



**HAL**  
open science

# New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species

Lucie Meyer, Romain Causse, Fanny Pernin, Romain Scalone, Geraldine Bailly, Bruno Chauvel, Christophe Délye, Valérie Le Corre

## ► To cite this version:

Lucie Meyer, Romain Causse, Fanny Pernin, Romain Scalone, Geraldine Bailly, et al.. New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species. PLoS ONE, 2017, 12 (5), pp.1-20. 10.1371/journal.pone.0176197 . hal-01606817

**HAL Id: hal-01606817**

**<https://hal.science/hal-01606817>**

Submitted on 26 May 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

# New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species

Lucie Meyer<sup>1</sup>, Romain Causse<sup>1</sup>, Fanny Pernin<sup>1</sup>, Romain Scalone<sup>3</sup>, Géraldine Bailly<sup>2</sup>, Bruno Chauvel<sup>1</sup>, Christophe Délye<sup>1</sup>, Valérie Le Corre<sup>1\*</sup>

**1** Agroécologie, AgroSup Dijon, INRA, Univ. Bourgogne Franche-Comté, Dijon, France, **2** Sustainable Agriculture Department, BASF France SAS, Ecully, France, **3** Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

\* [valerie.le-corre@dijon.inra.fr](mailto:valerie.le-corre@dijon.inra.fr)



**OPEN ACCESS**

**Citation:** Meyer L, Causse R, Pernin F, Scalone R, Bailly G, Chauvel B, et al. (2017) New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species. PLoS ONE 12(5): e0176197. <https://doi.org/10.1371/journal.pone.0176197>

**Editor:** Giovanni G Vendramin, Consiglio Nazionale delle Ricerche, ITALY

**Received:** September 22, 2016

**Accepted:** April 6, 2017

**Published:** May 10, 2017

**Copyright:** © 2017 Meyer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Sequence data for all SSR markers developed are available from GenBank (accession numbers KX867678 - KX867785).

**Funding:** This work was funded by BASF France SAS. BASF France SAS provided support in the form of a research grant and of salaries for LM and GB. We also acknowledge support from EU COST Action FA1203 "Sustainable management of *Ambrosia artemisiifolia* in Europe (SMARTER)" and

## Abstract

*Ambrosia artemisiifolia* L., (common ragweed), is an annual invasive and highly troublesome plant species originating from North America that has become widespread across Europe. New sets of genomic and expressed sequence tag (EST) based simple sequence repeats (SSRs) markers were developed in this species using three approaches. After validation, 13 genomic SSRs and 13 EST-SSRs were retained and used to characterize the genetic diversity and population genetic structure of *Ambrosia artemisiifolia* populations from the native (North America) and invasive (Europe) ranges of the species. Analysing the mating system based on maternal families did not reveal any departure from complete allogamy and excess homozygosity was mostly due the presence of null alleles. High genetic diversity and patterns of genetic structure in Europe suggest two main introduction events followed by secondary colonization events. Cross-species transferability of the newly developed markers to other invasive species of the *Ambrosia* genus was assessed. Sixty-five per cent and 75% of markers, respectively, were transferable from *A. artemisiifolia* to *Ambrosia psilostachya* and *Ambrosia tenuifolia*. 40% were transferable to *Ambrosia trifida*, this latter species being seemingly more phylogenetically distantly related to *A. artemisiifolia* than the former two.

## Introduction

Inferring recent demographic history and contemporary evolutionary processes are major goals in the field of population genetics. Climate change, human disturbances of natural habitats and human-aided dispersal can cause dramatic shifts in the distributions of natural species, and biological invasions are increasingly prevalent worldwide. Analyzing the genetic diversity and population genetic structure of native and introduced populations of an invasive

from the Carl Trygger Foundation for Scientific Research (CTS). Part of the work was achieved thanks to the Short Term Scientific Mission COST-STSM-FA1203-21009 and the grant CTS 14.425 attributed to RS. The funders did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of the authors supported by funders are articulated in the 'author contributions' section.

**Competing interests:** The authors of the manuscript have the following interests: during the realization of the study, GB was a paid employee of BASF France SAS and LM received a research grant and salary from BASF France SAS. The other authors have declared that no competing interests exist. Commercial affiliation for some of the authors does not alter our adherence to PLOS ONE policies on sharing data and materials.

species allows recovering pathways of invasion and identifying founding events and/or admixture events among invasive populations. All these processes affect the demographic success and future expansion of the invasive species and determine its potential for adaptation to new environmental conditions. Their understanding is invaluable for devising appropriate management strategies [1,2].

The molecular markers used for population genetics studies are currently essentially of two kinds: genome-wide Single Nucleotide Polymorphisms (SNPs) identified by next-generation-sequencing-based techniques such as Restriction site Associated DNA sequencing (RAD-seq) or genotyping-by-sequencing [3], and Simple Sequence Repeats (SSR) markers (microsatellites). Some limitations of SSR markers are low density throughout the genome, complex mutational patterns and possible presence of homoplasy and null alleles [4]. However, SSR markers are easy to score, highly polymorphic and thus highly informative and the theory and practice of SSR marker analysis and their inherent bias are well known [5], making them still the markers of choice for ecological and evolutionary studies. In comparison to SNPs, SSRs are especially well suited for analyzing processes occurring at small temporal or spatial scales and have proven highly relevant for revealing recent expansion and recent admixture or analyzing parentage and kinship [5–7]. Next-generation sequencing technologies now allow to rapidly develop large sets of SSRs [8]. In addition, the ever-increasing availability of transcriptome sequences (Expressed Sequence Tags, EST) in public databases enables fast and cost-effective development of genic SSRs marker (EST-SSRs). EST-SSRs are expected to be less polymorphic than gSSRs but also to display fewer null alleles and be more transferable among related species [9,10]. As their polymorphism may be influenced by selective processes, EST-SSRs may reveal somewhat different genetic patterns than gSSRs [11].

The genus *Ambrosia* in the *Asteraceae* family includes at least 51 species collectively known as “ragweeds” and mainly distributed in North America [12]. Four different species (*A. artemisiifolia* L., *A. trifida* L., *A. psilostachya* D.C. and *A. tenuifolia* Spreng.) occur in Europe but are native from America [13,14]. *A. artemisiifolia* is an annual herb mostly known as a successful invasive and a highly allergenic plant causing severe rhinitis and asthma [15,16]. It has been introduced in Europe in the 19<sup>th</sup> century by the import of contaminated grain and forage [17]. *A. artemisiifolia* has colonized different types of habitats such as railways, riversides and wastelands, as well as cultivated fields where it is now a noxious weed competing with several summer crops [17]. To investigate the population genetic structure in *A. artemisiifolia*, reliable and polymorphic molecular markers are needed. To date, only a few gSSR markers have been developed from French *A. artemisiifolia* populations [18,19]. These few gSSRs were used to assess the population genetic structure and patterns of colonization across continental and regional scales in Europe [20–25], North America [26] and China [27]. In addition, most of the gSSR markers available showed PCR amplification failures and excess homozygote genotypes [20–26]. Excess homozygosity can be caused by the presence of null alleles resulting from mutations at primer binding sites that preclude PCR amplification. Alternatively, excess homozygosity has sometimes been interpreted as evidence for partial selfing in a mostly out-crossing species. This issue was highly debated in several SSR-based population genetics studies conducted on *A. artemisiifolia* [20–23].

The present study had three purposes: (a) develop new nuclear SSR markers for *A. artemisiifolia* following three different approaches (whole-genome enrichment followed by 454 sequencing, whole-genome Illumina sequencing, and use of existing EST databases), (b) investigate the genetic diversity, population structure and mating system of *A. artemisiifolia* using populations sampled in North America and Europe, and (c) assess marker transferability to *A. trifida*, *A. psilostachya* and *A. tenuifolia*.

## Materials and methods

### Plant material

A total of 321 *A. artemisiifolia* individuals were sampled from 11 populations spanning the invasive range in Europe and 5 populations in North America (Table 1, Fig 1). Twenty individuals were sampled from two populations of *A. trifida*, 22 individuals from one population of *A. psilostachya* and 21 individuals from one population of *A. tenuifolia*. A 0.2-cm<sup>2</sup> leaf section was collected on each individual and DNA extracted as described in [28]. All three species studied are alien invasive, not protected species. Sampling locations were not localized within protected areas so that no specific permission was required. *Ambrosia artemisiifolia* is described as a diploid species (2n = 36, [13, 29, 30]). As the presence of triploid plants has sometimes been questioned [26], we counted nuclear chromosomes as described [31] in 10 plants randomly chosen from one French population. Results were in agreement with diploidy with 2n = 36. (S1 Fig). *Ambrosia trifida* is a diploid species with a different basic chromosome number (2n = 24) [29, 30], while *A. psilostachya* and *A. tenuifolia* have the same basic chromosome number as *A. artemisiifolia* but variable ploidy levels [13, 14, 29, 32].

