Research Article

Received: 17 December 2020 Revised: 12 February 2021 Accepted article published: 25 February 2021 Published online in Wiley Online Library: 18 March 2021

(wileyonlinelibrary.com) DOI 10.1002/ps.6336

Population structure and evolution of resistance to acetolactate synthase (ALS)-inhibitors in *Amaranthus tuberculatus* in Italy

Andrea Milani,[a] Ulrich Lutz,[b] Giulio Galla,[c] Laura Scarabel,[a] Detlef Weigel[b] and Maurizio Sattin[a]*

Abstract

BACKGROUND: Before 2010, *Amaranthus tuberculatus* (Moq.) J. D. Sauer was barely known to farmers and stakeholders in Italy. Since then, several populations resistant to acetolactate synthase (ALS)-inhibiting herbicides have been collected. In most populations, a known target site resistance-endowing mutation was found, a Trp to Leu substitution at position 574 of the *ALS* gene, but it was unclear whether they had evolved resistance independently or not. The aims of the work were (i) to elucidate the population structure of Italian ALS-resistant *A. tuberculatus* populations, and (ii) to analyze the *ALS* haplotypes of the various populations to determine whether resistance arose multiple times independently.

RESULTS: In order to determine the population structure of eight *A. tuberculatus* populations, eight previously described microsatellite loci were used. Two ancestors were found: three populations derived from one, and five from the other. In the 4-kb *ALS* region of the genome, including the 2-kb coding region, 389 single nucleotide polymorphisms were found. *In silico* haplotype estimation was used to reconstruct the sequence of three distinct haplotypes carrying the Trp574Leu mutation. In addition, no mutation was found in 83% of plants of a single population.

CONCLUSIONS: (i) Resistance must have arisen independently at least three times; (ii) at least one population was already resistant to ALS inhibitors when introduced in Italy; (iii) a single haplotype with a Trp574Leu mutation was shared among six populations, probably because of broad seed dispersal; and (iv) one population likely evolved nontarget site ALS inhibitors resistance.

Supporting information may be found in the online version of this article.

Keywords: Waterhemp; haplotype analysis; evolution; point mutations; resistance spread; resistance management

1 INTRODUCTION

Waterhemp [*Amaranthus tuberculatus* (Moq.) J.D. Sauer] is an invasive weed that is threatening the sustainability of important summer crops, especially in its native range in North America.[1,2] The major concern is related to its ability to expand its habitat and to evolve herbicide resistance. Its original habitat is typically riparian, infesting riverbanks and floodplains. It has been recorded as a casual alien in many countries all over the world and has become naturalized in some of these.[3] The first cases of *A. tuberculatus* resistant to acetolactate synthase (ALS)-inhibiting herbicides in the USA were recorded in 1993, and in neighboring Canada in 2002.[4] Since then, resistance cases as well as the number of involved herbicide sites of action (SoA) have substantially increased.

Herbicide resistance patterns, resistance mechanisms and their inheritance have been extensively investigated in the last 20 years.[5] It is only recently that population genetics has become sufficiently integrated to understand the dynamics of diffusion in agricultural habitats. In an early study, microsatellites (short sequence repeat, SSR) were used to trace habitat expansion of *A. tuberculatus* var. *rudis*, demonstrating its spread from western to eastern USA.[6] Subsequently, with a population genomics approach it was demonstrated that glyphosate resistance in Canada occurred because of both the introduction of resistant

* *Correspondence to: M Sattin, Istituto per la Protezione Sostenibile delle Piante (IPSP)-CNR c/o Agripolis, viale dell'Università 16, 35020, Legnaro, PD, Italy. E-mail: maurizio.sattin@cnr.it*

a *Institute for Sustainable Plant Protection (IPSP-CNR), Legnaro, Italy*

b *Max Planck Institute for Developmental Biology, Tübingen, Germany*

c *Edmund Mach Foundation, S. Michele all 'Adige, Italy*

populations from USA and independent selection for herbicide resistance.[7]

There is little information about the spread of *A. tuberculatus* outside its native North American range. In Italy, *A. tuberculatus* was first reported in the western Po Valley in the early 1980s, where it has become invasive and has started spreading along the Po river, forming very dense populations that now tend to displace the native flora.[8,9] Despite its invasiveness, weedy waterhemp in crop fields was discovered in the eastern Po Valley only in 2010. Since then, other field populations have been identified in the Emilia-Romagna (2014) and Veneto (2017) regions.[10] All of these populations were cross-resistant to thifensulfuron-methyl and imazamox due to a substitution at codon 574 of the *ALS* gene that is known to cause broad cross-resistance to ALS inhibitors in *A. tuberculatus*.[10,11] However, even though all populations had the same point mutation, it was unclear whether they originated from single or multiple selection events and how they have spread. Remarkably, all known populations were in soybean fields, a very different habitat from the original riparian one of this species in its native range. Only populations with ALS inhibitor resistance were found in the surroundings of the sampled populations and all stakeholders agreed that susceptible populations of this species had never been observed to infest crop fields before.

Post-emergence application of ALS inhibitors is the most used tool to control broadleaved, dicot weeds in soybean crops in Italy. The appearance of resistant biotypes of *A. tuberculatus* is of concern given the disruptive potential of this species. Herbicide resistance is a remarkable example of rapid plant evolution[12,13] following abrupt environmental changes – that is, the selective pressure exerted by herbicides. Origin and spread of resistance are mainly human-driven evolutionary processes,[14,15] and a deeper understanding of the evolutionary mechanisms could greatly benefit resistance risk assessment and management strategies.[16,17]

Haplotype analysis of loci under selection may provide information on whether a specific point mutation associated with resistance arose independently multiple times amongst different weed populations or derived from a single mutation event. If a point mutation gives a selective advantage in certain conditions, its frequency within a population will rise because of selective pressure. Neutral (not under selection) polymorphisms that are genetically linked with the point mutation would be inherited together with it because of genetic hitch-hiking,[18] thus creating an unique footprint of the selection event. In the case of selection of *de novo* mutations, the sequence of genetically linked polymorphisms that are inherited together (i.e. the haplotype) is expected to be very different among populations, especially if longer sequences are taken into consideration. After some generations, the variation (i.e. the polymorphism level) among nucleotides near the point mutation would be strongly reduced, or even completely eliminated (i.e. selective sweep).[19] Although this approach has been used often to study resistance to drugs and insecticides, there are only a few examples of haplotype analysis in studying herbicide resistance.[20] Haplotype analysis was used to infer the origin of resistance to Acetyl CoA Carboxylase (ACCase) inhibitors in *Alopecurus myosuroides*[21,22] and, more recently, the origin of glyphosate resistance in *A. tuberculatus*.[7]

The aim of this study was to unravel the origins and spread of resistance to ALS-inhibiting herbicides among eight populations of *A. tuberculatus* collected in Italian soybean fields.

