Theor Appl Genet (2010) 120:829–842
DOI 10.1007/s00122-009-1212-1

ORIGINAL PAPER

Pleiotropy of the branching locus (*B*) masks linked and unlinked quantitative trait loci affecting seed traits in sunflower

Eleni Bachlava · Shunxue Tang · Guillermo Pizarro · Gunnar Felix Schuppert · Robert K. Brunick · Doerthe Draeger · Alberto Leon · Volker Hahn · Steven J. Knapp

Received: 16 April 2009 / Accepted: 27 October 2009 / Published online: 18 November 2009

Abstract The discovery of unbranched, monocephalic natural variants was pivotal for the domestication of sunflower (*Helianthus annuus* L.). The branching locus (*B*), one of several loci apparently targeted by aboriginal selection for monocephaly, pleiotropically affects plant, seed and capitula morphology and, when segregating, confounds the discovery of favorable alleles for seed yield and other traits. The present study was undertaken to gain deeper insights into the genetics of branching and seed traits affected by branching. We produced an unbranched hybrid testcross recombinant inbred line (TC-RIL) population by crossing branched (*bb*) and unbranched (*BB*) RILs to an unbranched (*BB*) tester. The elimination of branching concomitantly eliminated a cluster of *B*-linked seed trait quantitative trait loci (QTL) identified by RIL per se testing. We identified a seed oil content QTL linked in repulsion and a 100-seed weight QTL linked in coupling to the *B* locus and additional unlinked QTL, previously masked by *B*-locus pleiotropy. Genomic segments flanking the *B* locus harbor multiple loci for domestication and post-domestication traits, the effects of which are masked by *B*-locus pleiotropy in populations segregating for branching and can only be disentangled by genetic analyses in unbranched populations. QTL analyses of NILs carrying wild *B* alleles substantiated the pleiotropic effects of the *B* locus. The effect of the *B* locus on branching was masked by the effects of wild alleles at independent branching loci in hybrids between monocephalic domesticated lines and polycephalic wild ecotypes; hence, the *B* locus appears to be necessary, but not sufficient, for monocephaly in domesticated sunflower.

Communicated by A. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1212-1) contains supplementary material, which is available to authorized users.

E. Bachlava · S. Tang · D. Draeger · S. J. Knapp (✉)
Institute of Plant Breeding, Genetics, and Genomics,
The University of Georgia, 111 Riverbend Road,
Athens, GA 30602, USA
e-mail: sjknapp@uga.edu

G. Pizarro · G. F. Schuppert · R. K. Brunick
Department of Crop and Soil Science, Oregon State University,
Corvallis, OR 97331, USA

A. Leon
Advanta Seeds, Balcarce Research Station, Ruta 226,
KM 60.3 (7620), Balcarce PCIA DE BS AS, Argentina

V. Hahn
State Plant Breeding Institute, University of Hohenheim,
70593 Stuttgart, Germany

Introduction

The discovery of unbranched, monocephalic natural variants and development of unbranched land races (primitive cultivars) by Native Americans was a pivotal event in the domestication of sunflower (*Helianthus annuus* L.). Unbranched sunflower emerged and were culled from a vast pool of strongly branched wild and weedy populations distributed throughout North America (Heiser 1945, 1951, 1976; Rogers et al. 1982). Hopi, Mandan and other land races are presumed to be modern-day descendants of the earliest unbranched types selected by Native Americans (Heiser 1976; Tang and Knapp 2003; Harter et al. 2004; Wills and Burke 2007). Wild ecotypes of *H. annuus* and other species of the genus *Helianthus* found throughout North America are profusely branched with indeterminate flowering (Heiser 1976; Rogers et al. 1982). Modern

oilseed and confectionery cultivars are monocephalic and rarely produce secondary branches and flowers (Putt 1964; Hockett and Knowles 1970). The unbranched, monocephalic phenotype maximizes seed yield, flowering synchrony and morphological uniformity, and has been critical for the development of high-yielding cultivars and hybrids to support the production of sunflower on 22 million hectares worldwide (Hockett and Knowles 1970; Fick et al. 1974b; Dedio 1980; Fick and Miller 1997; http://faostat.fao.org/).

