Selection on domestication traits and quantitative trait loci in crop–wild sunflower hybrids

ERIC J. BAACK,*§ YUVAL SAPIR,†§ MARK A. CHAPMAN,‡ JOHN M. BURKE‡ and LOREN H. RIESEBERG*§

Department of Botany, University of British Columbia, 6270 W University Blvd, Vancouver, BC, Canada V6T 1Z4, †Department of Life Sciences, Ben-Gurion University, PO Box 653, Beer Sheva, 84104 Israel, ‡Department of Plant Biology, University of Georgia, Athens, GA 30602, USA, §Department of Biology, Indiana University, 1001 E 3rd, Bloomington, IN 47405, USA

Abstract

The strength and extent of gene flow from crops into wild populations depends, in part, on the fitness of the crop alleles, as well as that of alleles at linked loci. Interest in crop–wild gene flow has increased with the advent of transgenic plants, but nontransgenic crop–wild hybrids can provide case studies to understand the factors influencing introgression, provided that the genetic architecture and the fitness effects of loci are known. This study used recombinant inbred lines (RILs) generated from a cross between crop and wild sunflowers to assess selection on domestication traits and quantitative trait loci (QTL) in two contrasting environments, in Indiana and Nebraska, USA. Only a small fraction of plants (9%) produced seed in Nebraska, due to adverse weather conditions, while the majority of plants (79%) in Indiana reproduced. Phenotypic selection analysis found that a mixture of crop and wild traits were favoured in Indiana (i.e. had significant selection gradients), including larger leaves, increased floral longevity, larger disk diameter, reduced ray flower size and smaller achene (seed) mass. Selection favouring early flowering was detected in Nebraska. QTLs for fitness were found at the end of linkage groups six (LG6) and nine (LG9) in both field sites, each explaining 11–12% of the total variation. Crop alleles were favoured on LG9, but wild alleles were favoured on LG6. QTLs for numerous domestication traits overlapped with the fitness QTLs, including flowering date, achene mass, head number, and disk diameter. It remains to be seen if these QTL clusters are the product of multiple linked genes, or individual genes with pleiotropic effects. These results indicate that crop trait values and alleles may sometimes be favoured in a noncrop environment and across broad geographical regions.

Keywords: crop-wild hybridization, hybrid zones, phenotypic selection, QTL analysis, recombinant inbred lines (RILs), sunflower

Received 29 June 2007; revision accepted 25 September 2007

Introduction

Introgression of crop traits and alleles into wild relatives can have important ecological consequences. Hybridization between the crop and wild plants has led to the loss of distinct wild genotypes in *Raphanus* in California and *Gossypium* in the Galapagos Islands (Panetsos & Baker 1968; Wendel & Percy 1990; Hegde *et al.* 2006), and has been implicated in the development of novel weed phenotypes,

Correspondence: Eric Baack, Fax: 563-387-1080; E-mail: baacer01@luther.edu

including the development of weed beets in European populations of sugar beets (Boudry *et al.* 1993; Ellstrand *et al.* 1999). While the exchange of genetic material between crops and wild relatives is likely as old as agriculture, the ecological effects of such exchanges on wild plant populations have received increased scrutiny following the introduction of transgenic technologies (Chapman & Burke 2006).

Introgression of crop alleles into wild plants depends upon several conditions. First, hybridization must occasionally occur; this has now been documented for more than 40 crops (Ellstrand 2003). Second, the hybrids must have

nonzero fitness, as has been extensively documented in many systems (Arriola & Ellstrand 1997; Snow *et al.* 2001; Allainguillaume *et al.* 2006; Mercer *et al.* 2006). So long as some backcrossing to the wild population occurs, even low-fitness hybrids will not significantly impede the spread of neutral or advantageous crop alleles, as independent assortment and recombination will rapidly free neutral or advantageous alleles from deleterious alleles with which they were initially associated. Third, the crop alleles must be neutral or advantageous in wild type background. Negatively selected alleles are unlikely to move beyond the crop margin. Note that the introgression of crop alleles into wild populations depends mostly on the fitness effects of the focal allele and that of alleles at tightly linked loci, rather than on the fitness levels of the early generation hybrids (Barton 1979).

Thus far, studies of crop–wild introgression have focused mainly on the fitness of crop–wild hybrid genotypes. Two recent studies demonstrate that information on genotypic fitness alone is insufficient for predicting the introgression of crop traits and alleles. For example, analyses of the effects of downy mildew resistance on *Lactuca serriola* found that experimental crop–wild hybrids had lower rates of infection, regardless of whether crop alleles conferring resistance were present or absent (Hooftman *et al.* 2007). Because heterosis appears to mask the effects of the resistance alleles, it is unclear whether they would be advantageous and spread in a wild genetic background. In contrast, studies on the effects of mildew resistance in crop–wild gooseberry hybrids found that late generation backcrosses to the wild had higher fitness than wild plants or early generation backcrosses, suggesting that linkage disequilibrium with other crop loci was reducing early hybrid fitness (Warren & James 2006). In this case, it is clear that the resistance alleles will spread if crop × wild hybridization occurs. In summary, to predict the evolutionary dynamics of crop alleles in wild populations, linkage relationships and the fitness effects of alleles and closely linked genes (i.e. quantitative trait loci or QTLs) must be understood.

The fitness effects of crop alleles may depend on environmental factors. The widespread evidence of local adaptation in plants (see Wright *et al.* 2006 and references therein) demonstrates that many alleles are beneficial in some, but not all environments. The average fitness of crop alleles, which may vary across space and time, will determine their eventual fate in the wild population. One promising idea to control transgene spread relies on variation in the fitness effects of crop alleles: if transgenes with positive effects in both the crop and noncrop environment are engineered to be adjacent to alleles that are unfavourable outside the crop environment, this should slow the spread of the transgene (tandem constructs; Gressel 1999). The success of this technology will depend upon consistent, strong negative fitness effects of the linked allele in the noncrop environment. Since many agronomic traits do not seem to introgress into wild relatives despite ongoing hybridization, domestication traits have been proposed as candidates. However, neither the average fitness effects of the alleles underlying domestication traits nor the variation between environments are known for many crop species.

In this study, we examine the fitness effects of crop QTLs in a hybrid sunflower population grown in two contrasting wild environments. The sunflower is an excellent experimental system for understanding of crop–wild gene flow. Prior work has documented short-term and long-term introgression of crop alleles into wild populations (Whitton *et al.* 1997; Linder *et al.* 1998), evaluated the potential for hybridization across the landscape (Burke *et al.* 2002a), examined the genetic basis of domestication (Burke *et al.* 2002b; Wills & Burke 2007) as well as its biogeography (Harter *et al.* 2004), and the fitness of crop–wild hybrids (Alexander *et al.* 2001; Cummings *et al.* 2002; Burke & Rieseberg 2003; Snow *et al.* 2003; Mercer *et al.* 2006; Mercer *et al.* 2007). We extend this work by combining the analysis of the fitness effects of different phenotypes with an examination of their genetic basis and the fitness effects of QTLs.

