

Article

Phylogeography and the Evolutionary History of Sunflower (*Helianthus annuus* L.): Wild Diversity and the Dynamics of Domestication

Brian Park * and John M. Burke 

Department of Plant Biology, University of Georgia, Miller Plant Sciences Bldg., Athens, GA 30602, USA; jmburke@uga.edu

* Correspondence: brian.park@uga.edu

Received: 30 January 2020; Accepted: 27 February 2020; Published: 29 February 2020



Abstract: Patterns of genetic variation in crops are the result of selection and demographic changes that occurred during their domestication and improvement. In many cases, we have an incomplete picture of the origin of crops in the context of their wild progenitors, particularly with regard to the processes producing observed levels of standing genetic variation. Here, we analyzed sequence diversity in cultivated sunflower (*Helianthus annuus* L.) and its wild progenitor (common sunflower, also *H. annuus*) to reconstruct phylogeographic relationships and population genetic/demographic patterns across sunflower. In common sunflower, south-north patterns in the distribution of nucleotide diversity and lineage splitting indicate a history of rapid postglacial range expansion from southern refugia. Cultivated sunflower accessions formed a clade, nested among wild populations from the Great Plains, confirming a single domestication event in central North America. Furthermore, cultivated accessions sorted by market type (i.e., oilseed vs. confectionery) rather than breeding pool, recapitulating the secondary development of oil-rich cultivars during its breeding history. Across sunflower, estimates of nucleotide diversity and effective population sizes suggest that cultivated sunflower underwent significant population bottlenecks following its establishment ~5000 years ago. The patterns inferred here corroborate those from previous studies of sunflower domestication, and provide a comprehensive overview of its evolutionary history.

Keywords: sunflower; phylogeography; domestication; demographic history; GBS

1. Introduction

Patterns of genetic variation in cultivated plants are the product of multiple processes that have occurred over their evolutionary histories. Gaining a comprehensive understanding of these patterns and the underlying processes requires reconstructing them within the broader context of the wild species from which they are derived. Such species-wide assessments can provide inferences into the ancestral lineages that gave rise to early domesticates and modern cultivars, and yield insights into the factors explaining the distribution of genetic diversity across gene pools. Such knowledge also has practical value, and can be applied to identify sources of novel alleles that should be preserved in germplasm collections and which may be of value in modern breeding programs [1–4]. Here, we characterize patterns of genetic variation in cultivated sunflower—a globally important oilseed crop—and its wild progenitor, common sunflower (both *Helianthus annuus* L.).

Common sunflower is a widely distributed annual herb whose native geographic range is centered in the Great Plains region of the United States and Canada [5]. Sunflower is thought to have been domesticated 3000–5000 years ago [6,7] by Native Americans who primarily used it as a source of edible seed [8]. Descendants of these early domesticates—the Native American landraces—were

introduced to Europe in the early 16th century, and eventually made their way to Russia [9], where the predecessors of modern oilseed cultivars were developed and grown at an industrial scale [10]. These Russian oilseed cultivars were reintroduced to North America in the mid-20th century, stimulating commercial sunflower production in the Americas, and ushering in the modern era of sunflower breeding. More recent breeding efforts have transitioned sunflower from primarily open-pollinated varieties (OPVs) into a hybrid crop comprising two major market classes (i.e., oilseed and confectionery) that are separated into two major heterotic groups: unbranched, female (i.e., male-sterile; HA) lines and recessively branched, male restorer (i.e., RHA) lines [11,12].

This historical account of the evolutionary history of sunflower serves as the basis of hypotheses concerning the expected population structure and levels of standing genetic variation across its gene pool. Previous studies have sought to infer these patterns in sunflower at various scales, and have shed light on certain aspects of its origin and subsequent evolution. For example, broad examinations across common and cultivated sunflower have provided strong evidence for a single domestication event [13–15], and indicate that there was a marked reduction in genetic variation that corresponds to a domestication bottleneck [16–21]. More detailed examinations within cultivated sunflower have shown that there was a secondary loss of variation following improvement, and that cultivars are more or less genetically differentiated by market type (i.e., oilseed and confectionery; [16,17]) and breeding pool (e.g., RHA vs. HA; [21,22]). These studies have provided valuable insight into the genetic consequences of domestication and improvement in sunflower.

To date, however, attempts to infer patterns of population divergence and population structure in common sunflower (particularly from a phylogeographic perspective) have been somewhat limited, and efforts to formally reconstruct the demographic history of sunflower domestication are lacking. As such, a number of important issues have not been addressed. For example, given that common sunflower populations currently span areas with widely varying climatic conditions during the Last Glacial Maximum (LGM; ca. 21,500 ybp; [23–25]), phylogeographic analyses across the latitudinal range of the species may provide insights into: the locations of refugial areas where it may have survived the LGM, and how it colonized its current distribution following glacial retreat. Furthermore, given the differences in levels of genetic diversity between cultivated sunflower and its wild progenitor, demographic reconstructions would be useful for determining the timing and sequence of demographic changes that may explain observed patterns of variation across gene pools.

In this study, we sought to characterize the phylogeographic history of sunflower. We used genotyping-by-sequencing (GBS) to examine a panel of accessions that span the latitudinal range of common sunflower and the breadth of cultivated sunflower diversity to: (1) reconstruct the postglacial migration history of common sunflower; (2) identify patterns of genetic relatedness and structure across the various breeding pools of cultivated sunflower; and (3) gain insights into the history of demographic changes associated with domestication and subsequent improvement.

2. Materials and Methods

2.1. Sampling and Sequencing

We obtained seeds of cultivated and common sunflower from the USDA North Central Regional Plant Introduction Station (Ames, IA, USA; Table 1). Thirty cultivated accessions were selected to represent each of the major types of cultivated sunflower (twelve “exotic” lines, including six Native American landraces and six open-pollinated varieties [OPV]; eighteen “elite” lines including ten HA and eight RHA lines). Sixteen accessions of common sunflower (hereafter referred to as wild sunflower) were selected, to provide full latitudinal coverage of the central portion of its native range in the United States (Figure 1). Note that these populations were selected so as to avoid the potential for recent crop-wild contact, which would negatively impact our interpretation of observed patterns of variation. We also sampled individuals from one accession for each of two related wild species (*Helianthus argophyllus* and *Helianthus petiolaris*), to root phylogenies and polarize SNPs for the downstream

analysis (see below). Seeds were sown in flats, and leaves were sampled from a single individual per cultivated sunflower line and up to 18 seedlings per wild sunflower accession. Genomic DNA was extracted from collected leaves using a modified CTAB protocol [26]. Resulting DNA extractions were checked for integrity on a 0.8% agarose gel and DNA quantity was measured with a Qubit 1.0 fluorometer (Invitrogen, Waltham, MA, USA).

Table 1. Accession numbers, improvement status, and geographic origins of wild and cultivated lines examined in this study. Cultivated accessions are categorized as exotic (i.e., Native American landraces and open-pollinated varieties [OPVs]), HA, or RHA lines. Market type (i.e., oilseed or confectionery) is denoted for HA/RHA lines. All seeds were obtained from the North Central Regional Plant Introduction Station (Ames, IA, USA).