While branching and indeterminate flowering are undesirable for agronomic production, both play a crucial role in sunflower breeding and hybrid seed production by creating a wider window for hybridization than can be achieved with unbranched, determinate flowering phenotypes (Fick et al. 1974b; Dedio 1980; Fick and Miller 1997). Modern single-cross hybrids are produced by hybridizing cytoplasmic-genic male-sterile (CMS), unbranched, determinate flowering female lines (A-lines) and fertile, branched, indeterminate flowering male lines (R-lines). Within the pool of modern parent lines, branching is controlled by a single locus (*B*) with significant pleiotropic effects on plant, seed and capitula morphology (Putt 1943; Hockett and Knowles 1970). The dominant allele (*B*) produces unbranched plants and is ubiquitous among A-lines and isogenic B-lines (fertility maintainer lines), whereas the recessive allele (*b*) produces branched plants and is nearly ubiquitous among R-lines (Fick and Miller 1997).

Branching increases the number of capitula and typically decreases capitula diameter and 100-seed weight and increases seed oil content (Putt 1943; Dedio 1980; Fick et al. 1974b; Tang et al. 2006). The effect of branching on seed oil content is presumably indirectly caused by changes in seed morphology, e.g., smaller seeds produced in smaller capitula have thinner pericarps. Seed oil content and seed morphology are genetically complex traits, although both are often moderately to highly heritable (Fick 1975; Leon et al. 1995, 2003; Mestries et al. 1998; Mokrani et al. 2002; Bert et al. 2003; Tang et al. 2006). Several of the seed trait quantitative trait loci (QTL) identified in sunflower have large effects and are tightly linked to hypodermis (*Hyp*) and phytomelanin (*P*) pigment loci (Leon et al. 1995, 2003; Tang et al. 2006). The *P* locus is tightly linked to the self-incompatibility (*S*) locus, and the genomic region surrounding the *P*–*S* interval harbors QTL for several domestication and post-domestication traits (Burke et al. 2002; Lexer et al. 2005; Tang et al. 2006; Wills and Burke 2007). Moreover, several large-effect QTL identified for seed oil content and seed morphology traits are tightly linked to the *B* locus and are likely caused by *B* locus pleiotropy (Mestries et al. 1998; Bert et al. 2003; Tang et al. 2006).

Branching and seed oil content are typically positively genetically correlated in sunflower (Fick et al. 1974b; Dedio 1980; Mestries et al. 1998; Tang et al. 2006); however, Bert et al. (2003) identified a QTL linked to the *B* locus in a population (XRQ × PSC8) where the allele transmitted by the branched parent decreased seed oil content. Could this result have been caused by the segregation of large-effect QTL linked to the *B* locus, which compensated for the large positive effect of the recessive allele (*b*) on seed oil content? Could the correlation between these traits be caused by both the pleiotropic effects of the *B* locus and effects of loci linked in repulsion to the *B* locus? To gain a deeper understanding of the pleiotropic effects of branching on seed traits, validate previously identified QTL, and search for QTL masked by the pleiotropic effects of the *B* locus, forward genetic analyses were completed in an unbranched testcross hybrid recombinant inbred line (TC-RIL) population, CMS-HA372 × (RHA280 × RHA801). The RILs were developed from a previously described cross between a branched (*bb*) oilseed R-line (RHA801) and an unbranched (*BB*) confectionery R-line (RHA280) (Tang et al. 2006). Several large-effect seed trait QTL identified in the RHA280 × RHA801 RIL population per se were found in clusters tightly linked to *B*, *Hyp* and *P* (Tang et al. 2006). Here, we report on a study and analysis of unbranched TC-RILs and the development and analysis of branched and unbranched near-isogenic lines (NILs) from hybrids between an unbranched elite inbred line (CMS-HA383) and branched wild donors (PI-AZ and PI-MT). The absence of branching among the TC-RILs facilitated testing for *B*-locus pleiotropy, validation of previously identified QTL linked to *Hyp* and *P*, screening for QTL previously masked by the pleiotropic effects of the *B* locus and a genome-wide search for QTL affecting seed yield.