In this study, we address whether: (i) wild trait values and wild QTL alleles are exclusively favoured in the noncrop environment; (ii) fitness QTLs have consistent effects in different geographical areas; and (iii) any QTL alleles have close to lethal effects that could be used to prevent transgene escape in tandem constructs.

Materials and methods

Study species

Sunflower (*Helianthus annuus* L.) is one of four crops domesticated in eastern North America (Harter *et al.* 2004; Smith 2006). Wild plants differ from the crop in many traits, including branching, number and size of flowering heads, leaf size and shape, and achene (called seeds hereafter) size (Burke *et al.* 2002b). The crop and its wild progenitor are completely interfertile, with no barriers to viability or fecundity in the F_1 when grown in the field (Snow *et al.* 1998). Wild sunflowers require outcrossing via pollinators, while the domesticated sunflower has been selected for self-pollination (Gandhi *et al.* 2005), although fitness increases when pollinators are present (Degrandi-Hoffman & Chambers 2006). Wild sunflowers occur in most of the range of the cultivated sunflower in the USA and Canada, mainly in disturbed soil, and can frequently be found along the edges of cultivated fields (Burke *et al.* 2002a). Wild populations growing adjacent to the crop experience ongoing gene flow, with neutral alleles introgressing into the wild genome (Linder *et al.* 1998) and persisting for

several generations (Whitton *et al*. 1997). However, the wild plants in these hybrid zones rarely display the crop phenotype (E. J. Baack, personal observation). This suggests that many crop traits are unfavourable in the environment adjacent to cultivated fields, despite the high fitness of F_1 plants under some conditions (Mercer *et al*. 2007).

Development of hybrid lines

Recombinant inbred lines (RILs) were developed from an initial cross between the cytoplasmic male sterile elite line cmsHA89 (USDA Ames 3963) and a wild plant grown from seed collected at Cedar Point Biological Station, Keith County, Nebraska, USA (Ann1238). The parental and F_1 generations were cultivated in greenhouses at Indiana University, Bloomington, Indiana, USA. A single F_1 plant was self-pollinated to generate F_2 plants, which were field-grown in Mexico. F_3 plants were greenhouse-grown at Indiana University and were used for a QTL mapping study of domestication traits (Burke *et al*. 2002b). The plants of the next three generations (F_4–F_6) were field-grown at Oregon State University, Corvallis, Oregon, USA. Plants were irrigated and fertilized to minimize effects of selection in the field, and weakly performing lines were hand-germinated and started in the greenhouse to minimize loss of RILs. Plants reproduced by self-pollination in each generation, leading to increasing levels of homozygosity in the RILs.

RILs offer several advantages for this study. First, the consistent genotypes allow replication within sites, allowing us to average genotypic effects across environmental variation within a site and, thus, to have increased power. The consistent genotypes also allow comparison of performance of genotypes between sites. The high levels of homozygosity maximize differences at additive loci, also increasing the ability to detect genetic effects on phenotypes and fitness. Finally, the same genotypes can be used in future studies. However, the use of RILs precludes direct comparisons with wild plants. Wild plants are self-incompatible and so we could not readily produce inbred lines. Thus, we do not have the exact genotype of the wild plant used to found the RIL population. Likewise, genotypic differences, both in terms of the alleles present and the level of heterozygosity, make fitness comparisons between RILs and wild plants of doubtful utility. Wild plants were thus not included in this experimental design.

Genotyping of recombinant inbred lines

Genomic DNA was extracted from each inbred line using the DNeasy plant mini kit (QIAGEN), quantified, and diluted to 10 ng/μL. Using published (Burke *et al*. 2002b; Tang *et al*. 2002) and unpublished (S. Tang and S. Knapp,

University of Georgia, Athens, Georgia, USA) linkage maps of sunflower, markers were selected to cover as much of the genome as possible. These markers were screened on a subset of eight RILs and those markers which were polymorphic were run on the entire set of 184 RILs. Initial genetic maps were constructed (see below) and further markers were selected from the above linkage maps to increase coverage and to reduce some intermarker distances.

The final map consisted of 109 markers, 101 of which were simple-sequence repeat (SSR) markers, seven were based on single-strand conformation polymorphism (SSCP), and one was a restriction fragment length polymorphism (RFLP) within a *cycloidea*-like gene (M.A. Chapman and J.M. Burke, unpublished data). For the SSR markers, polymerase chain reaction (PCR) was carried out following either the method of Burke *et al*. (2002b) with forward primers directly fluorescently labelled, or the M13-adapter M13-forward labelled primer method (Schuelke 2000; adapted for use in sunflower by Wills *et al*. 2005). The final annealing temperature was commonly 55 °C but varied from 50 to 60 °C. Primers were labelled with either FAM, HEX, VIC or TET, enabling PCR products to be resolved on a BaseStation Automated DNA Sequencer (MJ Research) or an ABI 3730xl DNA Analyser (Applied Biosystems). Amplicons were pooled such that multiple loci could be resolved per run. PCR for the SSCP and *cyc*-like gene was carried out using the same conditions as the SSR loci; however, the primers were not fluorescently labelled. SSCP products were resolved in MDE acrylamide (Cambrex) and silver-stained using standard protocols. The *cyc*-like gene was PCR-amplified and then digested using the restriction endonuclease *Mly*I. The resulting RFLPs resolved on agarose gels.

Map construction

The linkage map was produced from the full set of 184 RILs and 109 markers using MAPMAKER 3.0/EXP (Lander *et al*. 1987; Lincoln *et al*. 1992). Recombination fractions were translated into centiMorgan distances following Kosambi (1944). Initially, the 'group' command in MAPMAKER was used to identify linked markers with a LOD of > 10.0 and θ < 0.2. Further markers were successively added to these linkage groups (LGs) by relaxing the thresholds, and by following the known genomic locations of the loci where this information was available (Burke *et al*. 2002b; Tang *et al*. 2003). Marker orders within LGs were explored using the 'compare' command and verified using the 'ripple' command.