USDA PI Number	Improvement Status	Name	Geographic Origin	Sample Size
592304	wild	AB2	Alberta, CAN; 51.6, -112.7	16
592309	wild	AB1	Alberta, CAN; 49.9, -110.2	14
586816	wild	MT2	Montana, USA; 47.7, -104.2	16
531035	wild	MT1	Montana, USA; 46.6, -108.5	16
586837	wild	WY1	Wyoming, USA; 42.07, -104.18	9
435564	wild	CO3	Colorado, USA; 40.49, -106.83	13
468622	wild	CO2	Colorado, USA; 39.45, -108.05	12
435560	wild	CO1	Colorado, USA; 37.67, -104.83	16
586869	wild	NE2	Nebraska, USA; 41.37, -97.67	11
586866	wild	NE1	Nebraska, USA; 40.47, -96.37	14
586859	wild	KS2	Kansas, USA; 38.67, -96.67	17
664770	wild	KS1	Kansas, USA; 37.33, -95.79	14
468489	wild	OK1	Oklahoma, USA; 35.47, -98.36	13
435479	wild	NM1	New Mexico, USA; 35.32, -103.98	12
435366	wild	TX2	Texas, USA; 34.26, -99.52	16
649848	wild	TX1	Texas, USA; 32.01, -100.55	13
607510	elite – HA, confectionery	HAR7	USDA Breeding Program	1
599780	elite – HA, confectionery	HA285	USDA Breeding Program	1
599769	elite – HA, confectionery	HA008	USDA Breeding Program	1
552932	elite – HA, confectionery	HA286	USDA Breeding Program	1
509060	elite – HA, confectionery	HA350	USDA Breeding Program	1
639165	elite – HA, oilseed	HA442	USDA Breeding Program	1
632342	elite – HA, oilseed	HA433	USDA Breeding Program	1
599775	elite – HA, oilseed	HA124	USDA Breeding Program	1
599771	elite – HA, oilseed	HA061	USDA Breeding Program	1
561918	elite – HA, oilseed	HA378	USDA Breeding Program	1
664234	elite – RHA, confectionery	RHA325	USDA Breeding Program	1
552944	elite – RHA, confectionery	RHA282	USDA Breeding Program	1
599767	elite – RHA, oilseed	RHA299	USDA Breeding Program	1
597378	elite – RHA, oilseed	RHA400	USDA Breeding Program	1
597374	elite – RHA, oilseed	RHA397	USDA Breeding Program	1
578008	elite – RHA, oilseed	RHA386	USDA Breeding Program	1
531075	elite – RHA, oilseed	RHA362	USDA Breeding Program	1
531072	elite – RHA, oilseed	RHA359	USDA Breeding Program	1
294659	exotic – OPV	Peredovik	Russia, Asia	1
340790	exotic – OPV	VNIIMK8931	Russia, Asia	1
476853	exotic – OPV	Mammoth	Russia, Asia	1
496263	exotic – OPV	Damaya	China, East Asia	1
162454	exotic – OPV	Sunrise	USA, North America	1
650353	exotic – OPV	Guayacan	Uruguay, South America	1
369357	exotic – Native American landrace	Arikara	USA, North America	1

Table 1. Cont.

USDA PI Number	Improvement Status	Name	Geographic Origin	Sample Size
369360	exotic – Native American landrace	Seneca	USA, North America	1
432504	exotic – Native American landrace	Hopi Dye	USA, North America	1
600717	exotic – Native American landrace	Mandan	USA, North America	1
650646	exotic – Native American landrace	Maíz Negro	USA, North America	1
650761	exotic – Native American landrace	Maíz de Tejas	USA, North America	1
435624	outgroup – wild <i>H. argophyllus</i>	-	Texas, USA; 28.17, -97.00	2
613764	outgroup – wild <i>H. petiolaris</i>	-	North Dakota, USA; 46.86, -96.90	3

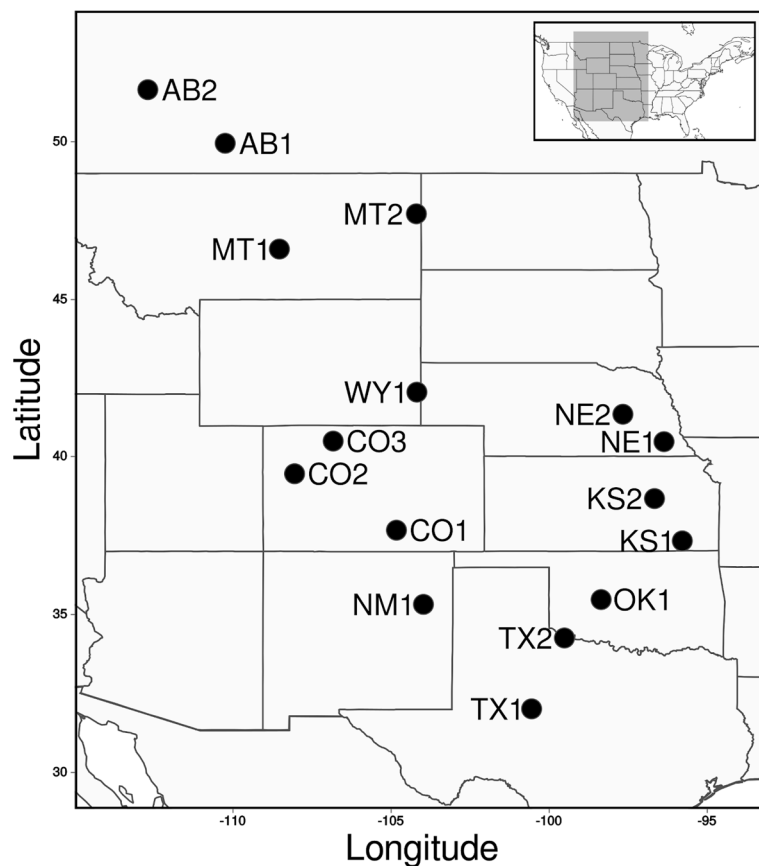


Figure 1. Sampling localities of the 16 populations of wild sunflower examined in this study.

DNA extractions were prepared for sequencing following a two-enzyme GBS protocol [27] using the restriction enzymes *HpaII* and *MseI*. Resulting libraries were pooled at 96-plex and sequenced on the Illumina Nextseq 500 sequencing platform (Illumina, San Diego, CA, USA) in high-output mode and set to produce 75 bp single-end reads. All library preparation and sequencing was performed at the Georgia Genomics and Bioinformatics Core (Athens, GA, USA).

2.2. Sequence Processing and Variant Calling

We used iPyrad version 0.9.16 [28] to process reads and call variants for downstream analysis. Briefly, raw demultiplexed reads were filtered with cutadapt 1.12 [29], to remove reads containing

adaptor sequences and > 5% low quality (phred score < 20) or ambiguous bases. Filtered reads were aligned to the HA412-HOv2 genome assembly [30] with BWA-MEM 0.7.17 [31] using default parameters, then sorted and indexed using samtools 1.10 [32]. Indexed reads were merged with BEDtools 2.29 [33], and bases were called for sites with ≥ 6 and $\leq 10,000$ reads. Merged reads were clustered across samples and aligned into GBS loci, and loci with > 20% shared heterozygous sites and > 10% variable sites were filtered to remove poorly aligned and paralogous loci. Remaining loci anchored to the 17 chromosome-level scaffolds on the HA412-HOv2 genome assembly were retained, and SNP filtering was conducted in VCFtools 0.1.16 [34].

Phylogenetic analyses were conducted on the “phylogenomics” dataset composed of biallelic SNPs present in $\geq 50\%$ of ingroup and outgroup samples (232 *H. annuus*; 2 *H. argophyllus*, 3 *H. petiolaris* samples) at a minor allele frequency greater than 1% (MAF > 0.01). The sample coverage threshold in this dataset was chosen because it allowed for the retention of lower coverage, high mutation rate sites which are useful in resolving recent divergences [35,36]. Population genetic analyses were conducted on three datasets, each consisting of both cultivated and wild sunflower samples (“ingroup_all”), and only wild (“ingroup_wild”) or cultivated (“ingroup_cultivated”) samples. The three datasets were composed of biallelic SNPs present in $\geq 80\%$ of samples at MAF > 0.01, thinned to include SNPs that were ≥ 1 kb apart to reduce non-independence amongst sites. Demographic reconstructions were conducted on an “ingroup_dadi” dataset consisting of two samples per wild population (32 total) and all 30 of the cultivated samples. The “ingroup_dadi” dataset consisted of biallelic SNPs present in $\geq 50\%$ of samples with a minor allele count of 2 (MAC = 2), spaced ≥ 1 kb apart, and polarized by alleles fixed in *H. petiolaris* and *argophyllus* (i.e., the ancestral state at each site was set to the alleles observed in *H. petiolaris* and *argophyllus*). The reduced sample size and sample coverage thresholds of the “ingroup_dadi” dataset were chosen as they increased the number of segregating sites available for demographic analysis. Furthermore, the relatively low MAF and MAC thresholds of the “ingroup” datasets were chosen to allow for the inclusion of rare variants which provide greater resolution of genetic structure and demographic events [37], while excluding singleton alleles that may reflect sequencing error.

2.3. Patterns of Genetic Diversity across Breeding Pools and Geographic Space

Estimates of mean nucleotide diversity (π) across all sites were calculated with VCFtools to identify patterns in the distribution of genetic diversity across breeding pools within *H. annuus* (i.e., wild, OPV, RHA, HA), as well as the wild populations, separately. Differences in π among breeding pools were assessed by computing the 95% confidence intervals from 1000 bootstrap replicates of per-site estimates of π using the “boot” and “boot.ci” functions in the R [38] package *boot* [39].

