ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Wild sunflower goes viral: Citizen science and comparative genomics allow tracking the origin and establishment of invasive sunflower in the Levant

Sariel Hübner[1] | Dana Sisou[1,2,3] | Tali Mandel[1] | Marco Todesco[4] | Maor Matzrafi[2] | Hanan Eizenberg[2]

[1]Galilee Research Institute (MIGAL), Tel-Hai Academic College, Upper Galilee, Israel

[2]Department of Phytopathology and Weed Research, Agricultural Research Organization, Newe Ya'ar Research Center, Ramat Yishay, Israel

[3]The Robert H. Smith Institute of Plant Sciences and Genetics, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel

[4]Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence
Sariel Hübner, Galilee Research Institute (MIGAL), Tel-Hai Academic College, Upper Galilee, Israel.
Email: sarielh@migal.org.il

Funding information
Israel Ministry of Agriculture, Grant/Award Number: 21-01-0034

Handling Editor: Ana Caicedo

Abstract

Globalization and intensified volume of trade and transport around the world are accelerating the rate of biological invasions. It is therefore increasingly important to understand the processes through which invasive species colonize new habitats, often to the detriment of native flora. The initial steps of an invasion are particularly critical, as the introduced species relies on limited genetic diversity to adapt to a new environment. However, our understanding of this critical stage of the invasion is currently limited. We used a citizen science approach and social media to survey the distribution of invasive sunflower in Israel. We then sampled and sequenced a representative collection and compared it with available genomic data sets of North American wild sunflower, landraces and cultivars. We show that invasive wild sunflower is rapidly establishing throughout Israel, probably from a single, recent introduction from Texas, while maintaining high genetic diversity through ongoing gene flow. Since its introduction, invasive sunflower has spread quickly to most regions, and differentiation was detected despite extensive gene flow between clusters. Our findings suggest that rapid spread followed by continuous gene flow between diverging populations can serve as an efficient mechanism for maintaining sufficient genetic diversity at the early stages of invasion, promoting rapid adaptation and establishment in the new territory.

KEYWORDS
ecological genetics, hybridization, invasive species, natural selection and contemporary evolution, population dynamics

1 | INTRODUCTION

Invasive species can significantly challenge the native ecological system and may pose a severe threat to ecosystem services, biodiversity, agriculture and human health (Pimentel et al., 2005).

Worryingly, the rate of invasions is increasing with expansion of global trade, transport and climate change (Hulme, 2009; Seebens et al., 2015), while efforts to control or prevent this fail to keep pace (Bradley et al., 2010). Despite their harmful impact, invasive species represent exceptional examples of contemporary evolution

and rapid adaptation, being able to establish themselves and flourish in new environments where conditions may be significantly different from their original habitat (Bock et al., 2015; Colautti & Lau, 2015). Moreover, successful invasive species are able to overcome the genetic barriers that are associated with dispersion bottlenecks and expansion load, and maintain high genetic and/or phenotypic diversity (Dlugosch & Parker, 2008). To explain these remarkable features, several mechanisms of diversity maintenance have been suggested, including admixture before or after invasion, repeated independent introductions, the size of the invading population, and continuous gene flow through bridgehead populations (Bock et al., 2015). Indeed, hybridization, introgression and local adaptation were shown to play a key role in boosting adaptation through accommodation of beneficial genetic variation (Hodgins et al., 2018; Todesco et al., 2020). For example, a recent study on teosinte invasion and establishment as a noxious weed in Europe highlighted the role of introgression of a herbicide resistance gene from maize in promoting adaptation to agricultural ecosystems within few decades (Le Corre et al., 2020). Other examples of rapid local adaptation have provided a more general understanding on how the interplay between demographic and adaptive processes lead to the establishment of an invasive species quickly across continents (van Boheemen et al., 2017; Colautti & Lau, 2015).

The process of invasion is well framed and can roughly be divided into three phases: introduction, initial establishment and expansion of the invading species in the new habitat. The early stages of invasion are critical for the chances of successful establishment (Bock et al., 2015). Massive or repeated introductions of invasive populations that harbour large standing genetic variation will allow a quicker transition to the establishment and expansion phases. Thus, deciphering the early stages of the invasion process is critical to understanding the fate of the invasion, and can guide the development of management and control practices (Hulme, 2006). Nevertheless, most studies of invasive species focus on ancient introductions, partially because monitoring the early stages of invasion and divergence from the source population remains challenging. Another challenge of studying an alien species is the limited information available on its distribution in the invaded region, especially for recently introduced species that are not included in the local flora documentation. Failure to properly represent the distribution of the target species may result in biased inferences on the level of genetic variation and divergence (Hübner & Kantar, 2021).

Common sunflower (*Helianthus annuus* L.) is an annual species native to North America that occurs across a wide ecogeographical range, from temperate climates in Canada to arid climates in northern Mexico (Kantar et al., 2015). Wild sunflower can also evolve into a weedy form in locations proximate to cultivated fields (Kane & Rieseberg, 2008), and crop–wild introgressions have been reported for breeding purposes, too (Hübner et al., 2019). The lack of crossing barriers between the cultivated sunflower and the weedy form were reported even outside the distribution range of wild sunflower, and promoted adaptation to the agricultural environment, including herbicide resistance (Ellstrand et al., 2013; Massinga et al., 2003;

Presotto et al., 2012). In some cases, weedy sunflower has developed into a noxious weed that poses a threat to agriculture in different regions around the world (Muller et al., 2011). There is therefore an urgent need to study the evolution of invasive sunflower outside of its native range, with important implications for agriculture, conservation and evolutionary biology.

The Mediterranean region, and particularly the Levant, has been a recipient area for biological invasions due to its location as a bridging land between three continents, and to intense human movement and activity over several millennia. Commerce and trade have increased dramatically the pace of invasion globally and specifically in Israel, which relies heavily on import of agricultural commodities (Rubin & Matzrafi, 2015). The first record of invasive wild sunflower in Israel dates to the 1970s, when limited spread was reported mainly in the northwest (Dafni & Heller, 1980). Previous attempts to identify the origin of invasive sunflower in Israel suggested a crop–wild ancestry (Lai et al., 2012) in accordance with other reports for weedy sunflower in Europe (Muller et al., 2011).

In this study, we explore the process of establishment and spread of invasive wild sunflower in Israel. To address this, we survey the distribution range of the invasive populations using a citizen science approach, which allowed us to explore a wide geographical range efficiently and a to directly communicate with citizens to promote awareness of invasive species. Next, we used genomic data to identify the population of origin in North America, and trace the route of its spread across the invaded region in Israel. We further address the contribution of crop–wild interactions and the demographic factors that facilitate the accumulation and maintenance of genetic diversity through rapid spread.

2 | MATERIALS AND METHODS

2.1 | Tracing and sampling invasive sunflower populations

To locate wild sunflower populations that represent the species current distribution across Israel, we took a citizen science approach. To reach out as broadly as possible, while targeting an audience with high engagement, we used the social media platform Facebook to communicate the project goals, guidelines and general information including photos to help to correctly identify wild sunflower plants. In addition, a WhatsApp group was created to further expand the outreach of the project and to record geographical coordinates of detected invasive populations and photos for validation (Figure S1). These platforms were selected because of their high popularity in Israel, specifically among 25–50 year-olds, who were anticipated to be the most committed and contributing group. Both platforms were managed in Hebrew to encourage direct and friendly communication with a broad local audience. Reported observations were validated by experts with available photos shared by participants, and recorded coordinates were used to draw geographical and environmental data for each

detected population. To select sites for sampling, a grid was projected on top of recorded coordinates and representing sites were selected at each 10-km^2 grid cell (Figure S2). This approach was previously shown to be effective for sampling in Israel due to the high correlation between geographical coordinates and environmental variables (Hübner et al., 2009) and was chosen to guarantee proper sampling without restriction to specific locations. Many of the recorded coordinates were located in private areas, urbanized or isolated locations with fewer than 10 different plants, and inaccessible spots, and thus were excluded from the survey. Following these steps, 22 sampling locations that represent the distribution range and environmental spectrum where invasive sunflower occurs in Israel were chosen for sampling (Figure S3, Table S1). At each location, mature seeds were collected from each of 10 plants and stored at 4°C.

