 907–913. <https://doi.org/10.1038/35016000>
- Hirsch, H., Richardson, D. M., & Le Roux, J. J. (2017). Introduction to the special issue: Tree invasions: Towards a better understanding of their complex evolutionary dynamics. *AoB PLANTS*, 9(3), 1–8. <https://doi.org/10.1093/aobpla/plx014>
- Hollingsworth, M. L., & Bailey, J. P. (2000). Evidence for massive clonal growth in the invasive weed *Fallopia japonica* (Japanese Knotweed). *Botanical Journal of the Linnean Society*, 133(4), 463–472. <https://doi.org/10.1111/j.1095-8339.2000.tb01589.x>
- Hoos, P. M., Whitman Miller, A., Ruiz, G. M., Vrijenhoek, R. C., & Geller, J. B. (2010). Genetic and historical evidence disagree on likely sources of the Atlantic amethyst gem clam *Gemma gemma* (Totten, 1834) in California. *Diversity and Distributions*, 16(4), 582–592. <https://doi.org/10.1111/j.1472-4642.2010.00672.x>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Keller, S. R., Fields, P. D., Berardi, A. E., & Taylor, D. R. (2014). Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species. *Journal of Evolutionary Biology*, 27(3), 616–627. <https://doi.org/10.1111/jeb.12330>
- Keller, S. R., & Taylor, D. R. (2008). History, chance and adaptation during biological invasion: Separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, 11, 852–866. <https://doi.org/10.1111/j.1461-0248.2008.01188.x>
- Keller, S. R., & Taylor, D. R. (2010). Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology*, 23(8), 1720–1731. <https://doi.org/10.1111/j.1420-9101.2010.02037.x>
- Keresztesi, B. (1983). Breeding and cultivation of black locust, *Robinia pseudoacacia*, in Hungary. *Forest Ecology and Management*, 6, 217–244. [https://doi.org/10.1016/S0378-1127\(83\)80004-8](https://doi.org/10.1016/S0378-1127(83)80004-8)
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Kurokouchi, H., & Hogetsu, T. (2014). Fine-scale initiation of non-native *Robinia pseudoacacia* riparian forests along the Chikumagawa River in central Japan. *Journal of Ecology and Environment*, 37(1), 21–29. <https://doi.org/10.5141/ecoenv.2014.003>
- Lavergne, S., & Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, 104(10), 3883–3888. <https://doi.org/10.1073/pnas.0607324104>
- Lawson Handley, L. J., Estoup, A., Evans, D. M., Thomas, C. E., Lombaert, E., Facon, B., ... Roy, H. E. (2011). Ecological genetics of invasive alien species. *BioControl*, 56(4), 409–428. <https://doi.org/10.1007/s10526-011-9386-2>
- Liesebach, H., & Schneck, V. (2012). Chloroplast DNA variation in planted and natural regenerated stands of black locust (*Robinia pseudoacacia* L.). *Silvae Genetica*, 61(1–2), 27–35.

- Liesebach, H., Yang, M., & Schneck, V. (2004). Genetic diversity and differentiation in a black locust (*Robinia pseudoacacia* L.) progeny test. *Forest Genetics*, 11(2), 151–161.
- Little, E. L. (1971). Atlas of United States trees, vol. 1. Conifers and important hardwoods. U.S. Department of Agriculture, Miscellaneous Publication 1146. Washington, DC. 9 p., 313 maps.
- Lu, M., Krutovsky, K. V., Nelson, C. D., Koralewski, T. E., Byram, T. D., & Loopstra, C. A. (2016). Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.). *BMC Genomics*, 17(1), 1–11. <https://doi.org/10.1186/s12864-016-3081-8>
- Merceron, N. R., Leroy, T., Chancerel, E., Romero-Severson, J., Borkowski, D., Ducouso, A., ... Kremer, A. (2017). Back to America: Tracking the origin of European introduced populations of *Quercus rubra* L. *Genome*, 790, 1–54. <https://doi.org/10.1139/gen-2016-0187>
- Michaux, A. (1813). Histoire des arbres forestiers de l'Amérique septentrionale. Paris, France.
- Michener, D. C. (1988). The introduction of black locust (*Robinia pseudoacacia* L.) to Massachusetts. *Arnoldia*, 48(4), 52–57.
- Miller, P. (1756–1759). *The gardeners dictionary*, 7th ed. London, UK: Nabu Press.
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19(4), 208–216. <http://doi.org/10.1016/j.tree.2004.01.009>
- Nadeau, S., Godbout, J., Lamothe, M., Gros-Louis, M.-C., Isabel, N., & Ritland, K. (2015). Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: A comparison between eastern and western North American postglacial colonization histories. *American Journal of Botany*, 102(8), 1342–1355. <https://doi.org/10.3732/ajb.1500160>
- Pairon, M., Petitpierre, B., Campbell, M., Guisan, A., Broennimann, O., Baret, P. V., ... Besnard, G. (2010). Multiple introductions boosted genetic diversity in the invasive range of black cherry (*Prunus serotina*; Rosaceae). *Annals of Botany*, 105(6), 881–890. <https://doi.org/10.1093/aob/mcq065>
- Pappert, R. A., Hamrick, J. L., & Donovan, L. A. (2000). Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany*, 87(9), 1240–1245.
