



The emergence of a new weed in maize plantations: characterization and genetic structure using microsatellite markers

A. Díaz · A. Taberner · L. Vilaplana

Received: 1 February 2019 / Accepted: 3 September 2019 / Published online: 13 September 2019
© The Author(s) 2019

Abstract The objective of this work was to assess the genetic variability and structure of a new weed in Spanish maize fields, and investigate its geographical patterns using 17 microsatellites. Commercial maize varieties (C), maize-like weeds (MLW), putative hybrids with C (WCH), and teosintes (Tm: *Zea mays* ssp. *mexicana* and Tp: *Z. mays* ssp. *parviglumis*) were analyzed. The weed genetic diversity (MLW and WCH: 0.52) was the lowest (C: 0.59, Tm: 0.66, and Tp: 0.71). Weeds (0.21) and teosintes (Tm: 0.27, Tp: 0.34) showed positive values for the inbreeding

coefficient (F_{IS}), which agrees with their low values for the observed heterozygosity (H_O), common in wild species; whereas C exhibited a negative F_{IS} value (-0.06 , excess of heterozygosity), common in domesticated species. Major clustering agreed with the different types of samples, even if some of the most hybridized weeds branched with the C cluster. Within the weeds, an evident tendency to group together depending on their geographical origin was perceived. Structure analyses confirmed the contribution of C to the genome of those weeds with the highest degree of hybridization. Consistently, the genetic variation (F_{ST}) was not negligible only when the teosintes were compared to the C group. Most of the molecular variance occurred within populations (51.83%) and not among populations (10.09%), with the highest value (32.33%) being found within the weed population. These new weeds seem to have a complex origin. Even if they are related to both, C and teosintes (Tm and Tp), they form an unidentified and genetically distinct group (F_{ST} : 0.13).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10722-019-00828-z>) contains supplementary material, which is available to authorized users.

A. Díaz (✉)
Unidad de Hortofruticultura, Centro de Investigación y Tecnología Agroalimentaria de Aragón, Instituto Agroalimentario de Aragón (IA2), CITA-Universidad de Zaragoza, Avenida de Montañana 930, 50059 Zaragoza, Spain
e-mail: adiazb@cita-aragon.es

A. Taberner · L. Vilaplana
Servicio de Sanidad Vegetal, Departamento de Agricultura, Ganadería, Pesca y Alimentación, Avenida Alcalde Rovira Roure 191, 25198 Lleida, Spain

Present Address:
L. Vilaplana
Nanobiotechnology for Diagnostics (Nb4D) Group, IQAC-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

Keywords *Zea mays* L. · Teosinte · *Zea mays* ssp. *mexicana* · *Zea mays* ssp. *parviglumis* · Weed · Microsatellites

Introduction

It is widely accepted that the wild ancestor of maize (*Zea mays* L.) is a Mexican annual teosinte. The term teosinte spans five species, two perennials: *Zea diploperennis* Iltis, Doebley and Guzman and *Zea perennis* (Hitchc.) Reeves and Mangelsdorf; and three annuals: *Zea luxurians* (Durieu and Ascherson) Bird, *Zea nicaraguensis* Iltis and Benz and *Zea mays* L., to which three teosinte subspecies belong (*Z. mays* ssp. *huehuetenangensis*, *Z. mays* ssp. *mexicana* and *Z. mays* ssp. *parviglumis*), apart from the cultivated maize (*Z. mays* ssp. *mays*). *Z. mays* ssp. *mexicana* and *Z. mays* ssp. *parviglumis* are the most closely related to domesticated maize and that is why they have been included in this study. Though morphologically maize resembles more *Z. mays* ssp. *mexicana*, the molecular studies carried out, from the first using isozymes (Doebley et al. 1984) to the most recent with microsatellite (Matsuoka et al. 2002a, b) and SNP (van Heerwaarden et al. 2011; Hufford et al. 2012) markers, prove that maize was domesticated from *Z. mays* ssp. *parviglumis*.

Maize has been the cereal with the highest production worldwide for at least the last 15 years for which data are available (FAO 2017). Actually, its production has experienced a 76% increase in that period. Spain is in 29th place in the world ranking of producer countries and 10th at European level (FAO 2017). These numbers are not negligible, especially if we take into account that, in terms of surface dedicated to the maize cultivation, Spain occupies the 59th and the 12th position in the world and in Europe, respectively. In this scenario, it is understandable that the discovery of a new weed in Spanish maize fields in summer 2014 has roused the sector to worry. The focal points of this new weed are mainly located in the Northeast of Spain, in the region of Aragon and, to a lesser extent, in Catalonia, and on a smaller scale, in the North of the French region New Aquitaine (EFSA 2016). Those two Spanish regions are among the most productive in maize cultivation, reaching the second and fourth positions, respectively, in the national ranking (Mapama 2016). Moreover, the provinces in which all the infested allotments have been detected are the first (Huesca, in Aragon) and the fourth (Lerida, in Catalonia) Spanish producers in terms of tons of maize (Mapama 2016). The Center for Plant Health and Certification of the Government of Aragon and the

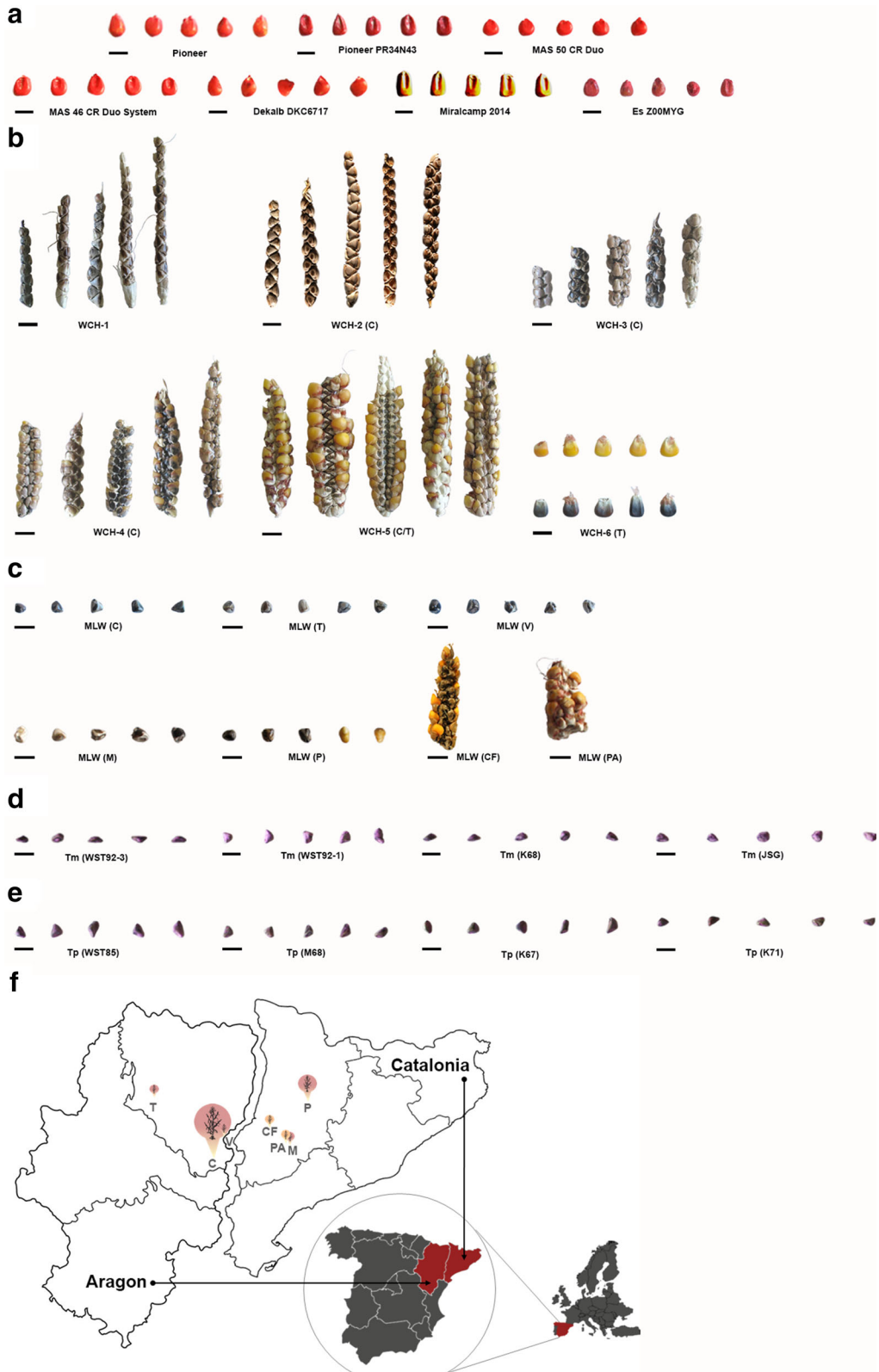
Plant Health Laboratory and Plant Protection Service of Catalonia have taken measurements aimed to restrain this weed or even eradicate it. The crop losses due to this new weed competition are so high that in Huesca (Aragon), the most affected area, some farmers have even plowed the maize plantations little after the sowing. Up to now, the chemical control in maize fields is not feasible as both plants are very similar, not only morphologically but also physiologically. There have been some advances in the description, the monitoring of the more severely affected areas, and the effectiveness of different control methods for this weed which have been called teosinte for its similarity with this maize wild relative (Pardo et al. 2016).