2.2 | DNA extraction and genotyping by sequencing

Seeds from all the collected plants were germinated, and genomic DNA was extracted from the young leaves of a single seedling for each mother plant. Leaves were collected directly into liquid nitrogen and ground. Prewarmed CTAB buffer (100 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 1.4 M NaCl, 2% w/v CTAB, 0.5% w/v NaHSO$_3$; Sarokin & Carlson, 1984) was added and the samples were incubated at 65°C for 60 min. After a short cool-down, cold phenol/chloroform/IAA (indole acetic acid) (25:24:1) mixture was added and the samples were vigorously vortexed and centrifuged for 5 min at 20,000 g at 4°C. The upper phase was transferred into a new tube, and chloroform/IAA (24:1) mixture was added. Samples were again vigorously vortexed and centrifuged for 5 min at 20,000 g at 4°C. Next, the upper phase was transferred again into a new tube and treated with 5 μl RNase A (10 mg ml^{-1}, abm-G117) for 30 min at 37°C. Once again chloroform/IAA (24:1) was added, vortexed, centrifuged and the upper phase was transferred to a new tube, followed by NaCl/isopropanol precipitation for 30 min at −20°C and two ice-cold 70% ethanol washes. Finally, samples were eluted in 50 μl low TE buffer (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA pH 8.0). Among samples, 22 did not germinate or yield enough high-quality DNA for library preparation and a total of 178 samples proceeded to the library preparation step.

For genotyping-by-sequencing (GBS) library preparation, 200 ng of genomic DNA was digested with two restriction enzymes (PstI, NEB-R3140; and MspI, NEB-R0106) for 5 h at 37°C (Poland et al., 2012). Each sample was ligated with a unique barcoded adapter and a common adapter for 2 hr at 22°C, and was cleaned and concentrated using 1.6× volumes of SPRI beads (Fisher Scientific; #09-981–123), prepared following Rohland and Reich (2012). Next, each sample was amplified using 2× KAPA HiFi HotStart ReadyMix (KAPABIOSYSTEMS; KK2600) with primers that produced complete Illumina adapters, and quantified using the Qubit dsDNA BR Assay Kit (Thermo; Q32850). From each

of the 178 libraries, 100 ng was pooled and concentrated, and quantified using the Qubit dsDNA BR Assay Kit. Pooled libraries were loaded on 1.5% agarose gels and DNA fragments in the range 400–600 bp were cut and extracted using Wizard SV Gel and PCR Clean-Up System (Promega; A9281) and quantified with the Qubit dsDNA BR Assay Kit. We used duplex-specific nuclease (DSN) normalization (Shagina et al., 2010) to reduce the portion of repetitive sequences in the samples, and re-amplified the pooled-depleted library. Fragment size was analysed using the Agilent High Sensitivity D1000 ScreenTape (Agilent; 5067–5584).

The pooled library was first sequenced on a MiSeq system (Illumina) using the V2 Nano kit, to validate and confirm normalization between samples, and then on four lanes of the Illumina NextSeq550 to generate 2× 150-bp reads in high-output mode.

2.3 | Variant calling and genotyping procedure

Raw sequence data were trimmed and demultiplexed using the "process_radtag" command in STACKS (Rochette & Catchen, 2017) with two mismatches allowed in the adapter sequence (Table S2). Demultiplexed reads from each sample were aligned to the XRQv1 reference genome (Badouin et al., 2017) using BWA MEM with default parameters (Li, 2013). Alignment files were processed and used to call variants across all invasive accessions in one batch using the HaplotypeCaller algorithm in GATK4 (Poplin et al., 2017) with "heterozygosity" and "heterozygosity-stdev" parameters set to 0.01 and 0.1, respectively, to account for the high level of heterozygosity expected in wild sunflower (Todesco et al., 2020). Raw variants were filtered for a minimum minor allele frequency of 5%, maximum of 30% missing data, minimum variant quality of 30, minimum genotype quality of 30 and minimum depth of 5 reads. Individuals with more than 70% missing data were excluded from downstream analysis.

To compare the invasive populations to domesticated germplasm (sunflower association mapping [SAM] population) and wild *Helianthus annuus* populations sampled in North America, the data set generated by Todesco et al. (2020) was obtained and filtered to include all *H. annuus* samples (wild and domesticated) and variants with VQSR tranches higher than 90. After each data set was filtered separately (see parameters above), the two vcf files were merged by coordinates using the vcf-merge command in VCFTOOLS (Danecek et al., 2011). Variants that were completely absent in at least one of the groups (invasive, domesticated or North American wild) were excluded from the merged data set and additional filtering for a maximum of 30% missing data across all samples was applied.

2.4 | Population genomics and statistical analyses

Population stratification analysis for the invasive sunflower samples was conducted with SNMF and principal components analysis

(PCA) as implemented in the LEA package in R (Frichot & François, 2015). For SNMF, different K values were tested between 2 and 10, with 10 replicates for each K followed by a cross entropy test that was performed across all K values to identify the best number of clusters. In the PCA, the Tracy–Widom test was performed to identify the number of eigenvectors that best represent the data set as an indication for the population structure among invasive sunflower samples. For validation, the program FAST-STRUCTURE (Raj et al., 2014) was used, followed by a "chooseK" analysis of K values of 2–10 to identify the number of clusters among invasive populations.

To identify the source of the invasive populations in Israel, a PCA was conducted with SMARTPCA as implemented in the EIGEN-SOFT package (Patterson et al., 2006) using the genetic data from the wild and domesticated samples obtained from Todesco et al. (2020). Invasive samples were projected on top of the generated PC space to avoid the bias introduced by the excess of missing data among invasive samples that were genotyped using GBS. To further reduce the stretching effect of projection, the shrink mode was activated in the analysis. This procedure allowed us to reduce the bias of using different data sets in a PCA. In addition, a neighbour-joining (NJ) network was constructed with SPIT-STREES4 (Huson, 1998) using all single nucleotide polymorphisms (SNPs) that passed the filtering procedure. As a complementary approach, a supervised machine learning method implemented in the package Locator (Battey et al., 2020) was used to predict the geographical source of the invasive populations based on genomic data. All wild samples obtained from Todesco et al. (2020) for which geographical coordinates are available ($n = 614$) were included in the analysis. Invasive samples were considered of unknown location and their geographical origin was predicted by the model. The analysis was replicated 10 times by bootstrapping over SNPs and results were summarized across all replicates.

To further study the differentiation and relatedness among invasive samples at each sampling location, global F_{ST} and the f_3-statistic were calculated using the ADMIXTOOLS package (https://github.com/uqrmaie1/admixtools) and a graph was constructed using the ggraph package in R (https://cran.r-project.org/web/packages/ggraph). Relatedness was calculated for each pair of samples using the level of identity by state calculated with SNPREL-ATE (Zheng et al., 2012). Signs of introgression from domesticated sunflower into invasive populations were examined using the f_3-statistic as implemented in ADMIXTOOLS. Population genetics statistics including Tajima's D, nucleotide diversity (π), heterozygosity and Tajima's θ were calculated globally for each chromosome in each population, and in 1-Mbp windows using the popGenome package in R (Pfeifer et al., 2014) and VCFTOOLS.