2 MATERIALS AND METHODS

2.1 Plant material

Seven of eight *A. tuberculatus* populations used in this study were described in a previous paper.[10] The original population codes were simplified as follows: G, 10–10L; F, 10–13L; A, 14–35L; B, 17–56; D, 17–60; H, 17–61; E, 17–65. Populations G, F, A, B, D and H were collected in northeastern Italy in soybean fields treated with ALS-inhibiting herbicides (G and F were collected in 2010; A in 2014; B, D and H in 2017). Those six field-collected populations were highly cross-resistant to imazamox and thifensulfuron-methyl due to a tryptophan to leucine mutation at codon 574 of *ALS*.[10] Population E was collected in the Po River floodplain in 2017, where it had very likely never been treated with herbicides and was confirmed to be completely susceptible to ALS-inhibiting herbicides.[10] Population C, not included in the previous study, was collected in 2017 from an ALS-treated soybean field and was highly resistant to thifensulfuron-methyl but susceptible to imazamox (A. Milani, L. Scarabel and M. Sattin, unpublished data). Populations G, F and A were maintained in nonwoven fabric cages (≥20 plants).

2.2 Microsatellite genotyping and analysis

Twenty seeds per population were germinated following a protocol described previously [23] and seedlings were grown in a glasshouse until flowering. Twelve plants per population (1:1 male: female ratio) were chosen randomly, and leaf samples were collected and stored at −80 °C.

2.2.1 *DNA extraction*

DNA was extracted from 100 mg frozen leaves using the CTAB (cetyltrimethylammonium bromide) method.[7] DNA integrity was estimated by electrophoresis on a 0.8% agarose/1× TAE gel containing 1× SYBR Safe DNA stain (Thermo Scientific, Pittsburgh, PA, USA). Both purity and quantity of DNA extracts were assessed with a NanoDrop 2000c UV–visible spectrophotometer (Thermo Scientific). Each DNA sample was used for microsatellite analyses and *ALS* amplicon sequencing.

2.2.2 *Amplification of microsatellite loci*

In order to genotype 96 *A. tuberculatus* samples, a modified version of a previously described protocol[2] was used. The same loci were analyzed, but primers were designed to be multiplexed in two multiplex PCR (mPCR) reactions. Four universal (tagged) primers were used: Hill (TGACCGGCAGCA-AAATTG),[8] Tail_D (CGGAGAGCCGAGAGGTG),[9] D8S1132 (GGCTAGGAAAGGTT AGTGGC)[10] and PAN3 (TGTAGAAAGACGAAGGGAAGG). Universal primers were 5′ labeled with different dyes (6FAM™, VIC™, NED™ and PET™, respectively) and all forward locus-specific primers were added with the tag sequences at the 5′ end. mPCR mixes were performed using GoTaq® G2 Hot Start Polymerase (Promega, Madison, WI, USA) in a 10-µL mixture including 2 µL of 5× GoTaq Flexi Buffer, dNTPs mix (0.2 mM each), MgCl2 (1.5 mM), forward (tailed) primers 0.2 µM, reverse (PIG-tailed) primers 0.3 µM, fluorescent primers 0.1 µM, 0.05 µL GoTaq DNA Polymerase and 50 ng DNA. Amplification conditions: 95 °C*2′, five cycles of [95 °C*30″, 61 °C*30″, 72 °C*20″], 10 cycles of [95 °C*30″, 64 °C*30″, 72 °C*20″], 20 cycles of [95 °C*30″, 58 °C*30″, 72 °C*20″], 72 °C*5′. mPCR mix 1 amplified loci C1140, C3695, AAC1, C4097 e C0745 and mPCR mix 2 amplified loci C4999, ATC9, C3561, TAG5 e C9333. mPCR products were then run into capillary electrophoresis with an ABI PRISM 3130xl

Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). LIZ500 was adopted as molecular mass standard. Primer combinations of multiple PCR mixes, primer sequences, repeat motif, and size of each locus and dyes used for visualization are reported in the Supporting Information, Table S4.

2.2.3 Microsatellite data analysis

Peak size was determined using Peak Scanner™ v1.0 software (Applied Biosystems). Descriptive statistics were obtained using GenAlEx 6.5.[24] The software POPGENE 1.32[25] was used to compute the dendrogram based on Nei's genetic diversity. The geographical distance matrix was calculated with the online tool Geographic Distance Matrix Generator[26] from the GPS coordinates of each population (due to privacy issues, GPS positions will be available only on confidential request). The pairwise population matrix of Nei's unbiased genetic distance and the geographical matrix were used to perform the Mantel test with GenAlEx 6.5. Model-based Bayesian analysis implemented in STRUCTURE v2.3.4[27] was used to explore population genetic structure. The analysis was performed using a burn-in of 500 000 iterations and a run length of 500 000 Markov Chain Monte Carlo (MCMC) replications in ten independent runs. Prior knowledge about the number of populations was not included. The number of clusters (K) in the dataset was determined by the averaged likelihood at each K and the variance between replicates was determined by running a continuous series of $K = 1$–9 to determine the optimal number of populations present within the 96 individuals. The optimum number of clusters was predicted following the *ad hoc* statistic ΔK[28] implemented in Structure Harvester v0.6.94.[29] Simulations were performed without *a priori* assumptions concerning the admixture model and correlation in the allele frequencies (all combinations were tested).

2.3 Amplicon sequencing of *ALS* gene

2.3.1 Primer design and PCR amplification

DNA samples were the same as those used for SSR genotyping. Concentrations were quantified with Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) and adjusted to 3 ng μL^{-1}. Primers were designed using the reference *A. tuberculatus* genome.[7] The *A. tuberculatus* complete *ALS* coding sequence (GenBank: EF157818.1) was used as a query to identify the contig containing the *ALS* locus in the reference genome with Blast.[30] Primers were designed with Benchling (https://benchling.com) to amplify a region of ≈4 kb including the *ALS* coding sequence (2 kb). Primer sequences are provided in Table S5. A pool of 36 random samples was used for initial primer tests and PCR optimization. Primers Fw_3 and Rev_4 were chosen because of superior specificity. Ten samples of 96 were amplified with primers Fw_4 and Rev_3 because amplification with Fw_3 and Rev_4 failed. PCR was performed using Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific) in 30-μL reactions including 6 μL of 5× Phire Green Reaction Buffer, dNTPs mix (0.2 mM each), forward and reverse primers 0.625 μM each, 0.4 μL Phire Hot Start II DNA Polymerase and 9 ng DNA. Amplification conditions: 1 min at 95 °C; 35 cycles of 5 s at 95 °C, 5 s at 60 °C, 60 s at 72 °C; 5 min at 72 °C. A model of *ALS* gene was obtained with R[31] genemodel[32] package.

2.3.2 Dual-indexed library preparation

Libraries were prepared using an in-house prepared Tn5 transposase and following a published tagmentation protocol.[33] In brief, the steps were: (i) Tn5 transposase conjugation to streptavidin magnetic beads (NEB, Ipswich, MA, USA); (ii) addition of conjugated beads to each DNA sample; (iii) incubation to allow tagmentation reaction; and (iv) Tn5 transposase stripping. Dual-indexed libraries were prepared using i7 and i5 index adapters (Illumina, San Diego, CA, USA), Q5® High-Fidelity DNA Polymerase (NEB) and tagmented DNA as template, following the manufacturer's instructions. After the indexing, PCR samples were pooled and size-selected (350–700 bp) on a BluePippin instrument (Sage Science, Beverly, MA, USA). Actual size and quality of libraries were evaluated using a Bioanalyzer High Sensitivity Chip and run on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The libraries then were sequenced to a coverage depth of 500× on an MiSeq™ (Illumina) instrument using MiSeq Reagent Micro Kit v2 (300-cycles) (Illumina).