The emergence of a new weed always raises numerous questions and challenges scientist to ascertain its origin. Panoply of explanations are possible in those cases, though they can be narrowed down especially when both, the crop and the weed, belong to the same genus, as it seems to be this case attending to their similar morphology. Some of them could be the transformation of cultivated forms into weed types, going through a transitional feral state; or the colonization of new habitats by a wild relative and its following hybridization with the crop (Ellstrand et al. 2010 and references therein), among others.

A first attempt to taxonomically classify this recently discovered weed, which has been called “Spanish teosinte”, using SNP markers has been made (Trtikova et al. 2017). The authors conclude that the samples collected in Aragon region (the area exhibiting the highest weed infestation) failed to group with any of the teosinte taxa presently defined.

As the identity of this recently discovered maize-like weed is uncertain, the objective of this work was to characterize it using microsatellite markers and to compare their genotypic profiles with those of cultivated and wild relatives to verify if any of them has originated it. This will foreseeably unveil the genetic similarities among groups and the genetic structure of the populations, aided by the information shed by *F*-statistics. A comparison among the samples collected from different infested areas will make possible to detect any existing geographic differentiation.

Microsatellites or Simple Sequence Repeat (SSR) markers have been extensively used to characterize modern maize varieties and landraces, as well as their wild relatives (Matsuoka et al. 2002b; Loáisiga et al.



2011; Warburton et al. 2011; Abakemal et al. 2015; Bedoya et al. 2017; Aci et al. 2018). In general, these markers have revealed themselves as very helpful tools for identification and diversity studies, as much as for establishing genetic relations among populations, due to their multi-allelic, hypervariable and codominant nature. Their robustness and reproducibility make them suitable for comparing different source results, which in this case raises their value as the maize scientific community has been working intensively in this field. Furthermore, the microsatellites developed in maize can be easily transferred to related species since, generally, microsatellite-flanking sequences are highly conserved. This becomes especially useful in those cases in which teosinte belonging to different species than *Z. mays* (i.e. *Z. diploperennis*, *Z. luxurians*, *Z. nicaraguensis*, and *Z. perennis* are objects of the study).

Up to date, this is the first attempt at characterizing this maize-like weed found in Spain using microsatellite markers.

Materials and methods

Plant material

Four different types of samples were used in this study (Fig. 1a–e; Table 1): (1) commercial maize (C); (2) putative hybrids between commercial maize and the weed (WCH); (3) maize-like weeds (MLW); teosinte: (4) *Z. mays* ssp. *mexicana* (Tm); and (5) *Z. mays* ssp. *parviglumis* (Tp). WCH are essentially MLW in which a gradation in terms of hybridization has been

established attending to morphological traits. In the case of C, seven different varieties cultivated in some of the farms in which the weed under study was detected were chosen (Fig. 1a). Only one or two grains from each of them were analysed (total of 10) due to the high homogeneity of commercial maize varieties.

In all the infested fields, an heterogeneous mix of plants supposedly coming from the natural spontaneous hybridization between the commercial maize and the weedy plant could be observed. A ranking from low (WCH-1) to high (WCH-5) degree of hybridization was established (Fig. 1b; Table 1). A gradation in color (from dark to yellow), in hardness of the glumes (from indurate to soft), and in ear architecture (from single to multiple grain-bearing spikelet per rachid), among others, can be easily noticed. Five ears within each category were collected and two kernels per ear were analyzed (up to 10 samples per hybrid). The putative hybrid 6 (WCH-6) consisted of a maize-like cob with dark purple and yellow grains, from which two kernels of each color were selected for further analyses. All the 50 samples of putative hybrids were collected from highly infested parcels in Candasnos and/or Torralba de Aragón (C, T, and C/T).

Regarding the maize-like weeds, up to ten grains were collected from infested fields in seven different locations in Northeastern Spain (Fig. 1c, f), three of them in Aragon: Candasnos (C), Torralba de Aragón (T), and Vencillón (V), and four in Catalonia: Mollerussa (M), Ponts (P), Castellón de Farfaña (CF), and Palau de Anglesola (PA). All the 66 samples were picked up from the ground as this weed has shattering ears, like teosinte, and quite the opposite to the nonshattering ears typical of cultivated maize. The detached grains collected in Aragon showed a hard and dark outer glume similar to the one present in teosinte, whereas the ones coming from Catalonia exhibited some degree of hybridization, as they consisted of a mix of dark and yellow kernels, these last ones with softer and lighter glumes, so more similar to modern corn (Fig. 1a–e). In the case of teosinte, four different accessions of *Z. mays* ssp. *mexicana* and *Z. mays* ssp. *parviglumis* were obtained from the International Maize and Wheat Improvement Center (CIMMYT, Mexico). Up to 20 samples (5 per accession) of each subspecies were used in this work (Fig. 1d, e; Table 1).

