Received: 15 June 2016 | Revised: 20 May 2017 | Accepted: 24 May 2017

DOI: 10.1111/mec.14202

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Genome-wide analysis of allele frequency change in sunflower crop–wild hybrid populations evolving under natural conditions

Jonathan Corbi[1,2] | Eric J. Baack[3] | Jennifer M. Dechaine[4] | Gerald Seiler[5] | John M. Burke[1]

[1]Department of Plant Biology, Miller Plant Sciences, University of Georgia, Athens, GA, USA

[2]CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, Univ Lyon, Université Lyon 1, Villeurbanne, France

[3]Department of Biology, Luther College, Decorah, IA, USA

[4]Department of Biological Sciences, Central Washington University, Ellensburg, WA, USA

[5]USDA-ARS, Northern Crop Science Lab., Fargo, ND, USA

Correspondence
John M. Burke, Department of Plant Biology, Miller Plant Sciences, University of Georgia, Athens, GA, USA.
Email: jmburke@uga.edu

Funding information
NSF Plant Genome Research Program, Grant/Award Number: IOS-1444522; USDA Biotechnology Risk Assessment Research Grants (BRAG) Program, Grant/Award Number: 2010-33522-21668, 2016-33522-25630

Abstract

Crop-wild hybridization occurs in numerous plant species and could alter the genetic structure and evolutionary dynamics of wild populations. Studying crop-derived alleles in wild populations is also relevant to assessing/mitigating the risks associated with transgene escape. To date, crop-wild hybridization has generally been examined via short-term studies, typically within a single generation, focusing on few traits or genetic markers. Little is known about patterns of selection on crop-derived alleles over multiple generations, particularly at a genome-wide scale. Here, we documented patterns of natural selection in an experimental crop × wild sunflower population that was allowed to evolve under natural conditions for two generations at two locations. Allele frequencies at a genome-wide collection of SNPs were tracked across generations, and a common garden experiment was conducted to compare trait means between generations. These data allowed us to identify instances of selection on crop-derived alleles/traits and, in concert with QTL mapping results, test for congruence between our genotypic and phenotypic results. We found that natural selection overwhelmingly favours wild alleles and phenotypes. However, crop alleles in certain genomic regions can be favoured, and these changes often occurred in parallel across locations. We did not, however, consistently observe close agreement between our genotypic and phenotypic results. For example, when a trait evolved towards the wild phenotype, wild QTL alleles associated with that trait did not consistently increase in frequency. We discuss these results in the context of crop allele introgression into wild populations and implications for the management of GM crops.

KEYWORDS
allele frequency, crop-wild hybridization, domestication traits, gene flow, *Helianthus annuus*, natural selection

1 | INTRODUCTION

Cultivated plants and their wild relatives often hybridize due to overlapping geographical distributions and phenologies (Ellstrand, 2003a). Introgressed, crop-derived alleles have been identified in at least one wild or weedy relative of 17 major cultivated species (reviewed in Ellstrand et al., 2013). Potential environmental consequences of crop-wild introgression include the reduction of genetic diversity in wild populations (Aerts et al., 2013) and the evolution of increasingly weedy or invasive species. In fact, crop-wild gene flow has affected

weed evolution in multiple taxa and has, in some cases, increased the fitness of weeds (Ellstrand et al., 2013; Muller, Latreille, & Tollon, 2011). Models examining the risk of crop allele introgression into wild or weedy relatives find that the probability of introgression depends on several factors, including the fitness effects of the allele, the effects of linked alleles and the probability of recombination (e.g., Ghosh, Meirmans, & Haccou, 2012; Haygood, Ives, & Andow, 2004). Predicting which crop-derived alleles are likely to introgress will thus require an understanding of their fitness effects under wild ecological conditions and in their genomic context.

Research in early-generation crop-wild hybrids grown under wild or weedy conditions has shown fitness advantages of hybrids over wild populations (Campbell & Snow, 2007; Kost, Alexander, Jason Emry, & Mercer, 2015; Mercer, Andow, Wyse, & Shaw, 2007) and that some crop-like traits are favoured over multiple generations (Campbell, Snow, & Sweeney, 2009). These findings support the view that crop-derived alleles have the potential to be advantageous in wild environments, but they did not attempt to identify underlying causal alleles or elucidate their genomic context. Furthermore, the long-term fitness effects of a particular allele may be difficult to determine from the performance of early-generation hybrids. Following the initial hybridization event, for example, potentially favourable crop-derived alleles will be linked to low-fitness alleles for certain crop-like traits. Only after there has been an opportunity for recombination will the fitness effects of an individual allele be clear (de Jong & Rong, 2013).

Quantitative trait locus (QTL) mapping studies in highly recombined (recombinant inbred line; RIL) sunflower and lettuce crop-wild hybrids have begun to associate crop-derived alleles with selectively advantageous phenotypes at a genome-wide scale (Baack, Sapir, Chapman, Burke, & Rieseberg, 2008; Dechaine et al., 2009; Hartman et al., 2012; Owart, Corbi, Burke, & Dechaine, 2014), but these studies did not monitor allele frequency change over time. The one study to examine genome-wide, year-to-year evolution of crop alleles under wild conditions found several genomic hotspots and cold areas for crop allele introgression in selfed and backcrossed crop-wild hybrid lettuce populations (Hooftman et al., 2011). The next step in predicting the long-term fate of crop-derived alleles in the wild requires monitoring genome-wide allele frequency changes in crop-wild hybrids that have had ample opportunity for recombination. Likewise, because studies have suggested that the fitness effects of crop-derived alleles may vary across environments, hybrids should be evaluated under multiple conditions (Hartman et al., 2012; Hovick, Campbell, Snow, & Whitney, 2012).

Identifying crop-derived alleles and traits that are advantageous in a variety of natural environments could help to anticipate the potential ecological risks of the introduction of transgenes into wild or weedy populations via hybridization. More specifically, this should provide insight into the types of transgenes that would be likely to invade wild/hybrid populations, and which crop-derived genes could contribute to transgene introgression through genetic hitchhiking. Moreover, crop-derived alleles that are generally disadvantageous are good candidates for use in transgene mitigation (Ellstrand, 2003b; Kwit, Moon, Warwick, & Stewart, 2011; Rong et al., 2010). For example, traits that arose during domestication, such as decreased competitive ability/height, greater susceptibility to herbivory (Chen, Gols, & Benrey, 2015), or loss of seed dormancy (Adler, Wikler, Wyndham, Linder, & Schmitt, 1993; Landbo & Jørgensen, 1997; Linder, Taha, Seiler, Snow, & Rieseberg, 1998) are generally thought to reduce fitness in the wild, and major loci affecting these traits could thus be linked to a transgene to mitigate its escape. However, some domestication traits are likely to confer a benefit in wild conditions. For example, a change in flowering time could facilitate the colonization of a new habitat (see Pyšek & Richardson, 2007 and references therein), or disease resistance introgressed from a cultivar to a wild relative could increase the fitness of the resulting hybrid (Warren & James, 2006). It is thus important to examine how domestication traits affect fitness across multiple wild environments and elucidate the genomic context of major loci affecting these traits.

Here, we describe the use of experimentally derived cultivated × wild sunflower (*Helianthus annuus* L.) hybrid populations to investigate the response of crop-derived alleles and traits to natural selection in two wild environments. Several phenotypes of cultivated sunflowers differ from those of the wild relative (Burke, Tang, Knapp, & Rieseberg, 2002), including earlier flowering, reduced branching and strong apical dominance resulting in a single flowering head (inflorescence) with many large seeds. Evidence indicates that the extant sunflower cultivars derive from a single domestication event ca. 4,000–5,000 BC (Harter et al., 2004; Wills & Burke, 2006) in east-central North America (Blackman, Scascitelli, et al. 2011) where it was grown as a source of edible seed and for ceremonial purposes (Heiser, Smith, Clevenger, & Martin, 1969; Putt, 1997; Soleri & Cleveland, 1993). After being introduced in Europe as an ornamental plant, cultivated sunflower was secondarily selected in Russia for its oil content, leading to the first "oilseed sunflowers" by the middle of the 18th century (Berville, Muller, Poinso, Serieys, & Gressel, 2005; Vrânceanu, 2000). Cultivated sunflower was then brought back to North America in the mid-20th century, where it has since been grown primarily as an oilseed crop (Putt, 1997).

Two-thirds of cultivated sunflower fields in the United States can be found in close proximity to, and flower coincidentally with, wild sunflower populations (Burke, Gardner, & Rieseberg, 2002). Introgression of cultivated sunflower alleles into wild or weedy populations (Linder et al., 1998) is commonly observed, even at distances exceeding several kilometres (Arias & Rieseberg, 1994; Greenleaf & Kremen, 2006), and these alleles can persist for multiple generations (Whitton, Wolf, Arias, Snow, & Rieseberg, 1997). Given this potential for hybridization and our extensive knowledge of the genetic basis of crop-like traits in sunflower (e.g., Blackman, Scascitelli, et al. 2011; Burke, Tang, et al., 2002; Chapman, Mandel, & Burke, 2013; Mandel, McAssey, Nambeesan, Garcia-Navarro, & Burke, 2014; Wills & Burke, 2007), sunflower is an excellent system for studying the fitness effects of crop-derived alleles in hybrid populations.

Our primary objectives were to (i) investigate allele frequency change for crop-derived alleles at a genome-wide scale over two

outcrossed generations; (ii) quantify the response of crop-like traits to natural selection over the same time period, and in the context of the genomic data; and (iii) determine whether or not selection has similar effects on allelic and/or phenotypic persistence across locations in sunflower crop–wild hybrid populations. These populations were initially derived from an existing collection of cultivated × wild sunflower RILs, thereby disrupting many of the linkage relationships among loci found in the parental lines.