2.3.3 Single nucleotide polymorphism (SNP)-calling pipeline and haplotype handling

After sequencing, reads of each sample were (i) de-multiplexed (removal of indexing primers); (ii) aligned to the reference contig with BWA-MEM[34]; (iii) sorted, indexed and then merged with SAMtools[35] to get a sorted multiple alignment file. Variants were called with the Bayesian genetic variant detector freebayes.[36] Complex variants (or multinucleotide variants, MNPs) were decomposed to primary SNPs with vcflib.[37] Genetic variants were annotated with SnpEff.[38] Haplotype estimation (phasing) was computed with SHAPEIT.[39,40] *In silico* phased haplotypes were aligned using MegaX,[41] and a neighbor-joining (NJ) tree was obtained with the same software (bootstrap: 1000). The tree was drawn with Interactive Tree Of Life (iTOL) v4.[42] The software POPART[43] was used to infer the TCS network[44] and to draw the georeferenced haplotype map from phased haplotypes and GPS data. To build the TCS haplotype network, two assumptions were made: (i) haplotypes with mutated Trp 574 codon should be highly conserved, because positively selected, whereas haplotypes with the wild-type (WT) Trp 574 codon should be less conserved, because not positively selected; and (ii) haplotypes with mutated Trp 574 codon, but with very low frequency, should be excluded from further analysis as they were likely a result of phasing errors. For these reasons, all sequences with WT Trp 574 codon and those with mutated Trp 574 codon observed with frequencies <4 were excluded from further analysis.

3 RESULTS

3.1 Genetic diversity and population structure

Eight of ten microsatellite loci gave consistent amplification profiles, whereas loci C9333 and C3561 did not, so these latter were excluded from further analyses. The mean number of observed alleles (N_a) across the investigated loci was 4.06, ranging from 1.4 to 6.1 (Table 1). The effective number of alleles ranged from 1.1 to 4.1. For all investigated loci, the effective number, which was estimated based on allele frequencies across all populations, was lower than the observed number. The Shannon information index across these loci was relatively low, ranging from 0.15 to 1.5. Comparable values for observed and expected heterozygosity were recorded for most loci, as also indicated by the F-statistics performed across the investigated loci. Accordingly, the inbreeding coefficient F was on average as low as 0.07. The only exception was the locus AAC1, which had fixation index (F) = 0.6 (Table 1).

At the population level, the mean number of observed alleles ranged from 3.4 to 5.6, whereas the effective number ranged from 1.8 to 3.9. The Shannon information index was on average 0.99,

Table 1. Descriptive statistics of genetic diversity calculated across markers (mean over populations) and *A. tuberculatus* populations (mean over loci), including sample size *N*, N_a (no. alleles), N_e (no. effective alleles), *I* (Shannon information index), H_o (observed heterozygosity), uH_e (unbiased expected heterozygosity) and *F* (fixation index calculated as $1 - (H_o/H_e)$)

	N		N_a		N_e		*I*		H_o		uH_e		*F*	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Locus														
AAC1	10.625	0.625	3.625	0.420	2.526	0.299	1.004	0.124	0.284	0.087	0.588	0.058	0.557	0.126
C3695	11.875	0.125	5.375	0.944	3.704	0.723	1.308	0.230	0.674	0.107	0.652	0.101	−0.087	0.061
C0745	11.875	0.125	5.375	0.420	3.691	0.491	1.397	0.126	0.687	0.071	0.713	0.055	0.008	0.046
C1140	11.750	0.164	6.125	0.611	4.122	0.492	1.536	0.094	0.712	0.053	0.772	0.023	0.032	0.078
C4097	11.750	0.164	3.250	0.559	1.698	0.177	0.660	0.116	0.401	0.108	0.384	0.064	−0.040	0.156
C4999	11.250	0.412	3.875	0.350	2.707	0.244	1.085	0.082	0.558	0.057	0.633	0.044	0.088	0.050
TAG5	11.375	0.324	3.500	0.327	1.797	0.182	0.748	0.101	0.399	0.055	0.419	0.064	−0.031	0.064
ATC9	10.750	0.250	1.375	0.183	1.138	0.081	0.153	0.083	0.093	0.058	0.099	0.056	0.011	0.120
Population														
G	11.375	0.183	2.500	0.423	1.808	0.282	0.603	0.164	0.263	0.095	0.367	0.099	0.228	0.140
F	11.125	0.125	3.750	0.559	2.577	0.422	0.979	0.182	0.508	0.102	0.542	0.095	0.036	0.077
A	11.250	0.366	3.500	0.567	2.189	0.331	0.849	0.169	0.373	0.089	0.479	0.091	0.204	0.111
E	12.000	0.000	5.625	1.051	3.864	0.895	1.285	0.254	0.604	0.109	0.628	0.107	−0.006	0.069
D	11.875	0.125	4.375	0.565	2.685	0.305	1.096	0.139	0.598	0.083	0.602	0.068	−0.051	0.084
H	11.000	0.423	4.375	0.800	2.855	0.562	1.009	0.250	0.411	0.109	0.514	0.127	0.112	0.083
C	11.500	0.378	5.000	0.655	3.184	0.598	1.178	0.206	0.504	0.106	0.597	0.098	0.097	0.111
B	11.125	0.639	3.375	0.460	2.223	0.248	0.890	0.122	0.546	0.117	0.531	0.058	0.021	0.188
Loci and population	11.406	0.121	4.063	0.250	2.673	0.183	0.986	0.068	0.476	0.037	0.533	0.033	0.075	0.040

Table 2. Pairwise population Nei unbiased genetic distance (above diagonal) and F_{ST} values (below diagonal)

	B	G	F	D	H	C	A	E
B		0.393	0.205	0.329	0.292	0.328	0.373	0.315
G	0.181		0.223	0.182	0.296	0.265	0.347	0.251
F	0.093	0.127		0.188	0.155	0.206	0.256	0.150
D	0.112	0.108	0.086		0.121	0.073	0.163	0.122
H	0.128	0.153	0.085	0.072		0.066	0.143	0.099
C	0.114	0.122	0.075	0.048	0.045		0.085	0.019
A	0.142	0.176	0.106	0.084	0.075	0.048		0.032
E	0.111	0.130	0.065	0.064	0.062	0.025	0.036	

ranging from 0.6 to 1.3. The average observed heterozygosity (H_o) was 0.48, but noticeably variable across populations. In detail, although heterozygosity exceeded 0.50 in most populations, which was consistent with the outcrossing reproductive system of *A. tuberculatus*, it was remarkably low in population G ($H_o = 0.26$). High variability also was found in gene diversity (expected heterozygosity, H_e), which ranged between 0.37 in population G to 0.63 in population E. As shown in Table 1, populations G and A displayed signs of moderate inbreeding, as indicated by fixation indices (F) > 0.2.