### Development of new nuclear SSR markers for *A. artemisiifolia*

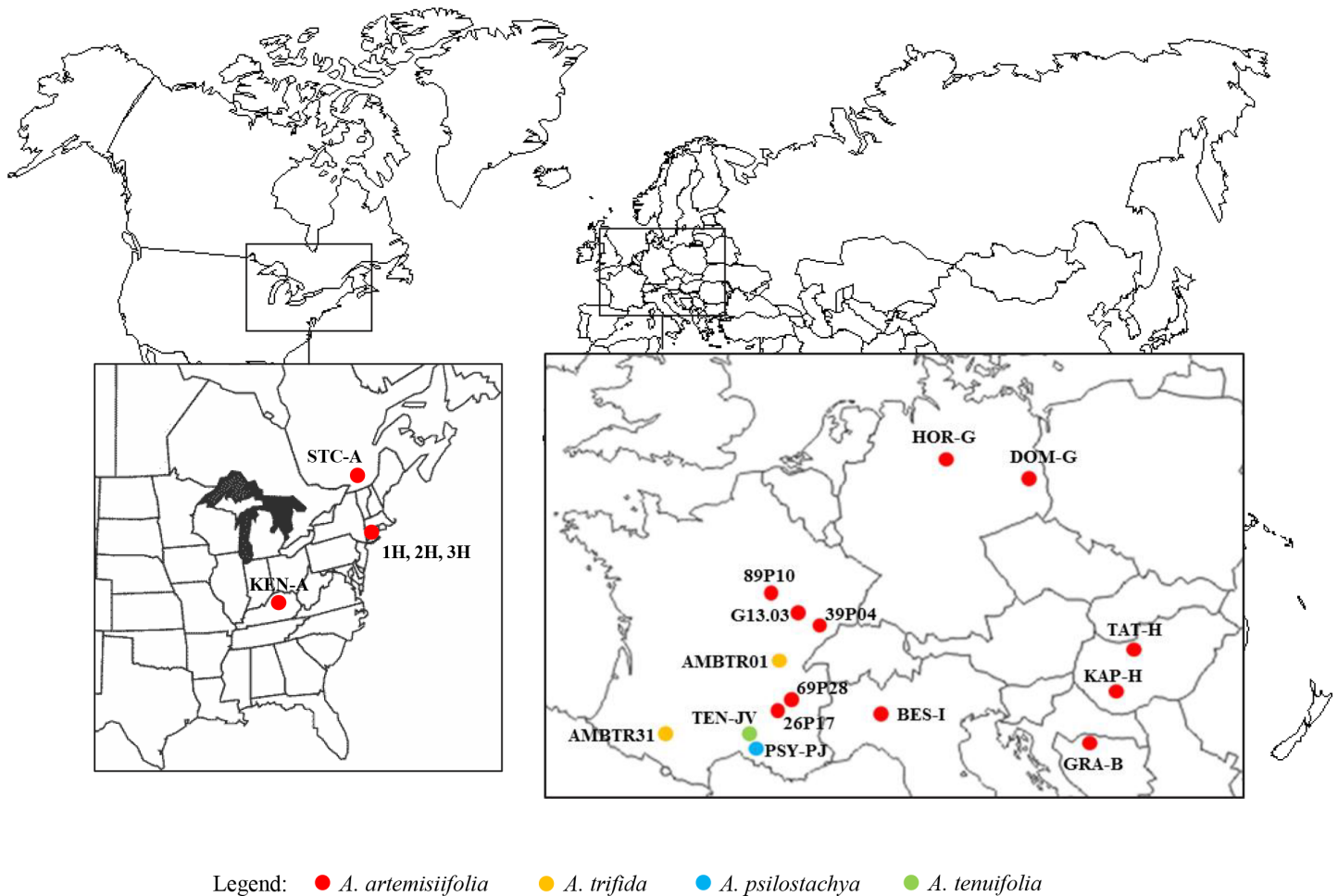
**Obtaining sequence data.** For the SSR-enriched gDNA library approach, total gDNA from 8 *A. artemisiifolia* individuals was isolated using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA) and processed by GenoScreen, (Lille, France). A SSR enriched DNA library was obtained as described in [33]. Briefly, total DNA was mechanically fragmented and enriched for AG, AC, AAC, AAG, AGG, ACG, ACAT and ATCT repeat motifs.

**Table 1. *Ambrosia* sp. populations analyzed.**

| Species                        | Population code | Country | Date of sampling | Nb of individuals analyzed | Geographic coordinates     |
|--------------------------------|-----------------|---------|------------------|----------------------------|----------------------------|
| <i>Ambrosia artemisiifolia</i> | 1H              | USA     | 2013             | 16                         | Not available <sup>a</sup> |
|                                | 2H              | USA     | 2013             | 11                         | Not available <sup>a</sup> |
|                                | 3H              | USA     | 2013             | 15                         | Not available <sup>a</sup> |
|                                | KEN-A           | USA     | 2010             | 24                         | N38°01'00", W84°33'10"     |
|                                | STC-A           | Canada  | 2010             | 20                         | N45°10'03", W73°40'50"     |
|                                | 26P17           | France  | 2005             | 24                         | N44°44'52", E04°55'07"     |
|                                | 39P04           | France  | 2011             | 24                         | N46°45'56"; E05°34'10"     |
|                                | 69P28           | France  | 2013             | 24                         | N45°44'49", E05°04'59"     |
|                                | 89P10           | France  | 2011             | 24                         | N48°10'40", E03°15'02"     |
|                                | GEN13.03        | France  | 2013             | 24                         | N47°11'29", E05°15'0"      |
|                                | BES-I           | Italy   | 2011             | 20                         | N45°18'25", E08°58'21"     |
|                                | HOR-G           | Germany | 2011             | 11                         | N52°17'28", E10°38'26"     |
|                                | DOM-G           | Germany | 2011             | 24                         | N51°38'21", E14°11'50"     |
|                                | TAT-H           | Hungary | 2009             | 20                         | N47°34'21", E18°27'18"     |
| KAP-H                          | Hungary         | 2011    | 20               | N46°22'12", E17°51'17"     |                            |
| GRA-B                          | Bosnia          | 2011    | 20               | N45°08'19", E17°15'51"     |                            |
| <i>Ambrosia trifida</i>        | AMBTR01         | France  | 2013             | 3                          | N46°24'35", E5°5'42"       |
|                                | AMBTR31         | France  | 2013             | 17                         | N43°15'48", E1°2'36"       |
| <i>Ambrosia psilostachya</i>   | PSI-PJ          | France  | 2014             | 22                         | N44°25'42" ; E04°42'44"    |
| <i>Ambrosia tenuifolia</i>     | TEN-JV          | France  | 2014             | 21                         | N43°49'54" ; E04°34'10"    |

<sup>a</sup>These populations were sampled in Connecticut, USA.

<https://doi.org/10.1371/journal.pone.0176197.t001>



**Fig 1. Map of the studied *Ambrosia* sp. populations.**

<https://doi.org/10.1371/journal.pone.0176197.g001>

Enriched fragments were subsequently amplified. Amplicons were sequenced on a 454 GS FLX Titanium system (454 Life Sciences, Branford, USA) following manufacturer’s protocols.

For the Illumina gDNA sequencing approach, total gDNA from one *A. artemisiifolia* individual was extracted using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). Sequencing was performed at GENTYANE (INRA, Clermont-Ferrand, France). An Illumina paired-end shotgun library was prepared by shearing 2 µg DNA using a Covaris M220 ultrasonicator (Covaris, Woburn, USA) and following the standard Illumina TruSeq DNA Library Kit protocol (Illumina, San Diego, USA). Sequencing was conducted on the Illumina MiSeq with 250 bp paired-end reads.

For the EST public data use approach, the two existing sets of *A. artemisiifolia* transcriptome 454 sequence data were downloaded from Genbank Sequence Read Archive [34]. They correspond to one individual sampled in the USA (accession SRX096892) and one sampled in Hungary (accession SRX098769). Both datasets were merged before analysis.

For each of the three sequence datasets, stringent sequence quality control and filtering were performed using the ShortRead package in the Bioconductor software [35]. Briefly, read ends were first trimmed by quality scores. Only sequences longer than 300 bp (454 reads) or 200 bp (Illumina reads) with a mean Phred quality score higher than 30% and less than 1% Ns



were retained. Exact sequence duplicates were discarded. In the Illumina dataset, only matching paired-end reads were kept after quality filtering and overlapping reads were merged using FLASH [36]. Detection of SSR motifs was conducted on the merged reads only, ensuring that the size of the flanking regions was large enough to design good-quality primers.

**SSR identification and primer design.** SSRs were identified with QDD version 3.1 [37]. Only 2- to 6-nucleotides motifs were considered. The minimum repeat unit was set to eight for di-nucleotides, six for tri-nucleotides, and five for longer motifs. Expected amplicon sizes were constrained to a 100–300 bp range. Primer pairs were thoroughly tested for clear, stable amplification on 12 *A. artemisiifolia* individuals from three populations (one French population from the Rhône Valley, the German population DOM and the American population KEN). PCRs were performed in 10- $\mu$ L as previously described [28]. Cycling parameters consisted in a first denaturation step (2 min at 95°C) followed by 39 cycles of 5 s at 95°C, 10 s at 60°C and 30 s at 72°C. Amplicons were visualised by electrophoresing five microliters of PCRs on 3% (wt/vol) agarose gels run for 25 min at 100V in Tris-Borate EDTA buffer.

## SSR marker validation and assessment of genetic polymorphism in *A. artemisiifolia*

**Genotyping.** SSRs successfully amplifying in *A. artemisiifolia* were used to genotype 384 individuals, including 321 *A. artemisiifolia* (16 populations), 20 *A. trifida* (two populations), 22 *A. psilostachya* (one population) and 21 *A. tenuifolia* (one population) individuals (Table 1). Genotyping was performed at GENTYANE (INRA, Clermont-Ferrand, France). PCR products were labelled with one fluorescent tag (6-FAM, NED, VIC or PET) and loaded on an ABI 3730XL capillary DNA analyzer (Applied Biosystem) with the size standard GS500 LIZ. Peak-canner version 1.0 (Applied Biosystems) and the R package MsatAllele were used to read allele sizes [38]. A Principal Component Analysis (PCA) was performed on genotype data using the package adegenet [39] in R 3.1.2 in order to examine the genetic relationship among the four species studied.

**Check for null alleles.** MicroChecker 2.2.0.3 was used to check for the presence of null alleles and scoring errors due to stuttering and large allele dropout for each marker in each *A. artemisiifolia* population [40]. The markers showing the overall lowest occurrence of null alleles and stuttering were retained for further analyses. Frequencies of null alleles at the retained loci in each population were estimated using INEST 2.1 [41].

**F<sub>ST</sub> outlier tests.** All SSR loci were screened for evidence of selection based on an F<sub>ST</sub> outlier test that identifies loci with an F<sub>ST</sub> value unexpectedly high (diversifying selection) or unexpectedly low (balancing or purifying selection). We used data from *A. artemisiifolia* and the software Bayescan [42]. This program implements a Bayesian method based on a multinomial Dirichlet distribution for allele frequencies. The Dirichlet distribution holds under a variety of demographic models when populations derive from a common gene pool. As a recent range expansion has been shown to increase the proportion of false selection event detection [43], we used a conservative prior value of 100 for the ‘odds of neutrality’ (only 1 locus out of 100 was under selection). For each locus, probability for selection was examined based on relative posterior probabilities for models with and without selection. We implemented 20 pilot runs of 5,000 iterations, a burn-in period of 50,000 iterations and 100,000 subsequent iterations with a sample size of 5,000 and thinning interval of 20.