2 | MATERIALS AND METHODS

2.1 | Plant materials and population maintenance

An experimental population of cultivated × wild sunflower RILs was established from an initial cross between the cytoplasmic male-sterile oilseed cultivar cmsHA89 (USDA Ames 3963, PI 650572) and a single wild *H. annuus* var. *annuus* individual (ANN1238, PI 659440) grown from seed collected at Cedar Point Biological Station (41°12.4′ N, 101°40.2′ W) in Keith County, Nebraska, USA (Baack et al., 2008; Burke, Tang, et al., 2002). Starting with the F_3 generation, plants were self-pollinated for 6–8 generations as has been previously described (Baack et al., 2008). This resulted in the production of 185 RILs, 169 of which were used in this study. The RIL population will be hereafter be referred to as generation 0 (G0).

Ten replicates of each G0 RIL were planted at the North Dakota Agricultural Experiment Station in Fargo, ND, in May 2007 in a fully randomized complete-block design (Dechaine et al., 2009). RILs were allowed to outcross naturally; estimated outcrossing rates were in excess of 80% based on comparisons of homozygosity between generations. To estimate reproductive output, all seeds produced by the G0 individuals (referred to as G1 seeds) were collected and weighed on a per-plant basis. The mass of an arbitrary subset of 50 filled seeds per plant was then determined and divided into total seed mass to estimate the number of seeds produced by each individual. To simulate natural selection, G1 seeds were then pooled in proportion to the total reproductive output (seed number) of each maternal (G0) plant.

In the spring of 2008, the ND field site and a second site in Decorah, IA were tilled prior to planting to simulate the disturbed environment typical of wild sunflowers growing along the margins of agricultural fields. A random sample of approximately 30 G1 seeds from the previously described pool was planted evenly into each of 6 rows in each of 10 blocks at each site. One seed head (corresponding to 20–50 G2 seeds) was collected from each surviving G1 plant. These seeds were pooled with each maternal head contributing seeds in proportion to the estimated whole-plant reproductive output. A random sample of this pool was selected for genotyping. The remaining G2 seeds were allowed to disperse naturally at each site. The sites were then tilled in the fall of 2008. G2 seeds began to germinate in May 2009 in both IA and ND. Seeds from the resulting G2 plants (i.e., G3 seeds) were collected at each site. As before, seeds were pooled with each maternal plant contributing seeds to the pool in proportion to the estimated whole-plant seed count. Samples from this pool were used for genotyping and the common garden experiments. All generations were grown under conditions that closely approximated the field conditions experienced by *H. annuus* crop-wild hybrid populations in the Midwestern USA. No irrigation, weed/pest control, or fertilizer was provided in any generation. The offspring of the G1 populations grown in the two field sites are hereafter referred to as IAG2 or NDG2 (IA or ND generation 2), and the next generation as IAG3 or NDG3 (IA or ND generation 3).

2.2 | Genotypic analyses

2.2.1 | DNA extraction

Total genomic DNA was isolated from 20 to 40 mg fresh leaf tissue collected from individual seedlings derived from seed of the relevant generation using the Qiagen DNeasy Plant Mini DNA extraction kit (Qiagen, Valencia, CA). DNA quantity and quality (260/280 and 260/230 ratios) was measured for each sample using a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Inc.; Wilmington, DE, USA). Genomic DNA was isolated from 169 G0 RILs and 144 randomly chosen individuals from each of the other populations/generations (i.e., G1, IAG2, IAG3, NDG2 and NDG3).

2.2.2 | SNP genotyping and mapping

A subset of 384 single nucleotide polymorphisms (SNPs) was selected from a larger set of 10,640 previously described SNPs (Bachlava et al., 2012). These SNPs were genotyped via an Illumina GoldenGate assay (Illumina Inc., San Diego, CA, USA) using the VeraCode technology. Of those, 358 SNPs could be successfully scored using GenomeStudio software V2011.1 (Illumina), and crop vs. wild allele origin could be determined for 278 of these SNPs. This latter subset of markers was used to analyse allele frequency change. Loci were rejected when the genotyped crop and wild parents both had null alleles, or due to heterozygosity of the crop parent (see Methods S1 for more details). The genomic locations of the full set of 358 SNPs were estimated relative to a framework set of 140 previously mapped SSRs (Chapman et al., 2008) via genotyping in the G0 and analysing the data using MAPMAKER/EXP 3.0 (Lander et al., 1987; Lincoln, 1992). QTL analyses were performed following the general methods of Owart et al. (2014).

2.3 | Common garden experiments

In May 2011, a random sample of G1, IAG3 and NDG3 seed pools, as well as the wild and crop RIL parents (ANN 1238 and HA 89), was planted into common garden experiments at the IA and ND sites. Six blocks with 12 rows × 16 columns were established at each site, each containing planting locations spaced at 0.5 m intervals for 50 G1, 40 IAG3, 40 NDG3, 6 ANN 1238 and 3 HA 89 plants. Four seeds were planted at each planting location in a completely randomized block design and then thinned to a maximum of one seedling per planting location. To ensure that

the wild parent (ANN1238) was represented in each block, greenhouse-grown seedlings (2–4 leaf stage) were used to supplement field-planted seeds.

We measured 12 phenotypic traits in the common gardens that are related to sunflower domestication and have been hypothesized to influence fitness in the wild. Flowering date was recorded for each individual when pollen was first visible on disc flowers. At senescence (i.e., when prereproductive leaves had begun to dry on >50% of the population), we recorded stem length, stem diameter immediately above the cotyledons, the number of primary branches (branch number), total number of leaves, leaf shape (estimated as length divided by width), the diameter of the largest flowering inflorescence and the total number of seed-producing inflorescences (head total). We estimated seed total per individual by counting the number of seeds across the diameter of a representative small and a large head on each individual, calculating average seed total per head and multiplying this by head total. Seed mass was determined by weighing 10 randomly chosen seeds from each plant. These seeds were also scanned on a flatbed scanner and the resulting images were used to estimate seed area using ImageJ (Schneider, Rasband, & Eliceiri, 2012). In addition, we rated each head for general insect head damage on a scale of 0–5: 0, 0% of heads and seeds within heads were damaged by herbivores; 1, 1–10% damaged; 2, 11–25% damaged; 3, 26–50% damaged; 4, 51–75% damaged; 5, 76–100% damaged.

2.4 | Data analyses

2.4.1 | Allele frequency change

For each SNP, allele frequency change from G1 to G3 in both locations (i.e., IAG3 and NDG3) was calculated using GENALEX v6.5 (Peakall & Smouse, 2006). As the SNP genotypes of both parents in the initial cross were known, frequency changes were scored such that a positive change indicated an increase in frequency of the crop allele. Significance was determined by comparing the observed change to that expected under neutral evolution (i.e., genetic drift only) using a custom Python script (data available from Dryad Digital Repository https://doi.org/10.5061/dryad.mp32f). Importantly, as the expected change due to drift is highly dependent on the initial allele frequency, the modelling of allele frequency change was determined as a function of the initial allele frequency at each locus. The bounds of allele frequency changes expected due to drift over two generations were established using a stochastic drift simulation with a population size of 900 and 30,000 replicates for each initial frequency. To account for multiple testing, the number of independent markers (i.e., the number of independent tests, M_{eff}) was estimated using the method described in Gao, Starmer, and Martin (2008) using R (R Development Core Team 2014). This approach uses composite LD among SNPs to capture the correlation between allele frequencies and derives the M_{eff} using the number of principal components that account for 99.5% of variation. Significance levels were then corrected for multiple comparisons using α/M_{eff} as the significance threshold.

2.4.2 | Hitchhiking analysis

To distinguish selection operating on multiple regions from genetic hitchhiking, we modelled selection and drift acting on linked loci for six linkage groups: 4, 5, 6, 7, 8 and 13. Only six linkage groups were modelled due to the computationally intensive nature of this work. The selected linkage groups contained seven of the nine loci where crop alleles were favoured in both Iowa and North Dakota, as well as at least one locus where estimates of selection without linkage found the wild allele to be strongly favoured ($s > 0.4$ on LG 4, 5, 6, 7 and 13; max $s = 0.2$ for LG 8). Chromosomal haplotype frequencies were obtained from the G0 RIL population, excluding heterozygous RILs at any given linkage group. A custom python script (Dryad Digital Repository https://doi.org/10.5061/dryad.mp32f) modelled selection, drift and recombination over three generations to obtain the minimum selection coefficients necessary to obtain the observed results for the IAG3 population, as similar patterns were observed in both the IA and ND G3. The direction and magnitude of selection were assumed to be constant over the three generations.

2.4.3 | Phenotypic trait means

Trait data for each population were non-normally distributed in most cases, violating ANOVA assumptions. We thus used 95% bootstrap confidence intervals to assess whether means differed between the G1 and G3 populations. Bias-corrected bootstrap confidence intervals were estimated in R using the boot procedure with 10,000 replicates (Canty & Ripley, 2015).

2.4.4 | QTL mapping

We used the SNP/SSR map described above to identify QTLs based on 11 phenotypic measures from G0 plants that were previously described and mapped using only the SSRs (Table S1) (Dechaine, Burger, & Burke, 2010; Dechaine et al., 2009). The QTLs were mapped using the composite interval mapping (CIM) procedure implemented in WinQTLCart (Wang, Basten, & Zeng, 2012). The default model (model 6) was used with the forward and backward stepwise regression ($p = .05$), a window size of 10 cM and a walk speed of 1 cM. To control for background noise, the number of cofactors was set to five. Significance thresholds were estimated based on 1,000 permutations (Churchill & Doerge, 1994). The additive effect (a) of the cultivar allele (HA89) and the amount of variance explained for each QTL were estimated in WinQTLCart. Additive effects were standardized to the standard deviation (i.e., a/SD) for each trait.