We then estimated patterns of genetic diversity across wild sunflower populations. Global population genetic parameters (e.g., Weir and Cockerham’s F_{ST} and F_{IS}) were estimated using all sites with the *hierfstat* R package [41]. Evidence of isolation-by-distance (IBD) was evaluated using a Mantel’s test with the “mantel.randtest” function as implemented in the R package *ade4* [42,43]. Pairwise estimates of Weir and Cockerham’s F_{ST} [40] were calculated using the “pairwise.WCfst” function in the *hierfstat* R package [41], and geographic distances between populations were calculated using the “dism” function in the *geosphere* R package [44]. Significance was assessed with 1×10^6 Monte Carlo simulations. Clinal trends in π were investigated with linear regression, treating latitude and longitude as fixed effects using the “lm” function in R. AMOVA was conducted using the *poppR* package in R [45] to determine how genetic diversity is distributed across the following scales: between genetic clusters identified through fastSTRUCTURE (described below); among populations within genetic clusters; among samples within each population; and within samples. Significance was assessed with 1000 Monte Carlo simulations.

2.4. Phylogenetic Relationships

We inferred phylogenetic relationships across wild and cultivated sunflower using RAxML 8.2.1 [46]. Analyses were conducted under GTR + CAT with ascertainment bias correction, with 20 tree searches and 100 bootstrap replicates to assess support. Trees were rooted with samples of *H. petiolaris*.

2.5. Population Clustering

We estimated individual ancestry coefficients with fastSTRUCTURE 1.0 [47] and ADMIXTURE 1.3 [48]. fastSTRUCTURE and ADMIXTURE analyses were run 5 times for $K = 1-20$ clusters using default parameters. The optimal number of clusters was determined using the “chooseK” tool and 10-fold cross-validation for fastSTRUCTURE and ADMIXTURE analyses, respectively. We then visualized samples in two-dimensional genetic space with principal component analysis (PCA) as implemented in the R package LEA [49].

2.6. Demographic History of Domestication

We modeled the divergence history between wild and cultivated sunflower using the diffusion approximation approach implemented in $\delta\text{a}\delta\text{i}$ 2.0.3 [50]. We formulated three models that vary with respect to the presence and directionality of gene flow. Model A (Figure S1A) describes a simple divergence without gene flow scenario, where ancestral populations of sunflower split at time T into wild and cultivated lineages. Following the split, the wild lineage undergoes an instantaneous size change to a current effective size of $N_{\text{wild-current}}$, while the cultivated lineage has a founding effective size of $N_{\text{cult-founder}}$, that grows or declines to a current effective size, $N_{\text{cult-current}}$. Model B (Figure S1B) expands on Model A, and describes a divergence with gene flow scenario, allowing for symmetrical gene flow ($M_{w \leftrightarrow c}$) between the wild and crop lineages. Similarly, Model C (Figure S1C) describes the same scenario as Model B, but allows for asymmetric gene flow ($M_{w \rightarrow c}$, migration from wild into cultivated; $M_{w \leftarrow c}$ migration from cultivated into wild) between the lineages.

An unfolded 2D site-frequency spectrum was generated using the program *easySFS* (<https://github.com/isaacovercast/easySFS>), sampling 24 and 30 haplotypes from the cultivated and wild lineages, respectively, to maximize the number of segregating SNPs for analysis [50]. Model fitting was performed using *dadi_pipeline* (https://github.com/dportik/dadi_pipeline), described in Portik et al. [51]. *dadi_pipeline* was run using custom settings (rounds = 4; replicates = 50, 50, 50, 100; algorithm steps = 3, 5, 15, 50; -fold parameters = 3, 2, 2, 1) and models were extrapolated to a grid size of 40, 50, 60 points and fitted with Nelder-Mead optimization. Maximum-likelihood parameter estimates from the best replicate run (i.e., highest log-likelihood) for each model were used to calculate the Akaike information criterion (AIC) scores for model testing [52] following Carstens et al. [53]. Standard deviations for parameter estimates were obtained using the FIM approach [54], which has been demonstrated to provide reasonable uncertainty estimates for datasets composed of effectively unlinked SNPs, compared to more computationally expensive bootstrapping. Parameter estimates and their associated 95% confidence intervals were converted to biological units assuming a mutation rate of 6.1×10^{-9} substitutions/site/generation [55], and an effective sequence length ([bases sequenced to derive SNPs]*[SNPs used in the frequency spectrum/total number of SNPs called]) of $L = 11.7 \times 10^6$ bp.

3. Results

3.1. Sample Sizes and SNP Datasets

1.53×10^9 reads were sequenced across 257 samples (222 wild, 30 cultivated, and 5 outgroup samples); on average, 5.96×10^6 reads were sequenced per sample (range = $3.27 \times 10^6-1.47 \times 10^7$ reads). Following quality filtering and processing, we assembled 1.08×10^6 loci; the number of SNPs recovered from these loci are listed in Table 2.

Table 2. The number of SNPs present in each dataset analyzed in this study. See text for details regarding the composition of the individual datasets.

Dataset	Samples	Sample Coverage	MAF/MAC Threshold	Thinning Interval	SNPs
phylogenomics	257	50%	MAF = 0.01	-	43,271
ingroup_all	252	80%	MAF = 0.01	1 kb	5745
ingroup_wild	222	80%	MAF = 0.01	1 kb	5571
ingroup_crop	30	80%	MAF = 0.01	1 kb	12,808
ingroup_dadi	62	50%	MAC = 2	1 kb	12,025

3.2. Patterns of Genetic Diversity across Breeding Pools and over Geographic Space

There were notable differences in nucleotide diversity (π) among breeding pools (Figure 2A), with a roughly two-fold difference in π between wild sunflower and both the exotic and elite lines (mean (95% CI): wild sunflower = 0.096 (0.094–0.098); exotic = 0.054 (0.051–0.058); HA = 0.046 (0.042–0.049); RHA = 0.036 (0.033–0.039)). Differences in π amongst cultivated samples were less pronounced, but the HA and RHA lines possessed 85% and 66% of the nucleotide diversity present in the exotic lines, respectively. The marked differences in diversity across breeding pools in *H. annuus* indicates that primitive domesticated lines and improved cultivars harbor progressively less genetic diversity than their wild progenitor.

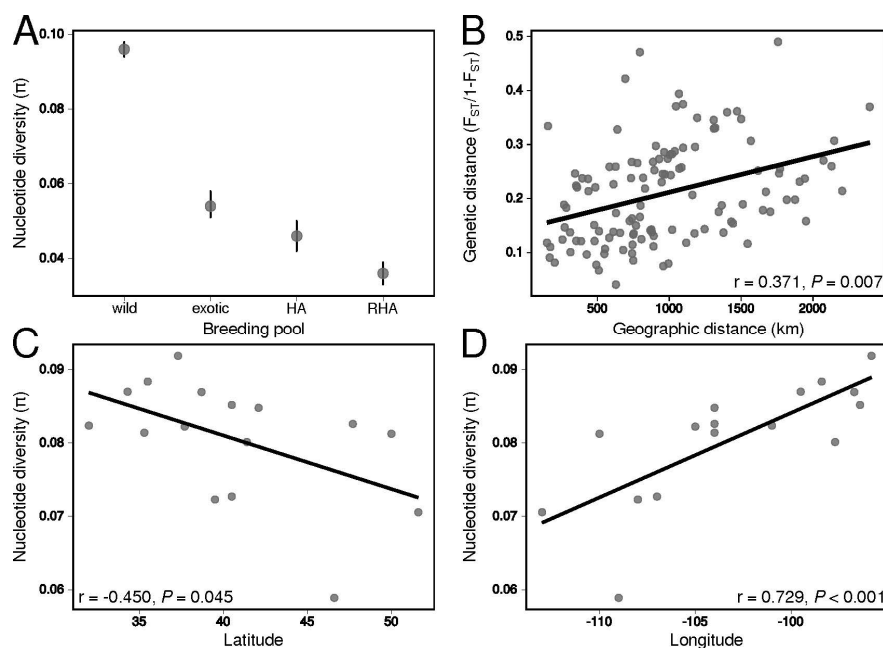


Figure 2. Patterns in the distribution of genetic diversity across breeding pools and over the geographic range of *H. annuus*. (A) Nucleotide diversity (π) varies significantly across breeding pools. In wild sunflower, pairwise genetic distances increase with geographic distances between populations (B). Furthermore, in wild sunflower, nucleotide diversity decreases with increasing latitude (C) and decreasing longitude (D).