Environmental data for each invasive population was obtained from the WorldClim database (Fick & Hijmans, 2017) based on recorded coordinates and included annual mean temperature (BIO1), maximum temperature in warmest month (BIO5), temperature annual range (BIO7), annual precipitation (BIO12), and precipitation seasonality (BIO15).

3 | RESULTS

3.1 | Tracing invasive sunflower populations

Common sunflower is native to North America and does not occur naturally in the Levant. Therefore, no official records are available for the distribution of wild sunflower in Israel. To detect "natural" populations of wild sunflower, we took a citizen science approach using social media platforms (Facebook and WhatsApp). People responded positively in both platforms and participants shared photos, comments and coordinates of wild sunflower populations from across the country. Interestingly, several reports on intentional seed transfer between regions due to a misidentification of wild sunflower as an ornamental plant indicated that people are actively spreading the species. After approximately one month of activity, national news channels reported about the project on traditional media (TV and newspapers), which significantly increased the activity and interest on the social media platforms (Figure 1a). After circa 100 days of activity, the project reached over 20,000 participants on Facebook, and a similar activity was achieved on the WhatsApp platform (Figure 1a). According to the profile of participants on Facebook (https://www.facebook.com/business/insights/) the main age group was 25–55 years (information from WhatsApp was unavailable), and higher participation was noticed among women (Figure 1b). A total of 156 locations were reported and cross-validated with photos and information shared by participants. Eventually, 22 locations were chosen for sampling, of which two sites (Ramla and Natanya, centre region) were later excluded because herbicide treatments were applied by local authorities before the populations could be sampled. Thus, a total of 20 sites were sampled and seeds from 10 different individual plants were collected (Figure 1c,d).

3.2 | The origin of invasive sunflower populations

From these 20 populations, 178 individuals, each germinated from seeds deriving from a different mother plant, were genotyped using a GBS protocol (Poland et al., 2012) modified for sunflower. Sequence data were processed and analysed, yielding a total of 3,850,685 variants across all individuals. The data set was extensively filtered for low-quality SNPs and five accessions with an excess of missing data were removed, leaving a total of 28,540 SNPs genotyped across 173 individuals. Next, the variants called in the invasive population were merged with an available SNP data set generated for wild sunflower populations from North America, landraces and cultivated varieties comprising the SAM population (Todesco et al., 2020). Following merging, a total of 20,169 SNPs genotyped across 1,200 individuals were kept for downstream analysis.

To explore the source of invasive sunflower in Israel, a PCA was conducted for the North American wild populations in addition to landraces and cultivars from the SAM population (Figure 2a). The invasive populations were projected on the PC space and were broadly clustered with the wild North American populations with tendency
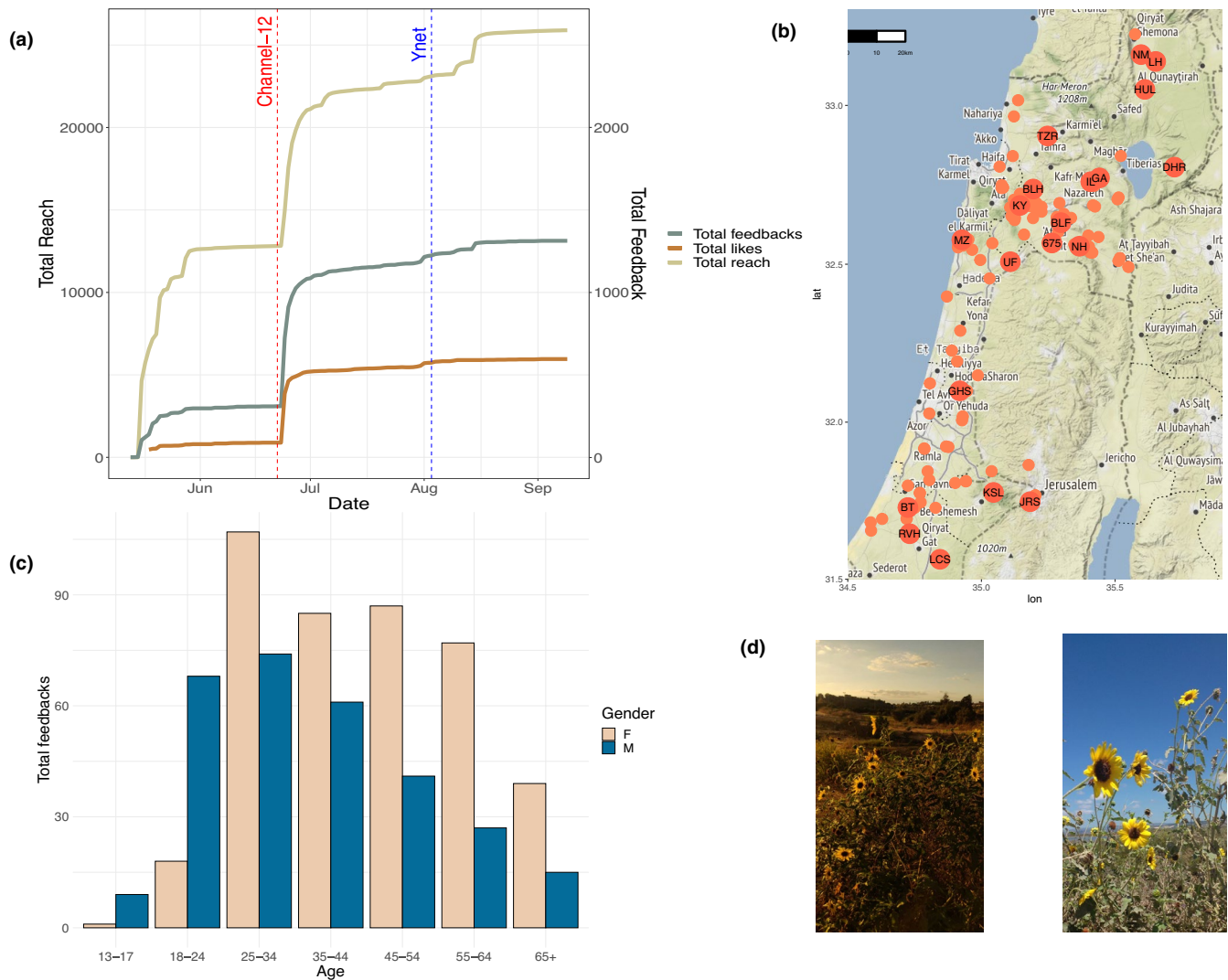
FIGURE 1 Tracing and sampling invasive sunflower populations. (a) Participation and engagement rate in the citizen survey of wild sunflower populations. Activity on the Facebook platform is shown across the duration of the project. Red and blue vertical lines mark the project coverage on the television channel and digital news website, respectively. (b) Distribution of participants' profile by age and gender. (c) Locations of all invasive sunflower records obtained from the survey based on coordinates provided by participants. Selected sampling sites are indicated with darker colour and the site name; a scale bar is indicated at the top left. (d) Photos of wild sunflower at two representative sampling sites

towards populations collected in southwest USA (Figure 2b). Due to the projection procedure, the invasive samples were scattered across the PC space despite the lower diversity captured in this population compared with the North American populations (see below). To further explore the similarity between the invasive samples and remaining populations, an NJ network was conducted with the same data set and further supported the PCA results, indicating that the invasive populations are monophyletic with a major single source in Texas (Figure 2c). To validate these results, a population differentiation test was conducted based on pair-wise F_{ST} between each wild North American population and the entire invasive population to maintain comparable sample sizes (Figure S4). Lowest F_{ST} scores with the invasive population were obtained for Texas (F_{ST} =0.010) followed by populations from Kansas (F_{ST} =0.016) and New Mexico (F_{ST} =0.018), suggesting that the source population is in southwest

USA. To refine the identification of the geographical source of the invasive populations, a machine learning analysis was applied using the genetic data from North American and invasive populations in addition to the geographical coordinates available for the North American populations. The analysis was replicated 10 times with good fit of the training runs (R^2 = .99), and pointed to Texas as the source for the invasive sunflower populations (Figure 2d).