2.4.5 | Congruence between genotypic and phenotypic evolution

We assessed the degree to which changes in phenotype were associated with the expected changes in genotype. For traits with significant differences in mean between the G1 and G3 generation in each site (G1 vs. G3IA in Iowa; G1 vs. G3ND in North Dakota; Table 1),

we examined the direction of change at associated QTLs. We scored the direction of change for each QTL using the markers within 2-LOD. Where markers within that interval showed different outcomes, we used the majority of markers to score the direction of change for the QTL. If the direction of change in frequency for the QTL matched the direction of change for the trait (e.g., trait more wild-like, wild allele increased in frequency), this was scored as "congruent"; if the change in frequency did not match (e.g., trait more wild-like, crop allele increased in frequency), this was scored as "incongruent"; and if there was no significant change in frequency for the QTL, it was scored as "n.s." Traits were determined to be more wild- or crop-like based on previously published work in sunflower (Burke, Tang, et al., 2002) and the phenotypes observed in the current study for the wild and crop parent lines. We tested whether the patterns of congruence and incongruence for the Iowa common garden differed from random using a chi-square goodness of fit.

TABLE 1 Composite interval mapping (CIM) results. Columns from left to right list the trait, linkage group, proximal flanking marker, 1-LOD interval in centimorgans, standardized additive effects and per cent variance explained. Within additive effects, parentheses indicate a negative effect of the crop allele and asterisks label QTL that were in the predicted direction for that phenotype; that is, the crop allele conferred a more crop-like phenotype

Trait name	LG	Marker	1-LOD (cM)	*a*/SD	PVE (%)
Branch number	3	M23M12	(31.3–45)	2.886	12.23
	4	SFW01149	(55.9–63)	(2.717)*	8.91
	9	CYC5A	(94.2–105.2)	2.928	9.89
	12	SFW00213	(59.7–80.8)	(3.481)*	17.73
Flowering time	1	SFW09467	(0–22.8)	1.877	7.48
	6	HT913	(96.7–100.1)	4.832	27.37
	7	ZVG29	(0–9)	(2.866)*	17.31
	8	SFW01442	(44.6–78.9)	2.607	13.93
	14	SFW02805	(21.3–30.3)	(3.496)*	8.71
	17	SFW02587	(73.2–78.4)	(5.080)*	15.99
Head damage	7	SFW01658	(34.6–45.9)	3.354*	10.11
	13	SFW05467	(15.4–16.4)	(3.423)	10.34
	17	SFW02587	(73.2–77)	7.356*	14.68
Head total	8	HT656	(32.3–48.6)	2.915	9.37
	12	SFW03117	(53–77.4)	(3.399)*	12.80
Leaf number	4	SFW00857	(105–121.4)	(3.560)*	11.75
	7	SFW01658	(33.6–45.9)	(2.737)*	7.70
	11	SFW05043	(113.7–118.7)	(2.337)*	5.97
	12	SFW09009	(54–60.6)	(3.183)*	11.14
	14	SFW03980	(58.81–77.91)	(2.415)*	5.87
	16	SFW04562	(113.6–144.2)	3.530	12.04
Leaf ratio	7	ZVG29	(0–8)	3.561	12.39
	12	SFW09009	(57–60.7)	3.746	14.44
Seed area	4	SFW06560	(89.8–104.9)	3.332*	19.56
Seed mass	3	M23M12	(33.3–45)	3.212*	8.65
	4	SFW03768	(0–8)	3.699*	10.27
	8	HT656	(34.0–46.6)	3.258*	8.66
	9	SFW04878	(116.1–118.8)	2.871*	6.87
Seed total	4	SFW03768	(0–4)	3.993	13.89
	6	SFW00099	(86.1–101.1)	(3.429)*	9.69
	13	HT568	(0–9.9)	(6.654)*	12.81
Stem diameter	1	ORS371	(49.7–55)	2.85*	8.89
	3	SFW01698	(47.9–60.8)	2.526*	7.71
	9	SFW04878	(116.1–118.1)	2.439*	7.06
	13	SFW05467	(5.9–16.4)	3.053*	10.89
Stem height	3	SFW07426	(16.5–17.5)	3.350	10.55
	8	HT656	(33.3–41.6)	3.308	12.49
	13	SFW05467	(15.4–16.4)	2.688	8.29

2.4.6 | Population differentiation

To visualize the differentiation between populations across generations vs. the proportion of variation contained within populations, we performed a principal coordinate analysis (PCoA). Based on genetic distance calculated with GenAlEx, each individual from each population was represented in a multidimensional space, with each dimension representing a linear model describing a percentage of the observed variance. We visually compared the G1 population to the initial G0 population and to the populations after one generation (IAG2 and NDG2) and two generations of natural selection (IAG3 and NDG3). Each of these comparisons (i.e., G0 vs. G1, G1 vs. IAG2 or NDG2 and G1 vs. IAG3 or NDG3) was made individually. For all three of the comparisons, the genotypes for individuals from each of the populations were represented by the first three components in a 3D plot (JMP® version 11, SAS institute Inc., Cary, NC), as even the third component explained more than 10% of the total variance.

3 | RESULTS

3.1 | Genome-wide allele frequency change

Here, we focus on the G1 vs. IAG3 and G1 vs. NDG3 comparisons (i.e., those covering two generations of selection; Figure 1). The primary reason for this is that the G1 population is the first outcrossed generation (the G0 were RILs) and was used in the common garden experiments. Overall, the frequency of the wild allele increased for 74.5% and 80.2% of the SNP loci in Iowa and North Dakota, respectively (Figure 1a, Table S2). Just 19.8% and 15.1% of the SNPs in IA and ND, respectively, exhibited allele frequency changes consistent with the effects of genetic drift (i.e., the 99.99% confidence intervals for drift simulations included the G3 allele frequency). The fraction of the crop alleles showing significant increase ($\alpha = 0.0001$ to control for multiple comparisons) represented 5.8% of SNP loci in IA and 4.7% in ND, with 7.2% showing an increase in at least one of the two sites. The vast majority of the SNPs exhibited the same pattern of evolution in both locations (Figure 1b; see detailed analyses in Figures S1 and S2), suggesting similar selective pressures at the two locations. Of 21 loci previously shown to be selected during

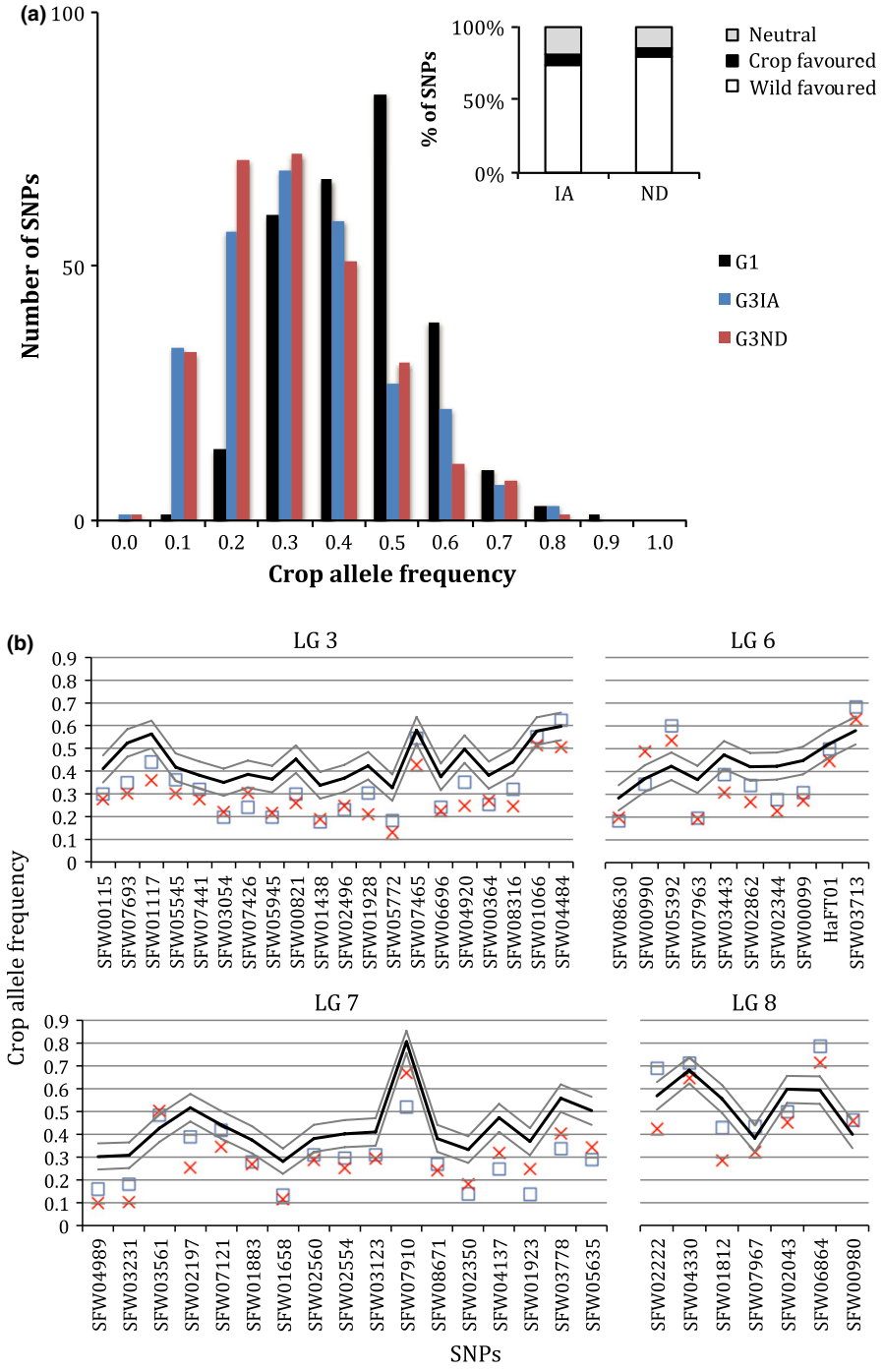


FIGURE 1 Genome-wide allele frequency change after two generations of selection. (a) Summary statistics. The main chart illustrates a clear downward shift in the frequency of crop alleles between the G1 and G3 generations (i.e., the distributions of the blue and red bars [G3IA and G3ND] shift to the left relative to the distribution of the black bars [G1]). The top right inset presents the proportion of SNPs with significant increases in frequency of the crop allele (crop favoured), the wild allele (wild favoured) or without significant changes in frequency (neutral; the observed changes are within those expected under genetic drift only) from G1–G3. (b) Crop allele frequencies in four representative linkage groups: 3, 6, 7 and 8. The crop allele frequencies are indicated for G1 (black line) and G3 in IA (blue squares) and ND (red crosses). The boundaries for expected allele frequencies after two generations of drift are indicated in grey lines (95% CI corrected for multiple testing)

sunflower domestication or improvement (Chapman et al., 2013; Mandel et al., 2014) that we assayed, 19 showed significant decreases in the frequency of the crop allele in both locations, with the remaining two showing significant decreases in North Dakota and nonsignificant decreases in Iowa (Table 2).