Pairwise Nei's unbiased genetic distance between populations and F_{ST} values indicated that populations had low genetic distances (mean uNei: 0.20, ranging from 0.02 to 0.39; Table 2). The estimated value of F_{ST} averaged across all comparisons was as low as 0.10, indicating that most genetic variation was found within populations. Genetic flow, which was estimated from F_{ST} in all pairwise comparisons, was on average 3.02, ranging from 1.13 to 9.62 for individual comparisons (Table 3). A high genetic flow was estimated by the pairwise comparisons C and E (9.62) as well as A and E (6.63). Both indices highlighted differences between populations B, F and G and the other populations, which were overall characterized by lower genetic differentiation. This scenario is graphically represented by the NJ tree and principal coordinate analysis [Fig. 1(A) and (B), respectively], which were generated from the pairwise estimates of uNei. Accordingly, both clustering approaches grouped A, D, H, C and E apart from G, F and B. A Mantel test performed by using genetic and geographical distances confirmed the lack of correlation between the two matrices and suggested no isolation-by-distance across the considered sampling range (Mantel correlation coefficient $R_{xy} = -0.257$, *P*-value = 0.280, $\alpha = 0.05$; see also Fig. S6).

By considering the geographical distribution and main reproductive strategy of *A. tuberculatus*, the genetic structure of the populations was investigated by adopting multiple computational

Table 3. Summary of gene flow (N_m) estimates for the investigated populations. Gene flow was estimated as $N_m = [(1/F_{ST}) - 1]/4$

	B	G	F	D	H	C	A	E
B								
G	1.134							
F	2.430	1.719						
D	1.976	2.073	2.648					
H	1.700	1.379	2.702	3.218				
C	1.941	1.797	3.081	4.943	5.332			
A	1.513	1.174	2.107	2.716	3.104	5.001		
E	1.995	1.672	3.571	3.667	3.795	9.622	6.633	

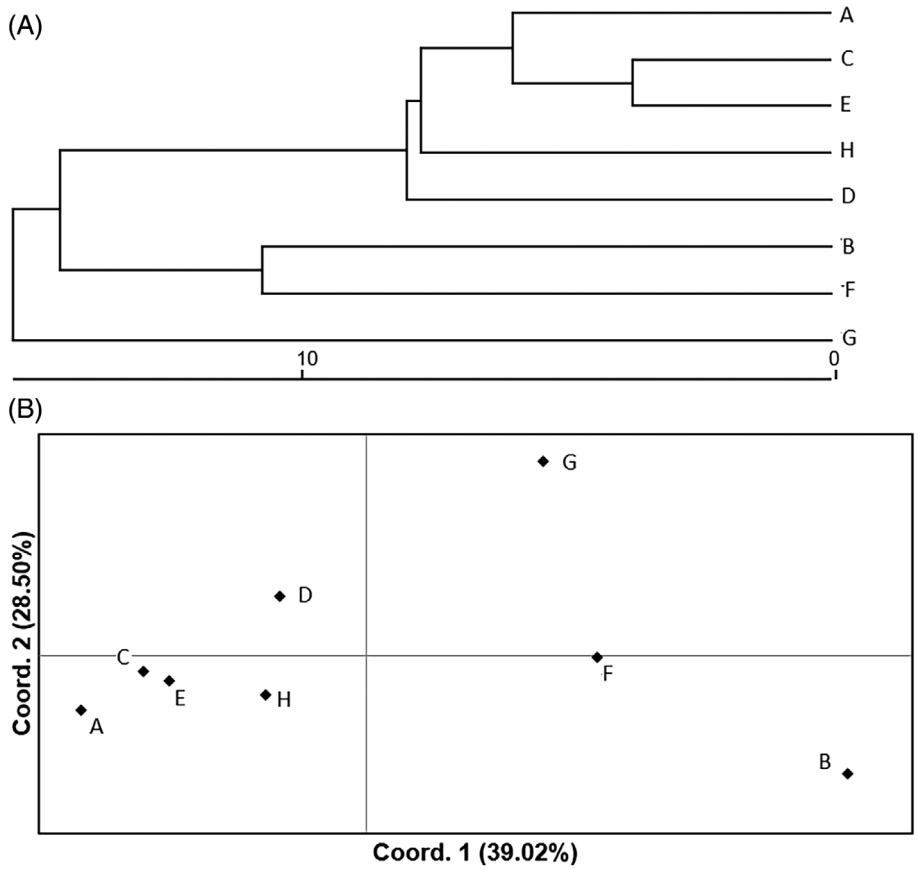

Figure 1. Neighbor-joining tree (A) and principal coordinate analysis (B), generated from the pairwise estimates of uNei between populations.

strategies concerning the admixture model and correlation of allele frequencies. All simulations estimated that the most likely number of ancestors (*K*) is 2 (Fig. S7). Although very low admixture levels (membership >80%) were detected for most individuals (Fig. 2), populations G, F, B and H included individuals with contrasting population assignments, indicating some admixture at population level. All individuals of the wild population (E) grouped within the same ancestral population.

Multiple simulations involving both ancestral and allele frequency models identified an additional level of genetic structure for $K = 4$. In this case, populations sampled in fields located in proximity (<1 km) such as G and F and, to a lesser extent, D and H, clustered apart. Seed dispersal caused by agricultural machinery is quite common within this range, thus proximal populations are expected to have the same genetic structure. For this reason, the scenario depicted for $K = 4$ is unlikely, and the most likely number of ancestral populations is two.

3.2 *ALS* analysis

Three-hundred and eighty-nine SNPs were called along the 3956-bp amplicon, after filtering for high-quality SNPs. One hundred and twelve SNPs were found in the upstream sequence (700 bp), 128 within the coding region (2016 bp) and 149 in the downstream sequence (1238 bp). A model of the gene is reported in Fig. S8, showing all of the mutations known to confer resistance to ALS-inhibitor herbicides. Within the coding region, 54 nonsynonymous

Pest Management Science published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

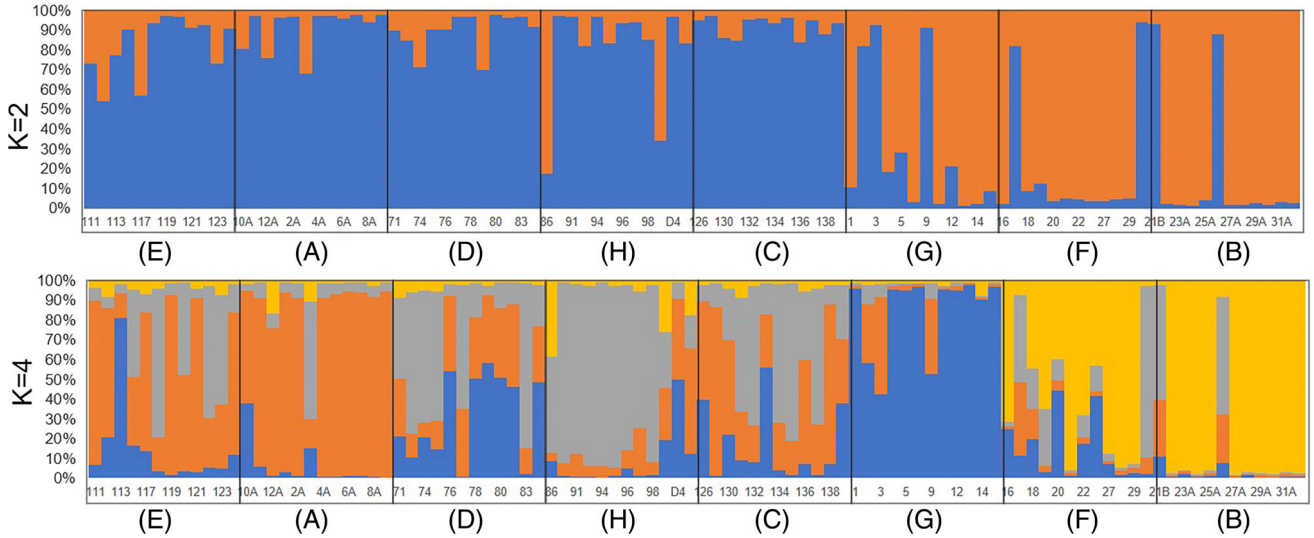

Figure 2. Population assignment by STRUCTURE of individuals for $K = 2$ (upper graph) and $K = 4$ (bottom graph), for the eight genotyped populations. Populations (A), (D), (H) and (C) clustered together with the nonresistant Italian population (E), separately from populations (G), (F) and (B). Population names and single plant codes are shown below the bar graph.