## Estimation of the mating system in *A. artemisiifolia*

The mating system of *A. artemisiifolia* was investigated using five gSSR markers (SSR10, SSR17, SSR47, SSR71 and SSR73) in six additional French populations sampled in 2014 and

located within a few kilometres around population GEN13.03 (Table 1). These gSSR markers showed less null alleles than others. Leaf tissue and mature seeds were collected on six to eight mother-plants per population. Eight to 16 progeny-plants per mother-plant were genotyped, yielding a total of 614 individuals. MLTR [44] was used to estimate the multi-locus outcrossing rate  $tm$ , the maternal inbreeding coefficient  $F$ , the outcrossing rates between related individuals  $tm-ts$  and the correlation of paternity  $rp$ .

## Genetic diversity and inbreeding

The allelic richness per locus and per population using a rarefaction method ( $A$ ), expected heterozygosity ( $H_s$ ) and the genetic differentiation ( $F_{ST}$ ) were calculated using Fstat [45]. Significance of  $F_{ST}$  values was based on 1000 bootstrap resampling over loci. Inbreeding coefficient ( $F_{IS}$ ) were estimated with INEST 2.1 [41] using a Bayesian procedure robust to the presence of null alleles. To assess the statistical significance of inbreeding we compared the model with inbreeding with the random mating model ( $F_{IS} = 0$ ) based on the Deviance Information Criterion (DIC). Genetic diversity and differentiation parameters for *A. artemisiifolia* were calculated over all populations, over North American populations and over European populations.

### *A. artemisiifolia* population structure

Population structure was assessed using Structure 2.2 [46]. The admixture model and correlated allele frequencies between populations were selected as specified [47] to determine the number of genetic clusters ( $K$ ) best fitting the data. The length of the burn-in period was 100,000 runs followed by 500,000 Markov Chain Monte Carlo. Ten iterations were performed for each value of  $K$  from 1 to 15.  $K$  was determined graphically based on log likelihood values as previously described [46] using the web-based program Structure Harvester [48]. In addition, the  $\Delta K$  method [49] was used to determine the best value of  $K$ . Finally, Clumpp 1.1.2 [50] and R 3.1.2 were used to produce graphical outputs for the inferred population structure.

## Genetic divergence and bottleneck tests

Genetic divergence is likely to vary across populations because of differences in population effective sizes and local migration rates. This is especially the case when recent founder effects have occurred, such as during range expansions. Patterns of genetic divergence were estimated for the invasive range (Europe) by calculating population-specific  $F_{ST}$  values based on the  $F$ -model [51]. We used the Bayesian method of Foll and Gaggiotti [52] implemented in the software GESTE v2. To assess geographical patterns in genetic divergence, we compared three models: one null model that simply estimate population-specific  $F_{ST}$  values, and two models that used either latitude or longitude as explanatory variables. In addition to the study of genetic divergence patterns, we investigated the signature of recent bottleneck events using the Wilcoxon test for excess expected heterozygosity implemented in INEST2.1 and based on the method of Cornuet and Luikart [53]. Analyses were run with the Two-Phase Mutation (TPM) model with default settings.

## Results

### Development of new nuclear SSR markers

Sequencing results, filtering and success rates of microsatellite loci development are summarized in Table 2. Most 454 reads were eliminated because of insufficient length, while most Illumina reads were eliminated because paired-end reads could not be merged. As expected, the proportion of quality reads containing a SSR motif was much higher among the 454 reads

**Table 2. Characteristics of sequence datasets used for development of new SSR markers.**

| Sequence dataset     | Total number of reads | Number of good quality reads (%) | PALs <sup>a</sup> | Number of tested loci | Number of loci successfully amplified (%) |
|----------------------|-----------------------|----------------------------------|-------------------|-----------------------|---|
| Enriched Genomic—454 | 556,018               | 18,218 (3.27%)                   | 270               | 96                    | 30 (31.2%)                                |
| Genomic—Illumina     | 12,000,000            | 923,229 (7.69%)                  | 2,720             | 110                   | 37 (33.6%)                                |
| EST—454              | 1,317,778             | 393,302 (29.85%)                 | 397               | 173                   | 41 (23.7%)                                |

<sup>a</sup> Potentially Amplifiable Loci (PALs).

<https://doi.org/10.1371/journal.pone.0176197.t002>

obtained from an enriched gDNA library (24%) than among the Illumina reads obtained from raw gDNA (0.3%) or among the transcriptome 454 reads (0.1%). The low rate of SSR motifs obtained from Illumina sequencing of raw gDNA was more than compensated for by the high amount of reads generated and this method allowed the identification of ten times more potentially amplifiable loci than 454 sequencing of enriched gDNA (Table 2). The distribution of motif length was very similar between the two methods used to develop gSSRs: on average the di-, tri-, tetra-, penta- and hexa-nucleotides accounted for 40.2%, 48.7%, 7.5%, 2.5% and 1% of gSSRs, respectively. By contrast, most EST-SSRs were tri-nucleotides (81.6%), and di-, tetra-, penta- and hexa-nucleotides accounted for 11.1%, 7%, 0.2% and 0% of EST-SSRs, respectively. Success rate of PCR amplification were quite similar among SSR sets, yielding 67 gSSRs (GenBank accession number KX867678—KX867743) and 41 EST-SSRs (GenBank accession number KX867744-KX867785) with consistent amplification in *A. artemisiifolia* (Table 2). Among these, 46 gSSRs and 32 EST-SSRs gave clear, easy to score patterns after capillary electrophoresis (S1–S3 Tables). Homology with known proteins was detected for 25 EST-SSRs (S3 Table).

### Genetic polymorphism at gSSRs and EST-SSRs loci in *A. artemisiifolia*

Genetic polymorphism was assessed using 16 *A. artemisiifolia* populations for a total of 321 individuals (Table 1). After checking for null alleles and stutters at each locus in each population, 14 gSSRs and 13 EST-SSRs were retained as best markers for population genetic analysis. Among the 27 loci tested, only one (SSR86, S1 Table) was unambiguously detected as being under selection (Bayesian probability = 1, S2 Fig). SSR86 showed less genetic differentiation ( $F_{ST} = 0.014$ ) than other markers, but a very high within-population genetic diversity ( $H_s = 0.776$ ), suggesting balancing selection at or near the locus. This locus was therefore discarded for further analyses.

All 26 retained loci were polymorphic and revealed high levels of genetic diversity (Table 3). The frequencies of null alleles estimated over all populations ranged between 0.06 and 0.19 and were on average similar between gSSRs and EST-SSRs (0.11; Table 3). Allelic richness per locus and population, calculated based on a minimum sample size of eight individuals (Fig 2), was slightly lower for EST-SSRs than for gSSRs (4.438 versus 4.748) but the difference was not significant (Wilcoxon test p-value = 0.778) (Fig 2). The mean expected heterozygosity within populations was high (0.635 for gSSRs and 0.625 for EST-SSRs) and not significantly different between the two types of markers (Wilcoxon test p-value = 0.778). The mean genetic differentiation  $F_{ST}$  was 0.072 for gSSRs and 0.058 for EST-SSRs (difference significant, Wilcoxon test p-value = 0.045).

### Insight into the mating system in *A. artemisiifolia*

Thirty-six maternal progenies sampled from six French populations were analysed with five gSSRs to estimate mating system parameters. In addition, direct evidence for the presence of



**Table 3. Genetic diversity and differentiation estimated at 13 gSSR and 13 EST-SSR loci in 16 populations of *A. artemisiifolia*.** Number of alleles ( $N_a$ ), mean observed ( $H_o$ ) and expected ( $H_s$ ) heterozygosity, genetic differentiation ( $F_{ST}$ ) and average across-populations frequency of null alleles ( $P_{null}$ ) are indicated for each marker.