3.2 | Hitchhiking analysis

Models incorporating linkage suggest that 11–30% of the genomic regions might be evolving neutrally, as compared to an estimate of 13% from models without linkage. For example, models of selection, recombination and drift using the RIL haplotypes for LG 6 found that six of the ten loci exhibited significant evidence of selection, with the wild allele being favoured at five loci and the crop allele favoured at one locus. Three additional loci exhibited significant changes in allele frequency due to hitchhiking, while the last locus did not exhibit a significant allele frequency change (see Table S3). In contrast, models that did not account for linkage found that nine of ten loci were under selection, with the wild favoured at seven loci and the crop allele favoured at two loci. Models incorporating

TABLE 2 Loci associated with sunflower domestication or improvement from Chapman et al. (2013) and Mandel et al. (2014) and their position in the HA 89 × ANN 1238 G0 population. Only loci that could be mapped in G0 are included below

Type	Locus	LG	Position (cM)	Nearest marker	Crop allele frequency change in IA (%)	Crop allele frequency change in ND (%)
N	c5456	12	45	SFW04499	−**0.71**	−**0.46**
N	c2016[b]	16	41.7	SFW03628	−0.10	−**0.15**
D	c1357	16	75.5	SFW01764	−0.09	−**0.29**
D	c1533[a]	7	15.1	SFW02197	−**0.24**	−**0.50**
D	c1666[a]	14	14.3	SFW01535	−**0.26**	−**0.24**
D	c2963[a]	14	4	SFW00087	−**0.46**	−**0.45**
D	c31150	12	79.8	SFW00213	−**0.51**	−**0.49**
D	c5898[a]	10	40	SFW06746	−**0.14**	−**0.34**
D	M23M12	3	42	SFW01928	−**0.29**	−**0.51**
I	c1144[a]	3	8.2	SFW00115	−**0.27**	−**0.33**
I	c1236[a]	15	99.5	SFW00608	−**0.56**	−**0.54**
I	c1406[a]	7	18.2	SFW02197	−**0.24**	−**0.50**
I	c1700	10	0	SFW03901	−**0.25**	−**0.20**
I	c1774	1	36.5	SFW05614	−0.08	−**0.25**
I	c1921[a]	7	46.9	SFW02560	−**0.18**	−**0.24**
I	c2588[a]	7	20.9	SFW02197	−**0.24**	−**0.50**
I	c5666[a]	14	16.6	SFW01535	−**0.26**	−**0.24**
I	L2K11[a]	10	34.9	SFW04216	−**0.22**	−**0.40**
I	*PHYB*[a,b]	1	30	SFW00509	−**0.13**	−**0.21**
I	*LPR*[a,b]	12	56	SFW02267	−**0.12**	−**0.20**
I	*PAL1*[a,b]	13	3	SFW05982	−**0.37**	−**0.45**
I	*RGL2*[a,b]	14	24.2	SFW01535	−**0.26**	−**0.24**
I	*MAX2*[a,b]	17	48.5	SFW04302	−**0.55**	−**0.65**

N: neutral (as compared to improvement); D: selected during domestication; I: selected during improvement; bold: significant change in frequency from G1 to G3.

[a]Significant evidence of selection during domestication or improvement (ML-HKA test—Chapman et al., 2013; ML-HKA test—Mandel et al., 2014). Loci from Chapman et al. (2013) except where marked with [b]. Positions were mapped on the Owart et al., 2014 map, except for the five loci associated with crop improvement in Mandel et al., 2014, where the map positions are reported from that paper.

linkage reduce the estimated strength of selection on each locus (average $s = 0.52$ without linkage to average $s = 0.25$ or 0.35, for models that minimize or maximize the effects of selection on linked loci, respectively) but still result in more than 35% of loci with estimated selection coefficients exceeding 0.25. Overall, while models incorporating hitchhiking found an increased fraction of neutral loci (up to 30%) compared to models of individual loci (13%), we estimated that selection was still acting on 70% or more of the markers.

3.3 | Phenotypic evolution

When G1 and G3 plants were assessed in common gardens, a total of ten phenotypic traits (of twelve) showed a significant change from G1 to G3 (IA or ND), eight traits at the IA site and five at the ND site (Table 3). Stem diameter and head total were the only traits that did not exhibit significant generational changes at either site. For all significant traits, the populations evolved to be more wild-like (Table 3; Figure S3).

Flowering time is one example of a trait with significant and consistent evolution towards a more wild-like phenotype at both sites. The crop parental line (HA89) flowered earlier than individuals from the wild parental population (ANN1238) at both sites (Table 3). Although all hybrid populations flowered earlier than the wild parent, the G3 populations flowered later than G1 at both sites, and this difference was significant for all but NDG3 at the ND site.

For most traits, differences between G1 and the two G3 populations were generally stronger and more consistent in the IA than the ND common garden site. The hybrid populations (G1, IAG3 and NDG3) presented an intermediate phenotype between the two parental populations for the majority of the traits in Iowa. However, the hybrid trait values exceeded both parents for several traits in ND, including branch number, leaf number, head total and seed total. Phenotypic values were higher in the hybrids than either parent for those traits, suggesting poor performance of the wild population in ND and transgressive segregation (Table 3, Figure S3). The heavy, moist soils of the ND site were quite different from the habitat of

TABLE 3 Summary statistics for 12 phenotypic traits measured in common garden experiments located in (a) Iowa and (b) North Dakota. Means (standard errors) are indicated for the crop parent (HA 89), the wild parent (ANN 1238), generation 1 hybrids (G1) and generation 3 hybrids from Iowa (G3IA) and North Dakota (G3ND). Traits showing significant evolution from G1 to either G3 population are in boldface; populations with different letters differ (non-overlapping bootstrap 95% confidence intervals). All size traits were measured in cm. Flowering time is days from seedling emergence, May 20 or May 31, in IA and ND, respectively

	HA 89	G1	G3IA	G3ND	ANN 1238
(a)					
Branch Number	1.045 (1.17)	4.254 (0.46)^{a}	5.526 (0.48)^{b}	6.695 (0.47)^{b}	8.345 (1.13)
Disc Diameter	42.11 (4.22)	48.01 (1.08)	51.94 (1.17)	49.14 (1.09)	37.65 (3.78)
Flowering Time	82.476 (2.77)	80.135 (0.77)^{a}	88.317 (0.85)^{b}	85.392 (0.81)^{b}	88.371 (2.95)
Head Damage	2.461 (0.35)	0.573 (0.09)^{a}	0.157 (0.09)^{b}	0.277 (0.09)^{b}	0.138 (0.18)
Head Total	0.899 (3.45)	12.59 (1.33)	14.347 (1.38)	15.322 (1.33)	24.159 (3.08)
Leaf Number	22.19 (13.17)	57.29 (5.27)^{a}	72.67 (5.46)^{b}	70.47 (5.30)^{b}	117.41 (12.27)
Leaf Ratio	1.643 (0.16)	1.646 (0.06)^{a}	1.444 (0.06)^{b}	1.517 (0.06)^{ab}	1.71 (0.14)
Seed Area	200.32 (32.88)	135.47 (2.75)^{a}	136.41 (2.8)^{ab}	127.95 (2.70)^{b}	88.98 (9.1)
Seed Mass (1sd)	18.21 (5.04)	14.12 (0.42)	13.17 (0.42)	12.34 (0.41)	6.89 (1.39)
Seed Total	251.8 (1102.9)	2796.3 (284.7)^{a}	3790.1 (307.8)^{ab}	4057.2 (287.1)^{b}	4653.6 (986.3)
Stem Diameter	8.797 (1.59)	13.36 (0.46)	16 (0.5)	16.27 (0.47)	16.031 (1.58)
Stem Height	80.7 (10.37)	124.6 (4.15)^{a}	158.6 (4.31)^{b}	155.1 (4.16)^{b}	169.3 (9.66)
(b)					
Branch Number	0.59 (3.03)	12.67 (1.2)	12.21 (1.33)	13.22 (1.31)	5.17 (2.88)
Disc Diameter	81.85 (4.89)	55.2 (1.94)^{a}	53.2 (2.14)^{ab}	50.24 (2.12)^{b}	30.22 (4.65)
Flowering Time	64.79 (3.81)	61.12 (1.62)^{a}	67.42 (1.78)^{b}	64.67 (1.76)^{ab}	91.45 (4.28)
Head Damage	1.74 (0.31)	1.02 (0.13)	0.684 (0.14)	0.633 (0.14)	0.255 (0.31)
Head Total	2.03 (10.25)	33.28 (4.06)	41.29 (4.49)	41.37 (4.42)	13.68 (13.68)
Leaf Number	23.69 (27.37)	91.98 (10.86)	126.12 (11.98)	108.54 (11.84)	51.12 (26.02)
Leaf Ratio	1.08 (0.12)	1.36 (0.045)	1.3 (0.047)	1.31 (0.05)	1.67 (0.09)
Seed Area	26.27 (1.55)	17.26 (0.57)^{a}	16.87 (0.61)^{a}	15.09 (0.6)^{b}	11.49 (1.45)
Seed Mass (1sd)	25.76 (3.34)	14.65 (1.15)^{a}	11.49 (1.22)^{b}	11.34 (1.21)^{ab}	2.89 (2.92)
Seed Total	567.13 (4729.5)	9638.7 (2006.47)	13620.98 (2197.89)	11268.79 (2173.99)	2541.97 (4889.8)
Stem Diameter	16.69 (2.76)	23.95 (1.09)	25.73 (1.21)	24.79 (1.19)	22.49 (2.63)
Stem Height	92.64 (14.63)	150.01 (5.78)^{a}	174.16 (6.4)^{b}	173.69 (6.31)^{b}	139.96 (13.9)

the wild parent in the sand hills of Nebraska and could have contributed to this observation.