Field sites used

One site, located at the Cedar Point Biological Station (University of Nebraska, Lincoln; www.unl.edu/cedarpt/;

41°12.3′N, 101°38.5′W), lies in the centre of the range of wild sunflowers and in the southern third of the region of sunflower cultivation. Plants were seeded into an experimental field adjacent to the laboratory facilities. Wild sunflowers grow at the edge of the field, as well in the surrounding short-grass prairie habitat. The second site, Indiana University Botany Experimental Fields (39°10.3′N, 86°30.2′W), is at the eastern edge of both the wild sunflower habitat and the area of sunflower cultivation. Presettlement vegetation in Indiana was deciduous forest, but the experimental fields have been used for cultivation for at least 50 years. Soils were a sandy loam (57% sand, 33% silt, 11% clay) in Nebraska, and a silt loam mixture (15% sand, 71% silt, 14% clay) in Indiana. Water retention in Indiana was much higher than in the well-drained soils of Nebraska. Both sites were ploughed prior to planting to mimic the disturbed conditions prevailing at crop edges and along roadsides where wild populations often occur, and to break up the existing sod.

Environmental data were obtained for each site. For the Indiana site, daily temperature and rainfall data were downloaded from the Indiana University Electronics Department station located 1 km from the field site (http://electron.electronics.indiana.edu/weather/). The Nebraska site weather observations were made approximately 2 km from the field site and were downloaded from the High Plains Climate Research Center site (www.hprcc.unl.edu).

Planting

Sufficient numbers of seed for the field experiment were available from 149 of the 184 fully genotyped lines. We used a mixture of F_6 and F_7 generation seeds. Nine blocks were established at each of the two sites. In each block, five pairs of rows were laid out, with 0.5 m separating the two rows in each pair and 1 m separating the pairs. Within each row, plants were spaced at 0.5 m. One block in Indiana had a 1 m uniform spacing for all plants. Each block was planted with an individual of the cultivated parent (HA89, the male-fertile version of the crop parent) plus each of 149 RILs in Indiana or each of 146 RILs in Nebraska (due to limited seeds for three RILs). Lines were assigned to random locations within each block, with each achene planted ~2.5 cm below the soil surface. Lines with poor germination or limited number of available achenes ($N = 18$) were hand-germinated by placing nicked achenes on moist filter paper overnight, then removing the achene coat and seed coat. The naked seed was then planted the following day.

We planted the Indiana site on 18–19 April 2005. Hand-watering of the site was carried out on the planting days; heavy rainfall after planting removed the need for further watering. We planted the Nebraska site in the following week, 26–27 April 2005. Seeds were watered three times weekly for the first 4 weeks following planting.

Within each block, we planted 1–4 achenes from each line to ensure adequate representation of all RILs. Where multiple seedlings emerged, the seedling nearest the centre of the planting site was retained and the others were removed. Thinning occurred on 18–19 May 2005 in Indiana, and 3 days later in Nebraska. Some thinned seedlings were transplanted to other blocks if that particular line was missing. Seedlings that did not survive until 28 May were scored as missing from the experiment. Emergence rates averaged ~75%, with a few lines showing much lower success. We were unable to determine whether low emergence rates were due to dormancy or reduced seed viability.

Traits measured

Morphological and phenological traits showing consistent differences between domesticated and wild populations (Burke *et al*. 2002b) were measured in this study. Surveys to track the date of first flowering and flowering duration of the primary head were made daily in Indiana and biweekly in Nebraska. We measured head diameter, disk diameter, and number of rays on the primary head of each plant on the second day of flowering. Ray length was calculated from the difference between head diameter and disk diameter. Primary flowering heads were bagged 16 days after flowering, as seed dispersal typically begins after 21 days. Bagging was delayed to allow seed predators such as American goldfinches to affect fitness. All heads were collected approximately 3 weeks after flowering: head diameter and depth were measured, as well as seed length, diameter, and average mass. Shattering results in part from the reflexed curvature of the head as it dries; head depth divided by head diameter was calculated as a proxy for shattering. Leaves were collected near the peak of flowering (15 July in Indiana, 22 July in Nebraska) and weighed when fresh and after drying. Fresh leaves were scanned immediately after collection; digital images were analysed using Scion Image for Windows (Scion Corporation) to obtain leaf area, leaf width and leaf length. Leaf shape was calculated as the ratio of leaf length divided by leaf width; leaf per cent moisture was calculated as wet mass divided by dry mass. At the end of the field season, all plants were harvested, with as much of the primary root collected as possible. After drying the plants, we measured stem diameter, stem height, leaf number, branch number, head number, heads per branch, root mass (dry weight), and length of primary root.

Due to the high level of mortality in Nebraska, several traits could not be consistently measured. Leaf collection occurred a week after a severe heat wave which killed the majority of plants. Therefore, leaf, flower and seed trait data were not collected for many lines in Nebraska. Floral data were collected but in most cases reflected very small heads that opened during the mid-July heat wave.

Data analysis

Most plants died in Nebraska prior to seed production, and the surviving plants set a reduced number of seeds. We therefore scored fitness as the proportion of plants that survived to produce seeds. In Indiana, however, most plants produced seeds, and fitness was scored as the total seed number.

In both sites, we observed substantial variation among blocks in both phenotypes and seed production (although not in survival). To account for this, all traits, except fitness, were standardized within each block in each site to have mean 0 and standard deviation 1, thereby correcting for differences in mean trait values between blocks. Values for each RIL were derived by averaging across blocks to have a single trait value for QTL analyses. Averaging the trait values across plants in different blocks within a site decreases the effect of environmental variation on phenotypes and increases the ability to detect genetic effects. Trait distributions were then checked for normality (PROC UNIFORM, SAS version 9.1, SAS Institute 2004) and transformed by either square-root or log transformations to improve their fit to model assumptions. Mean trait values were restandardized to ensure a mean of 0 and standard deviation of 1. For Indiana, relative fitness was calculated for each block by dividing by the highest fitness in that block, yielding values between 0 and 1. Mean fitness for each line in Indiana was calculated by averaging across the relative values for all blocks. Fitness in Nebraska was scored as the number of blocks in which an RIL survived to produce seeds out of the total number of blocks where it was present. Plants with zero fitness due to germination failure or extraneous factors (e.g. trampling by cows or vole herbivory) were treated as missing data. The number of plants per line ranged from 2 to 9, with a mean of 8.4 in Indiana and 8.1 in Nebraska.