To further track the entrance point and spread of wild sunflower in Israel, the f_3-statistic was calculated between the Texan population and each sampling site using the population from Utah as an outgroup. Conceptually, both f_3 and F_{ST} can measure the differentiation between populations, but f_3 is less prone to bias caused by differences in sample size. Low differentiation was observed between the Texan population and all invasive sampling sites as expected for the short time since the invasion (first records in the 1970s).
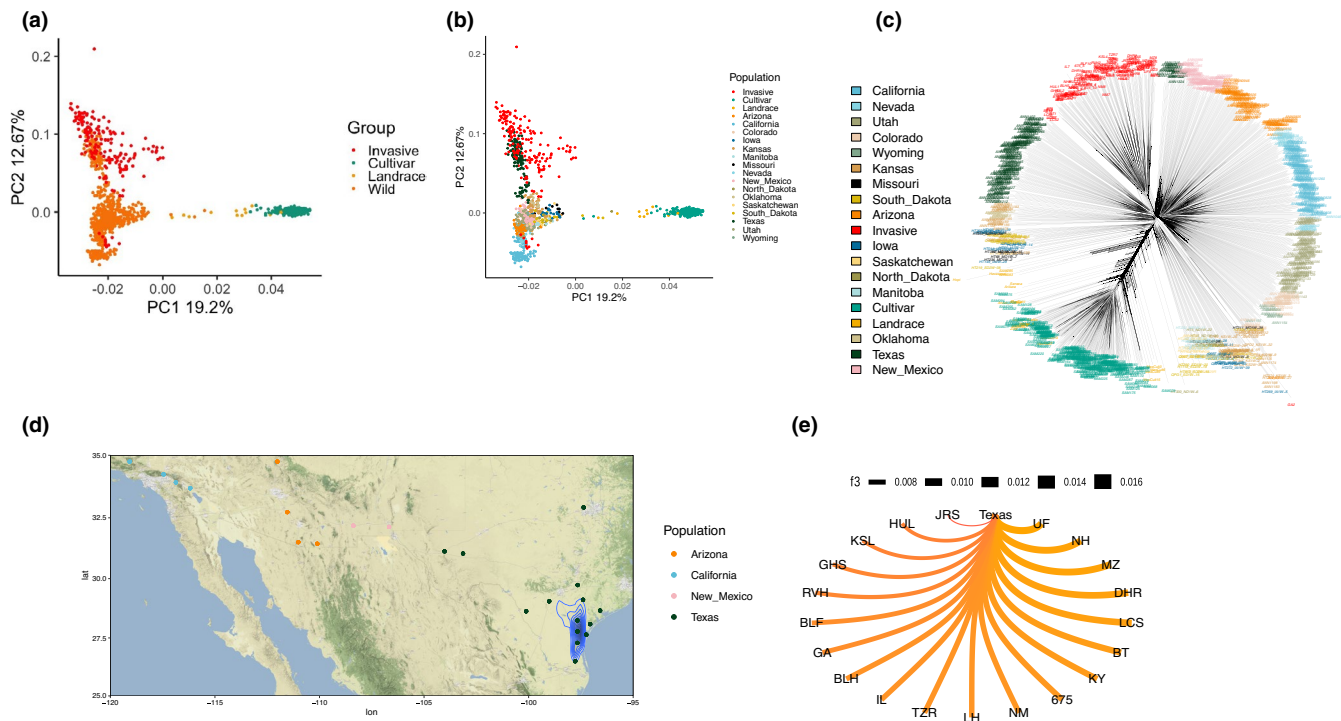
FIGURE 2 Identifying the source of invasive wild sunflower populations in Israel. (a) Principal component analysis coloured by type (wild, domesticated, invasive), and by country (b). (c) Neighbour-joining network of invasive, domesticated and wild North American populations. (d) Prediction of the geographical source of the invasive population using a machine learning approach. Coloured points on the map correspond to North American wild populations from different states (indicated in the key). The blue contour lines indicate the predicted source of invasive populations where the density of lines corresponds to the number of samples assigned to geographical position. (e) Graph calculated from the f_3-statistic between the Texan population and each invasive sunflower sampling site; thicker lines correspond to higher f_3 values

However, differences between sampling sites were observed and indicated that the closest populations (highest f_3 scores) to the Texan origin are Umm el Faheem (UF; $f_3 = 0.162$), Ma'ayan Zvi (MZ; $f_3 = 0.158$) and Nahal Harod (NH; $f_3 = 0.159$). These sites are located in northwest Israel, close to the Haifa harbour where most imported grains and other agricultural products are unloaded and distributed (Figures 1c and 2e).

3.3 | Tracking the dispersion and establishment of invasive populations

To further investigate the process of invasion and establishment, population dynamics were explored among sampling sites across Israel. To test for population stratification among invasive populations, a PCA was conducted followed by a Tracy–Widom test to identify the number of clusters represented in the data (Figure 3a; Figure S5). No significant signal of clear differentiation was identified analytically based on the Tracy–Widom test, although a deviation to three clusters in accordance with the geographical distribution was observed in a plot generated for the first two PCs, splitting the sampling sites in northwest, east and southwest Israel. To further examine the observed population structure, SNMF and FAST-STRUCTURE analyses were performed with K values ranging from 2 to 10 followed