3.4 | QTL mapping and genotype–phenotype congruence

We detected a total of 38 QTLs distributed across 13 of the 17 linkage groups (see Table 1 and Figure S4 for all linkage groups). As expected, QTL positions and additive effects were generally consistent with previous studies mapping these traits using only the SSR markers (Dechaine et al., 2009, 2010). We detected multiple QTLs affecting each trait for all but one trait (Table 1). The crop allele conferred a more crop-like phenotype at all QTLs for four traits. The six remaining traits showed a mixture of positive and negative additive effects of the crop allele across QTLs. We observed wild alleles increasing in frequency at most markers across the genome, and most phenotypes became more wild-like. However, changes in marker frequency within the 2-LOD support limits of a QTL were as likely to be incongruent with observed phenotypic changes as they were to be congruent ($\chi^2 = 0.22$, $p > .05$; Figure 2). In Iowa, ten QTLs were congruent while eight were incongruent, while in North Dakota (where fewer traits showed significant changes), one QTL was congruent and two were incongruent (Figure 2).

In discussing genotype–phenotype congruence in detail, we will focus on whether the observed allele frequency change at a marker matched the change in the trait across generations; for simplicity, we will omit the direction of the additive effects (but see Table 1). We report here the results of four linkage groups representative of different possible patterns (Figure 3): an increase in wild allele frequency at all SNPs (LG 3), incongruent (bottom of LG 6, top of LG 8) or congruent (bottom of LG 8) allele frequency changes between sites, and QTL likely involved in the domestication process (e.g., seed traits or flowering time; LGs 3, 6, 7 and 8).

We detected four QTLs on LG 3 (branch number, seed mass, stem height and stem diameter), spanning regions in which the wild

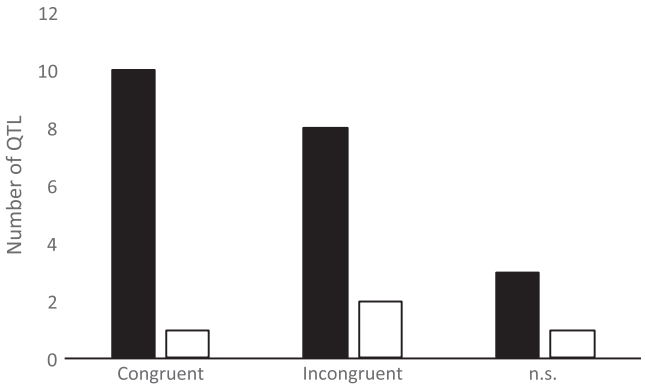


FIGURE 2 Congruence of phenotypic evolution and allele frequency changes at associated QTLs for Iowa (filled bars) and North Dakota (open bars). For each trait showing significant differences across generations in the common garden experiment (G1–G3IA in IA; G1–G3ND in ND; see Table 3), we scored the direction of allele frequency change of its associated QTLs. For QTLs with multiple markers within the 2-LOD limit, we scored the direction of change associated with the majority of markers (e.g., if three markers showed an increase of the wild allele and one marker had no significant change, then we scored the QTL change as wild). If the direction of change of both the trait and QTL was the same (e.g., flowering day became more wild-like and the wild allele significantly increased in frequency at a flowering day QTL), this was scored as congruent. Opposing changes were scored as incongruent. If the QTL frequency did not change significantly, then it was scored as "n.s."

allele increased in frequency. This linkage group is one of the six linkage groups that exhibited no instance in which the crop allele increased in frequency (along with LGs 2, 12, 14, 15 and 16; Figure S4).

Two QTLs were detected on the distal end of LG 6 (Figure 3): a flowering time QTL resulting in later flowering and a QTL for reduced seed total. The additive effect of this flowering QTL was among the highest we measured, explaining almost 30% of the variance in flowering time (Table 1). The right flanking markers of these QTLs were HaFT01, HaFT02 (previously reported by Blackman, Rasmussen, et al. 2011) and SWF03713. We observed a significant increase in frequency of the crop allele at SFW03713 in IA following two generations of selection, which is congruent with evolution towards later flowering time in IA. It is also worth noting that the crop allele conferred a more wild-like phenotype for this flowering time QTL. In contrast, the wild allele significantly increased at HaFT01 in ND (no significant change in IA). The wild allele also increased in frequency at the left flanking marker (SFW00099) for the seed total QTL on LG 6, which is congruent with the phenotypic evolution towards more seeds at both sites (significant only in ND). Although some genotype–phenotype congruence was observed for these QTL, it is difficult to interpret these results without a better understanding of the phenotypic effects of each marker.

A total of four QTLs were detected on LG 7, including leaf ratio and head damage. Interestingly, the head damage QTL was located in a region where markers showed an increase in the wild allele (Figure 3). Similarly to LG 6, the flowering time QTL on LG 7 spans markers exhibiting both congruent and incongruent evolution.

Finally, a total of four QTL were detected on LG 8, all of them colocalizing near the centre of this LG. The influence of this region on flowering time, with the crop allele conferring a wild-like later flowering phenotype, was previously reported (Dechaine et al., 2009). The QTLs on LG 8 were detected in a region where SNPs did not change significantly in frequency, or were selected for the wild allele in both locations. While this linkage group presented the highest proportion of markers showing an increase in the crop allele in at least one experimental site (3 of 7 SNPs, Figures 1b and 3), none of the detected QTL spanned these positions. Marker SFW02222 on LG 8 also provides one of only two examples (other is marker SFW06189 on LG 10) of significant allele frequency change towards the wild allele in one environment (ND) and the crop allele in the other (IA).

3.5 | Population differentiation

The PCoA plots, based on the global genetic distance between individuals, allow us to clearly observe the evolution from G1 to G3 (Figure 4). For all PCoAs, the cumulative percentage of the variance explained by the three-first components is quite high, ranging from 61% to 71% of the total variance. The distance between the G1 and the later generation populations undergoing selection increased with the number of generations (Figure 4b,c). At the G3 stage, both the IA and ND populations were beginning to separate from the G1 population but remained similar to each other.

4 | DISCUSSION

4.1 | Overall patterns of selection

Our most striking observation was that, over just two generations, the majority (80–85%) of markers surveyed exhibited evidence for allele frequency change due to selection. In most cases (74–80%), the wild allele increased in frequency. Similarly, all significant phenotypic changes and nonsignificant trends in phenotypic evolution (with the exception of leaf shape) at both field locations were in the direction of the wild parent. This apparent pervasive selective advantage of wild-derived alleles and traits suggests that crop-derived alleles/traits are generally maladapted to wild conditions. Indeed, it has been argued that alleles selected during domestication and improvement should be poorly fit to wild conditions and unlikely to introgress into wild genomes (e.g., Stewart, Halfhill, & Warwick, 2003). Consistent with this view, Hufford et al. (2013) found little evidence of introgression from maize into teosinte in genomic regions containing domestication genes such as *tga1*, *su1* and *bt2*. The consistent decreases in crop allele frequencies seen at the 21 loci previously shown to be selected during sunflower domestication or improvement (Chapman et al., 2013; Mandel et al., 2014) support the view that genomic regions associated with domestication-related traits are likely to be selected against in the wild.