SNPs were found. All sequences were checked for known resistance-endowing mutations (at the following codons: Ala 122, Pro 197, Ala 205, Asp 376, Arg 377, Trp 574, Ser 653, Gly 654).[45] Additionally, sequences were screened for a set of additional mutations, which were shown to be involved in changes in herbicide sensitivity in artificial selection experiments (codons Gly 121, Met 124, Val 196, Arg 199, Lys 256, Met 351, Met 352, Asp 375, Met 570, Val 571 and Phe 578).[46] From this set of nonsynonymous mutations potentially associated with herbicide resistance, a tryptophan-to-leucine change at codon 574 was the only resistance-endowing mutation found across the sequences (single bp substitution of TGG to TTG).

After *in silico* phasing of the sequenced region, all 192 haplotype sequences were used to build a NJ tree (Fig. 3). The NJ tree had three main branches: one mainly grouped haplotypes without the Trp574Leu point mutation (blue clades) and two separate branches mainly grouped haplotypes with the Trp574Leu point mutation (red clades). Alleles with the WT Trp 574 codon were found across the tree, thus indicating they were diverse. Focusing on alleles with the mutated Trp 574 codon that were found with frequencies of ≥4 revealed three different haplotypes. The most common haplotype, hap_1, was found 81 times, whereas the rarer ones hap_2 and hap_3 were found only 16 and four times, respectively. Hap_1 differed from hap_2 and from hap_3 by 45 and 58 SNPs, respectively, whereas hap_2 and hap_3 differed by 41 SNPs (see Fig. 4).

4 DISCUSSION

4.1Genetic diversity, population structure and amplicon sequencing of acetolactate synthase gene

Populations F and G were found in 2010, population A was found in 2014, whereas the others were found in 2017. Populations F, G, A, B, D and H had the same Trp574Leu codon at *ALS*, but it was unclear how they originated and spread. The structure of these populations and *ALS* haplotypes were investigated to clarify their evolutionary history.

Overall, our data on genome-wide genetic diversity are consistent with previous observations on *A. tuberculatus* within its native range and indicated at least two introductions of this species in northern Italy. Gene diversity indicated high within-population genetic diversity, that together with F_{ST} values suggested low genetic differentiation between populations, as had already been reported.[6] On average, observed and expected heterozygosity were similar, except for populations A and G, which had substantially lower *F* values.

Both the dioecious mating system and wind pollination are expected to promote extensive gene flow, allowing for genetic admixture and homogenization across large geographical areas,[6] limited by pollen dispersal. In *A. tuberculatus*, pollen typically disperses <1 km,[47] with pollen-mediated gene flow in glyphosate-resistant *A. tuberculatus* declining by 90% within 100 m.[48] As distances between the eight Italian sampling locations were greater than those reported for effective pollen mediated gene flow, it is reasonable to consider that genetic exchanges must have happened through seed dispersal. Populations D/H and G/F were the nearest accessions (3.5 and 1.4 km apart as the crow flies, respectively). Although uNei genetic distance between populations G and F was relatively high, low genetic differentiation was found for populations D and H. Furthermore, populations E and C had the lowest value of uNei distance despite their origins being 200 km apart. As a result, the genetic clustering of populations appeared to be largely independent of sampling location. In agreement, a Mantel test confirmed the lack of correlation between the two matrices and suggested no isolation by distance across the considered sampling range.

All simulations estimated the most likely number of ancestors (*K*) as two. Populations G, F and B clustered together, apart from a second cluster represented by the remaining populations (i.e. A, D, H, C and E). Although very low admixture levels were detected for most individuals, several populations (namely, G, F, B and H) included individuals with contrasting population assignments, indicating, to some extent, admixture at population level. This genetic population structure is consistent with the estimated high intrapopulation genetic diversity and low interpopulation genetic differentiation.

Amaranthus tuberculatus is an invasive alien species in Europe, native to North America, where the presence of two ancestral populations has been discussed widely.[49,50,6,51] The microsatellites used here were the same as the ones developed previously