◀ **Fig. 1** Phenotype of grains and/or kernels of the plant material studied. Samples are named according to Table 1: **a** commercial maize; **b** weed-commercial maize putative hybrid (WCH); **c** maize-like weed (MLW); **d** teosinte (*Z. mays* ssp. *mexicana*); **e** teosinte (*Z. mays* ssp. *parviglumis*). Locations, three in Aragon (Candasnos: C, Torralba de Aragón: T, and Vencillón: V) and four in Catalonia (Mollerussa: M, Ponts: P, Castellón de Farfaña: CF, and Palau de Anglesola: PA), in which the weedy samples were collected are indicated. Bars = 1 cm. **f** Map of the Iberian Peninsula with the Northeast of Spain zoomed in as the focal area of the new maize-like weed under study. The size of the icons (low, medium or high) represents the number of parcels surveyed and the color (orange: medium; red: high) reflects the level of weed infestation (A. Cirujeda, personal communication). (Color figure online)

Table 1 Description of the plant material used in the present study

Sample name	Name abbreviation	Type ^a	Location	No of samples	No of samples per type	
Weed	MLW (C)	Maize-like weed	Candasnos (Aragon)	10	66	
Weed	MLW (T)	Maize-like weed	Torralba de Aragón (Aragon)	10		
Weed	MLW (V)	Maize-like weed	Vencillón (Aragon)	10		
Weed	MLW (M)	Maize-like weed	Mollerusa (Catalonia)	10		
Weed	MLW (P) ^b	Maize-like weed	Ponts (Catalonia)	10		
Weed	MLW (CF) ^b	Maize-like weed	Castellón de Farfaña (Catalonia)	9		
Weed	MLW (PA) ^b	Maize-like weed	Palau de Anglesola (Catalonia)	7		
Hybrid 1	WCH-1 (C)	Weed-commercial maize putative hybrid	Candasnos (Aragon)	9		50
Hybrid 2	WCH-2 (C)	Weed-commercial maize putative hybrid	Candasnos (Aragon)	10		
Hybrid 3	WCH-3 (C)	Weed-commercial maize putative hybrid	Candasnos (Aragon)	9		
Hybrid 4	WCH-4 (C)	Weed-commercial maize putative hybrid	Candasnos (Aragon)	10		
Hybrid 5	WCH-5 (C/T)	Weed-commercial maize putative hybrid	Candasnos/Torralba de Aragón (Aragon) ^c	8		
Hybrid 6	WCH-6 (T)	Weed-commercial maize putative hybrid	Torralba de Aragón (Aragon)	4		
Accession WST 92-3	Tm (WST92-3)	Teosinte (<i>Z. mays</i> ssp. <i>mexicana</i>)	Mexico	5	20	
Accession WST 92-1	Tm (WST92-1)	Teosinte (<i>Z. mays</i> ssp. <i>mexicana</i>)	Mexico	5		
Accession K68-1	Tm (K68)	Teosinte (<i>Z. mays</i> ssp. <i>mexicana</i>)	Mexico	5		
Accession JS-G-311	Tm (JS6)	Teosinte (<i>Z. mays</i> ssp. <i>mexicana</i>)	Mexico	5		
Accession WST 85-3	Tp (WST85)	Teosinte (<i>Z. mays</i> ssp. <i>parviglumis</i>)	Mexico	5		19
Accession M68-C125	Tp (M68)	Teosinte (<i>Z. mays</i> ssp. <i>parviglumis</i>)	Mexico	4		
Accession K67-20	Tp (K67)	Teosinte (<i>Z. mays</i> ssp. <i>parviglumis</i>)	Mexico	5		
Accession K71-3	Tp (K71)	Teosinte (<i>Z. mays</i> ssp. <i>parviglumis</i>)	Mexico	5		
Pioneer Y72	C (P)	Commercial maize	Catalonia	1		10
Pioneer PR34N43	C (PR)	Commercial maize	Catalonia	1		
MAS 50 CR Duo System	C (CR50)	Commercial maize	Catalonia	2		
MAS 46 CR Duo System	C (CR46)	Commercial maize	Catalonia	1		
Dekalb DKC6717	C (D)	Commercial maize	Catalonia	2		
Miralcamp 2014 ^d	C (M)	Commercial maize	Catalonia	2		
Es Z00MYG	C (Z)	Commercial maize	Aragon	1		

^aThe putative hybrids between the maize-like weed and the commercial maize were ranked from low (WCH-1) to high degree of hybridization (WCH-6)

^bHeterogeneous mix of seeds with different degree of hybridization with maize

^cWCH-5 samples were collected from Candasnos and Torralba de Aragón indistinctively

^dUnknown commercial brand. It is named after the municipal boundary (Miralcamp) where it is grown and the year (2014) in which was collected. Miralcamp is only a few kilometers away from the highly infested areas Mollerusa and Castellón de Farfaña in Catalonia

All seeds were germinated on moistened filter paper in Petri dishes at 37 °C in dark into incubation chambers. After 4–5 days, the seedlings were transplanted to seedbeds filled with standard peat soil at room temperature in a greenhouse.

DNA extraction and microsatellite genotyping

DNA from young leaves was extracted according to protocols previously described (Doyle and Doyle 1990), with slight modifications (Díaz et al. 2017).

A set of 17 microsatellites (Online Resource 1) was chosen among those publicly available in the literature (Chin et al. 1996; Senior et al. 1998; Matsuoka et al. 2002a) and in the Maize Genetics and Genomics Database (MaizeGDB 2015) based in their degree of polymorphism and transferability following the expert's recommendations at CIMMYT. Microsatellites were the markers of choice for this study because neutral selective markers are required if the genotyping data are going to be used for population genetic structure analyses (Pritchard et al. 2000). Furthermore, when possible, microsatellites that map far apart were selected, so they were independent and more likely did not show linkage disequilibrium (LD), as this is also a prerequisite for the structure analyses (Pritchard et al. 2000; Corander et al. 2003). Amplifications were carried out in 20- μ l volume solutions containing 10 ng of genomic DNA, 75 mM Tris-HCl pH 9.0, 50 mM KCl, 20 mM $(\text{NH}_4)_2\text{SO}_4$, 2.5–3.0 mM MgCl_2 , 0.20 mM of each dNTP, 0.2 μ M of reverse and forward (fluorescently labelled with FAM, NED or HEX) primers and 2 U of Biotools DNA polymerase (Biotools, Madrid, Spain). PCRs were performed on a SimpliAmp Thermal Cycler (Applied Biosystems) programmed with an initial denaturation step at 94 °C for 5 min, followed by 35–40 cycles of 94 °C for 30 s, the annealing temperature of each primer pair (Online Resource 1) for 1 min, and 72 °C for 1 min, plus a final elongation step at 72 °C for 10 min. Subsequently, the labelled PCR products were separated on an ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Madrid, Spain) using the internal size standard GeneScanTM-LIZ500 (Applied Biosystems). The GeneMarker software version 2.7.0 (Softgenetics, State College, PA) was used to analyze the results, with peaks checked visually to detect possible errors in the size assignment.

Data analysis

Summary statistics coming from microsatellite genotyping data like the number of alleles and genotypes, major allele frequency (MAF), gene diversity (GD), observed heterozygosity (H_O), polymorphism information content (PIC), and inbreeding coefficient (F_{IS}) was calculated for each locus using PowerMarker 3.25 software (Liu and Muse 2005). Within each type of samples, the number of private alleles was calculated with the software GDA (Genetic Data Analysis) 1.0 (d10c) (Lewis and Zaykin 2001) and the proportion of shared alleles (PSA), as well as the test for Hardy–Weinberg Equilibrium (HWE), with the software GenAIEx 6.5 (Peakall and Smouse 2012), carrying out 1000 random permutations of the values among populations to calculate the upper and lower 95% confidence intervals to compare them to those expected under the null hypothesis, that is, when there are no differences among populations. The presence of null alleles was investigated with the software Micro-Checker 2.3.3 (Oosterhout et al. 2004).