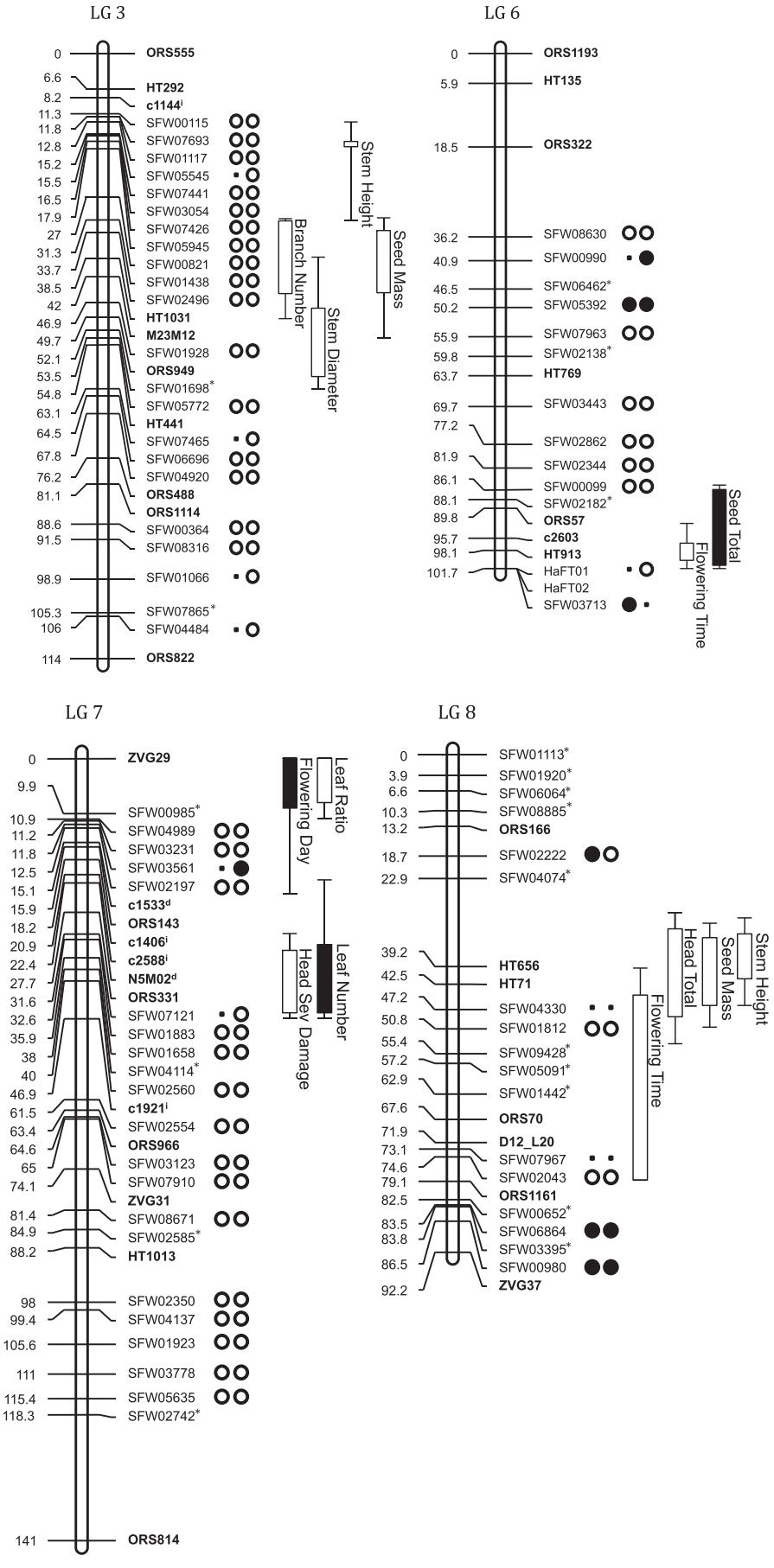
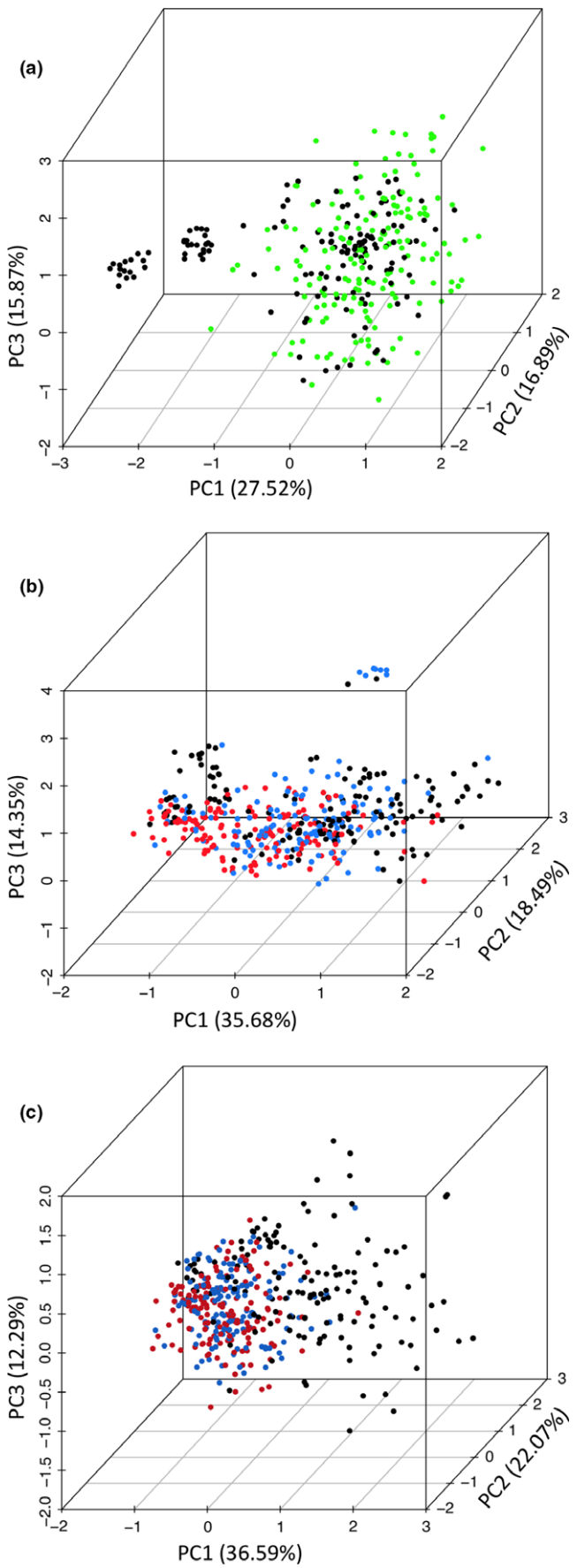


FIGURE 3 Graphical representation of allele frequency change and quantitative trait locus (QTL) mapping results for linkage groups 3, 6, 7, 8. The left of each linkage group reports the positions of (cM) consecutive markers, the first being 0 cM. Marker names are indicated on the right (SSRs in bold). Superscripts indicate whether SSR markers were selected during domestication (d) or improvement (i). Significant changes in SNP allele frequencies from G1 to G3 are shown by circles: increase of the crop allele (filled circle), increase of the wild allele (open circle) or nonsignificant change (dots) are indicated for IA (left) and ND (right). Significant QTL is shown to the right of each linkage group; filled and open bars indicate a negative and positive additive effect of the crop allele, respectively. 1-LOD (bar) and 2-LOD (tails) support limits are shown for each QTL. Asterisks (*) indicate markers for which genotyping was not successful or the differentiation between crop and wild allele was not possible for G1, G3 or both



The patterns of allele frequency change and phenotypic evolution that we observed were highly consistent between study locations (although significance levels differed). Only two loci (LG 8: SFW02222; LG 10: SFW06189; Figure S1) showed significant and opposing selection between the two sites, and the PCoA showed a convergence in genotypes across locations over time. This convergence likely resulted from the effects of selection, although genetic drift cannot be excluded as a contributing factor. Individuals in both the IA and ND populations largely overlapped in the G3 generation, suggesting a predominant role of common selective pressures rather than genetic drift. It is also notable that the distribution of G1 individuals was broader than the distribution of G3s. This is somewhat surprising, as the two sites clearly differed in several respects, and plant phenotypes differed as well, most strikingly in the 20-day difference in flowering time between each hybrid population when grown in ND vs. IA. The parental lines also responded differently to the two sites for several traits, including fitness. In the common gardens, the wild parent produced more seeds than any hybrid population and the crop parent (which produced the least) in IA, whereas both parental lines produced fewer seeds than the hybrid populations in ND. While the G0 generation was only grown at one site, and selection operating at this time might have eliminated certain alleles, all crop alleles started with a frequency of at least 0.19 in the G1 generation (average = 0.49), suggesting that selection on the G0 generation did not limit increases in the frequency of the crop allele between G1 and G3 in either site. Our results suggest that, despite differences in fitness correlates, selection favouring wild alleles was largely consistent between sites. This further supports the idea that most of the crop alleles studied herein are disadvantageous under natural conditions and would be unlikely to introgress into wild populations.

The design of this experiment limits our ability to precisely determine how much of the genome is under selection. The comparisons made here start with the G1, after one round of open pollination (estimated selfing rate <20%) between the initial RILs. This design meant that our population had very strong linkage disequilibrium, much higher than one would find in a wild population—although perhaps similar to what one would encounter in early-generation crop-wild hybrids. We detected selection acting on many chromosomal regions, but we cannot easily distinguish whether selection was acting on one or multiple loci within a region bounded by two markers. Likewise, selection could be acting directly on only one or a few regions of a chromosome, so much of the observed allele frequency change at different markers could be due to genetic hitchhiking. In that case, neutral wild alleles linked to an advantageous crop allele could appear as favoured. The results of our hitchhiking model

FIGURE 4 Principal component analysis (PCoA) based on the genetic distance using 344 SNPs of G1 with (a) G0; (b) G2IA and G2ND; (c) G3IA and G3ND. The three-first axes and the percentage of variance explained by each are reported. Green = G0; black = G1; blue = IA; red = ND

suggest that this is not the case—selection is acting directly on at least 70% of markers across the genome. Evidence of direct selection on much of the genome contrasts prior work in the sunflower hybrid RILs (G0 in this study), which found only a few QTLs for fitness (Baack et al., 2008; Dechaine et al., 2009). Limited power to detect QTLs in earlier studies is a possible explanation for this discrepancy. In addition, the earlier studies had lower planting densities and did not include overwinter seed survival.

4.2 | Potential introgression of crop alleles

Approximately 8% of the SNP markers in our study showed evidence of significant increases in the crop allele after two generations in at least one of the two locations, and models of selection and recombination support the inference of selection favouring the crop allele at these loci. High fitness crop alleles, such as these, could potentially introgress into wild sunflower populations. These results are consistent with those of several studies that have reported selection favouring crop alleles or traits under various conditions (*e.g.*, Dechaine et al., 2009; Mercer, Wyse, & Shaw, 2006; Mercer et al., 2007; Whitton et al., 1997). High fitness crop alleles seem initially puzzling; if the domesticated sunflower originated from wild progenitors in North America, most of its genes would be expected to be a subset of those found in the wild, and so favourable alleles would likely have already been present in the wild population.