Selection differentials were estimated for the two sites by taking the covariance between fitness (either seed number or survival to seed set) and different traits. Significant selection differentials (s) can result from selection on a trait or on correlated traits; to assess selection when trait correlations have been accounted for, selection gradients (β) were estimated via multiple regression (Lande & Arnold 1983; PROC REG or PROC LOGISTIC, SAS version 9.1, SAS Institute 2004). Due to different traits being measured at the two sites, the two analyses were conducted separately. Variance inflation factor (VIF) was used to detect highly correlated traits (VIF > 10; Neter *et al*. 1996) that can lead to inaccurate type 3 mean-square error estimates. No traits in Indiana were highly correlated, but four traits were excluded from the Nebraska selection gradient analysis due to high VIF. Linear and quadratic terms were analysed to test for the effects of directional and stabilizing selection. Separate analyses were carried out for linear and quadratic terms

due to high levels of correlation; if both linear and quadratic terms were significant for a factor, the quadratic term was used if it significantly improved model fit over the linear term. Nebraska survival data were re-analysed as binomial data (PROC LOGISTIC) with the resulting regression coefficients transformed to selection gradients (Janzen & Stern 1998). The significance of selection differentials and regression coefficients was assessed using 10 000 bootstraps of the mean RIL values (Jackboot macro, SAS Institute).

QTL analyses were carried out separately for the two study sites, again due to different suites of characters being measured at the two sites. QTL analyses employed composite interval mapping as implemented in WinQTL CARTOGRAPHER version 2.5 (Wang *et al*. 2005) with a significance level of 0.05. QTLs were added using forward regression with the standard model (model 6) for up to five control markers. A window size of 10 cm was used with a walk speed of 2 cm. Significant LOD scores were assigned for each trait following permutation tests with 1000 replicates (Churchill & Doerge 1994). Significance of QTLs for survival to set seed in Nebraska was confirmed using logistic regression of marker genotype on survival due to the binomial response variable (Xu & Atchley 1996).

Results

RIL production and genetic mapping

Genotypes were biased towards crop alleles, with 58% (± 15% SD) of the RILs having the crop allele at any given locus (range 10.2% crop to 84.2% crop). Some linkage groups had consistently high bias towards crop alleles, including LG 14 (74.7% crop) and LG 17 (79.1% crop), while LG13 was biased towards wild alleles (38.6% crop).

Environmental data

The summer of 2005 was moderate in temperature in Indiana but higher than average in precipitation. The maximum summer temperature was 36.1 °C, and the site received 57.4 cm of rainfall during the experiment. In contrast, Nebraska experienced a hotter than average summer, with temperatures reaching 42.1 °C on July 20 and 40.3 °C on July 23. Rainfall during the experiment totalled 23.6 cm.

Survival and reproduction

In Indiana, 1305 plants established (out of 1350 attempted). Seventy-eight established plants died due to stem chewing by voles (a cause of death never noted in the agronomic literature). Excluding these deaths, 79% of the plants survived to produce seeds, including eight of the nine individuals in the crop line (HA89). Surviving plants produced an average of 106 seeds, with a maximum of 1432.

Table 1 Selection differentials (*s*) and selection gradients (β) for fitness (seed production or survival) for plants in Indiana and Nebraska field plots. Selection gradients for Nebraska converted from logistic regression parameter estimates (Janzen & Stern 1998). Significance was determined by 10 000 bootstraps: significant values are shown in bold. Factors omitted from multiple regression due to multicollinearity are indicated by **. Abbreviations are used in Fig. 1 and Fig. S1

		Indiana		Nebraska	
Factor	Abbreviation	*s*	β	*s*	β
Flowering date	flwr_date	**−0.049**	−0.008	**−0.118**	−0.534
Stem diameter	stm_diam	**0.025**	−0.018	**0.032**	−0.766
Plant height	ht	0.009	0.000	**0.057**	**
Branch number	br_num	**0.033**	0.015	0.025	−0.423
Head number	hd_num	**0.026**	0.017	**0.072**	−0.030
Root mass	rt_mass	**0.022**	0.003	**0.044**	0.201
Root length	rt_len	**0.017**	−0.002	**0.028**	**
Leaf number	lf_num	**−0.015**	0.002	−0.018	−0.013
Leaf area	lf_area	**0.027**	**0.021**	−0.000	0.166
Leaf shape	lf_shape	**−0.025**	−0.002	0.028	−0.308
Leaf moisture content	lf_moist	−0.002	0.001	−0.007	0.718
Ray number	ray_num	0.011	**−0.025**	**0.026**	0.371
Ray length	ray_len	**0.024**	0.005	**0.128**	**
Disk diameter	disk_diam	**0.031**	**0.033**	**0.082**	1.159
Floral longevity	longevity	**0.028**	**0.017**	na	na
Shattering	shatter	0.000	0.005	**0.055**	0.348
Achene width	ach_wdth	0.012	0.014	0.023	**
Achene length	ach_len	0.009	−0.009	0.009	−0.026
Achene mass	ach_mass	**−0.049**	**−0.073**	0.008	0.060

Of the 1333 attempted, 1201 plants established in Nebraska. Of these, 114 survived to produce seeds (9.5%). None of the crop (HA89) plants survived to produce seeds in Nebraska. Three of the RILs had better than 50% survival, while 40 RILs had a single surviving plant. Those plants surviving to reproduce averaged 71.5 seeds, with a maximum of 781. A haphazard sample of 12 wild plants that volunteered in the plots averaged 1312 seeds per plant, with a maximum of ~7400.

Phenotypic selection analysis

Thirteen traits in Indiana had significant selection differentials, including flowering time, plant size, leaf size and shape, flowering head size and longevity, and achene mass (Table 1). Multiple regression on relative fitness in Indiana found five traits significantly correlated with fitness after correlated traits were accounted for. Several floral traits were under selection, including reduced ray number (selection differential, *s* = 0.011, selection gradient, β = −0.025), increased disk diameter (*s* = 0.031, β = 0.033) and increased floral longevity (*s* = 0.028, β = 0.017). Reduced achene mass (*s* = −0.37; β = −0.073) and increased leaf area (*s* = 0.027, β = 0.021) were also associated with increasing achene number (see Table 1). No quadratic or cross-product term led to increased explanatory power (see Table S1, Supplementary material for trait correlations).

Several traits had significant selection differentials in Nebraska, but none of these had significant linear or quadratic selection gradients in the full model regression model

(Table 1). Selection differentials favoured earlier flowering time, increased plant size (height, stem diameter, root mass, and root length), increased floral size (disk diameter, ray number, and ray length) and increased shattering. A reduced regression model examining only traits with significant selection differential found that early flowering time was significant (*P* = 0.011).

QTL analysis: Indiana

Two chromosomal blocks were associated with the number of seeds produced in Indiana, on linkage groups six (LG6) and nine (LG9), each explaining ~11% of the variation in seed number, with the wild genotype favoured on LG6 and the crop genotype on LG9. Each region changed fitness by 0.33–0.38 standard deviation units. Twenty-eight QTLs were detected for 15 other traits, with one to four QTLs inferred per trait (average = 1.9). The average magnitude of the effect size was 0.38 standard deviation units (0.29–0.64) and the average QTL explained 14% of the trait variation (Table 2; Fig. S1, Supplementary material).