We observed moderate genetic differentiation ($F_{ST} = 0.169$) and inbreeding ($F_{IS} = 0.177$) across and within populations of wild sunflower. F_{ST} between populations varied widely (range = 0.048–0.336; Table S1), and genetic differentiation was found to be spatially structured, as indicated by a significant pattern of IBD (Figure 2B; Mantel's $r = 0.371$, $P = 0.007$). Furthermore, clinal patterns in the distribution of diversity were observed, as indicated by significant declines in π with increasing latitude (Figure 2C;

$F_{1,14} = 4.82$, $P = 0.045$, $r = -0.450$) and decreasing longitude (Figure 2D; $F_{1,14} = 18.1$, $P < 0.001$, $r = 0.729$). AMOVA found that most genetic variation is partitioned within samples (68.8%, $P < 0.001$), with relatively little variation explained by differences between samples (13.2%, $P < 0.001$), populations (13.3%, $P < 0.001$), and genetic clusters (4.56%, $P < 0.001$). The sum of these results suggest that genetic diversity is continuously distributed across the range of wild sunflower, with the highest levels of diversity being concentrated in populations located in the southeastern portion of the range.

3.3. Phylogenetic Relationships

Phylogenetic analyses infer a clear geographic pattern of lineage splitting in wild sunflower (Figure 3A). Samples from each population were resolved as monophyletic (ML BS > 85), with the exception of the population in Wyoming (WY1), where one sample was resolved as sister to a clade of samples from a neighboring population in Montana. The earliest diverging lineage in our sample of wild sunflower was resolved as a population from south Texas (TX1; ML BS = 100), followed by a population from central Texas (TX2; ML BS = 97). Samples from outside of Texas form a well-supported clade (ML BS = 96), with populations from New Mexico and the central Great Plains region (Oklahoma, Kansas, and Nebraska) forming a grade (i.e., relationships resolved among these populations were resolved with ML BS < 50), with respect to a strongly-supported western clade (ML BS = 100), comprising populations from Colorado, Wyoming, Montana, and Alberta. Relationships in the western clade show a south-north pattern of lineage splitting, mirroring the patterns observed more broadly across wild sunflower. Taken together, these phylogenetic patterns suggest that range expansion in wild sunflower occurred along two separate, south–north migration fronts, with multiple genetic lineages colonizing the central portion of the Great Plains, and a single genetic lineage migrating into and diversifying over the western portion of its range.

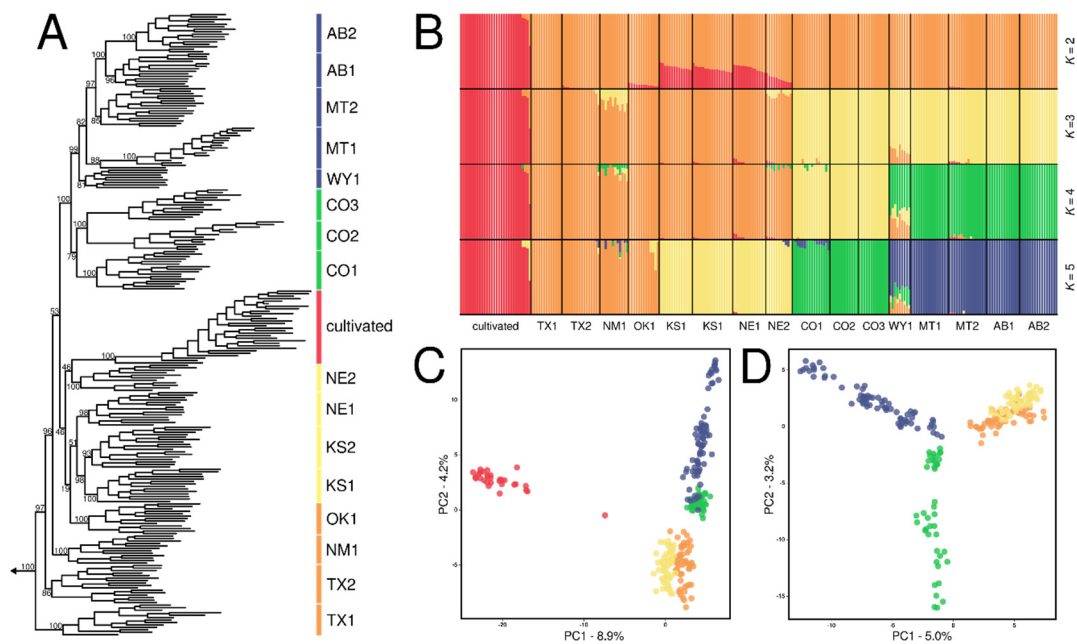


Figure 3. Phylogenetic relationships and population structure across sunflower. (A) Maximum likelihood phylogeny of wild and cultivated sunflower. ML BS support values are denoted for interior nodes and clades corresponding to populations. (B) Population assignments for $K = 2-5$ estimated through fastSTRUCTURE. (C) Positioning of wild and cultivated sunflower samples in two-dimensional genetic space along PCs 1 and 2. (D) Positioning of wild sunflower samples in two-dimensional genetic space along PCs 1 and 2. Colored bars in panel A and colored dots in panels C and D correspond to fastSTRUCTURE groups for $K = 5$ in panel B.

All cultivated sunflower accessions were resolved as a strongly supported clade (ML BS = 100) nested within the New Mexico–Central Great Plains grade (Figure 3A). Four Native American landraces (i.e., Hopi Dye, Arikara, Seneca, Maíz de Tejas, Maíz Negro) diverged early and form a grade at the base of the cultivated clade (Figure 4A). There is little phylogenetic structure after the early diverging Native American landraces, but cultivated accessions appear to sort largely by market type (i.e., oilseed vs. confectionery) rather than heterotic group. This is apparent in the resolution of all but two of the oilseed lines as a clade (the most-inclusive clade containing the oilseed lines PI 599775 (HA123) and two high-oil OPVs (Peredovik and VNIIMK 8931)), and the paraphyly of the confectionery lines. Two oilseed lines (PI 599771 (HA061) and PI 561918 (HA378)) cluster with the confectionery lines, which may be a result of introgressions rather than independently derived oilseed lines. Overall, relationships among cultivated accessions are decidedly complex, but the resolution of Native American landraces at the base of the cultivated sunflower phylogeny is consistent with the view that all modern cultivars of sunflower are descended from Native American landraces. Furthermore, the paraphyly of the confectionery lines, and the sorting of most oilseed lines into a clade, indicate that the oilseed lines were derived from a non-oilseed progenitor, consistent with the known breeding history of cultivated sunflower.

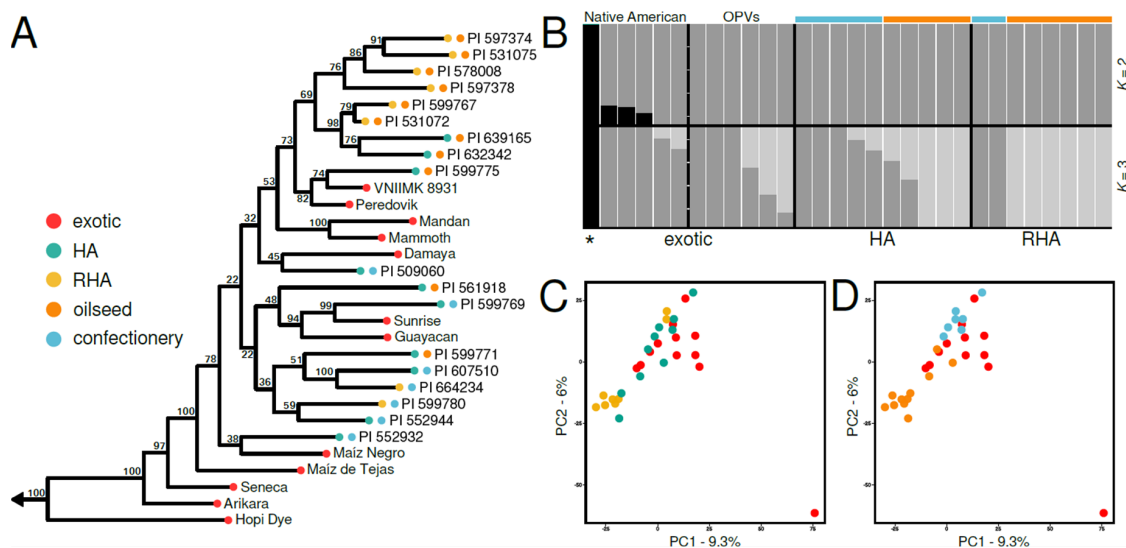


Figure 4. Phylogenetic relationships and population structure within cultivated sunflower. (A) Maximum likelihood phylogeny of cultivated accessions. Tip shapes are colored by breeding pool and market type (i.e., oilseed vs. confectionery), as shown in the inset. ML BS support values are noted for each node. (B) Population assignments for $K = 2$ – 3 , estimated through fastSTRUCTURE. The dashed line demarcates Native American landraces (left) from OPVs (right). The asterisk denotes the Native American landrace accession, Hopi Dye. Colored bars denote market type for each accession follow the coding scheme shown in the inset of panel A. Positioning of cultivated samples in two-dimensional genetic space along PCs 1 and 2 coded by breeding pool (C) and market type (D).