Several nonexclusive hypotheses may explain the high fitness of crop alleles. First, these alleles may have been introgressed into the crop from other *Helianthus* species. For example, disease resistance alleles were introgressed into cultivated sunflower from the hexaploid Jerusalem artichoke *H. tuberosus* (Charlet & Brewer, 1995; Hulke & Wyse, 2008; Kantar et al., 2014; Sujatha & Prabakaran, 2006). Introgressed alleles (via crop breeding) might have been novel in wild *H. annuus* and might confer a fitness advantage. A second possibility is that the crop parent may express phenotypes that are better adapted to the field sites used (IA, ND) than the wild parent, which was from Nebraska. Supporting this hypothesis, the wild parent in ND had lower average seed set than any of the hybrid lineages, suggesting that it was not well adapted to that environment. Third, the crop alleles may have been favoured because the wild parent had deleterious alleles at these loci. Wild sunflowers are obligate outcrossers and, as such, are expected to harbour a number of segregating deleterious alleles. Our hybrid population had a single wild parent; certain crop alleles might be favoured in this cross due to deleterious wild alleles at those loci, but the same crop alleles may not have been advantageous if we had used a different wild parent. Finally, it is possible that some increases in the crop allele are due to novel beneficial mutations that occurred following sunflower domestication, and are thus not widely represented in the wild. Our ability to distinguish among these hypotheses is limited by our knowledge of the genetic basis of many physiological traits. Many of the regions where the crop allele was favoured are not associated with any known function (*e.g.*, LG 8, 11, 13) in our QTL analysis. Ongoing genetic mapping of crop-related traits, particularly disease resistance, will help elucidate the function of these favourable alleles, thus improving our ability to predict transgene escape.

4.3 | Genotype–phenotype congruence

In addition to describing genome-wide patterns of selection, our data lend insight into the relationship between evolutionary change at the phenotypic and genotypic levels. Phenotypes almost exclusively evolved to be more wild-like. We would thus predict that alleles conferring a more wild-like phenotype would increase in frequency for markers associated with QTL. This predicted relationship was not consistently observed (Figure 2). Of the over 118 tested markers that fall within 2-LOD thresholds of mapped QTL, less than half follow the predicted relationship, in that the allele conferring the more wild-like phenotype for a QTL significantly increased in frequency in IA or ND. For example, the wild allele conferred a more wild-like phenotype for branch number, stem height and stem diameter QTLs on LG 3 and, as expected, wild alleles significantly increased in frequency for all loci within the 2-LOD intervals of these QTLs. Of the markers that followed predictions, only two involved an increase in the frequency of the crop allele: LG 4—SFW05141 for increased leaf number and LG 7—SFW03561 for later flowering time, both only in ND.

Genotype–phenotype congruence can be more deeply examined using flowering time; it exhibited significant and consistent evolutionary change across study sites, and we detected six QTLs affecting this trait. Half of the flowering time QTLs (on LGs 7, 14 and 17) followed the predicted pattern; the wild allele conferred a later flowering time, and all tested markers within the 2-LOD threshold of these QTL significantly increased in frequency at both sites (except LG 7—SFW03561). For the remaining three flowering time QTLs (on LGs 1, 6 and 8), the crop allele produced later flowering, but the wild allele increased in frequency for all SNPs (except LG 1—SFW08078 in IA) within the 2-LOD threshold of these QTL that exhibited significant allele frequency change (several SNPs did not change significantly at one or both sites). These results suggest that, even though the crop allele conferred the more advantageous phenotype at three QTLs, it was generally the wild alleles that increased in frequency.

A likely explanation for the lack of congruence between genotypic and phenotypic change is that unmeasured traits are the target of selection and are either linked to QTLs for measured traits, or share QTLs with some of these traits. Most of the genome, including several entire linkage groups, was not associated with any mapped QTLs. Also, our QTL analyses only detected loci of large effect; there are likely additional (many for some traits) unidentified loci affecting the studied phenotypic traits. Furthermore, our selection analyses only measured directional selection. Flowering time is likely under stabilizing selection in wild populations: for instance plants that flower too late will fail to mature seeds before being killed by hard frosts, while plants flowering very early might have limited resources to allocate to reproduction because of a small size. The mix of +/− QTL (i.e., QTL alleles with opposing effects on a phenotype) at loci controlling flowering time fits

this expectation, although other hypotheses (e.g., pleiotropic effects of alleles affecting flowering time on other traits) are also possible.

4.4 | Implications for GM traits

One application of this work is to identify the types of crop-like traits and their underlying genes that might be favoured in the wild if subjected to manipulation. However, the observed lack of congruence between our phenotypic and genotypic results indicates that phenotypic change may be an unreliable predictor of allele frequency change due to the possibility of selection on unobserved traits and/or limited knowledge of the true genetic basis of traits under observation. We showed that plants generally evolved to be more wild-like and crop alleles increased at relatively few regions of the genome. Some of these loci may merely mark segregating deleterious alleles in the wild population used in this cross. If any of these loci colocate with regions introgressed into other *H. annuus* wild populations or more distant species from crop sunflowers, these regions should be investigated further for phenotypic effects. Neutral crop alleles are also of potential concern, as they have been shown to persist in hybrid populations for at least 10 generations (Snow et al., 2010).

Finally, understanding crop allele frequency changes under wild conditions might also inform transgene mitigation. The many crop markers experiencing strong selection in both environments in our study suggests that many agronomic traits have severe fitness costs in the wild environment. In a few cases, we know the genes under selection during domestication; two were included in this study. Significant selection against the crop HaFT01 was observed here, but HaGA2OX (gibberellic acid 2 oxidase; Blackman, Rasmussen, et al. 2011; Chapman et al., 2008) showed little change from G1 to G3. Thus, understanding genes/alleles conferring a disadvantage under wild conditions (while having been selected for crop domestication/improvement) will be essential to making tandem constructs possible in sunflowers and thus preventing the spread of advantageous crop alleles to wild/feral populations.

ACKNOWLEDGMENTS

We would like to thank the many undergraduates who worked on this project from Central Washington University, Concordia College, the University of Georgia and Luther College, with particular thanks to Ellen Behrens, Katrina Freund, Anna Gudjonsdottir, Andrew Kingsriter, Sarah Nicla, Marjorie Olsen, Jeff Roeder and Molly Tulkki for their assistance in the laboratory and the field. This project was funded by a grant from the NSF Plant Genome Research Program (IOS-1444522) to JB and JD and by grants from the USDA Biotechnology Risk Assessment Research Grants (BRAG) Program (2010-33522-21668 to JB, EB and JD; 2016-33522-25630 to JB and EB).

DATA ACCESSIBILITY

All genotype and phenotype data, python scripts and R scripts are archived in the Dryad Data depository (https://doi.org/10.5061/dryad.mp32f).

AUTHOR CONTRIBUTIONS

E.B., J.B. and J.D. planned the field and genotyping portions of the experiment. E.B. carried out field work, analysed the field data, performed the drift and draft analyses and revised the manuscript; J.B. oversaw the genotyping and related analyses and revised the manuscript; J.C. carried out the genotyping, analysed the genetic data and drafted the manuscript; J.D. carried out field work, analysed the field data, performed QTL analyses and revised the manuscript; G.S. provided and planned the ND field site, carried out field work in ND and revised the manuscript.

REFERENCES

Adler, L., Wikler, K., Wyndham, F., Linder, C., & Schmitt, J. (1993). Potential for persistence of genes escaped from canola: Germination cues in crop, wild, and crop-wild hybrid *Brassica rapa*. *Functional Ecology*, *7*, 736–745.

Aerts, R., Berecha, G., Gijbels, P., Hundera, K., Van Glabeke, S., Vandepitte, K., … Honnay, O. (2013). Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications*, *6*, 243–252.

Arias, D. M., & Rieseberg, L. H. (1994). Gene flow between cultivated and wild sunflowers. *Theoretical and Applied Genetics*, *89*, 655–660.

Baack, E. J., Sapir, Y., Chapman, M. A., Burke, J. M., & Rieseberg, L. H. (2008). Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Molecular Ecology*, *17*, 666–677.

Bachlava, E., Taylor, C. A., Tang, S., Bowers, J. E., Mandel, J. R., Burke, J. M., & Knapp, S. J. (2012). SNPdiscovery and development of a high-density genotyping array for sunflower. *PLoS One*, *7*, e29814.

Berville, A., Muller, M.-H., Poinso, B. & Serieys, H. (2005). Ferality. Risks of gene flow between sunflower and other Helianthus species. In J. Gressel (Ed.), *Crop ferality and volunteerism* (pp. 209–230). Boca Ratón, FL: CRC Press.

Blackman, B. K., Rasmussen, D. A., Strasburg, D. A., Raduski, A. R., Burke, J. M., Knapp, S. J., … Rieseberg, L. H. (2011). Contributions of flowering time genes to sunflower domestication and improvement. *Genetics*, *187*, 271–287.

Blackman, B. K., Scascitelli, M., Kane, N. C., Luton, H. H., Rasmussen, D. A., Bye, R. A., … Rieseberg, L. H. (2011). Sunflower domestication alleles support single domestication center in eastern North America. *Proceedings of the National Academy of Sciences*, *108*, 14360–14365.

Burke, J. M., Gardner, K. A., & Rieseberg, L. H. (2002). The potential for gene flow between cultivated and wild sunflower (*Helianthus annuus*) in the United States. *American Journal of Botany*, *89*, 1550–1552.

Burke, J. M., Tang, S., Knapp, S. J., & Rieseberg, L. H. (2002). Genetic analysis of sunflower domestication. *Genetics*, *161*, 1257–1267.