by cross-entropy (SNMF) and "chooseK" (FAST-STRUCTURE) tests to identify the number of clusters that best represent the data. Similar to the Tracy–Widom test, no signal of clear stratification was obtained, indicating that divergence among sampling sites is too low to allow identification of the population structure analytically. Nevertheless, visualizing the SNMF and FAST-STRUCTURE results for different K values denoted interesting population dynamics in accordance with the PCA results (Figure 3b; Figures S6 and S7). At K = 2, individuals sampled in the southern region clustered separately from the remaining sampling sites and assignment of accessions was consistent with the identified cluster. At K = 3, accessions sampled in the north were further split in two, and at K = 4 accessions from most sampling sites were assigned to more than one cluster (Figures S6 and S7). These observations suggest that the southern sites are more genetically uniform and diverged from the remaining sites, while the split between the northern and eastern sampling sites is less pronounced and may indicate either a more recent split or higher rate of ongoing gene flow. The low level of divergence between invasive sunflower populations suggests that the process of differentiation is young; however, the presence of distinct clusters points to ongoing establishment at different geographical regions, with increasing level of genetic divergence (Figure 3c). To quantify the level of differentiation between clusters, a pair-wise F_{ST} was conducted after balancing sample size ($n_{north} = 29$, $n_{south} = 24$, $n_{east} = 28$) in the eastern cluster
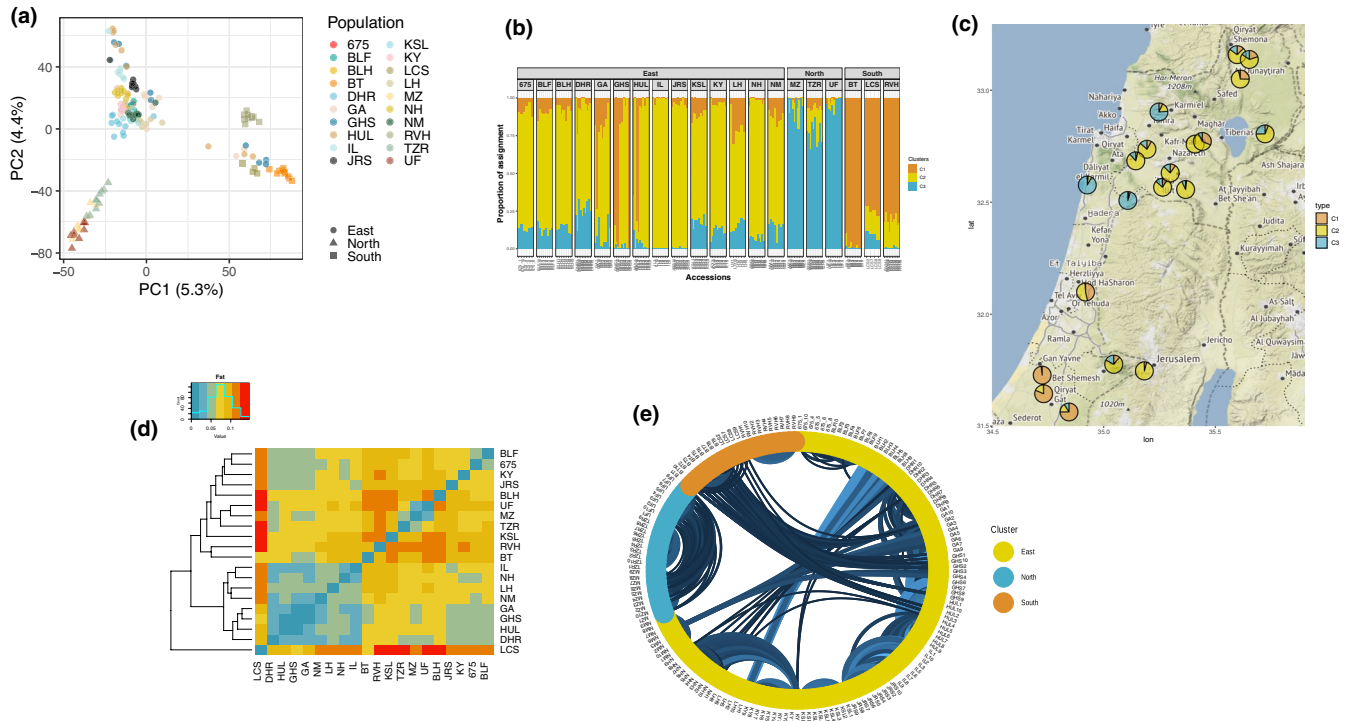
FIGURE 3 Population stratification among invasive wild sunflower in Israel. (a) Principal component analysis of invasive individuals coloured by sampling site. (b) Assignment of individuals and sampling sites to clusters at $K = 3$ based on the SNMF analysis. (c) Frequency of assignment to each cluster by sampling location. (d) Pair-wise F_{ST} between invasive sunflower sampling sites, where red and blue colours correspond to high and low F_{ST} values, respectively. (e) Relatedness among invasive individuals calculated from identity by state. Outer colours correspond to clusters and inner links correspond to the level of relatedness, with thicker links representing higher levels of relatedness

by excluding populations with low assignment (<80%). Overall, the F_{ST} scores supported the low differentiation among clusters with higher divergence between the southern and northern clusters (F_{ST} =0.06) than between each of them and the eastern cluster ($F_{ST:north-east}$ =0.04; $F_{ST:south-east}$ =0.05). To further evaluate the level of differentiation between sampling sites, a pair-wise F_{ST} was conducted. Overall, low differentiation (F_{ST} <0.15) was observed among sampling sites, yet higher values were obtained with increased geographical distance between sampling sites (r_{Mantel} = .29, p = .01), supporting the results of the PCA and SNMF (Figure 3d). To further test for correlation between population divergence and ecogeographical parameters, the pair-wise F_{ST} matrix was down-scaled to one vector using a multidimensional scaling (Figure S8). Significant correlations for population divergence were observed with longitude ($r = -.52$, $p = .01$) and precipitation ($r = -.57$, $p = .009$), although significant correlation was also observed among these two ecogeographical parameters ($r = .67$, $p = .001$). These results imply that the main factor contributing to population divergence is demographic and that the effect of environmental differences between regions is lower. Interestingly, signs of gene flow between geographically distant populations were also noted, including cases of accessions that were assigned to a different cluster than the remaining individuals from the same sampling site (Figure 3a–c). To further test for potential gene flow and seed transfer between sampling sites, a pair-wise relatedness matrix was constructed by calculating the

level of identity-by-state (IBS) across all individuals. Expectedly, high genetic similarity was observed among individuals from the same sampling site; however, extensive gene flow was also noted between sampling sites from the same cluster and, in some cases, even between clusters. The strongest cross-cluster gene flow was observed between Be'er Tuvia (BT) and Givat HaShlosha (GHS) which are geographically close (62 km), but also between BT and Givat Avni (GA) which are more distant (164 km). Other signs of misassignment or gene flow between geographically distant sampling sites were observed for Hulata (HUL) in the Upper Galilee and Jerusalem (JRS) in the centre, Drom Ha'rama (DHR) in the Golan Heights and Kfar Yehoshua (KY) near the coast, among others (Figure 3e). These signs of gene flow or misassignments are expected due to the intensive transport across Israel but also due to intentional transfer of seeds between regions, which was indeed reported by participants in the survey. Another potential source of gene flow is from domesticated sunflower, which is sparsely cultivated in fields close to some of the sampling locations. To test for potential gene flow from cultivated sunflower into wild populations, the f_3 test was conducted between domesticated accessions (represented as the entire SAM population) and each sampling site, where all possible rotations among invasive populations were examined. A negative f_3, which indicates admixture, was detected, albeit barely, only in the Givat HaShlosha (GHS) sampling site ($f_3 = -0.003$, $Z = -2.02$). Nevertheless, based on the GBS data alone, we cannot exclude that admixture between GHS

and other sampling sites may have spread some domesticated alleles also to other parts of the country, nor that introgressions were masked by repeated backcrosses in the wild.

3.4 | Maintenance of genetic diversity in invasive populations

A critical limitation for the establishment of an invasive population is the available genetic variation, which allows to efficiently respond to selection and avoid the genetic load of introduction to a new habitat. A lack of standing genetic variation or failure to retain it will risk the survival and establishment of the invading species. To quantify the extent of genetic diversity in the invasive population ($n = 173$) and compare it with the North American wild populations, a subset of ~20,000 variants were randomly sampled from each of the three largest North American populations in the data set: Texas ($n = 140$), Utah ($n = 100$) and California ($n = 110$). Population genetics statistics were calculated for each group and indicated that, overall, the diversity (θ_W) in the invasive population is significantly lower ($F = 31.1$, $p < .0001$) than in any other North American population, which is expected due to the recent invasion and the associated genetic bottleneck. Differences in the observed nucleotide diversity further supported this observation ($F = 87.9$, $p < .0001$); however, it was noted that the invasive population holds up to 70% of the diversity in the source population in Texas, indicating that the genetic bottleneck was not severe. A neutrality test statistic (Tajima's D) calculated for each population was significantly different between populations ($F = 130$, $p < .0001$) and suggested a recovery of genetic variation after a recent population contraction, as indicated by lower values in the invasive population compared with the native North American populations (Figure 4a,b; Figure S9).