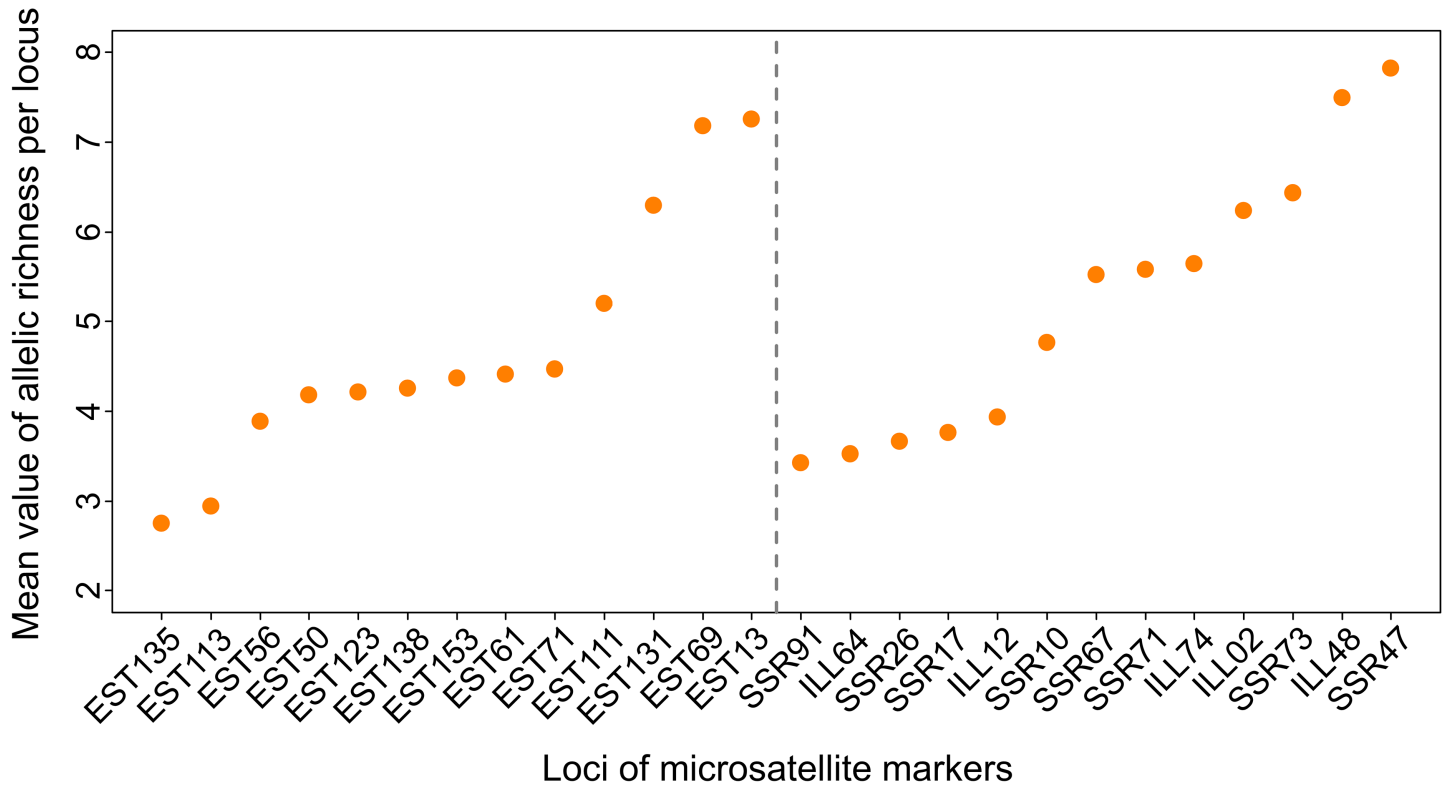
| Locus  | Repeated motif        | Size range | $N_a$ | $H_o$ | $H_s$ | $F_{ST}$ | $P_{null}$ |
|--------|-----------------------|------------|-------|-------|-------|----------|------------|
| ILL02  | (ACCACT) <sub>6</sub> | 265–297    | 12    | 0.495 | 0.605 | 0.054    | 0.108      |
| ILL12  | (AACAG) <sub>5</sub>  | 114–148    | 10    | 0.309 | 0.432 | 0.095    | 0.163      |
| ILL48  | (AGC) <sub>10</sub>   | 116–148    | 14    | 0.523 | 0.702 | 0.053    | 0.087      |
| ILL64  | (ACC) <sub>7</sub>    | 279–304    | 8     | 0.501 | 0.538 | 0.104    | 0.073      |
| ILL74  | (AGC) <sub>7</sub>    | 281–308    | 11    | 0.662 | 0.740 | 0.067    | 0.081      |
| SSR10  | (GATA) <sub>6</sub>   | 160–203    | 10    | 0.378 | 0.625 | 0.071    | 0.194      |
| SSR17  | (CATA) <sub>5</sub>   | 145–196    | 9     | 0.320 | 0.485 | 0.105    | 0.135      |
| SSR26  | (GAA) <sub>9</sub>    | 106–120    | 6     | 0.566 | 0.588 | 0.074    | 0.072      |
| SSR47  | (AG) <sub>9</sub>     | 96–136     | 20    | 0.488 | 0.762 | 0.066    | 0.127      |
| SSR67  | (GAA) <sub>7</sub>    | 183–230    | 14    | 0.580 | 0.707 | 0.091    | 0.105      |
| SSR71  | (TCC) <sub>7</sub>    | 127–161    | 12    | 0.715 | 0.706 | 0.065    | 0.071      |
| SSR73  | (AC) <sub>8</sub>     | 185–206    | 13    | 0.640 | 0.772 | 0.044    | 0.091      |
| SSR91  | (TC) <sub>8</sub>     | 108–118    | 6     | 0.537 | 0.592 | 0.072    | 0.159      |
| EST13  | (AGT) <sub>9</sub>    | 173–225    | 18    | 0.594 | 0.814 | 0.050    | 0.089      |
| EST50  | (AAAG) <sub>5</sub>   | 204–224    | 8     | 0.451 | 0.634 | 0.052    | 0.160      |
| EST56  | (AGC) <sub>7</sub>    | 208–223    | 5     | 0.647 | 0.688 | 0.052    | 0.070      |
| EST61  | (AAT) <sub>7</sub>    | 144–180    | 13    | 0.471 | 0.653 | 0.048    | 0.105      |
| EST69  | (AAT) <sub>7</sub>    | 116–138    | 12    | 0.614 | 0.761 | 0.070    | 0.090      |
| EST71  | (AAT) <sub>7</sub>    | 141–165    | 13    | 0.394 | 0.545 | 0.054    | 0.098      |
| EST111 | (ACC) <sub>7</sub>    | 134–159    | 9     | 0.491 | 0.625 | 0.099    | 0.087      |
| EST113 | (ACC) <sub>7</sub>    | 151–167    | 6     | 0.427 | 0.516 | 0.056    | 0.137      |
| EST123 | (ACC) <sub>7</sub>    | 114–142    | 8     | 0.374 | 0.589 | 0.064    | 0.165      |
| EST131 | (ACC) <sub>7</sub>    | 117–131    | 13    | 0.456 | 0.705 | 0.051    | 0.114      |
| EST135 | (ACG) <sub>7</sub>    | 132–148    | 6     | 0.419 | 0.456 | 0.043    | 0.073      |
| EST138 | (ACC) <sub>7</sub>    | 130–168    | 12    | 0.283 | 0.498 | 0.065    | 0.140      |
| EST153 | (AGG) <sub>6</sub>    | 172–193    | 8     | 0.641 | 0.645 | 0.052    | 0.058      |

<https://doi.org/10.1371/journal.pone.0176197.t003>

null alleles was sought by considering the progenies from maternal plants apparently homozygous at one locus. If a null allele was present, the maternal plant would actually be heterozygous (i.e., carrying one null allele and one detectable allele). Its progeny would thus contain some plants apparently homozygous for alleles different from the maternal one, but actually carrying one maternal null allele and one paternal detectable allele. Evidence for the presence of null alleles was obtained for all five markers. Depending on the marker considered, from 25% (3 out of 12) to 35% (8 out of 23) of the progenies from plants scored as homozygotes contained non-matching genotypes. Mating system parameters were estimated after excluding these progenies. The maternal inbreeding coefficient was non-significant ( $F = 0$ ). Multi-locus outcrossing rates ( $t_m$ ) were high and not significantly lower than 1 in all populations (Table 4). The rates of mating between related individuals ( $t_m$ - $t_s$ ) were not significant. Paternity correlations ( $r_p$ ) were weak and only significant for two populations. These results suggested complete outcrossing for *A. artemisiifolia* and large numbers of pollen donor parents.

### Patterns of genetic diversity and inbreeding in *A. artemisiifolia* populations

We compared patterns of genetic diversity between the native range (North America) and the invasive range (Europe). Allelic richness within population and mean expected heterozygosity



**Fig 2. Allelic richness for the 13 EST-SSRs and 13 gSSRs analyzed in *A. artemisiifolia*.** Markers are plotted by increasing mean value.

<https://doi.org/10.1371/journal.pone.0176197.g002>

within population were slightly higher in North America than in Europe (Table 5). However, for both parameters, the difference between the two ranges was not significant (Wilcoxon test p-values: 0.100 and 0.173 for allelic richness and expected heterozygosity, respectively). The inbreeding coefficient estimated taking null alleles into account was significantly higher than zero in only seven populations: four of the five North-American populations and three of the eleven European populations (S4 Table). Consequently,  $F_{IS}$  values were on average higher in the native range than in the invasive range (Table 5), although the difference was not significant (Wilcoxon test p-values: 0.2951).  $F_{ST}$  in the native range was low (0.042) but significant.  $F_{ST}$  in the invasive range was higher (0.071) and also significant. The difference in  $F_{ST}$  values between the two areas was significant based on 99% bootstrap confidence intervals.

**Table 4. Estimates of mating system parameters for six French *A. artemisiifolia* populations based on five SSR markers (SSR10, SSR17, SSR47, SSR71 and SSR73).** Multi-locus outcrossing ( $t_m$ ) and single-locus outcrossing ( $t_s$ ) rates, outcrossing rate between related individuals ( $t_m-t_s$ ), maternal inbreeding coefficient ( $F$ ) and correlation of paternity ( $r_p$ ) are indicated for each marker.

| Samples | $t_m$         | $t_s$         | $t_m-t_s$      | $F$            | $r_p$                 |
|---------|---------------|---------------|----------------|----------------|-----------------------|
| GEN02   | 1.025 (0.080) | 1.065 (0.083) | -0.040 (0.059) | -0.200 (0.058) | <b>0.151* (0.049)</b> |
| GEN05   | 0.882 (0.114) | 0.789 (0.115) | 0.093 (0.037)  | -0.037 (0.104) | 0.001 (0.179)         |
| GEN07   | 0.941 (0.077) | 0.895 (0.077) | 0.046 (0.046)  | -0.072 (0.152) | 0.106 (0.061)         |
| GEN10   | 0.963 (0.056) | 0.906 (0.085) | 0.057 (0.046)  | -0.200 (0.046) | 0.063 (0.048)         |
| GEN11   | 0.985 (0.081) | 0.933 (0.133) | 0.052 (0.073)  | -0.125 (0.130) | <b>0.175* (0.086)</b> |
| GEN17   | 0.762 (0.161) | 0.693 (0.118) | 0.069 (0.062)  | -0.110 (0.047) | 0.118 (0.074)         |

The values in brackets are S.D.

\* indicates significant values.

<https://doi.org/10.1371/journal.pone.0176197.t004>

### Population genetic structure of *A. artemisiifolia*

The posterior likelihood generated by Structure increased continuously with  $K$ , whereas the Evanno method (49) suggested that the most likely number of genetic clusters was six (S3 Fig). From  $K = 2$ , a west-east gradient of genetic variation was observed across Europe, showing that Central and Eastern European populations are different from Western European populations, and similar to the Northern American populations studied. At  $K = 6$ , a more detailed genetic structuring was revealed mainly within Europe (Fig 3). Most of the additional clusters were very specific to one or two populations (cluster 1: HOR-G, cluster 2: 89-P10, cluster 3: TAT-H, cluster 6: BES-I and DOM-G, Figs 3 and 4). The two main genetic clusters (clusters 4 and 5, Fig 4) were frequent in the western and south-eastern part of the invasive range, respectively, but only the first one (cluster 4) was observed at high frequencies in the native range.

Structure analyses were also performed separately using gSSR data only or EST-SSR data only. The most likely numbers of genetic clusters among the 16 populations were four (S4 Fig) and three (S5 Fig) for EST-SSRs and gSSRs, respectively. Overall, the same patterns were observed for both datasets, i.e., variation in cluster membership probabilities among populations and a west-east gradient of genetic variation across Europe.

### Patterns of local genetic divergence and bottlenecks in the invasive range

Population-specific  $F_{ST}$  values were best explained by the model that included latitude as a linear explanatory factor (posterior probability: 0.66) in comparison to the null model with no explanatory factor (posterior probability: 0.28) or the model including longitude (posterior probability: 0.04). Population-specific  $F_{ST}$  values increased with latitude (Fig 5A). While there was no linear relationship with longitude, a non-linear pattern was revealed with populations from Central Europe (Italy: BES-I and Germany: HOR-G, DOM-G) and two populations located in the western (89-P10) and eastern (TAT-H) parts of the range showing elevated  $F_{ST}$  values (Fig 5B). Noticeably, these populations were those harbouring specific genetic clusters under the most detailed Structure models (Figs 3 and 4). Increased genetic divergence may be due to recent founder events for these populations. However, no significant signatures of recent bottlenecks were detected based on the Wilcoxon test for expected heterozygosity excess in any of the studied populations.