A matrix of genetic distances was created from the genotypic data with the software GenAIEx 6.5 and a phylogenetic network was constructed with the NeighborNet method using the software SplitsTree 4.14.8 (Huson and Bryant 2006).

Principal Component Analysis (PCA) based on the correlation matrix of the allele frequencies was used to determine the associations among the samples. Three components were extracted (PC1, PC2 and PC3) which together accounted for 25.17% of the total variation in the dataset. The analysis was performed using JMP v11.1.1 software for Windows (SAS Institute Inc., Cary, NC, USA).

For the analysis of the population genetic structure, the method based on the unsupervised clustering algorithm implemented in STRUCTURE v2.3.4 (Pritchard et al. 2000) was performed. The admixture model was used to detect any underlying genetic structure across the whole set of individuals. A burning value of 100,000 followed by 100,000 MCMC (Monte Carlo Markov Chain) iterations were selected to run the analysis. Ten independent runs of Gibbs sampler were performed: (1) for the simulation K test ($2 \leq K \leq 6$); and (2) for each value of predefined clusters ($K = 2$, $K = 3$). The delta K (ΔK) method (Evanno et al. 2005) was applied to our simulation data in order to infer the optimal K value. The web

Table 2 Genetic variability measured as genetic diversity (GD) and observed heterozygosity (H_o), differentiation as number of private alleles (Np) and percentage of shared alleles (PSA) and inbreeding (F_{IS}) for the groups studied

Group	Np	PSA (95% C.I.)	GD [†]	$H_o^†$	$F_{IS}^†$
C	1	19.20 (19.58–32.58)	0.56 ^{ab}	0.61 ^a	– 0.04 ^b
MLW and WCH	6	20.58 (25.57–27.61)	0.52 ^b	0.43 ^b	0.17 ^a
Tm	16	29.79 (21.19–30.47)	0.66 ^{ab}	0.49 ^{ab}	0.28 ^a
Tp	31	33.01 (22.17–30.65)	0.71 ^a	0.49 ^{ab}	0.33 ^a

C, commercial; MLW, maize-like weed; WCH, weed-commercial maize putative Hybrid; Tm, teosinte (*Z. mays* ssp. *mexicana*); Tp, teosinte (*Z. mays* ssp. *parviglumis*)

[†]Different letters used to show statistically significant differences

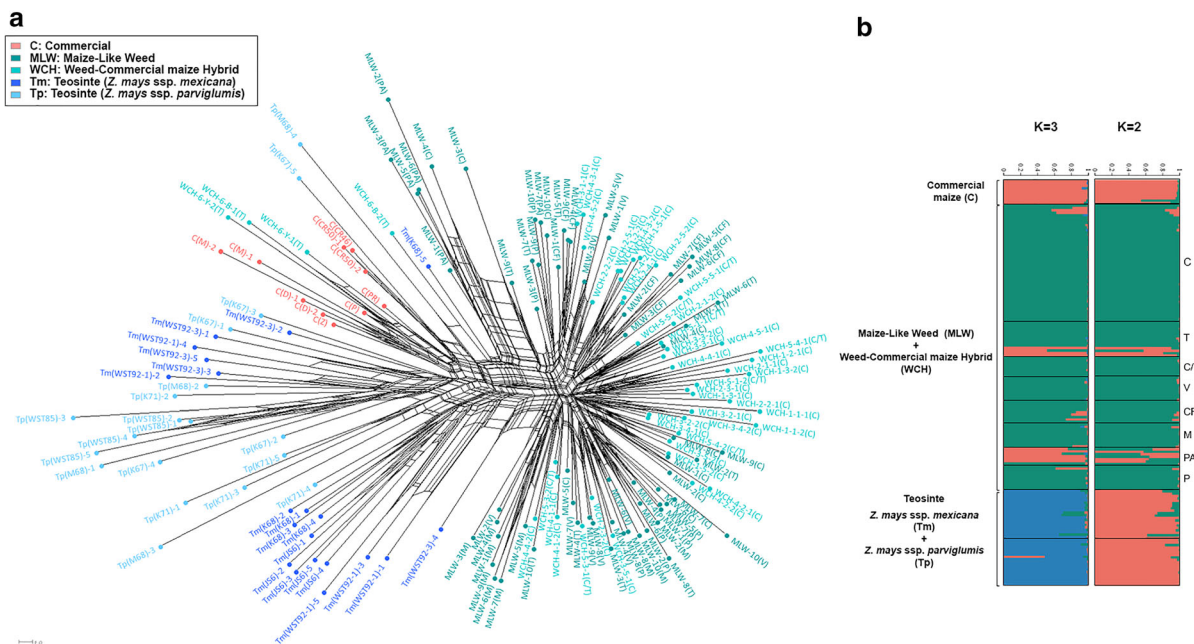


Fig. 2 Genetic structure of the 165 samples studied based on the genotypic data from 17 microsatellite markers. **a** Network built using the NeighborNet method (Fit = 90.13; LSFit = 98.58; Splits = 504). Samples names are according to Table 1; **b** population membership to each of the clusters inferred by the STRUCTURE analysis (K = 2, 3). Each bar represents an

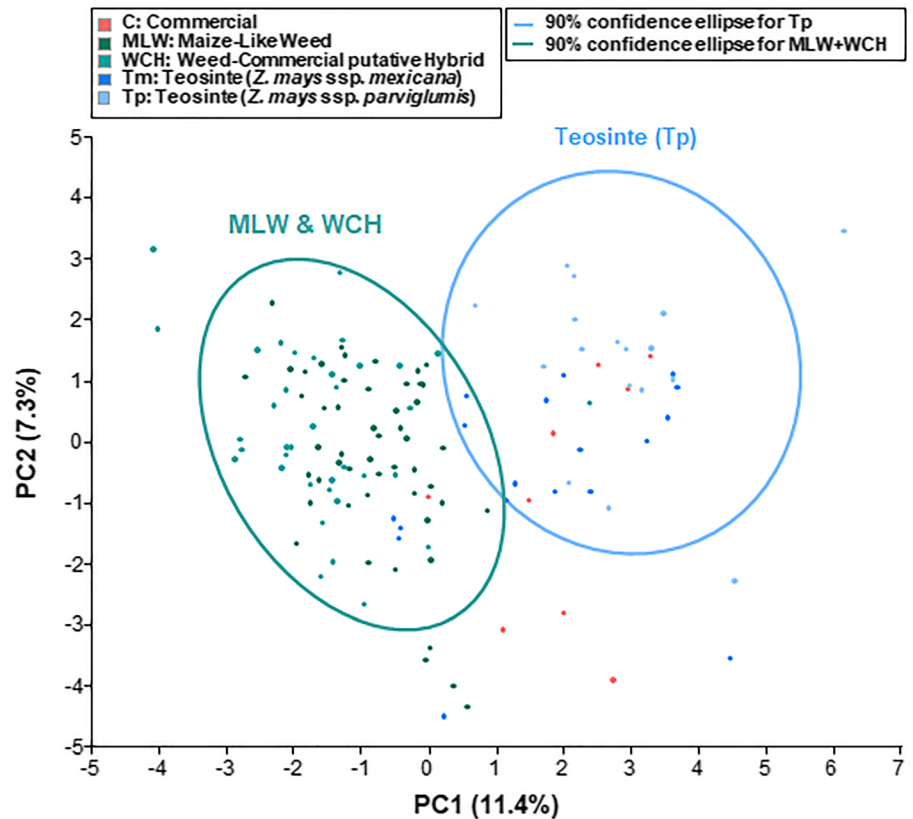
individual and the color reflects the proportion of membership to each cluster. The black boxes demarcate the limits of the different types of samples and the locations in which were collected [only for the maize-like weeds (MLW) and the weed-commercial maize putative hybrids (WCH)]. (Color figure online)

application Structure Plot V2.0 was used to generate the bar plots (Ramasamy et al. 2014). The genetic differentiation among groups (C, MLW and WCH, Tm and Tp) was also estimated by calculating Wright's F_{ST} statistic, as well as the overall fixation index (F_{IT}), at range interval of 95% with 10,000 bootstrapping using PowerMarker 3.25 software. Analysis of molecular variance (AMOVA) was performed with the software PowerMarker 3.25 to detect differentiation among populations and among individuals.