3.4. Population Clustering

fastSTRUCTURE and ADMIXTURE analyses infer diffuse, geographically defined population structure across wild sunflower. Both analyses disagree with respect to the optimal value of K (fastSTRUCTURE, $K = 3$; ADMIXTURE, $K = 8$), but consistently identified a distinct cluster of cultivated accessions while sorting wild samples into increasingly smaller, geographically defined clusters until $K = 17$, where all wild samples were sorted by their collecting locality. This fractal pattern of population clustering is consistent with IBD, so we present results from fastSTRUCTURE analyses for $K = 2$ – 6 , which circumscribe landscape-level, geogenetic clusters (Figure 3B). At $K = 2$, samples were sorted into cultivated and wild clusters. At $K = 3$, the wild samples split into two clusters corresponding to a

southern/eastern and a western cluster. At $K = 4$, the western cluster split into a southern-western cluster composed of samples from Colorado populations and a northern-western cluster composed of samples from Wyoming, Montana, and Alberta. Finally, at $K = 5$, the southern-eastern cluster split into a southern cluster composed of samples from Texas, New Mexico and Oklahoma and an eastern cluster composed of samples from Kansas and Nebraska. Instances of admixture were uncommon across populations, with most samples possessing $> 80\%$ ancestry in a given cluster. However, one Native American landrace accession (Hopi Dye) was consistently estimated to have ca. 50% membership in other clusters across all values of K , and samples from the Wyoming population were found to possess 30–40% admixed ancestry at $K = 4$ and 5 (Figure 3B). PCA recapitulates these patterns, with cultivated accessions positioned distantly from wild samples (Figure 3C), and wild samples from nearby populations grouping together in PC space (Figure 3D).

The population structure within cultivated sunflower is more complex. fastSTRUCTURE and ADMIXTURE analyses favored lower values of K (fastSTRUCTURE, $K = 2$; ADMIXTURE, $K = 1$) and, for both analyses, ancestry assignments at $K > 3$ were difficult to interpret. That being said, we present fastSTRUCTURE results for $K = 2–3$, which reveal subtle, biologically interpretable structure within the cultivated lines. At $K = 2$, a single Native American landrace accession (Hopi Dye) is inferred as its own cluster, with three other Native American landrace accessions (Arikara, Seneca, Maíz de Tejas) sharing some ancestry ($< 20\%$) with the Hopi Dye cluster (Figure 4B). At $K = 3$, Hopi Dye remains a unique cluster, and two additional clusters emerge to separate accessions largely by market type (i.e., oilseed vs. confectionery lines) rather than breeding pool (Figure 4B). PCA also infers subtle structure, with little separation of accessions by breeding pool (Figure 4C), and some differentiation occurring among market types (Figure 4D).

3.5. Demographic Reconstruction

Demographic reconstructions estimate that the wild and cultivated sunflower lineages diverged between 900–5400 ybp (Table 3). All models estimate current effective size of the wild lineage ($N_{\text{wild-current}}$) to be roughly 10- to 20-times greater than the current effective size of the cultivated lineage ($N_{\text{cult-current}}$). Similarly, all models estimate up to a 20-fold reduction in effective size between the founding ($N_{\text{cult-founder}}$) and contemporaneous ($N_{\text{cult-current}}$) cultivated lineage. These dramatic differences in current and historical population sizes between the cultivated and wild lineages are consistent with significant losses of genetic diversity during domestication and subsequent improvement.

Table 3. Maximum likelihood estimates for model parameters estimated using $\delta a \delta i$. 95% confidence intervals for parameter estimates are presented in parentheses. Migration rates (M) are presented in migrants per year.

Model	ll	AIC	ΔAIC	N_{ref} ($\times 10^3$)	$N_{\text{wild-current}}$ ($\times 10^3$)	$N_{\text{cult-founder}}$ ($\times 10^3$)	$N_{\text{cult-current}}$ ($\times 10^3$)	$M_{w \leftrightarrow c}$	$M_{w \rightarrow c}$	$M_{w \leftarrow c}$	T Years ($\times 10^3$)
A	-3303	6614	2719	8.51 (8.30–8.73)	14.7 (13.8–15.7)	8.75 (7.35–10.1)	0.865 (0.792–0.938)	-	-	-	0.912 (0.776–1.06)
B	-2123	4255	361	4.58 (4.31–4.86)	13.2 (12.2–14.1)	21.0 (10.8–31.3)	1.51 (1.39–1.63)	3.06 (2.96–3.16)	-	-	4.85 (4.32–5.37)
C	-1941	3895	0	4.23 (4.06–4.39)	14.5 (13.7–15.3)	9.34 (1.79–16.9)	0.773 (0.726–0.820)	-	3.81 (3.58–4.04)	0.353 (0.301–0.405)	5.37 (5.06–5.67)

AIC favored model C (Table 3; Figure 5; Figure S2), which models asymmetric migration between the cultivated and wild lineages. In this model, the cultivated and wild lineages diverged 5370 ybp, with the cultivated lineage undergoing sequential bottlenecks, ultimately resulting in a nearly 20-fold reduction in the effective size of the modern breeding pool. Migration from the wild lineage into the cultivated lineage ($M_{w \rightarrow c}$) was estimated at 3.81 migrants per year (95% CI: 3.58–4.04), which is an order of magnitude greater than migration from the cultivated into wild lineage ($M_{w \leftarrow c} = 0.353$ (0.301–0.405)).

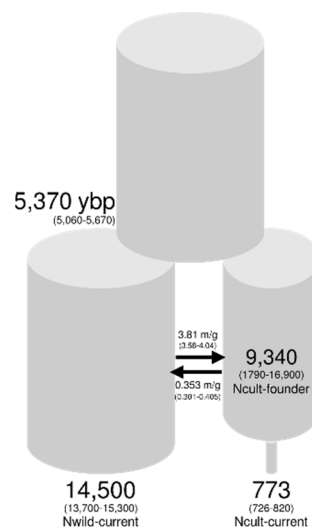


Figure 5. Parameter estimates and 95% confidence intervals (in parentheses) for a demographic model of sunflower domestication assuming divergence with asymmetric gene flow (Model C).

4. Discussion

4.1. Phylogeography of Wild Sunflower

Populations across the range of wild sunflower have diverged primarily along a south-north axis (Figure 3A), which is consistent with a scenario of postglacial range expansion from a southern refugium. The observed pattern of IBD (Figure 2B), partitioning of most genetic diversity at finer spatial scales, and relatively weak population structure (Figure 3B–D) indicates that genetic diversity is continuously distributed over the species range, and suggests that range expansion occurred rapidly following glacial retreat. Furthermore, linear declines in nucleotide diversity with increasing latitude and decreasing longitude (Figure 2C,D) indicate that range expansion likely occurred in a stepwise fashion [56], through sequential founding events as colonizing populations migrated north- and westward from refugial populations located in the southeastern portion of the range.

Together, these findings suggest that the dynamics of postglacial range expansion in wild sunflower are similar to those observed in numerous European plant species. Indeed, many such species have been found to have undergone dramatic range contractions into southern refugia during glacial periods and rapidly expanded northward as climates warmed following the LGM [23–25,57]. This general pattern of northward expansion from lower latitude refugial areas has also been observed in a number of widespread species, whose contemporary distributions span both glaciated and unglaciated North America: e.g., herbs (*Asclepias exaltata* [58], *Campanulastrum americanum* [59], *Trillium erectum* and *T. grandiflorum* [60,61], and *Symplocarpus foetidus* [62]); shrubs (*Dirca paulustris* [63], *Viburnum lantanoides* [64], *Viburnum nudum* complex [65], and the *Lentago* clade of *Viburnum* [66]); and trees (e.g., *Acer rubrum* and *sacharum* [67,68], *Carya cordiformis* and *ovata* [69,70], *Fagus grandifolia* [67,71], and *Pinus strobus* [72,73]).

Paleoecological studies provide a finer resolution to the approximate locations of refugial areas during the late Pleistocene, and suggest that many of the aforementioned species may have persisted in macrorefugia distributed along the Gulf Coast [74], with colder-adapted species surviving in smaller refugia along the Atlantic Coast [75], or in cryptic microrefugia at mid-latitudes [76,77]. In the case of wild sunflower, there are no fossils that place it in any of these regions during the LGM. However, there are records of composite pollen originating during the LGM from eastern Texas, the lower Mississippi River Valley, peninsular Florida, and the coastal Carolinas [74]. Given that the core of the wild sunflower distribution is centered in the Great Plains region [5], and that phylogenetic and genetic diversity is concentrated in the southern and eastern portions of its range, we postulate that its refugial

areas may have been located in adjacent areas such as eastern Texas and the lower Mississippi River Valley, both of which harbored grassland species during the LGM [74]. Future ecogeographic studies incorporating well-curated distributional data of wild sunflower populations and paleoclimatic niche modelling would serve as excellent tests of this hypothesis.