- Pessino, M., Chabot, E. T., Giordano, R., & DeWalt, R. E. (2014). Refugia and postglacial expansion of *Acronyria frisoni* Stark & Brown (Plecoptera:Perlidae) in North America. *Freshwater Science*, 33(1), 232–249. <https://doi.org/10.1086/675306>
- Petanidou, T., Godfree, R. C., Song, D. S., Kantsa, A., Dupont, Y. L., & Waser, N. M. (2012). Self-compatibility and plant invasiveness: Comparing species in native and invasive ranges. *Perspectives in Plant Ecology, Evolution and Systematics*, 14(1), 3–12. <https://doi.org/10.1016/j.ppees.2011.08.003>
- Porrás-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & ... M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, 4, 1–13. <https://doi.org/10.3389/fgene.2013.00098>
- Potter, K. M., Campbell, A. R., Josserand, S. A., Nelson, C. D., & Jetton, R. M. (2017). Population isolation results in unexpectedly high differentiation in Carolina hemlock (*Tsuga caroliniana*), an imperiled southern Appalachian endemic conifer. *Tree Genetics and Genomes*, 13(5), <https://doi.org/10.1007/s11295-017-1189-x>
- Pritchard, J. K. (2010). Documentation for structure software: Version 2.3, 6(3), 321–326. <https://doi.org/10.1002/spe.4380060305>
- Pritchard, J. K., Stephens, M., & Donnelly, P. J. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–949.
- R Development Core Team (2016). *R: A language and environment for statistical computing*. Version 3.3.1. Vienna Austria: R Foundation for Statistical Computing.
- Rambuda, T. D., & Johnson, S. D. (2004). Breeding systems of invasive alien plants in South Africa: Does Baker's rule apply? *Diversity and Distributions*, 10(5–6), 409–416. <https://doi.org/10.1111/j.1366-9516.2004.00100.x>
- Richardson, D. M., & Rejmánek, M. (2011). Trees and shrubs as invasive alien species – A global review. *Diversity and Distributions*, 17(5), 788–809. <https://doi.org/10.1111/j.1472-4642.2011.00782.x>
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from FStatistics under isolation by distance. *Genetics*, 145, 1219–1228.
- Saint-Jean de Crève, S.-J., & Coeur, J. H. (1786). Mémoire sur la culture et les usages du faux Acacia dans les États-Unis de l'Amérique septentrionale. Paris, France.
- Schütt, P. (1994). *Robinia pseudoacacia*. In *Enzyklopädie der Holzgewächse: Handbuch und Atlas der Dendrologie*. Weinheim, Germany: Wiley-VCH.
- Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., ... Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE*, 8(12), 1–14. <http://doi.org/10.1371/journal.pone.0084136>
- Stoekel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species – *Prunus avium* L. *Molecular Ecology*, 15(8), 2109–2118. <https://doi.org/10.1111/j.1365-294X.2006.02926.x>
- Stoekel, S., & Masson, J. P. (2014). The exact distributions of  $F_{IS}$  under partial asexuality in small finite populations with mutation. *PLoS ONE*, 9(1), 1–15. <https://doi.org/10.1371/journal.pone.0085228>
- Straker, K. C., Quinn, L. D., Voigt, T. B., Lee, D. K., & Kling, G. J. (2015). Black locust as a bioenergy feedstock: A review. *Bioenergy Research*, 8(3), 1117–1135. <https://doi.org/10.1007/s12155-015-9597-y>
- Swenson, N. G., & Howard, D. J. (2005). Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, 166(5), 581–591. <https://doi.org/10.1086/491688>
- Takezaki, N., & Nei, M. (1996). Genetic distances and the setting of conservation priorities. *Biological Conservation*, 75(3), 311. [https://doi.org/10.1016/0006-3207\(96\)85988-X](https://doi.org/10.1016/0006-3207(96)85988-X)
- Thompson, G. D., Bellstedt, D. U., Richardson, D. M., Wilson, J. R. U., & Le Roux, J. J. (2015). A tree well travelled: Global genetic structure of the invasive tree *Acacia saligna*. *Journal of Biogeography*, 42(2), 305–314. <https://doi.org/10.1111/jbi.12436>
- Uller, T., & Leimu, R. (2011). Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology*, 17(11), 3478–3485. <https://doi.org/10.1111/j.1365-2486.2011.02509.x>
- van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology*, 26(20), 5421–5434. <https://doi.org/10.1111/mec.14293>
- van Inghelandt, D., Melchinger, A. E., Lebreton, C., & Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics*, 120(7), 1289–1299. <https://doi.org/10.1007/s00122-009-1256-2>
- Verdu, C. F., Guichoux, E., Quevauvillers, S., De Thier, O., Laizet, Y., Delcamp, A., & Mariette, S. (2016). Dealing with paralogy in RADseq data: In silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecology and Evolution*, 6(20), 7323–7333. <https://doi.org/10.1002/ece3.2466>
- Vítková, M., Müllerová, J., Sádlo, J., Pergl, J., & Pyšek, P. (2017). Black locust (*Robinia pseudoacacia*) beloved and despised: A story of an

- invasive tree in Central Europe. *Forest Ecology and Management*, 384, 287–302. <https://doi.org/10.1016/j.foreco.2016.10.057>
- Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, 16(2), 97–159.
- Yang, M., He, Z., Huang, Y., Lu, L., Yan, Y., Hong, L., & Shi, S. (2017). The emergence of the hyperinvasive vine, *Mikania micrantha* (Asteraceae), via admixture and founder events inferred from population transcriptomics. *Molecular Ecology*, 26(13), 3405–3423. <https://doi.org/10.1111/mec.14124>
- Zenni, R. D., Bailey, J. K., & Simberloff, D. (2014). Rapid evolution and range expansion of an invasive plant are driven by provenance-environment interactions. *Ecology Letters*, 17(6), 727–735. <https://doi.org/10.1111/ele.12278>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Bouteiller XP, Verdu CF, Aikio E, et al. A few north Appalachian populations are the source of European black locust. *Ecol Evol*. 2019;9:2398–2414. <https://doi.org/10.1002/ece3.4776>