Results

The highest values of GD were exhibited by both teosintes subspecies, Tm and Tp, 0.66 and 0.71, respectively, and the lowest by the group of weeds under study, with the values of the C group falling in between (Table 2; Online Resource 2). The same pattern is shown by the PIC (Online Resource 2). The C group was the one with the highest values of H_o as maize commercial varieties usually consisted of

Fig. 3 Clustering performance of the different types of samples studied. Plot of the first two axes from a principal component analysis (PCA) of 165 samples genotyped with 17 microsatellite markers. Confidence ellipses at 90% are drawn for teosinte *Z. mays* ssp. *parviglumis* (Tp) and the group of weeds (MLW and WCH: maize-like weed and weed-commercial maize putative hybrids). (Color figure online)



hybrids. Apart from that, the group of weeds showed lower values than both teosintes, though those differences were not significant. In terms of parameters useful for identification, both teosinte subspecies, Tp and Tm, presented the highest number of private alleles, 31 and 16, respectively, followed by the group of weeds (6) and, finally, the C group (only 1). The same ordering was observed for the PSA (Tp: 33.01%, Tm 29.79%, MLW and WCH: 20.58%, C: 19.20%, Online Resource 3). The values of the F_{IS} for each group ranged from -0.06 in C to 0.27 and 0.34 in Tm and Tp, respectively. Again, the values of the group of weeds was intermediate (0.21).

The NeighborNet network built with the data coming from genotyping the 165 samples under study with 17 microsatellites markers showed that they were mainly grouped according to the type they belonged to, which were commercial maize (C), weeds (MLW and WCH), and accessions of teosinte (Tm and Tp) from Mexico (Fig. 2a; Table 1). One of the cluster brought together all the samples belonging to C, four putative WCH forms included in the most hybridized class (WCH-6), that is, the most similar to C, and two

teosinte samples from Mexico (*Z. mays* ssp. *parviglumis*, Tp). Virtually, all the weeds prospecting in Aragon and Catalonia (MLW and WCH) branched together in a clearly defined cluster. Within this group, an evident tendency to group together depending on their geographical origin could be perceived. In this sense, numerous subgroups were integrated almost exclusively by samples collected in a single location (like C and C/T, M, PA, CF,...) or in very close locations (like C, V and T). Only one case of samples coming from very diverse locations (up to five: C, T, V, M and P, from the two regions) forming a subgroup could be observed.

Most teosinte accessions fell clearly apart from the two groups described above. Two close subgroups constituted exclusively by members of Tp were easily distinguishable whereas the other teosinte subspecies (Tm) was spread in two more distant subgroups and especially one of them appeared interspaced with some accessions of Tp. The network also showed that the weeds under study and the commercial varieties were generally more closely related with other members within their respective groups than the two

teosinte subspecies when compared with samples within their groups.

The differentiation of the whole set of samples was investigated by PCA (Fig. 3). The first three eigenvectors accounted for 11.36, 7.26 and 6.55% of the genetic variation, respectively, adding up to 25.17%. As in the phylogenetic network, the PCA approach showed that the most clearly differentiated group was the one formed by the weedy plants (MLW and WCH). The teosinte subspecies appeared scattered along both axis, overlapping partially with the C varieties and with the MLW and WCH group, though Tp samples appeared forming a better defined group, just as seen before (Fig. 2a).

Without taking into account prior information about the groups of samples and using the ΔK method, two populations were identified ($K = 2$) and each individual was assigned to either population (Fig. 2b). The weak population structure signals, especially between the two teosinte subspecies, was probably due to their close relationship. That induced us to conduct the analysis specifying several numbers of populations ($K = 2$ and $K = 3$). The pattern of relationships revealed by the structure analysis (Fig. 2b) agreed with that observed in the phylogenetic network (Fig. 2a). These genetic clusters correspond to the three major groups: the C varieties, the weedy plants (MLW and WCH), and the teosintes (Tm and Tp). As previously observed in the NeighborNet network (Fig. 2a) and the PCA graph (Fig. 3a), the distinction between the two subspecies of teosinte is diffused.

The admixture or proportion of the genome of an individual that belongs to each inferred population is represented as K-colored segments within each vertical line (Fig. 2b; Online Resource 4). In general, considering three populations, the highest levels of admixture were observed in the group of weedy plants (MLW and WCH: 17 samples) and, to a lesser extent, in the teosintes (Tm: 3 samples and Tp: 1 sample), whereas no admixture was shown by the C varieties (morphological and genetically uniform). Those three Tm samples ((K68)-5, (WST92-1)-1, and (WST92-3)-4) shared up to 35% of their genetic background with the population of weeds. In the case of the Tp sample that exhibited a clear admixture ((K71)-3), 50% of its genome belong to its own cluster but the other 50% was shared with the C group (Fig. 2b and Online Resource 4).

Among the MLW samples that virtually did not exhibit any admixture at $K = 3$ (Fig. 2b; Online Resource 4), four samples coming from PA in Catalonia (MLW-2, 4, 5, 6 (PA)) shared most of their genetic background with the C varieties (in red in Fig. 2b), which completely agrees with their morphology (Fig. 1c) and reveals their high degree of hybridization with the crop. Specifically, these four MLW individuals exhibited a high membership ($> 96\%$) in the C cluster, whereas the one in their own cluster (MLW and WCH) was very low ($< 3.1\%$). Similarly, three out of the four WCH samples classified as the most hybridized with cultivated maize (WCH-6-B-1 (T), WCH-6-Y-1, 2 (T), Fig. 1b), which as well group together with the C varieties in the phylogenetic network (Fig. 2a), showed almost the same genetic structure as the C varieties, with a membership in the C cluster higher than 97%. On the other hand, their membership in the MLW and WCH cluster was lower than 1.2%. The remaining sample of the most hybridized class (WCH-6-B-2(T)) presented a 50% contribution to its genome from its own population and the other 50% from the C group (Online Resource 4). All these samples are noticeable in the genetic structure plot (in the locations PA and T) because of the high percentage in red within the green cluster of weeds (Fig. 2b).

All plants of C varieties had a high membership in their own cluster ($> 92\%$) for $K = 3$. Noteworthy, the whole C group seems to have derived from a small portion of the huge genetic variability present in the teosinte studied (half red in the Tp section of teosinte population, Fig. 2b). There are three particular individuals, C(D)-1, C(PR) and C(Z), in which the teosinte contribution to their genetic structure was 2.1, 2.1 and 3.6%, respectively.