### Transferability of SSRs and relationships among species

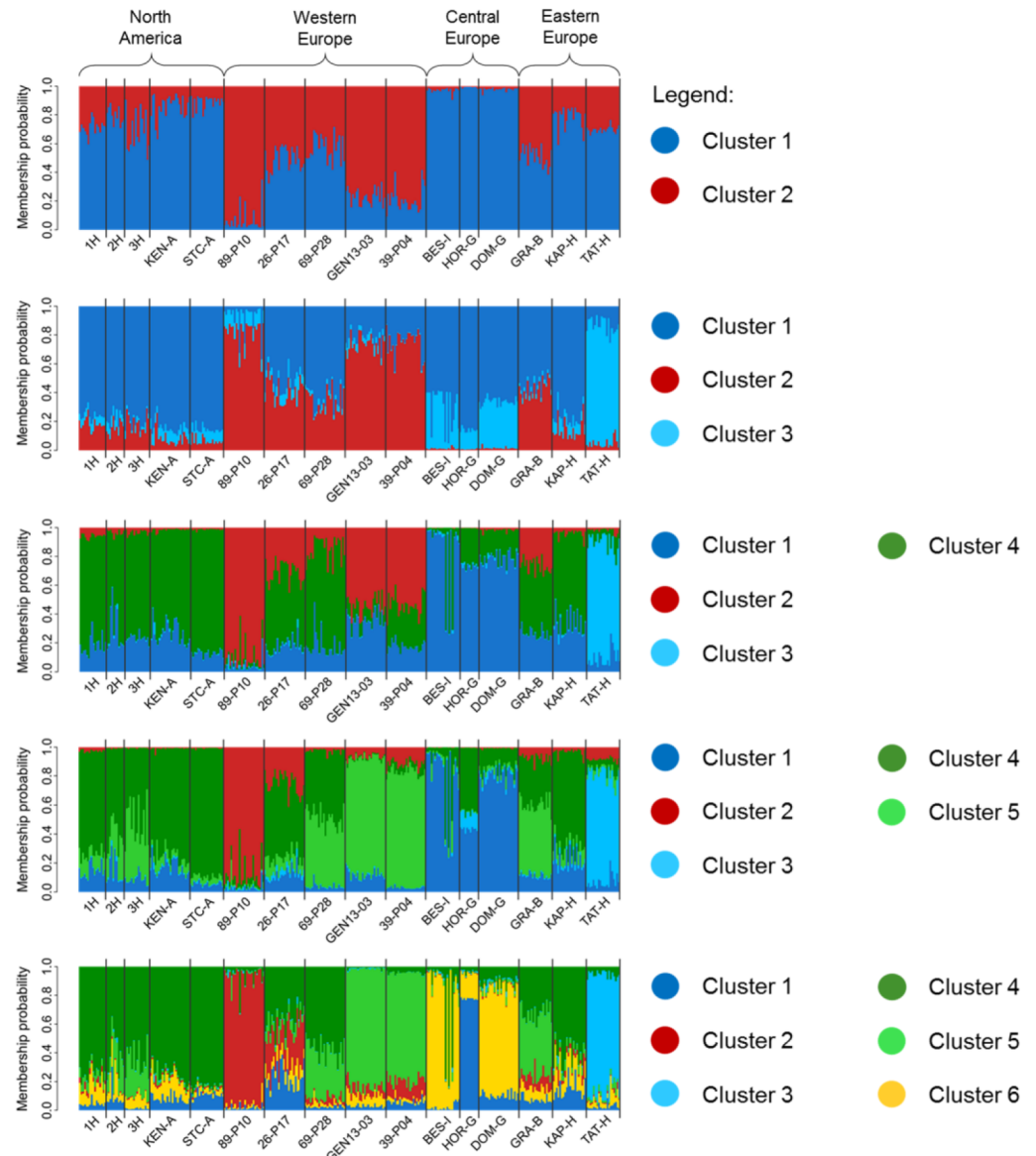
Cross-species transferability was tested for 31 gSSRs and 32 EST-SSRs. Among these markers, 32.2%, 54.8% and 67.7% of gSSRs and 46.9%, 75% and 81.2% of EST-SSRs gave consistent amplification and clear electrophoretic migration patterns in *A. trifida*, *A. psilostachya* and *A. tenuifolia*, respectively (S1–S3 Tables). Among the 26 markers used to analyse the genetic variation in *A. artemisiifolia* (Table 3), three gSSRs (SSR17, SSR26 and SSR73) and five EST-SSRs (EST-SSR13, EST-SSR61, EST-SSR69, EST-SSR111 and EST-SSR123) were scorable in all

**Table 5. Genetic diversity parameters across populations of *A. artemisiifolia* sampled within (i) North America and Europe, (ii) North America only and (iii) Europe only.**  $A$ : average allelic richness after rarefaction,  $H_O$ : observed heterozygosity,  $H_S$ : expected heterozygosity,  $F_{IS}$ : inbreeding coefficient estimated taking into account the presence of null alleles,  $F_{ST}$ : coefficient of genetic differentiation among populations.

| Group         | Number of populations | $A$   | $H_O$ | $H_S$ | $F_{IS}$ | $F_{ST}$ |
|---------------|-----------------------|-------|-------|-------|----------|----------|
| Overall       | 16                    | 3.989 | 0.544 | 0.630 | 0.078    | 0.064*   |
| North America | 5                     | 4.193 | 0.651 | 0.651 | 0.092    | 0.042*   |
| Europe        | 11                    | 3.896 | 0.496 | 0.620 | 0.065    | 0.071*   |

\*  $F_{ST}$  estimate significantly higher than zero based on 99% bootstrap confidence intervals.

<https://doi.org/10.1371/journal.pone.0176197.t005>



**Fig 3. Individual plant membership probabilities for the genetic clusters identified by the software Structure within 16 *A. artemisiifolia* populations sampled in Europe and in North America.** Populations are classified from a west (left) to east (right) gradient.

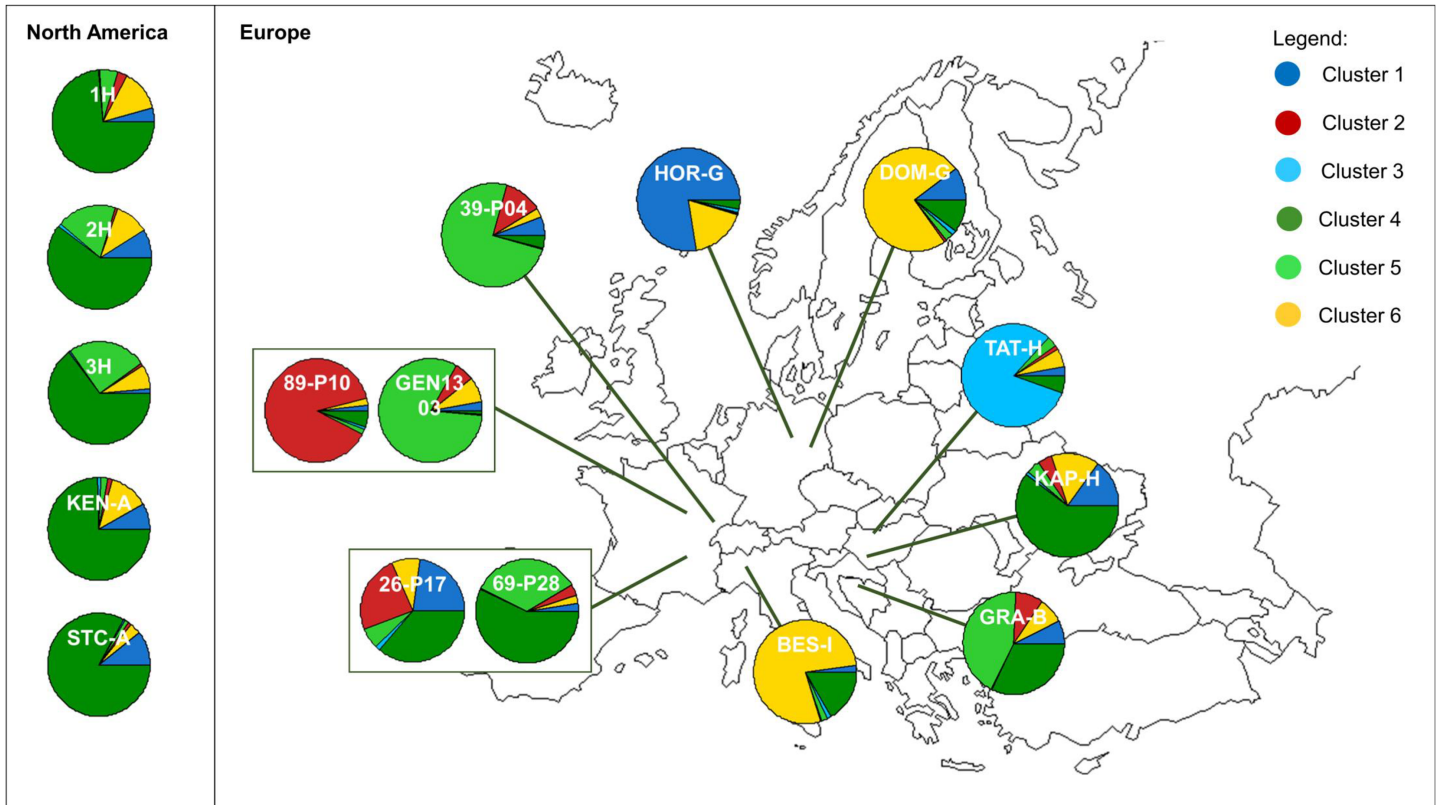
<https://doi.org/10.1371/journal.pone.0176197.g003>

three other species. Relationships among species were visualised by a PCA based on data at these eight markers (Fig 6). In coherence with the transferability of SSR markers, *A. trifida* was the most genetically divergent species, while *A. psilostachya* and *A. tenuifolia* appeared to be very genetically close to *A. artemisiifolia*.