Next, we quantified the genetic diversity among the three identified clusters of invasive sunflower in Israel. To reduce the statistical bias caused by the larger sample size in the "Eastern" cluster, a subset of three sampling sites with highest assignment scores based on the SNMF analysis was selected ("IL," "JRS," "NH") as representatives (Figure 3b). A Tajima's D test conducted for the three clusters designated a significantly higher score in the "Southern" and "Northern" clusters compared with the "Eastern" cluster ($F = 50.5$, $p < .0001$) indicating a stronger population contraction in the former clusters or a faster population expansion in the latter. Observed heterozygosity was significantly higher in the "Eastern" cluster while nucleotide diversity remained similar (Figure 4d; Figure S10). Interestingly, the "Northern" and "Southern" clusters harboured contrasting genetic variation, which is also reflected in the higher F_{ST} values between these two clusters compared to those between either of them and the "Eastern" cluster ($F = 23.6$, $p < .0001$). High rates of gene flow between each of the "Northern" and "Southern" clusters with the "Eastern" cluster is consistent with the higher levels of heterozygosity observed in the latter. Thus, the "Eastern" cluster serves as a bridge for gene flow between the "Northern" and "Southern" clusters and as a hub for maintenance of high standing genetic variation.

To test whether genetic differentiation between populations is uniformly distributed along the genome, a genome scan for F_{ST} score was calculated in windows of 1 Mbp. Differentiation between the "Northern" and "Southern" clusters was observed in 144 windows of high differentiation ($F_{ST} > 0.25$) compared with 81 windows observed between the "Southern" and "Eastern" clusters, and 45 windows between the "Northern" and "Eastern" clusters (Figure 4e). Next, a genome scan for Tajima's D statistic was conducted to identify genomic regions that deviate from neutrality and may indicate footprints of selection. Signals of selective sweeps in the "Eastern" population with overlaps between elevated F_{ST} and negative Tajima's D, were identified on chromosomes 10 and 14 (Tables S3–S5). In the "Southern" cluster, similar signals were observed on chromosomes 1, 4, 5, 8, 11 and 15, and in the "Northern" cluster no population-specific signal was detected (Figure 4f). Finer resolution of genotyping will be required to further explore genomic regions subject to soft selective sweeps, and identify candidate genes with high confidence.

4 | DISCUSSION

Biological invasions are becoming more frequent, mainly due to intensified anthropogenic activity (Bradley et al., 2010; Hulme, 2009; Seebens et al., 2015) and can cause substantial damage in the invaded territory (Pimentel et al., 2005). Despite their harmful potential, public awareness remains low, and authorities frequently fail to prevent biological invasions or control their rates (Hulme, 2006). Elucidating the process of invasion, from introduction to establishment, can assist in developing a more efficient protocol to reduce invasions or their effects on local habitats. Here we explore the distribution range of invasive wild sunflower in Israel using a citizen science approach and genomic data to infer the demographic course of spread and establishment.

Invasive species that were recently introduced are usually absent in formal records of the distribution of local flora, and therefore targeting populations across the distribution range may be prone to bias. To address this, we took a citizen science approach, which allowed us to expand the search to a broader range while reducing sampling bias (Hübner & Kantar, 2021). In addition, the citizen engagement also increased awareness to the threat posed by invasive species, which is an important aspect in prevention and control of spread. Similar surveys can be expanded to serve other purposes, including conservation of crop wild relatives, preservation of endangered species and monitoring climate adaptation (Ryan et al., 2018; Silvertown, 2009). Our main conclusions from the survey conducted in this study are: (i) defining the target audience is essential for the choice of outreach strategy, (ii) identifying a platform that is popular among the target audience is important to promote communication and engagement, and (iii) communicating the project continuously using local language and clear information can increase the interest and engagement of people in the project. The last point is particularly important in international efforts aiming to survey regions that
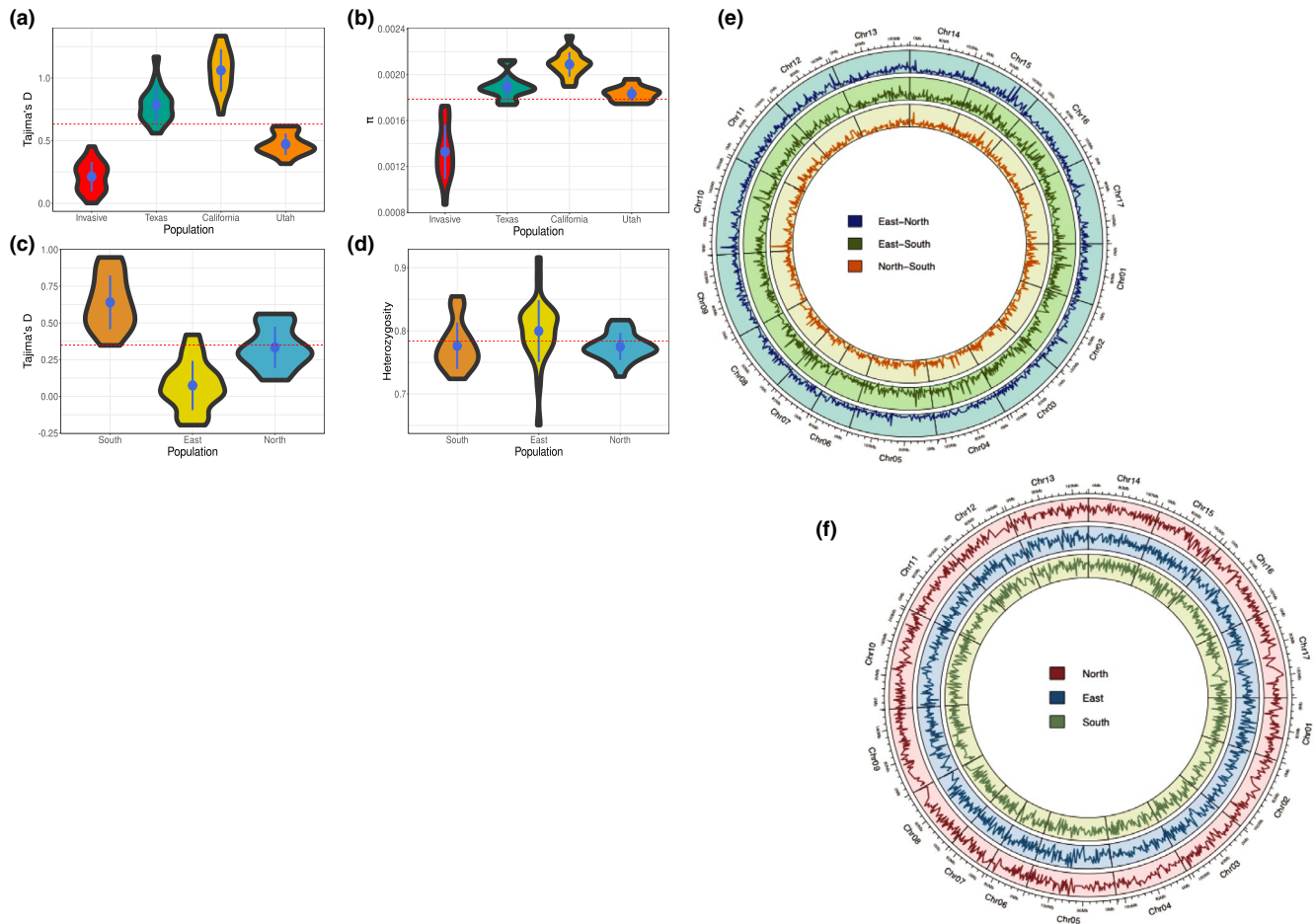
FIGURE 4 Population genomic statistics. (a) Comparison between the invasive population and three representative North American populations using Tajima's *D*, and (b) nucleotide diversity. (c) Comparison between the three identified invasive clusters using Tajima's *D* and (d) observed heterozygosity. Red dashed horizontal lines represent the average score across all populations. (e) Genome scans of differentiation between invasive cluster F_{ST} values and (f) Tajima's *D* in each cluster using 1-Mbp windows

cross countries and cultures. From our experience, addressing people in their local language and sharing scientific information encouraged participation, interest and engagement.