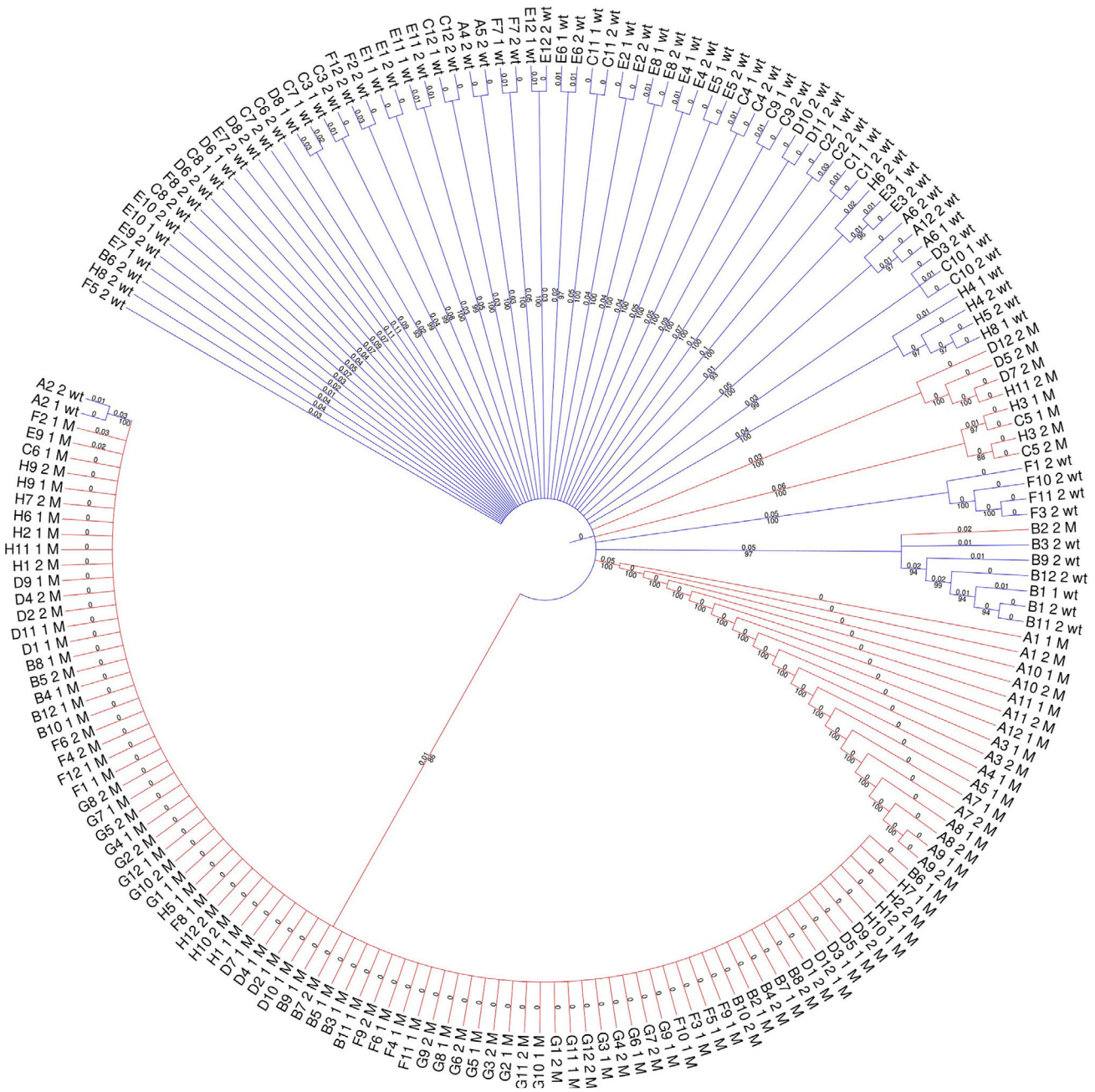

Figure 3. Neighbor-joining tree of *ALS* haplotypes (including singletons and rare haplotypes). Red clades refer to haplotypes carrying the Trp574Leu point mutation (indicated with M) and blue clades refer to the haplotypes without resistance-endowing mutations (indicated with wt). Note that branch lengths are arbitrary, but exact values are reported along with bootstrap value.

to study the structure of North American populations of *A. tuberculatus*.[6] Our microsatellite data may have identified the same two ancestors and therefore the introduction of this species in Italy might have involved both North American native populations at different times. Future investigations performed by using populations derived from both sites and a larger number of loci will possibly help in clarifying the relationship existing between North American and Italian populations.

4.1 Appearance and spread of resistance to ALS herbicides

If ALS-inhibitor resistance in Italy appeared roughly following the timeline of samplings, G and F were the first populations evolving resistance. They were found outside the original area where *A. tuberculatus* was present in Italy and no nonresistant populations were observed nearby. For this reason, it is likely that one or both populations were recently introduced already carrying the *ALS* resistance-endowing alleles. Whether they were both introduced in one event is currently unclear. Population G had the lowest genetic diversity and highest fixation index, possibly indicating that it underwent a drastic reduction in size (bottleneck event) or that it was recently introduced (founder effect). Population B was found geographically quite close to G and F, but seven years later: having the same ancestor and haplotype (hap_1) as population G and F, it likely originated from one of the two through seed dispersal.

Pest Management Science published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

All of the other resistant populations (A, C, D and H) had the same ancestor as the Italian nonALS inhibitor resistant population E. Within this cluster, population A evolved resistance independently, because it did not share *ALS* haplotypes with other populations. Only two plants of 12 of population C had the Trp574Leu point mutation and no further resistance-endowing mutations were found along the *ALS* sequence in other plants: in the other 10 plants the resistance mechanism must be nontarget site-mediated. Therefore, two resistance mechanisms co-exist in this population and resistance evolution must have been independent from the other populations. As stated before, this population is resistant to thifensulfuron-methyl only, but soybean in Italy is normally treated with thifensulfuron-methyl and imazamox, which are tank-mixed. How can thifensulfuron-methyl resistance arise in this situation? Populations A and C were genetically similar to the nonresistant population E, therefore they are possibly derived from it. It is currently not possible to determine whether the single Trp574Leu allele found within population E produced the haplotypes found in A or C (through selection from standing genetic variation) or whether instead seeds of population A or C were dispersed into population E, because the Trp574Leu mutation in population E was found only in a single haplotype. Introduction of ALS-resistant populations from outside remains the most likely event, but our data cannot exclude that selection occurred in Italy for these two cases.

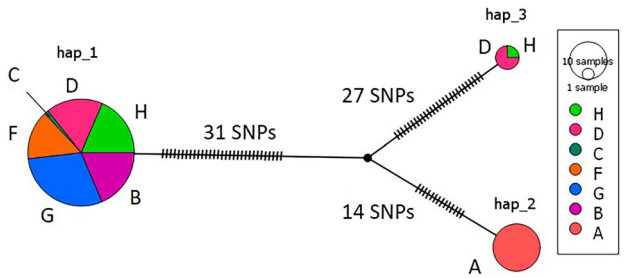

Figure 4. Haplotype network at *ALS* locus (coding and noncoding sequences) considering only haplotypes found at least four times and having the point mutation Trp574Leu. Colors refer to populations where the sequences were found. The circle size is proportional to the frequency of each haplotype. Tick marks along a branch indicate the number of mutations between two neighboring haplotypes.

Populations D and H had the same ancestor as populations A, C and E, but they shared the resistant allele hap_1 with the northern populations B, F and G (which had a different ancestor). Another resistant haplotype, hap_3, was found in these two populations, and therefore population D and H might have evolved hap_3 independently, but also received seed from the northern populations. Intriguingly, seed dispersal caused little, if any, effect on population admixture. This might be due to microsatellite markers being not linked to loci that were subject to selection, whereas the *ALS* haplotypes were selected: genotypes from different populations had mixed without creating evident admixture, because microsatellite loci had neutral fitness. Instead, the *ALS* gene underwent herbicide selective pressure and resistant *ALS* alleles completely replaced the diversity at the *ALS* locus through selective sweeps.

The results support the hypothesis that resistance to ALS-inhibiting herbicides in *A. tuberculatus* populations occurred because of both independent selection and spread of resistant haplotypes. Having excluded gene flow and local selection, the spread of resistance alleles can only be due to seed dispersal, although the underlying mechanisms remain unclear. Seeds of invasive species can be spread in several ways, both human- and/or environment-mediated. There are two scales of spreading: a long-distance dispersal, likely responsible for the introduction of populations G and F from outside Italy, and a short/medium-distance dispersal, responsible for the spreading of the mutated