Materials and methods

Testcross hybrid RIL population development and phenotyping

We developed 173 unbranched (*BB* or *Bb*) TC-RILs by crossing 173 RHA280 × RHA801 RILs (F_7) to CMS-HA372 (PI 534658), a highly inbred CMS line (A-line). The RILs were developed from a hybrid between CMS fertility restorer (R) lines (RHA280 × RHA801) and have been previously described (Tang et al. 2006). RHA280 is an unbranched (*BB*) low-oil R-line, whereas RHA801 is a branched (*bb*) high-oil R-line (Fick et al. 1974a; Roath et al. 1981). CMS-HA372 is an unbranched (*BB*) oilseed A-line (Miller and Gulya 1990). HA372, the fertile isogenic CMS sterility maintainer (B) line, was field tested

and phenotyped in lieu of the sterile A-line (CMS-HA372), as were the RIL parents (RHA280 and RHA801).

The parents, tester and 173 TC-RILs were field tested in Corvallis, Oregon in 2001 and 2002 in a randomized complete block design with two replications per year. Planting, cultivation and harvesting were conducted as previously described (Tang et al. 2006). Capitula were combine-harvested at the R9 stage during physiological maturity (Schneiter and Miller 1981). Threshed seeds were oven-dried and cleaned of debris before measuring 100-seed weight and estimating seed yield. Seed oil contents were measured from 5 g of oven-dried seed per experimental unit by nuclear magnetic resonance, as previously described (Tang et al. 2006).

Statistical and QTL mapping analyses of the TC-RIL population

Restricted maximum likelihood (REML) estimates of year, TC-RIL, TC-RIL × year and residual variance components were estimated using SAS PROC VARCOMP (SAS Institute 2004). Broad-sense family mean heritabilities (h^2) were estimated by

$$h^2 = \sigma_G^2 \Big/ \left(\sigma_G^2 + \frac{\sigma_{GY}^2}{r} + \frac{\sigma_E^2}{ry}\right),$$

where σ_G^2 is the between TC-RIL variance component, σ_{GY}^2 is the between TC-RIL × year variance component, σ_E^2 is the residual variance component, r is the number of replications and y is the number of years (Bernardo 2002). Type III F statistics and least square means for TC-RILs within and across years were estimated using PROC MIXED. Spearman rank correlations between years (year one × year two rank correlations for each trait) and Pearson correlations between traits across years were estimated using PROC CORR. Heterosis percentages were estimated for each trait by $[(y_T - y_P)/y_P] \times 100$, where y_T is the least square mean for HA372 and y_P is the least square mean for a CMS-HA372 × (RHA280 × RHA801) testcross hybrid RIL.

QTL were identified and statistics were estimated in two stages. First, composite interval mapping (CIM) analyses (Zeng 1993) were performed on TC-RILs least square means within and across years using WinQTL Cartographer (Wang et al. 2007). Cofactors were selected using stepwise regression and LOD scores and other statistics were estimated by testing the null hypothesis in 2 cM intervals across the genome. Second, multilocus genetic models were developed and tested using DNA and phenotypic marker loci as independent variables in mixed linear model analyses. Loci selected for inclusion in multilocus genetic models were centered or nearly centered on LOD maxima of QTL estimated by CIM. Type III F statistics, least square means for DNA and phenotypic marker locus genotypes and other statistics were estimated using PROC MIXED in SAS 9.1. The null hypothesis of no intralocus testcross hybrid effect (H_0:$y_{A*A} \neq y_{A*a}$) was tested for each locus using linear contrasts among least square means by ESTIMATE statements in PROC MIXED. The testcross effect was estimated as $a^* = (y_{A*A} - y_{A*a})/2$, where y_{A*A} is the CMS-HA372/RHA280 heterozygote mean, y_{A*a} is the CMS-HA372/RHA801 heterozygote mean, and A^*, A, and a correspond to alleles transmitted by the tester (CMS-HA372), RHA280 and RHA801, respectively (Bernardo 2002). We tested for aliasing of independent variables using the ALIASING statement in PROC GLM. Coefficients of determination (R^2) were also estimated for the different multilocus genetic models using PROC GLM.