QTL analysis: Nebraska

Two chromosomal blocks were associated with survival to seed production in Nebraska, on LG6 and LG9, each explaining ~12% of the variation in survival, with each region affecting fitness by 0.35–0.40 standard deviation units. On LG6, the wild genotype was favoured, while on LG9, crop alleles were favoured. Twenty-eight QTLs were

Table 2 Inferred QTL positions, additive effects of a wild allele in standard deviation units, and per cent variation explained (PVE) for 18 traits using composite interval mapping in a recombinant inbred population derived from cultivated (cmsHA89) × wild (*Helianthus annuus* var. *annuus*) sunflower grown in Indiana and Nebraska. QTLs for a given trait with overlapping positions in both field populations are shown on the same line. 'Prior map' indicates concordance with QTL inferred by Burke *et al.* (2002b), with 'C' indicating overlapping 1-LOD positions, 'L' indicating the same linkage group (LG) but different position, and 'NA' indicating a trait that was not measured in the prior study, while blanks indicate that no QTL was found on that linkage group in the prior study

| Trait	Indiana						Nebraska						Prior map
	LG	Position	Marker	1-LOD	Effect (wild)	PVE	LG	Position	Marker	1-LOD	Effect (wild)	PVE	
Achene number	6	70.7	HT913	69.2–70.7	0.33	11							NA
	9	52.0	HT978	47.6–52.0	–0.38	11							NA
Survival							6	70.5	HT913	66.9–70.5	0.35	12	NA
							9	52.0	HT978	48.3–52.0	–0.40	12	NA
Flowering date	6	70.7	HT913	70.0–70.7	–0.64	36	6	64.5	HT918	58.9–68.5	–0.505	26	L
	9	51.9	HT978	48.4–51.9	0.32	11	9	52.0	HT978	38.9–52.0	0.40	15	C
	14	12.9	ORS398	9.0–19.6	0.437	10	14	14.8	ORS398	12.0–17.6	0.707	30	
	17	37.5	ORS735	33.3–39.5	0.41	9							C
Stem diameter	3	29.0	HT441	18.5–37.6	–0.403	15							C
Plant height	3	8.9	HT292	4.9–15.0	–0.363	13							C
	6	68.5	ORS57	61.7–70.5	–0.368	14							C
	7	8.4	ORS331	3.9–14.3	0.296	9							C
							8	50.7	ZVG37	48.5–50.7	–0.40	13	
							10	10.2	HT347	2.8–14.4	–0.332	8	L
							10	30.7	HT419	26.7–32.7	0.347	9	L
Head number							6	70.5	HT913	69.8–70.5	0.31	9	
							9	52.0	HT978	47.0–52.0	–0.33	9	C
Root length	17	0.0	ORS172	0–10.7	0.293	10							NA
Root mass	3	19.6	ORS949	10.8–29.7	–0.349	13	3	12.9	HT1031	5.3–20.4	–0.341	12	NA
							9	52.0	HT978	46.2–52.0	–0.29	8	NA
Leaf number	6	70.7	HT913	69.2–70.7	–0.55	25	6	70.5	HT913	64.1–70.5	–0.41	14	L
	7	12.4	ORS331	8.4–16.5	0.39	16							L
Leaf shape							6	46.0	HT769	39.1–58.0	–0.457	18	
	16	20.0	ORS899	8.4–31.9	0.415	19							
Leaf area	3	29.0	HT441	19.7–41.0	–0.337	11							L
Leaf moisture content	6	70.7	HT913	68.9–70.7	–0.51	22	6	68.5	ORS57	61–70.5	–0.458	23	
							9	32.1	CY05A	18.6–40.9	–0.58	31	
							12	0	ZVG54	0–14.1	–0.392	18	
	13	50.6	ORS799	42.5–50.6	–0.35	8							
	15	31.8	HT716	21.6–47.3	0.319	10							
Ray number	1	12.0	HT446	6–22.9	–0.356	12							L
							4	12.3	HT664	6.2–18.1	–0.37	13	
							14	12.8	HT319	7.8–19.0	–0.47	12	
							16	74.8	HT52	54.5–83.9	–0.27	9	
Ray length	13	0.0	HT568	0–6.0	0.314	10							
	16	70.8	HT52	60.1–86.0	–0.322	11							
							6	70.7	HT913	69.9–70.7	–0.512	28	L
							9	52.0	ORS176	43.5–52.0	–0.37	14	C
							14	14.8	HT319	10.8–20.4	–0.473	13	
							16	74.8	HT52	63.8–84.4	–0.255	8	
Disk diameter							4	16.3	ORS366	10.1–23.0	–0.37	11	L
							6	70.4	HT913	69.4–70.4	0.33	11	
							9	52.0	HT978	48.4–52.0	–0.43	16	
	14	20.6	HT319	11.5–30.7	–0.325	9	14	20.6	HT319	14.4–29.9	–0.381	13	
Shattering (head reflex)	9	52.0	HT978	35.8–52.0	0.29	9							
							13	0	HT568	0–6.7	0.381	13	
							16	37.3	ORS656	29.2–46.2	0.498	23	
Achene length	2	0.0	ORS925	0–7.0	–0.3451	13							
	4	65.4	HT221	56.0–70.2	–0.40	17							
	5	6.7	ORS547	3.7–12.8	–0.39	15							L
	13	0.0	HT568	0–24.3	0.33	15							
Achene mass	6	70.7	HT913	69.8–70.7	–0.58	28							L
	10	20.2	ORS565	14.2–26.1	–0.395	12							L
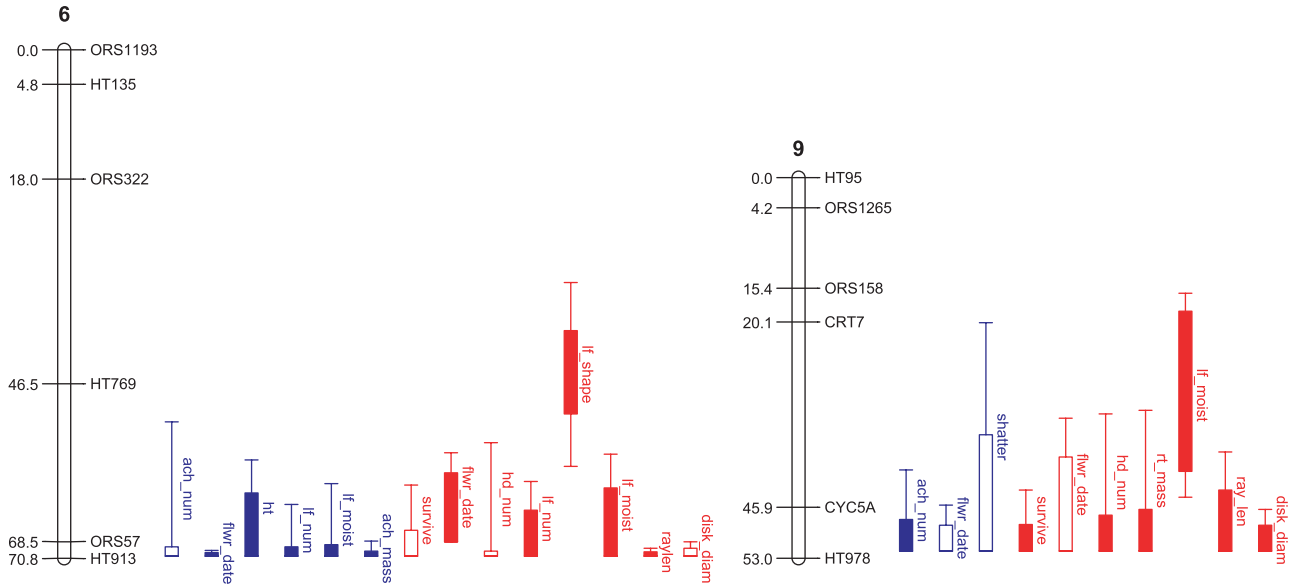
Fig. 1 Map of linkage groups 6 and 9 derived from the cultivated (cmsHA89) × wild (*Helianthus annuus* var. *annuus*) sunflower recombinant inbred population grown in Indiana and Nebraska. Marker names are listed to the left of each linkage group, and boxes to the right of each linkage group indicate QTL positions. A vertical bar marks the 1-LOD support interval for each QTL, while the lines extending from the box mark the 2-LOD support intervals. Blue bars show QTLs detected in Indiana, while red indicate those detected in Nebraska. Open bars indicate QTLs where the crop allele decreases the trait value. See Table 1 for trait abbreviations.