## Discussion

### New, highly polymorphic nuclear SSRs in *A. artemisiifolia*

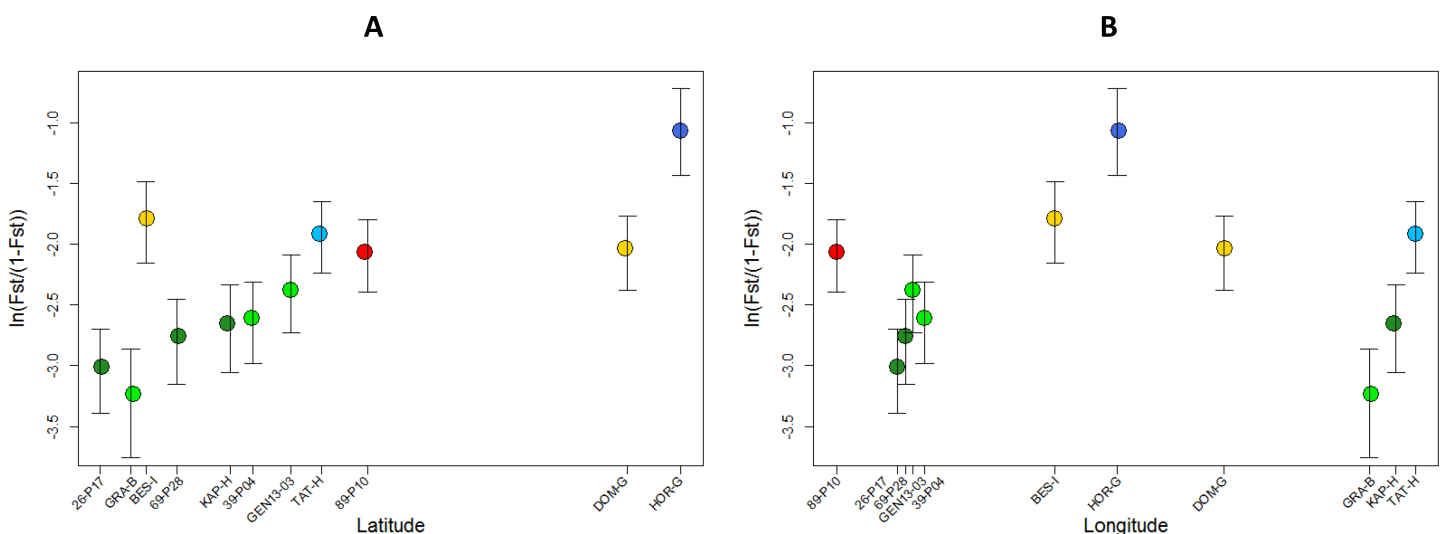
Next-generation sequencing technologies have considerably facilitated the development of SSRs for non-model organisms. Until recently, the method of choice was 454 sequencing of



**Fig 4. Genetic structure of 16 populations of *Ambrosia artemisiifolia* analyzed using 26 SSR markers.** Proportions of the six genetic clusters within 16 *A. artemisiifolia* populations sampled in Europe and in North America.

<https://doi.org/10.1371/journal.pone.0176197.g004>

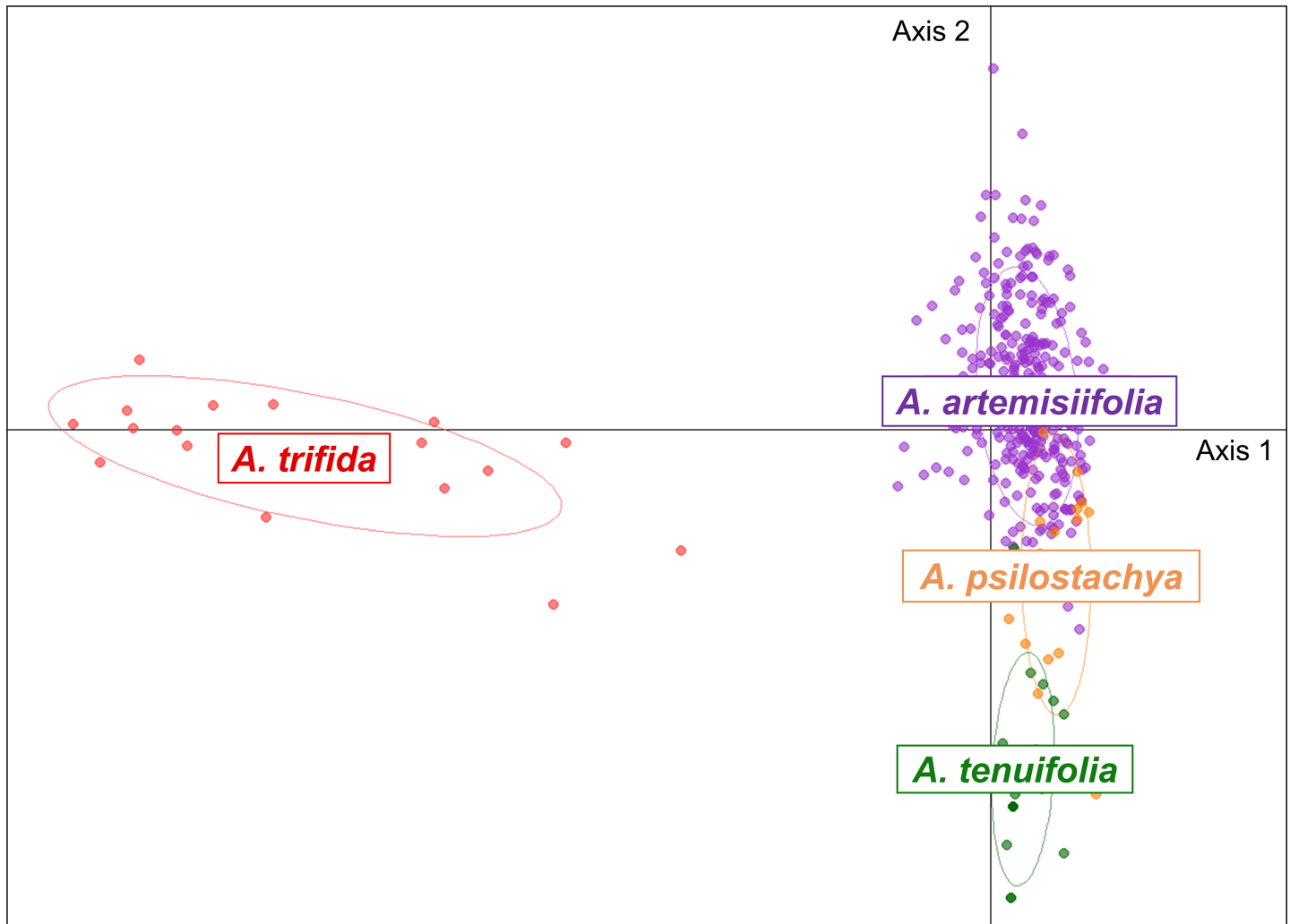
enriched gDNA libraries. 454 sequencing generates relatively long reads, which facilitates the design of primers within the regions flanking SSRs. However, the more recent Illumina sequencing technique provides much higher numbers of reads at a lower cost, and can now



**Fig 5. Population-specific genetic divergence (expressed as  $F_{ST}/(1 - F_{ST})$ ) of *Ambrosia artemisiifolia* in Europe, as a function of (A) latitude and (B) longitude.** Dots are coloured according to the most frequent genetic cluster identified by Structure (at  $K = 6$ ) in each population (see Figs 3 and 4).

<https://doi.org/10.1371/journal.pone.0176197.g005>





**Fig 6. Relationships among *Ambrosia* species assessed with PCA based on eight SSR markers.** Three gSSRs (SSR 17–26 and 73) and five EST-SSRs (EST-SSRs 13–61–69–111 and 123) were used. Each color represents one *Ambrosia* species.

<https://doi.org/10.1371/journal.pone.0176197.g006>

generate paired reads of 2×250 bp or longer [8]. Here, we implemented a rigorous initial quality filtering of reads. Further, we merged Illumina paired-end reads and kept only long-enough sequences, which facilitated primer design. A similar amplification success of potentially amplifiable loci was obtained from Illumina and from 454 data. This, together with the Illumina technology yielding one to two orders of magnitude more reads than 454, highlights Illumina as the currently most efficient sequencing technology for developing SSRs, provided reads are carefully checked for quality and paired-end reads are merged.

All markers developed were highly polymorphic in *A. artemisiifolia*. As compared to gSSRs, EST-SSRs are expected to be less polymorphic and more prone to the influence of selective processes: divergent selection may increase the estimation of genetic differentiation among populations at these loci, while purifying or balancing selection may have the opposite effect [9–11]. Here, allelic richness and expected heterozygosity were similar between the two kinds of markers, while genetic differentiation was slightly lower for EST-SSRs. Most EST-SSRs were tri-nucleotide repeats, for which length polymorphism does not result in any frameshift in the coding sequence. No influence of selection could be detected for any of the EST-SSRs analysed.

The high level of polymorphism observed at both non-genic and genic locations in the genome of *A. artemisiifolia* likely reflects very large effective population sizes in a plant species known to have recently undergone a demographic expansion both in its native range [26] and in invasive ranges [15]. Similarly, a large variation for life traits is known in *A. artemisiifolia* [54].

### Null alleles rather than partial selfing explain excess homozygotes in *A. artemisiifolia*

One undesirable counterpart of high nucleotide polymorphism is the presence of null alleles resulting from mutations at primer binding sites. Here, null alleles were observed for both gSSRs and EST-SSRs, with overall estimated frequencies of about 10%. This is consistent with a literature survey indicating that null alleles frequencies are often below 20% but can in some cases range from 40% to 75% [55]. Analyzing progenies in French populations provided direct evidence for the presence of null alleles, but no evidence for selfing or biparental inbreeding. Our results were consistent with a previous study of invasive populations from China [27] that also indicated complete allogamy and no shift towards partial selfing during invasion. Significant  $F_{IS}$  values were observed in only seven populations out of sixteen, indicating that null alleles are the main cause for excess homozygosity. Significant  $F_{IS}$  values were observed in a small minority of populations from Europe but in the majority of populations from the native range. As any evolution of the mating system towards loss of selfing during invasion seems unlikely, it remains to be investigated whether this might be due to a different functioning of the populations in the two ranges, with populations from the native range showing some Wahlund effect [26].