First reports of invasive wild sunflower populations in Israel date to the 1970s (Dafni & Heller, 1980). Since then, this species has spread, based on the citizen survey, across the entire Mediterranean climate region in Israel (Figure 1). In accordance with the first reports, the demographic inferences conducted here indicated that invasion started in the northwest close to the port of Haifa and from there spread to other parts of the country (Figures 2 and 3). Genomic comparisons between the invasive population in Israel and native North American wild sunflower populations indicated that its origin is in the southwest USA, from where it was presumably introduced with imported agricultural products. Previous studies on other invasive species in Israel suggested southwest USA as a source of the invading population (Abu-Nassar & Matzrafi, 2021; Yair et al., 2017). Southwest USA, and specifically Texas, is located on the same latitude as Israel (31°N) and shares some of the climatic conditions, although both regions are characterized by a wide spectrum of environmental conditions. The similarity between regions supports the

benefit of niche conservatism and abiotic pre-adaptation requirement for a successful invasion (Bock et al., 2015; Petitpierre et al., 2012). Therefore, authorities should increase precautions to avoid similar introductions from this region in the future.

Following its introduction, wild sunflower has spread rapidly throughout the country (except in the desert region) with mild demographic footprints of divergence, although indications of deviation into three clusters were detected (Figure 3a–c). Reduced genetic divergence among invasive populations is expected due to the recent introduction and spread; however, additional factors that could limit the rate of divergence were detected including seed transfer and hybridization. Continuous seed transfer across the invaded geographical range can increase the rate of spread and reduce divergence. Agriculture and other commercial activities were previously reported as major drivers of biological invasion and spread (Hulme, 2009), also across Israel (Matzrafi et al., 2021). The observed distribution of invasive populations along roads and in proximity to agricultural centres suggest that these factors contributed substantially to the spread of wild sunflower (Figures 1 and 3). Interestingly, the attractiveness of sunflower also contributed to its spread by people

mistakenly identifying it as an ornamental plant. Another key factor in reducing divergence is the extensive gene flow detected between each of the "Northern" and "Southern" clusters with the "Eastern" cluster, which serves as a bridge for gene flow between regions. Intraspecific hybridization can potentially accelerate adaptation by (i) increasing the available genetic variation for selection, (ii) increasing the chances for heterotic effects and transgressive phenotypes, and (iii) reducing the effect of genetic load by expanding the genetic variation and masking of deleterious alleles (Bock et al., 2015; Dlugosch et al., 2015; Rius & Darling, 2014; Todesco et al., 2016). The high heterozygosity observed in the "Eastern" population supports the role of this region in maintaining high genetic diversity across all regions.

The role of wild–crop interactions in promoting invasion and weediness was also examined. Previous studies found evidence for gene flow from cultivated fields into a counterpart invading species, and indicated that weedy traits, including herbicide resistance, have contributed to establishment of the invading species around cultivated fields (Lai et al., 2012; Le Corre et al., 2020; Presotto et al., 2017). In Israel, sunflower is grown for confectionary purposes and only a few varieties that are characterized by a very large achene are cultivated. In the past, sunflower was grown across all regions but today the crop is declining due to broomrape infestation. Few invasive individuals collected in Israel were previously investigated, but only moderate support for crop–wild interactions was detected based on an EST (expressed sequence tage) data set (Lai et al., 2012). Here, we used genomic data and f_3-statistics to detect interactions with higher confidence. No widespread crop–wild interactions were detected, similarly to previous reports for invasive sunflower in Argentina, where no introgressions between cultivated and wild (invasive) *Helianthus annnuus* were detected (Mondon et al., 2018). Here, crop–wild introgression was detected only in one sampling site (GHS), although we cannot exclude that introgressions from cultivated sunflower have enhanced adaptation in the invasive populations and that repeated backcrosses in the wild impede their detection with GBS data.

Invasive species have a remarkable ability to efficiently maintain and exploit genetic diversity. Standing genetic variation is viewed as the main driver of rapid adaptation (Barrett & Schluter, 2008), although other mechanisms of diversity maintenance should be further explored (Bock et al., 2015). Recent studies have investigated the source of genetic variation in invasive species and emphasized the importance of precolonization admixture and multiple introductions followed by hybridization at the invaded territory (Barker et al., 2019; van Boheemen et al., 2017). Here we provide an intriguing example for genetic diversity maintenance in an invasive population through a process of rapid spread followed by extensive gene flow. This mechanism preserves high heterozygosity in the "Eastern" cluster, which serves as a bridge for gene flow across the distribution range. Single introduction of an invasive species may be more sensitive than multiple introductions to the impact of inbreeding depression and expansion load. The potential environmental similarity between the source region (Texas) and Israel may have

compensated, at least partially, for this negative impact. Moreover, several mechanisms were previously suggested as a means to overcome potential genetic bottlenecks including large founding population size and diversity, conversion of dominance to additive genetic variation (Bryant & Meffert, 1988; Whitlock et al., 1993) and fitness-dependent recombination, which can increase the rate of adaptation specifically under stressful conditions and in populations with high inbreeding (Roze & Lenormand, 2005; Rybnikov et al., 2021). Thus, rapid adaptation in invasive species may be promoted also by demographic processes.

The rate of biological invasions is increasing with expansion of globalization and accelerated climate change. Prevention efforts and regulation should target commodities imports especially from regions detected as source of invasive species, and in increasing public awareness to the risk of biological invasions. Wild sunflower has been introduced from the southwest USA and has spread rapidly throughout Israel, and thus eradication at early stages of the invasion may have been the most effective means to avoid its establishment in the Levant.

AUTHOR CONTRIBUTIONS

S.H. planned and designed the research, D.S. performed the citizen survey and sampled invasive populations, T.M. and M.T. performed library preparation and sequencing, S.H. conducted the analyses and wrote the first draft, and M.T., D.S., M.M., H.E. and T.M. provided comments and edited the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Sequence data are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive under Bioproject PRJNA748828. Seeds from invasive populations are available upon request. Benefits generated: benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Sariel Hübner https://orcid.org/0000-0003-3660-4634

REFERENCES

Abu-Nassar, J., & Matzrafi, M. (2021). Effect of herbicides on the management of the invasive weed solanum rostratum dunal (Solanaceae). *Plants*, *10*(2), 1–9. https://doi.org/10.3390/plants10020284

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N. C., Bowers, J. E., Hübner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., … Langlade, N. B. (2017). The

sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 546(7656), 148–152.

Barker, B. S., Cocio, J. E., Anderson, S. R., Braasch, J. E., Cang, F. A., Gillette, H. D., & Dlugosch, K. M. (2019). Potential limits to the benefits of admixture during biological invasion. *Molecular Ecology*, 28(1), 100–113. https://doi.org/10.1111/mec.14958

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, 23(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Battey, C. J., Ralph, P. L., & Kern, A. D. (2020). Predicting geographic location from genetic variation with deep neural networks. *Elife*, 9, 1–22. https://doi.org/10.7554/eLife.54507

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., Turner, K. G., Whitney, K. D., & Rieseberg, L. H. (2015). What we still don't know about invasion genetics. *Molecular Ecology*, 24(9), 2277–2297. https://doi.org/10.1111/mec.13032

Bradley, B. A., Blumenthal, D. M., Wilcove, D. S., & Ziska, L. H. (2010). Predicting plant invasions in an era of global change. *Trends in Ecology and Evolution*, 25(5), 310–318. https://doi.org/10.1016/j.tree.2009.12.003

Bryant, E. H., & Meffert, L. M. (1988). Effect of an experimental bottleneck on morphological integration in the housefly. *Evolution*, 42(4), 698. https://doi.org/10.1111/j.1558-5646.1988.tb02488.x

Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24(9), 1999–2017. https://doi.org/10.1111/mec.13162

Dafni, A., & Heller, D. (1980). The threat posed by alien weeds in Isreal. *Weed Research*, 20, 277–283.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, 24(9), 2095–2111. https://doi.org/10.1111/mec.13183

Dlugosch, K. M., & Parker, I. M. (2008). Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters*, 11(7), 701–709. https://doi.org/10.1111/j.1461-0248.2008.01181.x

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T. J., Haccou, P., Lu, B.-R., Snow, A. A., Neal Stewart, C., Strasburg, J. L., van Tienderen, P. H., Vrieling, K., & Hooftman, D. (2013). Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution, and Systematics*, 44, 325–345. https://doi.org/10.1146/annurev-ecolsys-110512-135840

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Hodgins, K. A., Bock, D. G., & Rieseberg, L. H. (2018). Trait evolution in invasive species. *Annual Plant Reviews Online*, 1, 459–496.