inferred for 11 other traits, with one to four QTLs inferred per trait (average = 2.2). The average magnitude of the effect size was 0.40 standard deviation units (from 0.26 to 0.71), and on average 15% of the variation was explained by each QTL (Table 2; Fig. S1).

Concordance of QTL detected

The QTLs for several traits were approximately concordant between Indiana and Nebraska, including three for flowering date, one for root mass, one for leaf number, one for leaf moisture content, and one for disk diameter. The direction of effect was the same for all of these concordant QTLs in both field sites (Table 2; Fig. S1). Fitness, measured as survival to seed set (for Nebraska) or number of seeds produced (for Indiana), mapped to the same position on two linkage groups in both populations. Combining the results of both field sites, 41 distinct QTLs were found in one or both field sites for traits previously mapped (Burke *et al*. 2002b). Of these, eight had overlapping positions on linkage groups, 12 were on the same linkage group but in a different position, and 21 were on different linkage groups in our study. Seven of the eight concordant QTLs were in agreement with the prior study in terms of the direction of effect. The exception was a locus for number of heads on LG9 in Nebraska, where the wild allele had the effect of decreasing the number of flowering heads in this study. Nine of the 12 QTLs that were located on the same

linkage group but at different locations agreed in effect direction with the previous study (Table 2).

Discussion

Traits under selection

Despite the contrasting environments and resulting differences in survival and reproduction, phenotypic selection analyses found many traits with significant correlations with fitness in both Indiana and Nebraska. No trait had opposing selective effects in the two environments. Surprisingly, crop traits were favoured for three of these, including earlier flowering time, larger stem diameter, and larger flowering disk diameter, while the number of flowering heads was not under selection at either site. After accounting for correlated traits, significant selection gradients in Indiana favoured wild trait values for ray and seed size but crop trait values were favoured for leaf size, floral longevity, and disk size. The number of heads was not significant in multiple regression analyses despite high levels of variation (range one to 32 in Indiana, mean 3.8).

Some traits under selection in Nebraska seem puzzling at first: why would larger floral disks and longer ray flowers be favoured when selection was measured as survival to seed production? Selection on these traits is likely an artefact of the timing of mortality in Nebraska. Many plants produced a very small floral head just prior to death. Plants

that flowered prior to the extreme temperatures produced larger floral heads and were also more likely to set seed, leading to the correlation of floral traits with survival.

The selection differentials favouring early flowering in both sites are puzzling. Wild plants consistently flower later than the crop, yet early flowering plants (or those with correlated traits) were consistently favoured in the two sites. The high levels of hybridization seen between crop and wild plants should allow this trait to introgress into wild populations: that wild plants retain a late-flowering phenotype suggests that either flowering time has different effects on fitness in a wild-type genetic background, or that selection against correlated traits in the wild counteracts selection for early flowering. Intriguingly, similar patterns were seen in studies of flood tolerance in irises, in which an allele from one species (which is not flood tolerant) was favoured in the other species' background (Martin *et al.* 2006).

Alleles under selection

The same chromosomal segments (the distal ends of LG6 and LG9) were under selection in both sites, with the crop allele consistently favoured at one locus and the wild allele at the other. The consistent selection on these two loci in the two different environments is surprising given the differences in the environment, survivorship, and fitness measures. In Indiana, the position of the fitness QTL on LG6 aligns with QTL for two traits under selection where the wild phenotype was favoured: fewer days to flowering and smaller achene mass. Although crop sunflowers flower earlier than wild sunflowers, and crop alleles at most QTLs affecting flowering time lead to earlier flowering (Burke *et al.* 2002b), the crop allele at the flowering time QTL on LG6 produces *later* flowering. Reduced achene mass has a significant selection gradient in Indiana, making it a second candidate for the target of selection. A QTL for flowering time also maps to the distal end of LG6 in Nebraska, as do QTLs for disk diameter and ray length. Flowering time is the best candidate trait for the target of selection in Nebraska, as a significant selection gradient was found in the reduced regression model, and the wild allele leads to earlier flowering time.

On LG9, it is the crop allele that is favoured at the QTL in both environments. In Indiana, this QTL overlapped with QTLs for flowering time, which had a significant selection differential, although not a significant selection gradient, and shattering, which was not under selection. In Nebraska, the QTL overlaps with five traits with significant selection differentials, including disk diameter, ray flower length, head number, root mass and flowering time. For all of these traits, crop alleles result in changes in the favoured direction. Of these, flowering time is the most likely target of selection as it had a significant selection gradient in the reduced model.