### Genetic diversity, population structure and population-specific genetic divergence in *A. artemisiifolia*

Genetic diversity within population was similar in North America (the native range) and in Europe, but genetic differentiation among populations ( $F_{ST}$ ) was greater in the invasive range. A similar trend had also been observed previously [23]. This difference in  $F_{ST}$  values may arise simply because only a small area of the native range was sampled (five American populations) in our study and in [23]. This pattern is also consistent with a scenario involving multiple introduction events, as previously proposed [20,22,23]. Rare alleles initially present in American populations may have shifted to high frequencies in different European populations after invasion and local demographic expansion [22]. The maintenance of high levels of genetic diversity within invasive populations, a trend opposite to that found in many other biological invasions processes [56], can be attributed to high numbers of introduced seeds in multiple events [57], high gene flow and possibly genetic admixture among introduced populations [21,58].

The main pattern of population structure we observed in Europe was a west-east gradient. Differentiation between the western and eastern parts of the European invasive range had previously been observed [22] and attributed to two main, distinct invasion sources. Here, we also observed that populations from central Europe (Germany to Italy) were genetically distinct from both Western and Eastern European populations. Several genetic clusters predicted by Structure were not observed or were very infrequent in North American populations, suggesting that we may have sampled only a fraction of the native sources. Alternatively, the additional clusters revealed by Structure under the more refined model ( $K = 6$ ) could simply reflect the elevated genetic divergence of some populations. Indeed, Structure analyses are known to be biased towards inferring extra genetic clusters when some populations have undergone strong recent genetic drift; in that case, and contrary to the assumptions of the admixture

model, not all genetic clusters are ancestral sources for the present populations [59]. Population-specific genetic divergences as estimated based on the F-model largely varied among populations, a pattern not expected if all populations similarly derived from a number of ancestral sources [51]. This, together with the outputs of Structure for increasing genetic partitioning (Fig 3), suggests secondary founding events associated with genetic drift. This hypothesis was not supported by signatures for recent bottlenecks; however, it is well known that bottleneck tests have a very limited power [60]. Populations from Italy, Germany and one population from Hungary (TAT) likely had their genetic sources in the South-Eastern part of Europe, whereas one population from France (89-P10) likely originated from eastern France. Although this would need to be validated based a more extensive sampling, genetic patterns revealed here are overall consistent with two main distinct colonization events in Europe (in South-Eastern France: the Rhone valley, and South-Eastern Europe: the Pannonian plain), with secondary colonization events arising northwards and towards Central Europe.

### Genetic variation among species of the genus *Ambrosia*

Most SSR markers developed for *A. artemisiifolia* (65% and 75%, respectively) were transferable to *A. psilostachya* and *A. tenuifolia*, whereas only 40% were transferable to *A. trifida*. The genus *Ambrosia* is composed of many, not clearly delineated species [15] for which there is no well-established phylogeny. Former morphological classification considered *A. artemisiifolia*, *A. psilostachya* and *A. tenuifolia* as related species belonging to one same group, while *A. trifida* was classified in a separate group [30]. This is consistent with differences in gametophytic chromosome numbers ( $n = 18$  for *A. artemisiifolia*, *A. psilostachya* and *A. tenuifolia* but  $n = 12$  for *A. trifida*, [13,14,29,30,32]) and with a chloroplast DNA phylogeny [61]. The success of SSR marker transfer among species fully confirms these previous data. In addition, some degree of hybridization between *A. artemisiifolia* and *A. psilostachya* was suggested [32]. This, or homoplasmy at SSR markers, may explain the overlapping genetic variation between the two species.

*A. psilostachya* and *A. tenuifolia* are perennial species that reproduce both sexually and clonally [13, 14, 32, 62]. Although these species are of less concern than the annual *A. artemisiifolia* and *A. trifida*, being less widespread and invasive, our SSR markers will be useful for assessing vegetative *versus* sexual reproduction, as well as for identifying colonization sources and relatedness among populations. However, given that several ploidy levels were reported in these species [13, 14, 29, 32], we recommend that ploidy is carefully checked before markers developed from *A. artemisiifolia* are transferred to *A. psilostachya* and *A. tenuifolia*. *A. trifida* is a noxious annual weed, very widespread in its native area [63] and introduced in several European countries including, for instance, France, Italy, Slovenia and Serbia [64,65]. Despite the potential threat set by this species on both human health and agriculture, population genetics studies are still lacking. Given its distant relation to *A. artemisiifolia* and the low success rate of markers transferability, we recommend that additional SSR markers are specifically developed for *A. trifida*.

### Conclusions

Large sets of genomic SSRs and EST-SSRs were developed and validated in *A. artemisiifolia*, providing useful new resources for genetic studies of this highly noxious invasive weed. All markers were highly polymorphic. EST-SSRs revealed as many alleles as gSSRs and yielded similar genetic diversity estimates. The genetic patterns revealed for a set of American and European populations confirmed results from previous studies by showing a high within-population genetic diversity in both the native and invasive ranges. A geographical gradient of

genetic variation in Europe was consistent with at least two major colonization events in Western and Eastern Europe, respectively. Secondary founding events were identified, especially in Central Europe. In addition, we settled a former controversy by demonstrating that inbreeding observed within populations is attributable mostly to the presence of null alleles rather than to selfing. Last, most SSRs were transferable to three other *Ambrosia* species. These SSRs can readily be used for studying key aspects of the biology and population dynamics of the two species most closely related to *A. artemisiifolia* (*A. psilostachya* and *A. tenuifolia*).

## Supporting information

**S1 Fig. Karyotype showing the 36 chromosomes of *A. artemisiifolia* in a telophase (left) and in the end of a prophase (right).** The red arrows indicate the chromosomes. Gx100. (TIF)

**S2 Fig. Results of Bayescan  $F_{ST}$  outlier analysis on 14 gSSR (a) and 13 EST-SSR loci (b).** The vertical bars correspond to threshold P-values of 0.05 (solid line) and 0.01 (dashed line) for the neutral model. (a) Data from all 16 populations. (b) Data from 11 European populations. (TIF)

**S3 Fig. Data maximum likelihood value (left) and deltaK method (right) results used to determine the most likely number of genetic clusters with 13 EST-SSR and 13 gSSR markers.** (TIF)

**S4 Fig. Genetic structure of 16 populations of *A. artemisiifolia* analyzed using the 13 EST-SSR markers.** (TIF)

**S5 Fig. Genetic structure of 16 populations of *A. artemisiifolia* analyzed using the 13 gSSR markers.** (TIF)

**S1 Table. gSSR markers obtained by 454 sequencing of enriched *A. artemisiifolia* gDNA and showing consistent PCR amplifications and clear electrophoretic migration patterns.** Loci in bold represent the markers selected in *A. artemisiifolia*. (DOCX)

**S2 Table. gSSR markers obtained by Illumina sequencing of raw *A. artemisiifolia* gDNA and showing consistent PCR amplifications and clear electrophoretic migration patterns.** Loci in bold represent the markers selected in *A. artemisiifolia*. (DOCX)

**S3 Table. EST-SSR markers obtained by 454 sequencing of *A. artemisiifolia* ESTs and showing consistent PCR amplifications and clear electrophoretic migration patterns.** Loci in bold represent the markers selected in *A. artemisiifolia*. (DOCX)

**S4 Table. Genetic diversity parameters at 26 nuclear SSR markers for 16 populations of *Ambrosia artemisiifolia*.** *N*: number of individuals genotyped, *A*: average allelic richness after rarefaction,  $H_O$  observed heterozygosity,  $H_S$  expected heterozygosity,  $F_{IS}$  inbreeding coefficient estimated taking into account the presence of null alleles. \*  $F_{IS}$  estimates significantly greater from zero (using the Bayesian model comparison based on Deviance Information Criterion implemented in INEST2.1). (DOCX)

## Acknowledgments

We thank Charles Poncet and the GENTYANE team (INRA Clermont-Ferrand) for performing the genotyping and Dr Mona Abirached-Darmency (INRA Dijon) for producing the karyotype of *A. artemisiifolia*.

## Author Contributions

**Conceptualization:** VLC.

**Data curation:** VLC.

**Formal analysis:** LM VLC.

**Funding acquisition:** GB CD.

**Investigation:** LM RC FP RS.

**Project administration:** CD GB.

**Supervision:** VLC CD BC GB.

**Visualization:** LM.

**Writing – original draft:** LM.

**Writing – review & editing:** LM VLC GB BC CD.

## References

1. Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, et al. The population biology of invasive species. *Annu Rev Ecol Syst.* 2001; 32: 305–332.
2. Lee CE. Evolutionary genetics of invasive species. *Trends Ecol Evol.* 2002; 17: 386–391.
3. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol.* 2013; 22: 2841–2847. <https://doi.org/10.1111/mec.12350> PMID: 23711105
4. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol.* 2014; 4: 4399–4428. <https://doi.org/10.1002/ece3.1305> PMID: 25540699
5. Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, et al. Current trends in microsatellite genotyping. *Mol Ecol Resour.* 2011; 11: 591–611. <https://doi.org/10.1111/j.1755-0998.2011.03014.x> PMID: 21565126
6. Haasl RJ, Payseur BA. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb).* 2011; 106: 158–171.
7. Hodel RGJ, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, et al. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st Century. *Appl Plant Sci.* 2016; 4: 1600025.
8. Castoe TA, Poole AW, de Koning APJ, Jones KL, Tomback DF, Oyler-McCance SJ, et al. Rapid microsatellite identification from illumina paired-end genomic sequencing in two birds and a snake. *PLoS One.* 2012; 7: e30953. <https://doi.org/10.1371/journal.pone.0030953> PMID: 22348032
9. Ellis JR, Burke JM. EST-SSRs as a resource for population genetic analyses. *Heredity.* 2007; 99: 125–132. <https://doi.org/10.1038/sj.hdy.6801001> PMID: 17519965
10. Postolache D, Leonarduzzi C, Piotti A, Spanu I, Roig A, Fady B, et al. Transcriptome versus Genomic Microsatellite Markers: Highly Informative Multiplexes for Genotyping *Abies alba* Mill. and Congeneric Species. *Plant Mol Biol Report.* 2014; 32: 750–760.
11. Defaveri J, Viitaniemi H, Leder E, Merilä J. Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Mol Ecol Resour.* 2013; 13: 377–392. <https://doi.org/10.1111/1755-0998.12071> PMID: 23356957
12. The Plant List. Version 1.1. [Internet]. 2016. Available from: <http://www.theplantlist.org/>
13. Bassett IJ, Crompton CW. The biology of Canadian Weeds—*Ambrosia Artemisiifolia* L. and *Ambrosia psilostachya* DC. *Can J Plant Sci.* 1975; 11: 463–476.