The average values of the F_{ST} index (Table 3) revealed that the two teosinte subspecies (Tp and Tm) were the populations showing the highest local genetic differentiation (0.32 and 0.27, respectively). In the case of the C group, no structure or differentiation was appreciated (F_{ST} was virtually 0). In between, the weedy group (MLW and WCH) presented a differentiation F_{ST} value of 0.13.

The pair-wise F_{ST} index allowed us to establish comparative measurements of genetic variation between two particular populations (Table 4). The only cases in which such a variation was not negligible

Table 3 Average and per locus genetic differentiation

Locus	F_{ST} (C)	F_{ST} (MLW and WCH)	F_{ST} (Tm)	F_{ST} (Tp)	F_{IT}
phi014	0.00	0.15	0.41	0.25	0.22
phi029	- 0.13	- 0.02	0.41	0.19	0.07
phi031	0.08	- 0.08	0.04	0.41	0.03
phi046	- 0.19	0.24	0.34	0.62	0.26
phi059	0.13	0.15	0.34	0.31	0.19
phi062	- 0.37	- 0.02	0.44	0.50	0.08
phi069	- 0.05	0.15	0.12	0.40	0.18
phi072	0.11	0.10	0.20	0.27	0.14
phi083	- 0.21	0.10	0.48	0.22	0.16
phi084	- 0.57	0.14	0.57	0.31	0.21
phi085	0.00	0.18	0.40	0.36	0.22
phi115	- 0.11	0.17	0.24	0.30	0.17
phi102228	- 0.19	- 0.15	- 0.23	0.18	- 0.11
phi108411	- 0.18	0.49	0.11	0.49	0.32
phi227562	0.24	0.46	- 0.13	0.49	0.36
phi308707	0.15	0.56	0.46	0.29	0.42
phi331888	0.07	0.09	0.18	0.09	0.10
Mean	- 0.05	0.13	0.27	0.32	0.17
SD	0.05	0.04	0.05	0.03	0.03

F_{ST} values among the groups described in Table 1 and F_{IT} in the whole set of samples
C, commercial; MLW, maize-like weed; WCH, weed-commercial maize putative Hybrid; Tm, teosinte (*Z. mays* ssp. *mexicana*); Tp, teosinte (*Z. mays* ssp. *parviglumis*)

Table 4 Pairwise F_{ST} values for group differentiation

Group	F_{ST} (C)	F_{ST} (MLW and WCH)	F_{ST} (Tm)	F_{ST} (Tp)
F_{ST} (C)	0.00			
F_{ST} (MLW and WCH)	0.02	0.00		
F_{ST} (Tm)	0.15	- 0.13	0.00	
F_{ST} (Tp)	0.14	- 0.16	0.04	0.00

C, commercial; MLW, maize-like weed; WCH, weed-commercial maize putative hybrid; Tm, teosinte (*Z. mays* ssp. *mexicana*); Tp, teosinte (*Z. mays* ssp. *parviglumis*)

was when the teosinte populations (Tm and Tp) were compared to the C group.

Two levels of classification were considered in the AMOVA analysis (Table 5): (1) the populations obtained in the NeighborNet network and the genetic structure analyses but considering the two teosinte subspecies separately (C, MLW and WCH, Tm and Tp), and (2) the individuals. The highest percentage of variation (32.33%) was found within the population composed by the weedy plants (MLW and WCH), whereas the lowest value (2.49%) was observed within the C group. This pattern could also be perceived at the individual level, as MLW and WCH presented the highest percentage of variation (25.19%) and C the lowest (3.07%). The genetic differentiation among populations (10.09%) was lower than the sum within

both, populations (51.83%) and individuals (38.08%). The two subspecies of teosinte (Tm and Tp) showed very similar values of percentage of variation at both levels, within population (8.28 and 8.73%, respectively) and within individual (4.98 and 4.74%, respectively). Moreover, those values were intermediate between those exhibited by MLW and WCH and C within populations (32.33 and 2.49%, respectively) and within individuals (25.29 and 3.07%, respectively).

Discussion

MLW and WCH clustering together is a reasonable result as this is an artificial division, done only with

Table 5 Genetic variation at population and individual levels

Source of variation	Sum of squares	Variation (%)
Among populations	338.30	10.09
<i>Within population</i>		
C	83.66	2.49
MLW and WCH	1084.25	32.33
Tm	277.62	8.28
Tp	292.65	8.73
<i>Within individuals</i>		
MLW and WCH	848.00	25.29
C	103.00	3.07
Tm	167.00	4.98
Tp	159.00	4.74
Total	3353.47	100.00

Analysis of molecular variance (AMOVA) using the genotyping data coming from 17 SSR markers

C, commercial; MLW, maize-like weed; WCH, weed-commercial maize putative hybrid; Tm, teosinte (*Z. mays* ssp. *mexicana*); Tp, teosinte (*Z. mays* ssp. *parviglumis*)

classification purposes during sample collection. In fact, there were weed samples (i.e. MLW (PA)) that looked more like the maize crop than some of the hybrids (Fig. 1a–c), and coincidentally, they form a subgroup within the weed cluster closer to the C group. The geographical grouping within this cluster proves the limited dispersion of the different genotypes, what is always desirable in the case of weeds. The few cases of exchange between the different crop zones could be explained by the propagation through the combine harvesters, as those farmers frequently contract the same equipment.

The genetic variability of the group of weeds studied here is lower than the observed in the teosinte accessions from both subspecies, especially Tp, which seems to be the most diverse among our whole set of samples, including also the C group. These higher levels of gene diversity found in teosinte subspecies compared to maize cultivars have also been reported before (Matsuoka et al. 2002a). The number of private alleles present in both teosintes is also high (especially in Tp) considering that Tm, Tp and C belongs to the same species. Interestingly, this weed group is the one in which the PSA value fell further out from the 95% confidence interval (below the lower limit), that is, the relatedness within that group is lower than the expected under the null hypothesis which assumes

no difference among populations (Online Resource 3). Both the weed group and the teosintes showed positive values for F_{IS} (deficiency of heterozygous), typically observed in wild species. Those values were higher than the ones reported before for these same two subspecies of teosinte (Aguirre-Liguori et al. 2017), probably due to the fact that in our study all the samples within each subspecies belonged only to four different accessions (Table 1). Similarly, in the case of the weeds under study, several seeds were collected from the same kernel (especially in the case of WCH), so they happened to be half-siblings. Quite the opposite, the C group exhibited a negative value for the F_{IS} (6% excess of heterozygous), as it is common in domesticated species, mainly those consisting of hybrids coming from inbred parental lines.

In agreement with their higher genetic diversity, the teosinte accessions in the PCA graph were less tightly clustered than the group of weeds under study, confirming their higher diversity. The vague separation between some samples of the two teosinte subspecies in the network (Fig. 2a), the PCA graph (Fig. 3a), and the structure analysis (Fig. 2b) could be also due to the fact that the prospected samples are originally classified in the germplasm banks attending mainly to morphological criteria. In fact, some taxonomic errors in teosinte accessions from genebanks and herbaria have been recently reported (Sánchez González et al. 2018). Another possible explanation is the putative hybridization between the two subspecies of teosinte given that their native habitats overlap in Central and Northern Mexico (Sánchez González et al. 2018). This has been suggested before in a phylogenetic study in which 237 teosinte plants were genotyped with 93 microsatellite markers, resulting in the individuals of *Z. mays* ssp. *parviglumis* and *Z. mays* ssp. *mexicana* largely but not completely separated (Fukunaga et al. 2005). The possibility of the lack of resolution due to the low number of markers and teosinte samples cannot be either dismissed. In this sense, in an exhaustive study carried out with 646 teosinte individuals belonging to the same two teosinte subspecies (*Z. mays* spp. *parviglumis* and *Z. mays* spp. *mexicana*) and 33,464 SNPS, they were separated in different clusters though, interestingly, the individuals in the boundary between both subspecies showed clear genetic similarity (Aguirre-Liguori et al., 2017).