The selection on flowering time and seed mass might explain the fitness of the chromosomal segments seen here. However, unmeasured traits that are correlated with flowering time could potentially explain these results (Lande & Arnold 1983). For example, physiological traits are challenging to measure on the scale of this experiment but might well be subject to different selection in the crop and wild environments.

Consistency of QTLs

Many of the inferred QTLs differed between Indiana and Nebraska. Some of the differences in the QTLs detected at each site may be explained by the truncation of the growing season in Nebraska. Due to the early mortality of many plants there, our ability to sample leaf, floral, and seed traits was limited, and potentially biased. QTLs that are concordant across environments are excellent candidates for predicting and analysing introgression from crops to wild populations. Nonconcordant QTLs may be due to genotype by environment interactions, and so are likely to have unpredictable effects on introgression across the range of the wild plant.

This study detected several QTLs found in a previous study (Burke *et al.* 2002b), as well as novel QTLs, but in general found many fewer QTLs per trait. Our effect size estimates were much larger than Burke *et al.*'s (2002b) study, which was performed in a greenhouse. These differences are easily explained by the smaller sample size (149 RILs vs. 374 F_3 individuals in the earlier study), which would lead to fewer QTLs detected and larger estimated effect sizes. In addition, the larger environmental variation in this field study would likely decrease the number of QTLs detected. When QTLs were detected in both Indiana and Nebraska, they were concordant with, or on the same linkage group as, QTLs found in the previous study, with the single exception of a flowering time QTL on LG14. This suggests that the concordant QTLs have consistent effects in a wide range of environmental conditions.

RILs, selection, and limitations

Several aspects of the cross used to generate the RILs may explain the observed bias towards crop alleles. First, the RILs were propagated by self-fertilization, which is typical in crop sunflowers but very rare in wild populations, which are typically self-incompatible. This selection for self-fertilization should have led to a preponderance of crop alleles on linkage groups associated with self-pollination. The *S* (self-incompatibility) locus is found on LG17 (Burke *et al.* 2002b; Gandhi *et al.* 2005) and we did indeed see a preponderance of crop alleles on this LG (see Results). Other modifier loci for self-incompatibility may also have biased

the genotypes resulting in the observed higher than 50% overall frequency of crop genotypes. In general, the inbred elite crop-line has probably been purged of many deleterious alleles, while out-crossing wild plants typically harbour many recessive deleterious alleles. Selection during the creation of RILs probably eliminated some lines that were homozygous for deleterious wild alleles with major effects on fitness, biasing the allele frequencies, and may explain the overall higher than expected percentage of crop alleles. On the other hand, the crop used in the initial cross was the male-sterile cmsHA89, indicating that all RILs must have inherited the wild restorer allele from the wild parent so as to be pollen fertile. Mapping of the restorer allele suggests that the major restorer locus is on LG 13 (Yu *et al.* 2003), which was biased towards wild alleles in this study (61.4% wild alleles).

Biasing the allele frequencies on certain linkage groups, either towards the crop or wild alleles, should decrease the power to detect QTL due to decreased variation. However, QTLs were detected on the three most-biased linkage groups, including date of flowering and disk diameter on LG14, achene length and leaf moisture content on LG 13, and flowering date and root mass on LG 17.

Conclusions and further work

We discovered that some crop QTL alleles are favoured in a noncrop environment and genetic background. This was predicted from earlier genetic studies because the cultivated sunflower is known to harbour alleles with effects in the direction of the wild parent (Burke *et al.* 2002b). An example of this is the QTL for flowering time on linkage group 6, where the cultivar allele causes later rather than earlier flowering. Further evaluation of the fitness effects of these QTLs will require introgressing them into the wild background to test whether their fitness effects are consistent when fewer crop alleles are present.

Our work provides limited support for the concept of tandem constructs as a means for limiting crop-wild gene flow in that the fitness QTLs detected have consistent effects in different geographical areas. Unfortunately, the effects are not large enough to contain strongly favourable transgenes. Transgenes with strong ecological consequences, such as the *Bt* transgene, can lead to an increase in fecundity of as much as 55% in BC1 hybrids compared to wild plants (Snow *et al.* 2003). Also, the fitness QTLs that we detected were highly pleiotropic, which would limit their practical value to breeders. Nonetheless, our study has important implications for the investigation of crop-wild introgresion. The long-term dynamics of escaped crop alleles will depend upon their fitness effects and those of tightly linked genes. As far as we are aware, this is the first such genome-wide analysis of the fitness effects of crop QTLs in a noncrop environment and genetic background.

Alleles have moved between crops and wild plants for millennia. Existing crop–wild hybrid zones and populations can be used to examine patterns of gene flow, but when coupled with studies of genetically characterized populations (e.g. recombinant inbred lines, near isogenic lines, or back-cross mapping populations), it becomes possible to analyse the relationships among loci, phenotypes, and fitness (e.g. Lexer *et al.* 2003; Weinig *et al.* 2003; Martin *et al.* 2006). The consistent QTLs for fitness across two disparate environments identified here provide a framework for examining the introgression of linked traits in sunflowers. For example, the crop allele favoured on LG9 is linked to, or pleiotropic with, an allele that decreases seed shattering, which we would expect to be unfavourable in the noncrop environment (although it had no effect on fitness in the conditions of this experimental study). Identifying such combinations of favourable and unfavourable alleles will allow the use of natural experiments to better understand the dynamics of crop-wild gene flow. Escaped transgenes in wild populations have been rare so far (but see Reichman *et al.* 2006), allowing few opportunities to study the effects of fitness and linkage on their spread. The knowledge of the fitness effects of crop traits will allow ecologists and evolutionary biologists to make use of crop-wild gene flow occurring over the past millennia to better predict the outcome of escapes when they do occur.

References

Alexander HM, Cummings CL, Kahn L, Snow AA (2001) Seed size variation and predation of seeds produced by wild and crop-wild sunflowers. *American Journal of Botany*, **88**, 623–627.

Allainguillaume J, Alexander M, Bullock JM *et al.* (2006) Fitness of hybrids between rapeseed (*Brassica napus*) and wild *Brassica rapa* in natural habitats. *Molecular Ecology*, **15**, 1175–1184.

Arriola PE, Ellstrand NC (1997) Fitness of interspecific hybrids in the genus *Sorghum*: persistence of crop genes in wild populations. *Ecological Applications*, **7**, 512–518.

Barton NH (1979) Gene flow past a cline. *Heredity*, **43**, 333–339.

Boudry P, Moerchen M, Saumitou-Laprade P, Vernet P, Van DH (1993) The origin and evolution of weed beets: consequences for

the breeding and release of herbicide-resistant transgenic sugar beets. *Theoretical and Applied Genetics*, **87**, 471–478.