14. Insausti P, Soriano A, Sanchez R a. Effects of flood-influenced factors on seed germination of *Ambrosia tenuifolia*. *Oecologia*. 1995; 103: 127–132. <https://doi.org/10.1007/BF00328433> PMID: 28306953
15. Essl F, Bir K, Brandes D, Broennimann O, Bullock JM, Chapman DS, et al. Biological Flora of the British Isles: *Ambrosia artemisiifolia*. *J Ecol*. 2015; 103: 1069–1098.
16. Ziska LH, Gebhard DE, Frenz DA, Faulkner S, Singer BD, Straka JG. Cities as harbingers of climate change: Common ragweed, urbanization, and public health. *J Allergy Clin Immunol*. 2003; 111: 290–295. PMID: 12589347
17. Smith M, Cecchi L, Skjøth CA, Karrer G, Šikoparija B. Common ragweed: a threat to environmental health in Europe. *Environ Int*. 2013; 61: 115–126. <https://doi.org/10.1016/j.envint.2013.08.005> PMID: 24140540
18. Genton BJ, Jonot O, Thévenet D, Fournier E, Blatrix R, Vautrin D, et al. Isolation of five polymorphic microsatellite loci in the invasive weed *Ambrosia artemisiifolia* (*Asteraceae*) using an enrichment protocol. *Mol Ecol Notes*. 2005; 5: 381–383
19. Molecular Ecology Resources Primer Development Consortium (2009) Permanent genetic resources added to Molecular Ecology Resources database 1 January 2009–30 April 2009. *Mol Ecol Resour* 9: 1375–1429. <https://doi.org/10.1111/j.1755-0998.2009.02746.x> PMID: 21564911
20. Genton BJ, Shykoff JA, Giraud T. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Mol Ecol*. 2005; 14: 4275–4285. <https://doi.org/10.1111/j.1365-294X.2005.02750.x> PMID: 16313592
21. Chun YJ, Fumanal B, Laitung B, Bretagnolle F. Gene flow and population admixture as the primary post-invasion processes in common ragweed (*Ambrosia artemisiifolia*) populations in France. *New Phytol*. 2010; 185: 1100–1107. <https://doi.org/10.1111/j.1469-8137.2009.03129.x> PMID: 20028474
22. Gladioux P, Giraud T, Kiss L, Genton BJ, Jonot O, Shykoff JA. Distinct invasion sources of common ragweed (*Ambrosia artemisiifolia*) in Eastern and Western Europe. *Biol Invasions*. 2010; 13: 933–944.
23. Gaudeul M, Giraud T, Kiss L, Shykoff JA. Nuclear and chloroplast microsatellites show multiple introductions in the worldwide invasion history of common ragweed, *Ambrosia artemisiifolia*. *PLoS One*. 2011; 6: e17658. <https://doi.org/10.1371/journal.pone.0017658> PMID: 21423697
24. Kocis-Tubic N, Djan M, Velickovic N, Anackov G, Obreht D. Gradual loss of genetic diversity of *Ambrosia artemisiifolia* L. populations in the invaded range of central Serbia. *Genetika*. 2014; 46: 255–268.
25. Ciappetta S, Ghiani A, Gilardelli F, Bonini M, Citterio S, Gentili R. Invasion of *Ambrosia artemisiifolia* in Italy : Assessment via analysis of genetic variability and herbarium data. *Flora*. 2016; 223:106–113.
26. Martin MD, Zimmer EA, Olsen MT, Foote AD, Gilbert MTP, Brush GS. Herbarium specimens reveal a historical shift in phylogeographic structure of common ragweed during native range disturbance. *Mol Ecol*. 2014; 23: 1701–1716. <https://doi.org/10.1111/mec.12675> PMID: 24450363
27. Li X- M, Liao W- J, Wolfe LM, Zhang D-Y. No evolutionary shift in the mating system of North American *Ambrosia artemisiifolia* (*Asteraceae*) following its introduction to China. *PLoS One*. 2012; 7: e31935. <https://doi.org/10.1371/journal.pone.0031935> PMID: 22384104
28. Délye C, Matějček A, Gasquez J. PCR-based detection of resistance to acetyl-CoA carboxylase-inhibiting herbicides in black-grass (*Alopecurus myosuroides* Huds) and ryegrass (*Lolium rigidum* gaud). *Pest Manag Sci*. 2002; 58: 474–478. <https://doi.org/10.1002/ps.485> PMID: 11997974
29. Mulligan GA. Chromosome numbers of Canadian weeds. *Can J Bot*. 1957; 35: 779–789.
30. Payne WW. A re-evaluation of the genus *Ambrosia* (*Compositae*). *J Arnold Arboretum* 1964; 45: 401–430.
31. Baltisberger M, Widmer A. Chromosome numbers and karyotypes within the genus *Achillea* (*Asteraceae: Anthemideae*). *Willdenovia*. 2016; 46: 121–35.
32. Wagner WH, Beals TF. Perennial ragweeds (*Ambrosia*) in Michigan, with the description of a new intermediate taxon. *Rhodora* 1958; 60:177–204.
33. Malausa T, Gilles A, Megléc E, Blanquart H, Duthoy S, Costedoat C, et al. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Resour*. 2011; 11: 638–644. <https://doi.org/10.1111/j.1755-0998.2011.02992.x> PMID: 21676194
34. Sequence Read Archive. National Center for Biotechnology Information, National US Library of Medicine. Available: <https://www.ncbi.nlm.nih.gov/sra>.
35. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015; 12: 115–121. <https://doi.org/10.1038/nmeth.3252> PMID: 25633503
36. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27: 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507> PMID: 21903629

37. Megléc E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A, et al. QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Mol Ecol Resour.* 2014; 14: 1302–1313. <https://doi.org/10.1111/1755-0998.12271> PMID: 24785154
38. Alberto F. MsatAllele\_1.0: An R package to visualize the binning of microsatellite alleles. *J Hered.* 2009; 100: 394–397. <https://doi.org/10.1093/jhered/esn110> PMID: 19126639
39. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008; 24: 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129> PMID: 18397895
40. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Mol Ecol Notes.* 2004; 4: 535–538.
41. Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered.* 2009; 100: 106–113. <https://doi.org/10.1093/jhered/esn088> PMID: 18936113
42. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics.* 2008; 180: 977–93. <https://doi.org/10.1534/genetics.108.092221> PMID: 18780740
43. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol.* 2014; 23: 2178–2192. <https://doi.org/10.1111/mec.12725> PMID: 24655127
44. Ritland K. Extensions of models for the estimation of mating systems using n independent loci. *Heredity.* 2002; 88: 221–228. <https://doi.org/10.1038/sj.hdy.6800029> PMID: 11920127
45. Goudet J. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *J Hered.* 1995; 86: 485–486.
46. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155: 945–959. PMID: 10835412
47. Falush D, Stephens M, Pritchard JK. Inference of Population Structure Using Multilocus Genotype Data : Linked Loci and Correlated Allele Frequencies. *Genetics.* 2003; 164: 1567–1587. PMID: 12930761
48. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2011; 4:359–361.
49. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005; 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739
50. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007; 23: 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233> PMID: 17485429
51. Gaggiotti OE, Foll M. Quantifying population structure using the F-model. *Mol. Ecol. Res.* 2010; 10: 821–830.
52. Foll M, Gaggiotti OE. Identifying the environmental factors that determine the genetic structure of Populations. *Genetics* 2006; 174: 875–891. <https://doi.org/10.1534/genetics.106.059451> PMID: 16951078
53. Cornuet JM and Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics.* 1996; 144: 2001–2014. PMID: 8978083
54. Ortmans W, Mahy G, Chauvel B, Monty A. Performance variation of common ragweed (*Ambrosia artemisiifolia* L.) across invasion levels in Western Europe. *Flora.* 2016; 220: 134–141.
55. Dakin EE, Avise JC. Microsatellite null alleles in parentage analysis. *Heredity.* 2004; 93: 504–509. <https://doi.org/10.1038/sj.hdy.6800545> PMID: 15292911
56. Facon B, Genton BJ, Shykoff J, Jarne P, Estoup A, David P. A general eco-evolutionary framework for understanding bioinvasions. *Trends Ecol Evol.* 2006; 21: 130–105. <https://doi.org/10.1016/j.tree.2005.10.012> PMID: 16701488
57. Chapman DS, Makra L, Albertini R, Bonini M, Páldy A, Rodinkova V, et al. Modelling the introduction and spread of non-native species: International trade and climate change drive ragweed invasion. *Glob Chang Biol.* 2016; 22: 3067–3079. <https://doi.org/10.1111/gcb.13220> PMID: 26748862
58. Chauvel B, Dessaint F, Cardinal-Legrand C, Bretagnolle F. The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. *J Biogeogr.* 2006; 33(4): 665–763.
59. Falush D, van Dorp L, Lawson DJ. A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. Preprint. Available: bioRxiv 2016;
60. Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Béer E, Robinson S, Vasquez-Carrillo C, Pauli JN, Palsboll PJ. Reliability of genetic bottleneck tests for detecting recent population declines. *Mol Ecol.* 2012; 21: 3403–3418. <https://doi.org/10.1111/j.1365-294X.2012.05635.x> PMID: 22646281

61. Miao B, Turner B, Simpson B, Mabry T. Chloroplast DNA study of the genera *Ambrosia* s.l. and *Hymenoclea* (*Asteraceae*): Systematic implications. *Plant Syst Evol.* 1995; 194: 241–255.
62. Insausti P, Grimoldi AA. Gap disturbance triggers the recolonization of the clonal plant *Ambrosia tenuifolia* in a flooding grassland of Argentina. *Austral Ecol.* 2006; 31: 828–836.
63. Bassett I, Crompton C. The biology of Canadian weeds—*Ambrosia trifida* L. *Can J Plant Sci.* 1982; 55: 1003–1010.
64. Follak S, Dullinger S, Kleinbauer I, Moser D, Essl F. Invasion dynamics of three allergenic invasive *asteraceae* (*Ambrosia trifida*, *Artemisia annua*, *Iva xanthiifolia*) in central and eastern Europe. *Preslia.* 2013; 85: 41–61.
65. Chauvel B, Rodriguez A, Moreau C, Martinez Q, Bilon R, Fried G. Développement d'*Ambrosia trifida* L. en France : connaissances historiques et écologiques en vue d'une éradication de l'espèce. *J Bot Soc Bot France* 2015; 71: 25–38.