In terms of genetic distance, major clustering agreed with the different types of samples, maize varieties, weeds and teosintes, being the three of them clearly separated. Even though, some of the most hybridized weeds (WCH-6) were precisely included in the same cluster as C varieties or in their proximity (MLW(PA), Fig. 2a). These results are compatible with the two hypotheses pointed out before, the gradual transformation of the crop into weed types and/or the hybridization between the crop and the wild forms. Regarding the teosinte samples grouped with the C maize varieties, similar results have been reported in a previous work (Matsuoka et al. 2002b). The authors found a closer relationship between maize grown nowadays and teosinte belonging to *Z. mays* ssp. *parviglumis* in comparison to *Z. mays* ssp. *mexicana* also using microsatellite markers (both, our and their study, have genotyped samples from five teosinte accessions in common).

Unlike previous observations (Trtikova et al. 2017), the C maize varieties and the weedy (MLW and WCH) plants partially overlapped in the PCA graph. This can be explained by the fact that the C varieties chosen in this study were precisely those cultivated in the regions affected by these new weeds. Indeed, in many cases, they were the varieties cultivated in the same farms in which the weedy plants were prospected; whereas in the work above (Trtikova et al. 2017), maize varieties generally grown in Spain were used. So, it is not preposterous to think that the C maize varieties chosen for the present study are the ones that have most probably participated in the hybridization with the maize-like weeds to render the weedy hybrids.

Even if the Tp samples here studied may not contain the exact putative ancestor of the cultivated maize, the virtual lack of admixture observed in the group of C varieties and the total coincidence of the genetic structure with a small portion present in that Tp population, support the idea of a genetic relationship between maize and *Z. mays* ssp. *parviglumis*. This is also backed by the grouping of a few Tp samples with all the C varieties in the phylogenetic network here constructed (Fig. 2a).

The low genetic variation between different populations could explain the fact that the genetic structure simulation analysis followed by the ΔK method for calculating the number of populations were unable to distinguish more than two populations

(Fig. 2b). All this is also in agreement with the AMOVA results, which show that most of the molecular variance occurred principally within populations (a total of 51.83%), secondly within individuals (a total of 38.08%), and residually among populations (10.09%). Similar results were obtained in a study in which the lowest values of molecular variance were found among races and/or clusters of American maize accessions. Likewise, the highest molecular variation percentages were observed within races and/or clusters and within plants, though in this case, the latter was higher (Vigouroux et al. 2008).

Conclusions

The high genetic similarity between three C samples from two varieties developed by the same seed breeding company (CR46 and CR50) and one accession of teosinte (*Z. mays* ssp. *parviglumis*), could suggest the use of this kind of exotic material in breeding programs, though it does not seem a common practice (Warburton et al. 2017). There is a relationship between the C maize varieties and some of the weed under study (those with the highest degree of hybridization). However, its origin must be somehow more complex than the derivation from the crop exclusively, otherwise the genetic background of both whole populations should be very similar (and not only for a few MLW and WCH plants, as we observed in this study, Fig. 2b).

Many crosses between maize and teosinte have been reported since several decades ago (Collins and Kempton 1920; Mangelsdorf and Reeves 1939; Mangelsdorf 1947; Doebley et al. 1990). Maize-related weeds, like those recently emerged in Spain, are able to efficiently pollinize the domesticated maize plants, whereas the fertilization in the reciprocal direction does not seem to be so successful (Trtikova et al. 2017). The gene called *teosinte crossing barrier1* (*Tcb1*) has been identified and reported to play a restrictive role in the crossability of teosinte with maize (Evans and Kermicle 2001), among others (Evans and Kermicle 2001; Kermicle and Evans 2010). This explains the asymmetrical gene flow between *Z. mays* ssp. *parviglumis* and maize (Baltazar et al. 2005), as well as between *Z. mays* ssp. *mexicana* and maize (Ellstrand et al. 2007) previously observed. According to this, it is reasonable to expect that after a

few generations from the first hybridization between the weed and the domesticated maize, the population quickly drifts towards weediness. Actually, the hybrids resulting from applying teosinte pollen on maize silks are described to be vigorous and highly fertile (Evans and Kermicle 2001). Under this hypothesis, in the gradation of the putative hybrids collected for our study (Fig. 1b), the most similar to the cultivated maize (WCH-6) would be the closest to the original cross between the crop and the weed. As the pollination by the weedy plants seems to be favored to the detriment of the fertilization by the domesticated maize (propagation of the crossing barrier strong allele), on-going crosses with weeds would render more and more weedy plants (from WCH-5 to WCH-3), as it has also been described in teosinte (Kermicle 2006). At the same time, maintaining morphological similarities with the crop (as is the case of these MLW found in Spain) decreases the chances of being removed by the farmers, especially in the case of maize plantations as the presence of weeds is one of the main causes of yield reduction. However, other samples (i.e. WCH-1, 2; MLW(C, T, V, M, P) in Figs. 1, 2a) are better differentiated from both, C varieties and teosintes. Concurrently, attending to their morphology, some plants resembled more teosinte as they produced minute kernels with hard, stony fruitcases like those found on them, whereas others bore bigger kernels with yellow soft outer glumes similar to maize. So, multiple origins seems to be underlying these heterogeneous array of individuals which, due to its recent emergence (Pardo et al. 2016), still needs to be unambiguously elucidated.

The microsatellite profile of this new weed emerged in Europe does not match any of the teosinte samples analyzed. Due to the complexity of teosinte diversity, with the existence of different species, subspecies and races, it becomes necessary to broaden the spectrum of samples to genotype in order to establish more comparisons. In this sense, accessions from *Z. mays* ssp. *huehuetenangensis*, the remaining subspecies within the same species as the cultivated maize and the teosintes here studied (*Z. mays* ssp. *mexicana* and *Z. mays* ssp. *parviglumis*) could be included as it has not been compared to this new weed before. Any of the not sampled clades of teosinte or even some extinct species could be in the origin of this maize-related weed.

Valuable tools are being developed to protect and preserve teosinte genetic diversity not only in its centers of origin but also in its dispersion areas (Sánchez González et al. 2018). At the same time, great efforts are made to eradicate it from maize plantations. There is no doubt that it is a challenging situation for scientists that requires coordinated actions to come up with a satisfactory solution.