Our analysis of wild sunflower adds to the growing body of work demonstrating the general trend of south-north postglacial migration inferred in wide-ranging North American plant species. Interestingly, there is a dearth of phylogeographic studies that have examined a wide-ranging plant species that spans the entirety of the Great Plains (but see [78]). As such, our study provides interesting insights into central North American plant phylogeography. One pattern inferred in wild sunflower is the resolution of separate south-north patterns of lineage splitting in the western and central Great Plains region (Figure 3A), with multiple genetic lineages colonizing the central Great Plains, and a single genetic lineage migrating and diversifying over the western Great Plains. This pattern may be explained by the physiography of the western Great Plains, which is at higher elevation and possesses a cool, arid climate. This region—the High Plains—has been shown to be an important biogeographic break for many animals in North America [79], where previously isolated taxa/populations situated on either side of the region have been shown to have come into secondary contact as climates warmed during the Holocene [80–82]. Indeed, in our study, samples from a population collected in eastern Wyoming were resolved as paraphyletic (Figure 3A), possessing slightly greater nucleotide diversity compared to other populations at similar latitudes (Figure 2C) and having > 30% admixed ancestry at higher values of K (Figure 3B). Our findings in wild sunflower suggest that the High Plains played an important role in generating contemporary patterns of divergence and genetic structure in not just animals, but wide-ranging plants distributed across central North America.

4.2. Insights into the Domestication and Breeding History of Cultivated Sunflower

Patterns of divergence and population structure in cultivated sunflower are complex, but largely reflective of its domestication and breeding history. Cultivated accessions form a strongly supported clade, nested among wild sunflower populations from the central Great Plains (Figure 3A), and estimated to have arisen into an independently evolving entity ca. 5370 ybp (Table 3; Figure 5). Within the cultivated clade, five Native American landraces (Hopi Dye, Arikara, Seneca, Maíz de Tejas, and Maíz Negro) were resolved as the earliest diverging lineages, which split in succession from a single founding lineage and eventually gave rise to the modern cultivars (Figure 4A). These findings agree well with those from previous studies, demonstrating that extant sunflower cultivars trace back to a single origin of domestication [13–15] ca. 5000 ybp [7] in east-central North America [13,16]. Propagules of this initial domestication were then presumably dispersed between different Native American cultures who used it for food and cultural purposes [8].

Patterns of phylogenetic and population structure outside of the early diverging landraces become apparent when the cultivars are coded by market type (i.e., oilseed vs. confectionery) rather than breeding pool (i.e., exotic vs. HA vs. RHA) (Figure 4). The sorting of cultivars by market type is not unexpected [17,21,22], as early breeding efforts in Eastern Europe were focused on increasing oil content [10], which likely resulted in substantial genetic differentiation. The development of inbred lines and accompanying transition to hybrid breeding occurred much more recently [83]. Our results reflect this history, where nearly all oilseed lines were resolved as a clade in relation to a grade composed of confectionery lines, with two Russian developed high-oil OPVs (Peredovik and VNIIMK 8931) splitting early within the oilseed clade's history. Given the limited sample of accessions included in our study, future studies examining a greater number and diversity of both wild and domesticated lineages will be useful in confirming the patterns inferred in this study, and gaining more pointed insights into the origin of domesticated sunflower and the effects of historical breeding efforts in generating observed patterns of relatedness and genetic structure. Of particular value might be an expansion of the wild sunflower sampling to provide better coverage of the eastern and western portions of its range.

4.3. Domestication and Its Effects on Polymorphism

Domestication and improvement have generated large differences in observed levels of genetic diversity in the wild and cultivated sunflower breeding pools. Unsurprisingly, exotic and elite lines were found to harbor ca. 60% and 50% of the nucleotide diversity (π) present in wild sunflower, respectively (Figure 2A). These results compare favorably with those from previous surveys of SSR and SNP diversity in sunflower, which have consistently estimated up to a 50% reduction in various measures of diversity (primarily gene diversity (H_e)) between wild and cultivated sunflower [16–21,84]. The consistency across studies and marker types suggest that the effects of domestication and improvement were dramatic, affecting both SSR and SNP variation across the sunflower genome.

Demographic reconstructions provide some insight into the timing and order of these changes in genetic diversity (Table 3; Figure 5). For example, we inferred a dramatic 12-fold reduction in effective size over the history of the cultivated lineage (i.e., $N_{\text{cult-founder}}$ vs. $N_{\text{cult-current}}$) and a more subtle 1.5-fold difference in effective sizes between the wild lineage and the founding population of the cultivated lineage (i.e., $N_{\text{wild-current}}$ vs. $N_{\text{cult-founder}}$). These results support the notion that genetic diversity in cultivated sunflower was lost progressively, with a moderate loss of diversity during the initial domestication bottleneck, and more severe reductions in diversity following strong directional selection and additional bottlenecks during improvement (reviewed in [85–87]). Losses in genetic diversity following domestication and improvement are a common feature of many cultivated plant species [87,88]. However, the patterns observed in sunflower contrasts with those reconstructed in other annual crops such as common bean and maize, where current effective sizes are much larger than their inferred domestication bottleneck sizes, possibly due to rapid population expansion or ongoing gene flow with wild relatives [89–91].

In sunflower, moderate rates of gene flow from the wild into cultivated breeding pools do not appear to have strongly influenced current effective sizes (Table 3; Figure 5), which may be reflective of the targeted nature of introgression events in cultivated sunflower breeding (e.g., the introduction of disease resistance loci from wild donors [84,92,93]). Overall, these results are consistent with the known history of cultivated sunflower, but many issues remain unresolved: specifically, the duration of the domestication bottleneck, and the tempo and mode of bottleneck-induced population declines. A genomic analysis of contemporary and archeological specimens (e.g., [94]) with recently developed methods designed to infer more granular changes in effective population size through time [95] may be useful in generating richer insights into the broad demographic patterns observed in this study.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/3/266/s1>, Table S1: Pairwise F_{ST} estimates between the cultivated accessions and wild sunflower populations, Figure S1: Schematic representations of demographic models fitted using $\delta a \delta i$. Estimated parameters are noted with text for each model, Figure S2. $\delta a \delta i$ analysis of model C. Upper panels are the observed and expected site frequency spectra. Lower panels are a heat map and histogram of residuals.

Author Contributions: Conceptualization, B.P. and J.M.B.; Methodology, B.P.; Formal Analysis, B.P.; Investigation, B.P. and J.M.B.; Resources, J.M.B.; Data Curation, B.P. and J.M.B.; Writing—Original Draft Preparation, B.P.; Writing—Review and Editing, B.P. and J.M.B.; Visualization, B.P. and J.M.B.; Supervision, J.M.B.; Project Administration, B.P. and J.M.B.; Funding Acquisition, J.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agriculture and Food Research Initiative (award no. 2016-33522-25630) of the USDA National Institute of Food and Agriculture and by the NSF Plant Genome Research Program (award no. IOS-1444522).

Acknowledgments: We are grateful to Magdy Alabady, Casey Morrow, and Noah Workman at the Georgia Genomics and Bioinformatics Core (GGBC) for their help with library preparation and sequencing. We also thank Andrea Dixon for assistance with procuring samples and laboratory work and Eric Baack for providing tissue samples and useful discussions. Finally, we would like to acknowledge members of the Burke Lab and two anonymous reviewers for providing helpful comments that improved this manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data Accessibility: Illumina sequence data generated in this study are accessible from the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under BioProject number PRJNA609331.