Burke JM, Rieseberg LH (2003) Fitness effects of transgenic disease resistance in sunflowers. *Science*, **300**, 1250–1250.

Burke JM, Gardner KA, Rieseberg LH (2002a) The potential for gene flow between cultivated and wild sunflower *(Helianthus annuus)* in the United States. *American Journal of Botany*, **89**, 1550–1552.

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002b) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257–1267.

Chapman MA, Burke JM (2006) Letting the gene out of the bottle: the population genetics of genetically modified crops. *New Phytologist*, **170**, 429–443.

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971.

Cummings CL, Alexander HM, Snow AA *et al.* (2002) Fecundity selection in a sunflower crop-wild study: can ecological data predict crop allele changes? *Ecological Applications*, **12**, 1661–1671.

Degrandi-Hoffman G, Chambers M (2006) Effects of honey bee (Hymenoptera: Apidae) foraging on seed set in self-fertile sunflowers (Helianthus annuus L.). *Environmental Entomology*, **35**, 1103–1108.

Ellstrand NC (2003) *Dangerous Liaisons? When Cultivated Plants Mate with Their Wild Relatives*. John Hopkins University Press, Baltimore, Maryland.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. In: *Annual Review of Ecology and Systematics* (eds Fautin DG Futuyma DJ James FC), pp. 539–563.

Gandhi SD, Heesacker AF, Freeman CA *et al.* (2005) The self-incompatibility locus (*S*) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theoretical and Applied Genetics*, **111**, 619–629.

Gressel J (1999) Tandem constructs: preventing the rise of super-weeds. *Trends in Biotechnology*, **17**, 361–366.

Harter AV, Gardner KA, Falush D *et al.* (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature*, **430**, 201–205.

Hegde SG, Nason JD, Clegg JM, Ellstrand NC (2006) The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution*, **60**, 1187–1197.

Hooftman DAP, Nieuwenhuis BPS, Posthuma KI, Oostermeijer JGB, den Nijs H (2007) Introgression potential of downy mildew resistance from lettuce to *Lactuca serriola* and its relevance for plant fitness. *Basic and Applied Ecology*, **8**, 135–146.

Janzen FJ, Stern HS (1998) Logistic regression for empirical studies of multivariate selection. *Evolution*, **52**, 1564–1571.

Kosambi DD (1944) The estimation of map distance from recombination values. *Annals of Eugenics*, **12**, 172–175.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Lander ES, Green P, Abrahamson J *et al.* (1987) MAPMARKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.

Lexer C, Randell RA, Rieseberg LH (2003) Experimental hybridization as a tool for studying selection in the wild. *Ecology*, **84**, 1688–1699.

Lincoln ES, Daly MJ, Landers ES (1992) Constructing genetic maps with MAPMARKER/EXP 3.0. Whitehead Institute Technical Report, Cambridge, MA.

Linder CR, Taha I, Seiler GJ, Snow AA, Rieseberg LH (1998) Long-term introgression of crop genes into wild sunflower populations. *Theoretical and Applied Genetics*, **96**, 339–347.

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics*, **172**, 2481–2489.

Mercer KL, Andow DA, Wyse DL, Shaw RG (2007) Stress and domestication traits increase the relative fitness of crop-wild hybrids in sunflowers. *Ecology Letters*, **10**, 383–393.

Mercer KL, Wyse DL, Shaw RG (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.

Neter J, Kutner M, Nachtsheim C, Wasserman W (1996) *Applied Linear Statistical Models*. Irwin, Chicago.

Panetsos CA, Baker HG (1968) Origin of variation in wild *Raphanus sativus* (Cruciferae) in California. *Genetica*, **38**, 243–&.

Reichman JR, Watrud LS, Lee EH *et al.* (2006) Establishment of transgenic herbicide-resistant creeping bentgrass (*Agrostis stolonifera* L.) in non-agronomic habitats. *Molecular Ecology*, **15**, 4243–4255.

SAS Institute I (2004) *The SAS System for Windows*, Version 9.1. The SAS Institute, Cary, North Carolina.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Smith BD (2006) Eastern North America as an independent center of plant domestication. *Proceedings of the National Academy of Sciences, USA*, **103**, 12223–12228.

Snow AA, Moran-Palma P, Rieseberg LH, Wszelaki A, Seiler G (1998) Fecundity, phenology, and seed dormancy of F1 wild-crop hybrids in sunflower (*Helianthus annuus,* Asteraceae). *American Journal of Botany*, **85**, 794–801.

Snow AA, Pilson D, Rieseberg LH *et al.* (2003) A *Bt* transgene reduces herbivory and enhances fecundity in wild sunflowers. *Ecological Applications*, **13**, 279–286.

Snow AA, Uthus KL, Culley TM (2001) Fitness of hybrids between weedy and cultivated radish: implications for weed evolution. *Ecological Applications*, **11**, 934–943.

Tang SX, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theoretical and Applied Genetics*, **107**, 6–19.

Tang SYuJ-K, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124–1136.

Wang S, Basten CJ, Zeng ZB (2005) Windows QTL CARTOGRAPHER 2.5. Department of Statistics, North Carolina State University, Raleigh, North Carolina.

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

Weinig C, Dorn LA, Kane NC *et al.* (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, **165**, 321–329.

Wendel JF, Percy RG (1990) Allozyme diversity and introgression in the Galapagos-Islands endemic *Gossypium darwinii* and its relationship to continental *Gossypium barbadense*. *Biochemical Systematics and Ecology*, **18**, 517–528.

Whitton J, Wolf DE, Arias DM, Snow AA, Rieseberg LH (1997) The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. *Theoretical and Applied Genetics*, **95**, 33–40.

Wills DM, Burke JM (2007) QTL analysis of the early domestication of sunflower. *Genetics*.

Wills DM, Hester ML, Liu AZ, Burke JM (2005) Chloroplast SSR

polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. *Theoretical and Applied Genetics*, **110**, 941–947.

Wright JW, Stanton ML, Scherson R (2006) Local adaptation to serpentine and non-serpentine soils in *Collinsia sparsiflora*. *Evolutionary Ecology Research*, **8**, 1–21.

Xu SZ, Atchley WR (1996) Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics*, **143**, 1417–1424.

Yu JK, Tang S, Slabaugh MB *et al*. (2003) Towards a saturated genetic linkage map for cultivated sunflower. *Crop Science*, **43**, 367–387.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Linkage maps and QTL positions derived from the cultivated (cmsHA89) × wild (*H. annus var. annus*) sunflower recombinant inbred population grown in Indiana and Nebraska.

Table S1 Correlation coefficients between phenotypic traits for plants grown in the Indiana and Nebraska field plots.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/
10.1111/j.1365-294X.2007.03596.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.