Acknowledgements The authors thank J. A. Aranjuelo for technical support; A. Cirujeda, G. Pardo, J. M. Montull, J. M. Llenes for prospecting the maize-like weeds; International Maize and Wheat Improvement Center (CIMMYT, Mexico) for supplying the teosinte accessions; and D. L. Goodchild for reviewing the English language. This work was funded by the Spanish Institute for Agricultural and Food Research and Technology (INIA) through the project E-RTA-2014-00011-C02.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abakemal D, Hussein S, Derera J, Semagn K (2015) Genetic purity and patterns of relationships among tropical highland adapted quality protein and normal maize inbred lines using microsatellite markers. *Euphytica* 204:49–61. <https://doi.org/10.1007/s10681-014-1332-9>
- Aci MM, Lupini A, Mauceri A, Morsli A, Khelifi L, Sunseri F (2018) Genetic variation and structure of maize populations from Saoura and Gourara oasis in Algerian Sahara. *BMC Genet* 19:51. <https://doi.org/10.1186/s12863-018-0655-2>
- Aguirre-Liguori JA, Tenailon MI, Vázquez-Lobo A, Gaut BS, Jaramillo-Correa JP, Montes-Hernandez S, Souza V, Eguiarte LE (2017) Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol Ecol* 26:4226–4240. <https://doi.org/10.1111/mec.14203>
- Baltazar BM, Sanchez-Gonzalez JJ, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theor Appl Genet* 110:519–526. <https://doi.org/10.1007/s00122-004-1859-6>
- Bedoya CA, Dreisigacker S, Hearne S, Franco J, Mir C, Pransanna BM, Taba S, Charcosset A, Warburton ML (2017)

- Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PLoS ONE* 12:e0173488. <https://doi.org/10.1371/journal.pone.0173488>
- Chin E, Senior ML, Shu H, Smith JS (1996) Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39:866–873. <https://doi.org/10.1139/g96-109>
- Collins GN, Kempton JH (1920) A teosinte-maize hybrid. *J Wash Acad Sci* 19:1–37
- Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163:367–374
- Díaz A, Martín-Hernández AM, Dolcet-Sanjuan R, Garcés-Claver A, Álvarez JM, García-Mas J, Picó B, Monforte AJ (2017) Quantitative trait loci analysis of melon (*Cucumis melo* L.) domestication-related traits. *Theor Appl Genet* 130:1837–1856. <https://doi.org/10.1007/s00122-017-2928-y>
- Doebley J, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (Gramineae). *Syst Bot* 9:203–218. <https://doi.org/10.2307/2418824>
- Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F₂ population: implications for the origin of maize. *Proc Natl Acad Sci USA* 87:9888–9892. <https://doi.org/10.1073/pnas.87.24.9888>
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- EFSA (European Food Safety Authority) (2016) Relevance of new scientific evidence on the occurrence of teosinte in maize fields in Spain and France for previous environmental risk assessment conclusions and risk management recommendations on the cultivation of maize events MON810, Bt11, 1507 and GA21. EFSA supporting publication. EN-1094. <https://www.efsa.europa.eu/en/supporting/pub/en-1094>
- Ellstrand NC, Garner LC, Hegde S, Guadagnuolo R, Blancas L (2007) Spontaneous hybridization between maize and teosinte. *J Hered* 98:183–187. <https://doi.org/10.1093/jhered/esm002>
- Ellstrand NC, Heredia SM, Leak-García JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evol Appl* 3:494–504. <https://doi.org/10.1111/j.1752-4571.2010.00140.x>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Evans MMS, Kermicle JL (2001) Teosinte crossing barrier1, a locus governing hybridization of teosinte with maize. *Theor Appl Genet* 103:259–265. <https://doi.org/10.1007/s001220100>
- FAO (2017) FAOSTAT: Statistics Division of Food and Agriculture Organization of the United Nations. <http://faostat.fao.org/>
- Fukunaga K, Hill J, Vigououx Y, Matsuoka Y, Sanchez GJ, Liu K, Buckler ES, Doebley J (2005) Genetic diversity and population structure of teosinte. *Genetics* 169:2241–2254. <https://doi.org/10.1534/genetics.104.031393>
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44:808–811. <https://doi.org/10.1038/ng.2309>
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267. <https://doi.org/10.1093/molbev/msj030>
- Kermicle JL (2006) A selfish gene governing pollen-pistil compatibility confers reproductive isolation between maize relatives. *Genetics* 172:499–506. <https://doi.org/10.1534/genetics.105.048645>
- Kermicle JL, Evans MM (2010) The *Zea mays* sexual compatibility gene *ga2*: naturally occurring alleles, their distribution, and role in reproductive isolation. *J Hered* 101:737–749. <https://doi.org/10.1093/jhered/esq090>
- Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet. <http://lewis.eeb.uconn.edu/lewishome/software.html>
- Liu K, Muse SV (2005) Powermarker: integrated analysis environment for genetic marker data. *Bioinformatics* 21:2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>
- Loáisiga CH, Brantestam AK, Díaz O, Salomon B, Merker A (2011) Genetic diversity in seven populations of Nicaraguan teosinte (*Zea nicaraguensis* Iltis et Benz) as estimated by microsatellite variation. *Genet Resour Crop Evol* 58:1021–1028. <https://doi.org/10.1007/s10722-010-9637-6>
- MaizeGDB (2015) Maize genetics and genomics database. <https://www.maizegdb.org/>
- Mangelsdorf PC (1947) The origin and evolution of maize. In: Demerec M (ed) *Advances in genetics*, vol 1. Carnegie Institution, Cold Spring Harbor, New York, pp 161–207
- Mangelsdorf PC, Reeves RG (1939) The origin of Indian corn and its relatives. *Tex Agric Exp Station Bull* 574:1–315
- MAPAMA (2016) Ministry of Agriculture and Fisheries, Food and Environment. <https://www.mapama.gob.es/es/>
- Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002a) Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. *Theor Appl Genet* 104:436–450. <https://doi.org/10.1007/s001220100694>
- Matsuoka Y, Vigououx Y, Goodman MM, Sanchez J, Buckler E, Doebley J (2002b) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* 99:6080–6084. <https://doi.org/10.1073/pnas.052125199>
- Oosterhout CV, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Pardo G, Cirujeda A, Marí AI, Aibar J, Fuertes S, Taberner A (2016) El teosinte: descripción, situación actual en el valle del Ebro y resultados de los primeros ensayos. *Vida Rural* 408:42–48

- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus* 3:431. <https://doi.org/10.1186/2193-1801-3-431>
- Sánchez González JDJ, Ruiz Corral JA, García GM, Ojeda GR, Larios LDLC, Holland JB, Romero GEG (2018) Eco-geography of teosinte. *PLoS ONE* 13:e0192676. <https://doi.org/10.1371/journal.pone.0192676>
- Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci* 38:1088–1098. <https://doi.org/10.2135/cropsci1998.0011183X003800040034x>
- Trtikova M, Lohn A, Binimelis R, Chapela I, Oehen B, Zemp N, Widmer A, Hilbeck A (2017) Teosinte in Europe—searching for the origin of a novel weed. *Sci Rep* 7:1560. <https://doi.org/10.1038/s41598-017-01478-w>
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Sanchez Gonzalez JJ, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci USA* 108:1088–1092. <https://doi.org/10.1073/pnas.1013011108>
- Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez González J, Doebley J (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am J Bot* 95:1240–1253. <https://doi.org/10.3732/ajb.0800097>
- Warburton ML, Wilkes G, Taba S, Charcosset A, Mir C, Dumas F, Madur D, Dreisigacker S, Bedoya C, Prasanna BM, Xie CX, Hearne S, Franco J (2011) Gene flow among different teosinte taxa and into the domesticated maize gene pool. *Genet Resour Crop Evol* 58:1243–1261. <https://doi.org/10.1007/s10722-010-9658-1>
- Warburton ML, Rauf S, Marek L, Hussain M, Ogunola O, Sanchez Gonzalez J (2017) The use of crop wild relatives in maize and sunflower breeding. *Crop Sci* 57:1227–1240. <https://doi.org/10.2135/cropsci2016.10.0855>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.