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., & Rieseberg, L. H. (2019). Hybridization altered gene content and disease resistance. *Nature Plants*, 5, 54–62.

Hübner, S., Höffken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A., Schmid, K., & Fridman, E. (2009). Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Molecular Ecology*, 18(7), 1523–1536.

Hübner, S., & Kantar, M. B. (2021). Tapping diversity from the wild: From sampling to implementation. *Frontiers in Plant Science*, 12, 1–14. https://doi.org/10.3389/fpls.2021.626565

Hulme, P. E. (2006). Beyond control: Wider implications for the management of biological invasions. *Journal of Applied Ecology*, 43(5), 835–847. https://doi.org/10.1111/j.1365-2664.2006.01227.x

Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*, 14(1), 68–73. https://doi.org/10.1093/bioinformatics/14.1.68

Kane, N. C., & Rieseberg, L. H. (2008). Genetics and evolution of weedy Helianthus annuus populations: Adaptation of an agricultural weed. *Molecular Ecology*, 17(1), 384–394.

Kantar, M. B., Sosa, C. C., Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. A., Bernau, V., Kane, N. C., Marek, L., Seiler, G., & Rieseberg, L. H. (2015). Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.). *Frontiers in Plant Science*, 6, 1–11. https://doi.org/10.3389/fpls.2015.00841

Lai, Z., Kane, N. C., Kozik, A., Hodgins, K. A., Dlugosch, K. M., Barker, M. S., Matvienko, M., Yu, Q., Turner, K. G., Pearl, S. A., Bell, G. D. M., Zou, Y. I., Grassa, C., Guggisberg, A., Adams, K. L., Anderson, J. V., Horvath, D. P., Kesseli, R. V., Burke, J. M., … Rieseberg, L. H. (2012). Genomics of compositae weeds: EST libraries, microarrays, and evidence of introgression. *American Journal of Botany*, 99(2), 209–218. https://doi.org/10.3732/ajb.1100313

Le Corre, V., Siol, M., Vigouroux, Y., Tenaillon, M. I., & Délye, C. (2020). Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 117(41), 25618–25627. https://doi.org/10.1073/pnas.2006633117

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv*:1303.3997.

Massinga, R. A., Al-Khatib, K., St. Amand, P., & Miller, J. F. (2003). Gene flow from imidazolinone-resistant domesticated sunflower to wild relatives. *Weed Science*, 51(6), 854–862. https://doi.org/10.1614/WS-03-032R

Matzrafi, M., Raz, H., Rubin, B., Yaacoby, T., & Eizenberg, H. (2021). Distribution and biology of the invasive weed Parthenium hysterophorus L. in Israel. *Frontiers in Agronomy*, 3, 1–12. https://doi.org/10.3389/fagro.2021.639991

Mondon, A., Owens, G. L., Poverene, M., Cantamutto, M., & Rieseberg, L. H. (2018). Gene flow in Argentinian sunflowers as revealed by genotyping-by-sequencing data. *Evolutionary Applications*, 11(2), 193–204. https://doi.org/10.1111/eva.12527

Muller, M. H., Latreille, M., & Tollon, C. (2011). The origin and evolution of a recent agricultural weed: Population genetic diversity of weedy populations of sunflower (Helianthus annuus L.) in Spain and France. *Evolutionary Applications*, 4(3), 499–514.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335, 1344–1348. https://doi.org/10.1126/science.1215933

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52, 273–288. https://doi.org/10.1016/j.ecolecon.2004.10.002

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing

approach. *PLoS One*, *7*(2), e32253. https://doi.org/10.1371/journ al.pone.0032253

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., & Shakir, K. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.

Presotto, A., Hernández, F., Díaz, M., Fernández-Moroni, I., Pandolfo, C., Basualdo, J., Cuppari, S., Cantamutto, M., & Poverene, M. (2017). Crop-wild sunflower hybridization can mediate weediness throughout growth-stress tolerance trade-offs. *Agriculture, Ecosystems and Environment*, *249*, 12–21. https://doi.org/10.1016/j.agee.2017.08.003

Presotto, A., Ureta, M. S., Cantamutto, M., & Poverene, M. (2012). Effects of gene flow from IMI resistant sunflower crop to wild Helianthus annuus populations. *Agriculture, Ecosystems and Environment*, *146*(1), 153–161. https://doi.org/10.1016/j.agee.2011.10.023

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology and Evolution*, *29*(4), 233–242. https://doi.org/10.1016/j.tree.2014.02.003

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Roze, D., & Lenormand, T. (2005). Self-fertilization and the evolution of recombination. *Genetics*, *170*(2), 841–857. https://doi.org/10.1534/genetics.104.036384

Rubin, B., & Matzrafi, M. (2015). Weed management in Israel-challenges and approaches. In *Weed science in the Asian-Pacific region* (253–270). Indian Society of Weed Science.

Ryan, S. F., Adamson, N. L., Aktipis, A., Andersen, L. K., Austin, R., Barnes, L., & Dunn, R. R. (2018). The role of citizen science in addressing grand challenges in food and agriculture research. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1891), 20181977.

Rybnikov, S., Weissman, D. B., Hübner, S., & Korol, A. B. (2021). Fitness dependence preserves selection for recombination across diverse mixed mating systems. *Journal of Theoretical Biology*, *528*, 110849.

Sarokin, L., & Carlson, M. (1984). Nucleic acids research nucleic acids research. *Methods*, *12*(21), 8235–8251.

Seebens, H., Essl, F., Dawson, W., Fuentes, N., Moser, D., Pergl, J., Pyšek, P., van Kleunen, M., Weber, E., Winter, M., & Blasius, B. (2015). Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology*, *21*(11), 4128–4140. https://doi.org/10.1111/gcb.13021

Shagina, I., Bogdanova, E., Mamedov, I. Z., Lebedev, Y., Lukyanov, S., & Shagin, D. (2010). Normalization of genomic DNA using

duplex-specific nuclease. *BioTechniques*, *48*(6), 455–459. https://doi.org/10.2144/000113422

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, *9*, 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of Ambrosia artemisiifolia in Europe and Australia. *Molecular Ecology*, *26*(20), 5421–5434.

Whitlock, M. C., Phillips, P. C., & Wade, M. J. (1993). Gene interaction affects the additive genetic variance in subdivided populations with migration and extinction. *Evolution*, *47*(6), 1758–1769. https://doi.org/10.1111/j.1558-5646.1993.tb01267.x

Yair, Y., Sibony, M., & Rubin, B. (2017). Four Ambrosia species in Israel: Invasive, naturalized and casual alien plants. *Israel Journal of Plant Sciences*, *64*(1–2), 93–98.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.