Campbell, L. G., & Snow, A. A. (2007). Competition alters life history and increases the relative fecundity of crop–wild radish hybrids (*Raphanus* spp.). *New Phytologist*, *173*, 648–660.

Campbell, L. G., Snow, A. A., & Sweeney, P. M. (2009). When divergent life histories hybridize: Insights into adaptive life-history traits in an annual weed. *New Phytologist*, *184*, 806–818.

Canty, A., & Ripley, B. (2015). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-17.

Chapman, M. A., Mandel, J. R., & Burke, J. M. (2013). Sequence validation of candidates for selectively important genes in sunflower. *PLoS One*, *8*, e71941.

Chapman, M. A., Pashley, C. H., Wenzler, J., Hvala, J., Tang, S., Knapp, S. J., & Burke, J. M. (2008). A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell*, *20*, 2931–2945.

Charlet, L. D., & Brewer, G. J. (1995). Resistance of native sunflowers (Asterales: Asteraceae) to the banded sunflower moth (Lepidoptera: Cochylidae). *Environmental Entomology*, *24*, 1224–1228.

Chen, Y. H., Gols, R., & Benrey, B. (2015). Crop domestication and its impact on naturally selected trophic interactions. *Annual Review of Entomology*, *60*, 35–58.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, *138*, 963–971.

Dechaine, J. M., Burger, J. C., & Burke, J. M. (2010). Ecological patterns and genetic analysis of post-dispersal seed predation in sunflower (*Helianthus annuus*) crop-wild hybrids. *Molecular Ecology*, *19*, 3477–3488.

Dechaine, J. M., Burger, J. C., Chapman, M. A., Seiler, G. J., Brunick, R., Knapp, S. J., & Burke, J. M. (2009). Fitness effects and genetic architecture of plant–herbivore interactions in sunflower crop–wild hybrids. *New Phytologist*, *184*, 828–841.

Ellstrand, N. C. (2003a). *Dangerous liaisons?: When cultivated plants mate with their wild relatives*. Baltimore, MD: JHU Press.

Ellstrand, N. C. (2003b). Current knowledge of gene flow in plants: Implications for transgene flow. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 1163–1170.

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T. J., … Stewart, C. N. Jr (2013). Introgression of crop alleles into wild and weedy populations. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 325–345.

Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*, *32*, 361–369.

Ghosh, A., Meirmans, P. G., & Haccou, P. (2012). Quantifying introgression risk with realistic population genetics. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 4747–4754.

Greenleaf, S. S., & Kremen, C. (2006). Wild bees enhance honey bees' pollination of hybrid sunflower. *Proceedings of the National Academy of Sciences*, *103*, 13890–13895.

Harter, A. V., Gardner, K. A., Falush, D., Lentz, D. L., Bye, R. A., & Rieseberg, L. H. (2004). Origin of extant domesticated sunflowers in eastern North America. *Nature*, *430*, 201–205.

Hartman, Y., Uwimana, B., Hooftman, D. A., van de Wiel, C. C. M., Smulders, M. J. M., Visser, R. G. F., & van Tienderen, P. H. (2012). Genomic regions in crop-wild hybrids in lettuce are affected differently in different environments: Implications for crop breeding. *Evolutionary Applications*, *5*, 629–640.

Haygood, R., Ives, A. R., & Andow, D. A. (2004). Population genetics of transgene containment. *Ecology Letters*, *7*, 213–220.

Heiser, C. B., Smith, D. M., Clevenger, S. B., & Martin, W. C. (1969). The North American Sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club*, *22*, 1–218.

Hooftman, D. A. P., Flavell, A. J., Jansen, H., den Nijs, H. C. M., Syed, N. H., Sørensen, A. P., Orozco-ter Wengel, P., & van de Wiel, C. C. M. (2011). Locus-dependent selection in crop-wild hybrids of lettuce under field development conditions and its implication for GM crop development. *Evolutionary Applications*, *4*, 648–659.

Hovick, S. M., Campbell, L. G., Snow, A. A., & Whitney, K. D. (2012). Hybridization alters early life-history traits and increases plant colonization success in a novel region. *The American Naturalist*, *179*, 192–203.

Hufford, M. B., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., & Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, *9*, e1003477.

Hulke, B. S., & Wyse, D. L. (2008) Using Interspecific hybrids with *Helianthus tuberosus* L. to transfer genes for quantitative traits into cultivated sunflower, *H. annuus* L. *Proceedings of 17th International Sunflower Conference, June 8–12, 2008*, Cordoba, Spain. Pp. 729–731.

de Jong, T. J., & Rong, J. (2013). Crop to wild gene flow: does more sophisticated research provide better risk assessment? *Environmental Science & Policy*, *27*, 135–140.

Kantar, M. B., Betts, K., Michno, J.-M., Luby, J. J., Morrell, P. L., Hulke, B. S., … Wyse, D. L. (2014). Evaluating an interspecific *Helianthus annuus*×&Helianthus tuberosus population for use in a perennial sunflower breeding program. *Field Crops Research*, *155*, 254–264.

Kost, M. A., Alexander, H. M., Jason Emry, D., & Mercer, K. L. (2015). Life history traits and phenotypic selection among sunflower crop–wild hybrids and their wild counterpart: Implications for crop allele introgression. *Evolutionary Applications*, *8*, 510–524.

Kwit, C., Moon, H. S., Warwick, S. I., & Stewart, C. N. Jr (2011). Transgene introgression in crop relatives: Molecular evidence and mitigation strategies. *Trends in Biotechnology*, *29*, 284–293.

Landbo, L., & Jørgensen, R. B. (1997). Seed germination in weedy *Brassica campestris* and its hybrids with *B. napus*: Implications for risk assessment of transgenic oilseed rape. *Euphytica*, *97*, 209–216.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., & Newberg, L. A. (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, *1*, 174.

Lincoln, S. (1992). Constructing genetic maps with MAPMAKER/EXP 3.0. *Whitehead Institute Technical Report*, Cambridge, MA.

Linder, C. R., Taha, I., Seiler, G. J., Snow, A. A., & Rieseberg, L. H. (1998). Long-term introgression of crop genes into wild sunflower populations. *Theoretical and Applied Genetics*, *96*, 339–347.

Mandel, J. R., McAssey, E. V., Nambeesan, S., Garcia-Navarro, E., & Burke, J. M. (2014). Molecular evolution of candidate genes for crop-related traits in sunflower (*Helianthus annuus* L.). *PLoS One*, *9*, e99620.

Mercer, K., Andow, D., Wyse, D., & Shaw, R. (2007). Stress and domestication traits increase the relative fitness of crop-wild hybrids in sunflower. *Ecology Letters*, *10*, 383–393.

Mercer, K. L., Wyse, D. L., & Shaw, R. G. (2006). Effects of competition on the fitness of wild and crop-wild hybrid sunflower from a diversity of wild populations and crop lines. *Evolution*, *60*, 2044–2055.

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*, 499–514.

Owart, B. R., Corbi, J., Burke, J. M., & Dechaine, J. M. (2014). Selection on crop-derived traits and QTL in sunflower (*Helianthus annuus*) crop-wild hybrids under water stress. *PLoS One*, *9*, e102717.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288–295.

Putt, E. D. (1997). Early history of sunflower. In A. A. Schneiter (Ed.), *Sunflower technology and production* (pp. 1–19). Madison, WI: American Society of Agronomy.

Pyšek, P., & Richardson, D. M. (2007). Traits associated with invasiveness in alien plants: Where do we stand? In W. Nentwig (ed.) *Biological Invasions* (pp. 97–125). Berlin: Springer.

R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Rong, J., Song, Z., De Jong, T. J., Zhang, X., Sun, S., Xu, X., … Lu, B.-R. (2010). Modelling pollen-mediated gene flow in rice: Risk assessment and management of transgene escape. *Plant Biotechnology Journal*, *8*, 452–464.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*, 671–675.

Snow, A. A., Culley, T. M., Campbell, L. G., Sweeney, P. M., Hegde, S. G., & Ellstrand, N. C. (2010). Long-term persistence of crop alleles in weedy populations of wild radish (*Raphanus raphanistrum*). *New Phytologist*, *186*, 537–548.

Soleri, D., & Cleveland, D. A. (1993). Hopi crop diversity and change. *Journal of Ethnobiology*, *13*, 203–231.

Stewart, C. N., Halfhill, M. D., & Warwick, S. I. (2003). Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics*, *4*, 806–817.

Sujatha, M., & Prabakaran, A. J. (2006). Ploidy manipulation and introgression of resistance to *Alternaria helianthi* from wild hexaploid *Helianthus* species to cultivated wunflower (*H. annuus* L.) aided by anther culture. *Euphytica*, *152*, 201–215.

Vrânceanu, A.-V. (2000). Floarea-soarelui hibridă.Bucharest, Romania: Editura Ceres.

Wang, S., Basten, C., & Zeng, Z. (2012). *Windows QTL cartographer 2.5*. Raleigh, NC: Department of Statistics, North Carolina State University.

Warren, J., & James, P. (2006). The ecological effects of exotic disease resistance genes introgressed into British gooseberries. *Oecologia*, *147*, 69–75.

Whitton, J., Wolf, D. E., Arias, D. M., Snow, A. A., & Rieseberg, L. H. (1997). The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. *TAG Theoretical and Applied Genetics*, *95*, 33–40.

Wills, D. M., & Burk, J. M. (2006). Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). *Journal of Heredity*, *97*, 403–408.

Wills, D. M., & Burke, J. M. (2007). Quantitative trait locus analysis of the early domestication of sunflower. *Genetics*, *176*, 2589–2599.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Corbi J, Baack EJ, Dechaine JM, Seiler G, Burke JM. Genome-wide analysis of allele frequency change in sunflower crop–wild hybrid populations evolving under natural conditions. *Mol Ecol*. 2018;27:233–247. https://doi.org/10.1111/mec.14202