References

1. Harris, D.R. Vavilov's concept of centres of origin of cultivated plants: Its genesis and its influence on the study of agricultural origins. *Biol. J. Linn. Soc.* **1990**, *39*, 7–16. [[CrossRef](#)]
2. Schoen, D.J.; Brown, A.H.D. The conservation of wild plant species in seed banks. *BioScience* **2001**, *51*, 960. [[CrossRef](#)]
3. Takeda, S.; Matsuoka, M. Genetic approaches to crop improvement: Responding to environmental and population changes. *Nat. Rev. Genet.* **2008**, *9*, 444–457. [[CrossRef](#)]
4. Dempewolf, H.; Baute, G.; Anderson, J.; Kilian, B.; Smith, C.; Guarino, L. Past and future use of wild relatives in crop breeding. *Crop. Sci.* **2017**, *57*, 1070–1082. [[CrossRef](#)]
5. Marek, L.F. Crop wild relatives of sunflower in North America. In *North American Crop Wild Relatives*; Greene, S., Williams, K., Khoury, C., Kantar, M., Marek, L., Eds.; Springer: Berlin, Germany, 2019; Volume 2, pp. 453–483.
6. Heiser, C.B.; Smith, D.M.; Clevenger, S.; Martin, W.C. The North American sunflowers (*Helianthus*). *Mem. Torrey. Bot. Club* **1969**, *22*, 1–218.
7. Smith, B.D. Eastern North America as an independent center of plant domestication. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12223–12228. [[CrossRef](#)] [[PubMed](#)]
8. Heiser, C.B. The sunflower among the North American Indians. *Proc. Am. Phils. Soc.* **1951**, *95*, 432–448.
9. Putt, E.D. Early history of sunflower. In *Sunflower Production and Technology*; Schneiter, A.A., Ed.; American Society Agron: Madison, Wisconsin, 1997; pp. 1–19.
10. Gavrilova, V.A.; Anisimova, I.N. Genealogy of the sunflower lines created on the basis of Russian varieties. *Helia* **2017**, *67*, 1–12. [[CrossRef](#)]
11. Korell, M.; Moosges, G.; Friedt, W. Construction of a sunflower pedigree map. *Helia* **1992**, 7–16.
12. Baute, G.J.; Kane, N.C.; Grassa, C.J.; Lai, Z.; Rieseberg, L.H. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol.* **2015**, *206*, 830–838. [[CrossRef](#)]
13. Harter, A.V.; Gardner, K.A.; Falush, D.; Lentz, D.L.; Bye, R.A.; Rieseberg, L.H. Origin of extant domesticated sunflowers in eastern North America. *Nature* **2004**, *430*, 201–205. [[CrossRef](#)]
14. Wills, D.M.; Burke, J.M. Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). *J. Hered.* **2006**, *97*, 403–408. [[CrossRef](#)] [[PubMed](#)]
15. Blackman, B.K.; Scascitelli, M.; Kane, N.C.; Luton, H.H.; Rasmussen, D.A.; Bye, R.A.; Lentz, D.L.; Rieseberg, L.H. Sunflower domestication alleles support single domestication center in eastern North America. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14360–14365. [[CrossRef](#)] [[PubMed](#)]
16. Cronn, R.; Brothers, M.; Klier, K.; Bretting, P.K.; Wendel, J.F. Allozyme variation in domesticated annual sunflower and its wild relatives. *Theor. Appl. Genet.* **1997**, *95*, 532–545. [[CrossRef](#)]
17. Tang, S.; Knapp, S.J. Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. *Theor. Appl. Genet.* **2003**, *106*, 990–1003. [[CrossRef](#)] [[PubMed](#)]
18. Burke, J.M.; Knapp, S.J.; Rieseberg, L.H. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* **2005**, *171*, 1933–1940. [[CrossRef](#)]
19. Liu, A.; Burke, J.M. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **2006**, *173*, 321–330. [[CrossRef](#)]
20. Kolkman, J.M.; Berry, S.T.; Leon, A.J.; Slabaugh, M.B.; Tang, S.; Gao, W.; Shintani, D.K.; Burke, J.M.; Knapp, S.J. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* **2007**, *177*, 457–468. [[CrossRef](#)]
21. Mandel, J.R.; Dechaine, J.M.; Marek, L.F.; Burke, J.M. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theor. Appl. Genet.* **2011**, *123*, 693–704. [[CrossRef](#)]
22. Mandel, J.R.; Nambeesan, S.; Bowers, J.E.; Marek, L.F.; Ebert, D.; Rieseberg, L.H.; Knapp, S.J.; Burke, J.M. Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* **2013**, *9*, e1003378. [[CrossRef](#)]

23. Hewitt, G.M. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* **1999**, *68*, 87–112. [[CrossRef](#)]
24. Hewitt, G.M. The genetic legacy of the Quaternary ice ages. *Nature* **2000**, *405*, 907–913. [[CrossRef](#)] [[PubMed](#)]
25. Hewitt, G.M. Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. B* **2004**, *359*, 185–195; discussion 195. [[CrossRef](#)] [[PubMed](#)]
26. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)] [[PubMed](#)]
27. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253. [[CrossRef](#)]
28. Eaton, D.A.R. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **2014**, *30*, 1844–1849. [[CrossRef](#)] [[PubMed](#)]
29. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10. [[CrossRef](#)]
30. Todesco, M.; Owens, G.L.; Bercovich, N.; L egar e, J.S.; Soudi, S.; Burge, D.O.; Huang, K.; Ostevik, K.L.; Drummond, E.B.M.; Imerovski, I.; et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *BioRxiv* **2019**, 790279.
31. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
32. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup the Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
33. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 1–34. [[CrossRef](#)] [[PubMed](#)]
34. 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.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
35. Huang, H.; Knowles, L.L. Unforeseen Consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Syst. Biol.* **2016**, *65*, 357–365. [[CrossRef](#)]
36. Eaton, D.A.R.; Spriggs, E.L.; Park, B.; Donoghue, M.J. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering Plants. *Syst. Biol.* **2017**, *66*, 399–412. [[CrossRef](#)]
37. Linck, E.; Battey, C.J. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol. Ecol. Resour.* **2019**, *19*, 639–647. [[CrossRef](#)]
38. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
39. Canty, A.; Ripley, B.D. *boot: Bootstrap R (S-Plus) Functions*, R package version 1.3-24; R Foundation for Statistical Computing: Vienna, Austria, 2019.
40. Weir, B.S.; Clark Cockerham, C. Estimating F-statistics for the analysis of population structure. *Evolution* **1984**, *38*, 1358. [[PubMed](#)]
41. Goudet, J. hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [[CrossRef](#)]
42. Dray, S.; Dufour, A.B. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **2007**, *22*. [[CrossRef](#)]
43. Bougeard, S.; Dray, S. Supervised multiblock analysis in R with the ade4 package. *J. Stat. Softw.* **2018**, *86*. [[CrossRef](#)]
44. Hijmans, R.J. *Geosphere: Spherical Trigonometry*, R package version 1.5-10; R Foundation for Statistical Computing: Vienna, Austria, 2019.
45. Kamvar, Z.N.; Tabima, J.F.; Gr unwald, N.J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281. [[CrossRef](#)]
46. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
47. Raj, A.; Stephens, M.; Pritchard, J.K. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **2014**, *197*, 573–589. [[CrossRef](#)] [[PubMed](#)]

48. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)] [[PubMed](#)]
49. Frichot, E.; François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **2015**, *6*, 925–929. [[CrossRef](#)]
50. Gutenkunst, R.N.; Hernandez, R.D.; Williamson, S.H.; Bustamante, C.D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **2009**, *5*, e1000695. [[CrossRef](#)]
51. Portik, D.M.; Leaché, A.D.; Rivera, D.; Barej, M.F.; Burger, M.; Hirschfeld, M.; Rödel, M.O.; Blackburn, D.C.; Fujita, M.K. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* **2017**, *26*, 5245–5263. [[CrossRef](#)]
52. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer Science & Business Media: Berlin, Germany, 2007.
53. Carstens, B.C.; Brennan, R.S.; Chua, V.; Duffie, C.V.; Harvey, M.G.; Koch, R.A.; McMahan, C.D.; Nelson, B.J.; Newman, C.E.; Satler, J.D.; et al. Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Mol. Ecol.* **2013**, *22*, 4014–4028. [[CrossRef](#)]
54. Coffman, A.J.; Hsieh, P.H.; Gravel, S.; Gutenkunst, R.N. Computationally efficient composite likelihood statistics for demographic inference. *Mol. Biol. Evol.* **2016**, *33*, 591–593. [[CrossRef](#)]
55. Sambatti, J.B.M.; Strasburg, J.L.; Ortiz-Barrientos, D.; Baack, E.J.; Rieseberg, L.H. Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. *Evolution* **2012**, *66*, 1459–1473. [[CrossRef](#)]
56. Slatkin, M.; Excoffier, L. Serial founder effects during range expansion: A spatial analog of genetic drift. *Genetics* **2012**, *191*, 171–181. [[CrossRef](#)]
57. Petit, R.J.; Aguinagalde, I.; de Beaulieu, J.-L.; Bittkau, C.; Brewer, S.; Cheddadi, R.; Ennos, R.; Fineschi, S.; Grivet, D.; Lascoux, M.; et al. Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **2003**, *300*, 1563–1565. [[CrossRef](#)] [[PubMed](#)]
58. Broyles, S. Postglacial migration and the loss of allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae). *Am. J. Bot.* **1998**, *85*, 1091. [[CrossRef](#)] [[PubMed](#)]
59. Barnard-Kubow, K.B.; Debban, C.L.; Galloway, L.F. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *Am. J. Bot.* **2015**, *102*, 1842–1853. [[CrossRef](#)] [[PubMed](#)]
60. Griffin, S.R.; Barrett, S.C.H. Genetic variation in *Trillium erectum* (Melanthiaceae), a widespread forest herb in eastern North America. *Can. J. Bot.* **2004**, *82*, 316–321. [[CrossRef](#)]
61. Griffin, S.R.; Barrett, S.C.H. Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: Inferences from phylogeography. *Am. J. Bot.* **2004**, *91*, 465–473. [[CrossRef](#)]
62. Kim, S.H.; Cho, M.S.; Li, P.; Kim, S.C. Phylogeography and ecological niche modeling reveal reduced genetic diversity and colonization patterns of skunk cabbage (*Symplocarpus foetidus*; Araceae) from glacial refugia in Eastern North America. *Front. Plant Sci.* **2018**, *9*, 648. [[CrossRef](#)]
63. Peterson, B.J.; Graves, W.R. Chloroplast phylogeography of *Dirca palustris* L. indicates populations near the glacial boundary at the Last Glacial Maximum in eastern North America. *J. Biogeogr.* **2016**, *43*, 314–327. [[CrossRef](#)]
64. Park, B.; Donoghue, M.J. Phylogeography of a widespread eastern North American shrub, *Viburnum lantanoides*. *Am. J. Bot.* **2019**, *106*, 389–401. [[CrossRef](#)]
65. Spriggs, E.L.; Eaton, D.A.R.; Sweeney, P.W.; Schlutius, C.; Edwards, E.J.; Donoghue, M.J. Restriction-Site-Associated DNA Sequencing Reveals a Cryptic *Viburnum* Species on the North American Coastal Plain. *Syst. Biol.* **2019**, *68*, 187–203. [[CrossRef](#)]
66. Spriggs, E.L.; Schlutius, C.; Eaton, D.A.; Park, B.; Sweeney, P.W.; Edwards, E.J.; Donoghue, M.J. Differences in flowering time maintain species boundaries in a continental radiation of *Viburnum*. *Am. J. Bot.* **2019**, *106*, 833–849. [[CrossRef](#)]
67. McLachlan, J.S.; Clark, J.S.; Manos, P.S. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* **2005**, *86*, 2088–2098. [[CrossRef](#)]
68. Saeki, I.; Dick, C.W.; Barnes, B.V.; Murakami, N. Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): Impacts of habitat specialization, hybridization and glacial history. *J. Biogeogr.* **2011**, *38*, 992–1005. [[CrossRef](#)]

69. Bemmels, J.B.; Dick, C.W. Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species. *J. Biogeogr.* **2018**, *45*, 1739–1750. [[CrossRef](#)]
70. Bemmels, J.B.; Lacey Knowles, L.; Dick, C.W. Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8431–8436. [[CrossRef](#)]
71. Morris, A.B.; Graham, C.H.; Soltis, D.E.; Soltis, P.S. Reassessment of phylogeographical structure in an eastern North American tree using Monmonier’s algorithm and ecological niche modelling. *J. Biogeogr.* **2010**, *37*, 1657–1667. [[CrossRef](#)]
72. Nadeau, S.; Godbout, J.; Lamothe, M.; Gros-Louis, M.C.; Isabel, N.; Ritland, K. Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: A comparison between eastern and western North American postglacial colonization histories. *Am. J. Bot.* **2015**, *102*, 1342–1355. [[CrossRef](#)]
73. Zinck, J.W.R.; Rajora, O.P. Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (*Pinus strobus*) in North America: Single refugium, multiple routes. *BMC Evol. Biol.* **2016**, *16*, 56. [[CrossRef](#)]
74. Jackson, S.T.; Webb, R.S.; Anderson, K.H.; Overpeck, J.T.; Webb III, T.; Williams, J.W.; Hansen, B.C.S. Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quat. Sci. Rev.* **2000**, *19*, 489–508. [[CrossRef](#)]
75. Williams, J.W.; Shuman, B.N.; Webb, T.; Bartlein, P.J.; Leduc, P.L. Late-quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecol. Monogr.* **2004**, *74*, 309–334. [[CrossRef](#)]
76. Delcourt, P.A.; Delcourt, H.R.; Brister, R.C.; Lackey, L.E. Quaternary vegetation history of the Mississippi embayment. *Quat. Res.* **1980**, 111–132. [[CrossRef](#)]
77. Delcourt, H.R.; Delcourt, P.A. Ice age haven for hardwoods. *Nat. Hist.* **1984**, 22–28.
78. Avendaño-González, M.; Morales-Domínguez, J.F.; Siqueiros-Delgado, M.E. Genetic structure, phylogeography, and migration routes of *Bouteloua gracilis* (Kunth) Lag. ex Griffiths (Poaceae:Chloridoideae). *Mol. Phylogenetics Evol.* **2019**, *134*, 50–60. [[CrossRef](#)] [[PubMed](#)]
79. Remington, C.L. Suture-zones of hybrid interaction between recently joined biotas. *Evol. Biol.* **1968**, *321*, 428.
80. Swenson, N.G.; Howard, D.J. Do suture zones exist? *Evolution* **2004**, *58*, 2391–2397. [[CrossRef](#)] [[PubMed](#)]
81. Swenson, N.G.; Howard, D.J. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *Am. Nat.* **2005**, *166*, 581–591. [[CrossRef](#)] [[PubMed](#)]
82. Swenson, N.G. GIS-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *J. Evolution. Biol.* **2006**, *19*, 717–725. [[CrossRef](#)]
83. Fick, G.N.; Miller, J.F. Sunflower breeding. In *Sunflower Production and Technology*; Scheiter, A.A., Ed.; American Society of Agronomy: Madison, WI, USA, 1997; pp. 395–440.
84. Hübner, S.; Bercovich, N.; Todesco, M.; Mandel, J.R.; Odenheimer, J.; Ziegler, E.; Lee, J.S.; Baute, G.J.; Owens, G.L.; Grassa, C.J.; et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* **2019**, *5*, 54–62. [[CrossRef](#)]
85. Burke, J.M.; Burger, J.C.; Chapman, M.A. Crop evolution: From genetics to genomics. *Curr. Opin. Genet. Dev.* **2007**, *17*, 525–532. [[CrossRef](#)]
86. Burger, J.C.; Chapman, M.A.; Burke, J.M. Molecular insights into the evolution of crop plants. *Am. J. Bot.* **2008**, *95*, 113–122. [[CrossRef](#)]
87. Moyers, B.T.; Morrell, P.L.; McKay, J.K. Genetic Costs of Domestication and Improvement. *J. Hered.* **2018**, *109*, 103–116. [[CrossRef](#)]
88. Miller, A.J.; Gross, B.L. From forest to field: Perennial fruit crop domestication. *Am. J. Bot.* **2011**, *98*, 1389–1414. [[CrossRef](#)] [[PubMed](#)]
89. Schmutz, J.; McClean, P.E.; Mamidi, S.; Wu, G.A.; Cannon, S.B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, *46*, 707–713. [[CrossRef](#)] [[PubMed](#)]
90. Beissinger, T.M.; Wang, L.; Crosby, K.; Durvasula, A.; Hufford, M.B.; Ross-Ibarra, J. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2016**, *2*, 16084. [[CrossRef](#)] [[PubMed](#)]
91. Wang, L.; Beissinger, T.M.; Lorant, A.; Ross-Ibarra, C.; Ross-Ibarra, J.; Hufford, M.B. The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* **2017**, *18*, 1–13. [[CrossRef](#)] [[PubMed](#)]
92. Seiler, G.J. Utilization of wild sunflower species for the improvement of cultivated sunflower. *Field Crop Res.* **1992**, *30*, 195–230. [[CrossRef](#)]

93. Seiler, G.J.; Qi, L.L.; Marek, L.F. Utilization of sunflower crop wild relatives for cultivated sunflower improvement. *Crop Sci.* **2017**, *57*, 1083–1101. [[CrossRef](#)]
94. Wales, N.; Akman, M.; Watson, R.H.B.; Sánchez Barreiro, F.; Smith, B.D.; Gremillion, K.J.; Gilbert, M.T.P.; Blackman, B.K. Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3000 years of sunflower domestication and improvement. *Evol. Appl.* **2019**, *12*, 38–53. [[CrossRef](#)]
95. Li, H.; Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **2011**, *475*, 493–496. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).