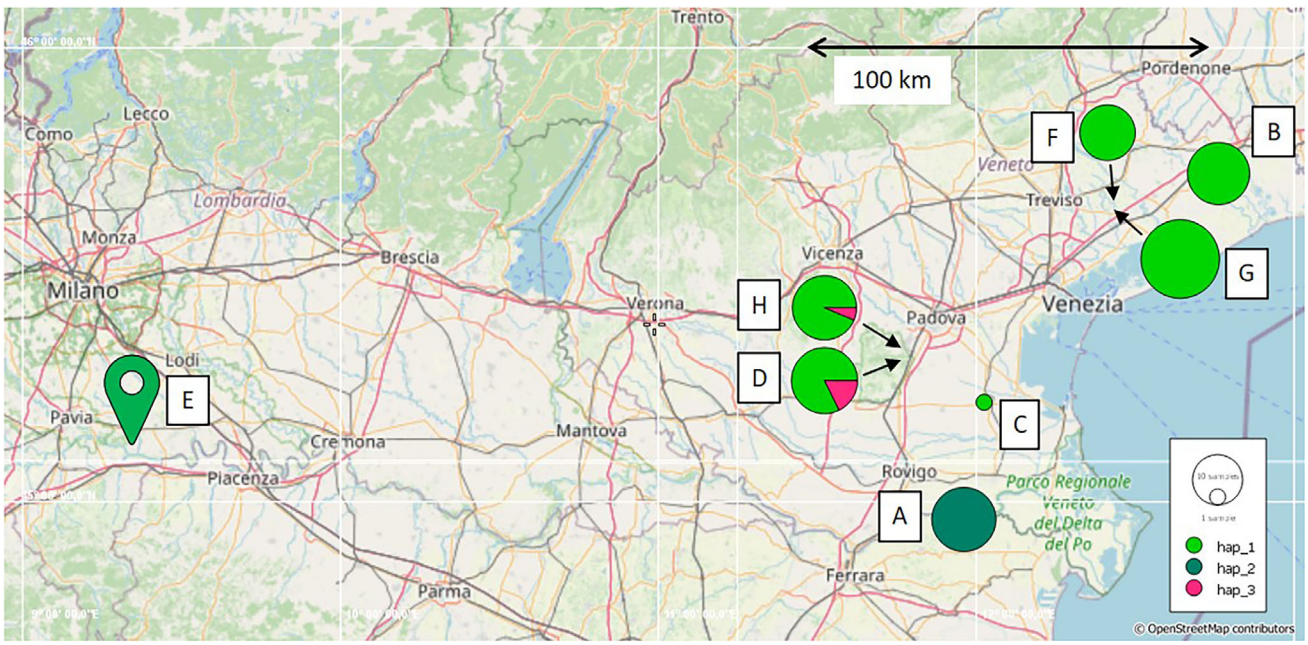

Figure 5. Georeferenced map of haplotype diversity. Each color corresponds to different haplotypes. Note that population E (the dark green pinpoint on the left) had a plant with the point mutation Trp574Leu, but the haplotype was removed from the analyses because of a singleton and it was added to the map for completeness of information. The position of some populations is indicated with arrows to avoid overlapping of pie charts.

Pest Management Science published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

haplotype across northeastern Italy. Long-distance seed dispersal could have occurred through contaminated commercial seeds or animal feed. Short-distance dispersal (0–30 km) could have resulted from contaminated machinery or manures,[52] whereas medium-distance dispersal is likely the consequence of irrigation and rainfall events[53] as well as transportation through migrating wildlife such as ducks and geese.[54] Seeds of *A. retroflexus* also were found in *Perdix perdix* (grey partridge) and *Emberiza schoeniclus* (reed bunting) droppings in Poland.[55,56] All of these species live in the Po Valley[57] and might account for seed dispersal of *A. tuberculatus* in Italy.

Soybean is a genetically stable cleistogamous species and thus some farmers self-produce their next-season seeds. The self-production and the (exclusive) self-use of soybean seeds is allowed in Italy, although not encouraged, but can only explain the dispersal of weed seeds within the boundaries of the farmer's properties. An illegal practice that could explain the introduction of alien weeds is to sow noncertified soybean seeds (e.g. feedstock). A recent report[58] stated that 30% of soybean production in Italy has uncertain origin.

5 CONCLUSIONS

By combining information on the population structure and *ALS* gene sequencing it was possible to formulate a hypothesis of evolutionary history of resistance to ALS-inhibiting herbicides in *A. tuberculatus* in Italy (Fig. 5). Our data support the hypothesis that herbicide resistance evolution started from at least three geographically separated populations. Six of seven populations had the point mutation Trp574Leu, and five of six had the same mutated haplotype, indicating a common origin of resistance. Among these five populations with a common mutated haplotype, the first collected was likely the origin of the mutated haplotype, whereas the others evolved resistance because of seed dispersal from that one. We suspect that this original resistant population was already resistant before its introduction to Italy. One of six resistant populations had a different mutated haplotype, indicating separate evolution. Another population did not have any known endowing-resistance mutation along the whole *ALS* gene (excluding two individuals with the Trp574Leu) and clearly evolved resistance independently from the others. For both of the last two populations, there were no elements to exclude their derivation from a susceptible Italian population.

As general remarks, the self-production of next-season soybean seeds should be avoided, even if allowed, if herbicide resistance is suspected. The presence of even a few plants of *A. tuberculatus* in fields should be promptly recognized and should be controlled at any cost. It also is strongly advised to implement integrated weed management strategies, such as crop rotation, mechanical control and use of pre-emergence herbicides.

ACKNOWLEDGEMENTS

This work was supported by a PhD grant to Andrea Milani funded by the National Research Council of Italy. Work at the MPI for Developmental Biology was supported by the Max Planck Society. Thanks to Alison Garside for revising the English text.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

BIBLIOGRAPHY

1 Tranel PJ, Riggins CW, Bell MS and Hager AG, Herbicide resistances in *Amaranthus tuberculatus*: a call for new options. *J Agric Food Chem* **59**:5808–5812 (2011).
2 Wychen V, Survey of the most Common and Troublesome Weeds in Broadleaf Crops, Fruits & Vegetables in the United States and Canada (2016). http://wssa.net/wp-content/uploads/2016-Weed-Survey_Broadleaf-crops.xlsx [10 July 2019].
3 CAB International, *Amaranthus tuberculatus* (rough-fruited waterhemp), *Invasive Species Compend* (2019). https://www.cabi.org/isc/datasheet/112200 [25 June 2020].
4 Heap I, The International Herbicide-Resistant Weed Database, 2020. http://www.weedscience.org/ [1 October 2020].
5 Tranel PJ, Herbicide resistance in *Amaranthus tuberculatus*†. *Pest Manag Sci* **77**:43–54 (2021).
6 Waselkov KE and Olsen KM, Population genetics and origin of the native North American agricultural weed waterhemp (*Amaranthus tuberculatus*; Amaranthaceae). *Am J Bot* **101**:1726–1736 (2014).
7 Kreiner JM, Giacomini DA, Bemm F, Waithaka B, Regalado J, Lanz C *et al.*, Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc Natl Acad Sci U S A* **116**:21076–21084 (2019).
8 Soldano A, Naturalizzazione in val Padana di « *Amaranthus rudis* » Sauer (Amaranthaceae) esotica nuova per la flora italiana. Segnalazione di altre specie di importazione nuove per alcune regioni dell'Italia settentrionale o per qualche provincia del Piemonte. *Riv Piemont di Stor Nat* **3**:61–70 (1982).
9 Iamonico D, Taxonomic revision of the genus *Amaranthus* (Amaranthaceae) in Italy. *Phytotaxa* **199**:1–84 (2015).
10 Milani A, Scarabel L and Sattin M, A family affair: resistance mechanism and alternative control of three *Amaranthus* species resistant to acetolactate synthase inhibitors in Italy. *Pest Manag Sci* **76**:1205–1213 (2020).
11 Foes MJ, Lui L, Tranel PJ, Wax LM and Stoller EW, A biotype of common waterhemp (*Amaranthus rudis*) resistant to triazine and ALS herbicides. *Weed Sci* **46**:514–520 (1998).
12 Powles SB and Yu Q, Evolution in action: plants resistant to herbicides. *Annu Rev Plant Biol* **61**:317–347 (2010).
13 Délye C, Jasieniuk M and Le Corre V, Deciphering the evolution of herbicide resistance in weeds. *Trends Genet* **29**:649–658 (2013).
14 Palumbi SR, Humans as the world greatest evolutionnary force. *Science* **293**:1786–1790 (2001).
15 Neve P, Busi R, Renton M and Vila-Aiub MM, Expanding the eco-evolutionary context of herbicide resistance research. *Pest Manag Sci* **70**:1385–1393 (2014).
16 MacLean RC, Hall AR, Perron GG and Buckling A, The population genetics of antibiotic resistance: Integrating molecular mechanisms and treatment contexts. *Nat Rev Genet* **11**:405–414 (2010).
17 Resistance is … complex. Nat Ecol Evol **2**:405 (2018). https://doi.org/10.1038/s41559-018-0495-5.
18 Maynard Smith J and Haigh J, The hitch-hiking effect of a favourable gene. *Genet Res* **89**:391–403 (2008).
19 McVean G, The structure of linkage disequilibrium around a selective sweep. *Genetics* **175**:1395–1406 (2007).
20 Hawkins NJ, Bass C, Dixon A and Neve P, The evolutionary origins of pesticide resistance. *Biol Rev* **94**:135–155 (2019).
21 Delye C, Nucleotide variability at the acetyl coenzyme A carboxylase gene and the signature of herbicide selection in the grass weed *Alopecurus myosuroides* (Huds.). *Mol Biol Evol* **21**:884–892 (2004).
22 Menchari Y, Délye C and Le Corre V, Genetic variation and population structure in black-grass (*Alopecurus myosuroides* Huds.), a successful, herbicide-resistant, annual grass weed of winter cereal fields. *Mol Ecol* **16**:3161–3172 (2007).
23 Scarabel L, Varotto S and Sattin M, A European biotype of *Amaranthus retroflexus* cross-resistant to ALS inhibitors and response to alternative herbicides. *Weed Res* **47**:527–533 (2007).
24 Peakall R and Smouse PE, GenALEx 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**:2537–2539 (2012).

25 Yeh FC and Boyle TBJ, Population genetic analysis of codominant and dominant markers and quantitative traits. *Belgian J Bot* **129**:157–163 (1997).
26 Ersts PJ, Geographic Distance Matrix Generator(version 1.2.3), American Museum of Natural History, Center for Biodiversity and Conservation. http://biodiversityinformatics.amnh.org/open_source/gdmg [3 September 2020].
27 Pritchard JK, Stephens M and Donnelly P, Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959 (2000).
28 Evanno G, Regnaut S and Goudet J, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**:2611–2620 (2005).
29 Earl DA and vonHoldt BM, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**:359–361 (2012).
30 Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ, Basic local alignment search tool. *J Mol Biol* **215**:403–410 (1990).
31 R Core Team, R: A Language and Environment for Statistical Computing, 2020. https://www.r-project.org/ [12 December 2020].
32 Monroe JG, genemodel: Gene Model Plotting in R, 2017. https://cran.r-project.org/package=genemodel [1 December 2020].
33 Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G and Sandberg R, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**:2033–2040 (2014).
34 Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv:1303.3997 [q-bio.GN], 1–3 (2013).
35 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079 (2009).
36 Garrison E and Marth G, Haplotype-based variant detection from short-read sequencing, arXiv:1207.3907 [q-bio.GN], 1–9 (2012).
37 Garrison E, A simple C++ library for parsing and manipulating VCF files, MIT. https://github.com/vcflib/vcflib#vcflib [2 September 2019].
38 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**:80–92 (2012).
39 Delaneau O, Marchini J and Zagury JF, A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**:179–181 (2012).
40 Delaneau O, Howie B, Cox AJ, Zagury JF and Marchini J, Haplotype estimation using sequencing reads. *Am J Hum Genet* **93**:687–696 (2013).
41 Kumar S, Stecher G, Li M, Knyaz C and Tamura K, MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**:1547–1549 (2018).
42 Letunic I and Bork P, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**:W256–W259 (2019).
43 Leigh JW and Bryant D, POPART: full-feature software for haplotype network construction. *Methods Ecol Evol* **6**:1110–1116 (2015).
44 Clement M, Snell Q, Walke P, Posada D and Crandall K, TCS: estimating gene genealogies, in *Proceedings of the 16th International Parallel Distribition Process Symposium*, 7 pp, IEEE, New York, NY (2002).
45 Tranel PJ, Wright TR and Heap I, Mutations in herbicide-resistant weeds to ALS inhibitors. http://weedscience.org/mutations/mutationdisplayall.aspx [22 July 2019].
46 Duggleby RG, McCourt JA and Guddat LW, Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiol Biochem* **46**:309–324 (2008).
47 Liu J, Davis AS and Tranel PJ, Pollen biology and dispersal dynamics in waterhemp (*Amaranthus tuberculatus*). *Weed Sci* **60**:416–422 (2012).
48 Sarangi D, Tyre AJ, Patterson EL, Gaines TA, Irmak S, Knezevic SZ *et al.*, Pollen-mediated gene flow from glyphosate-resistant common waterhemp (*Amaranthus rudis* Sauer): consequences for the dispersal of resistance genes. *Sci Rep* **7**:44913 (2017).
49 Sauer J, Revision of the dioecious amaranths. *Madroño* **13**:5–46 (1955).
50 Pratt DB and Clark LG, *Amaranthus rudis* and *A. tuberculatus*, one species or two? *J Torrey Bot Soc* **128**:282 (2001).
51 Kreiner JM, Giacomini DA, Bemm F, Waithaka B, Regalado J, Lanz C *et al.*, Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *Proceedings of the National Academy of Sciences* **116**:21076–21084 (2019). http://doi.org/10.1073/pnas.1900870116.
52 Schwartz LM, Norsworthy JK, Young BG, Bradley KW, Kruger GR, Davis VM *et al.*, Tall waterhemp (*Amaranthus tuberculatus*) and Palmer amaranth (*Amaranthus palmeri*) seed production and retention at soybean maturity. *Weed Technol* **30**:284–290 (2016).
53 Norsworthy JK, Griffith G, Griffin T, Bagavathiannan M and Gbur EE, In-field movement of glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*) and its impact on cotton lint yield: evidence supporting a zero-threshold strategy. *Weed Sci* **62**:237–249 (2014).
54 Webb EB, Farmer JA, Webb EB, Ii AP and Bradley KW, Evaluating the potential for weed seed dispersal based on waterfowl consumption and seed viability. *Pest Manage Sci* **73**:2592–2603 (2017).
55 Orłowski G and Czarnecka J, Granivory of birds and seed dispersal: viable seeds of *Amaranthus retroflexus* L. recovered from the droppings of the grey partridge *Perdix perdix* L. *Polish*. *J Ecol* **57**:191–196 (2009).
56 Orłowski G and Czarnecka J, Winter diet of reed bunting *Emberiza schoeniclus* in fallow and stubble fields. *Agric Ecosyst Environ* **118**:244–248 (2007).
57 International Union for Conservation of Nature and Natural Resources. The IUCN Red List of Threatened Species. Version 2019–2, 2019. https://www.iucnredlist.org/ [1 September 2019].
58 Saggio L, Soia, la tracciabilità tradita del seme, Terra e Vita https://terraevita.edagricole.it/featured/soia-la-tracciabilita-tradita-del-seme/ [1 September 2019].