We used an identical approach to re-analyze seed oil content and 100-seed weight QTL in the RHA280 × RHA801 RIL population using a combination of DNA and phenotypic marker loci identified by RIL per se (Tang et al. 2006) and TC-RIL testing. The additive (a) effects of the B locus, and DNA and phenotypic marker loci linked or unlinked to the B locus were estimated using Type III F statistics and ESTIMATE statements in PROC MIXED. In the RIL population, additive effects were equal to $a = (y_{AA} - y_{aa})/2$, where y_{AA} is the RHA280/RHA280 homozygote mean, and y_{aa} is the RHA801/RHA801 homozygote mean. We tested for aliasing of independent variables using the ALIASING statement in PROC GLM. DNA and phenotypic marker loci selected for inclusion in multilocus genetic models were either centered or nearly centered on LOD maxima of QTL identified by CIM.

Development and analyses of elite × land race and elite × wild hybrids

Single-cross hybrids were produced between CMS-HA372 and 11 wild ecotypes, CMS-HA383 and 11 wild ecotypes, CMS-HA372 and four unbranched Native American land races, and CMS-HA383 and two unbranched Native American land races. Seeds of CMS-HA383 (PI 578872), an elite oilseed A-line (Miller and Gulya 1995), were obtained from Dr. Jerry F. Miller (USDA-ARS, Fargo, ND, USA). Seeds of Arikara (PI 369357), Havasupai (PI 369358), Seneca (PI 369360) and the 13 wild ecotypes were obtained from the USDA-ARS National Plant Germplasm System, Ames, Iowa (http://www.ars-grin.gov/). Wild ecotypes are identified by the prefix PI (plant introduction) and abbreviations for the states where they were originally collected. The 11 wild ecotypes were PI-AZ (PI 468575), PI-CA (PI 435593), PI-CO (PI 468625), PI-MT (PI 531022), PI-MX (PI 413123), PI-ND (PI 468439), PI-NV (PI 468596), PI-OK (PI 435619), PI-OR (PI 531015), PI-SD (PI 413039), PI-UT (PI 468619), PI-WA

(PI 531018) and PI-WY (PI 413019). Seeds of Tarahumara were obtained from Seeds of Change (http://www.seedsofchange.com/). Four replicate single-cross hybrids were produced for each elite × exotic parent combination using independent individuals. We planted a single plot per replicate per hybrid, 6.1 m in length, in Corvallis, Oregon. Forty seeds were initially planted per plot and were later thinned to 20 plants per plot. Plants were phenotyped for the presence or absence of branching at the R5 reproductive stage of development (Schneiter and Miller 1981). Capitula of elite × wild hybrids were bagged pre-anthesis and fertile hybrids were manually selfed during anthesis to screen for self-compatibility.

Development and analyses of wild introgression lines

We developed replicate BC_3S_1 and BC_4S_1 populations by phenotypic selection for profuse whole plant branching among S_1, BC_1S_1, BC_2S_1 and BC_3S_1 progeny from hybrids between an elite unbranched oilseed A-line (CMS-HA383) and two branched wild ecotypes (PI-AZ and PI-MT). CMS-HA383 × PI-AZ and CMS-HA383 × PI-MT were manually selfed to produced F_2 seeds. F_2 populations ($n = 400$ each) were grown in the field and phenotyped for branching in Corvallis, Oregon in 2000. Two branched individuals were selected in the CMS-HA383 × PI-AZ and CMS-HA383 × PI-MT F_2 populations, backcrossed to CMS-HA383, and manually selfed to produce four BC_1S_1 populations. We self-pollinated a BC_1 individual per replicate per hybrid to produce four BC_1S_1 populations. The latter were field tested in Corvallis, Oregon in 2001 and phenotyped for branching. One branched BC_1S_1 individual within each lineage was selected and backcrossed to CMS-HA383 to produce four BC_2 populations. Summer–winter cycles of field testing, phenotyping, selection, backcrossing, and selfing were repeated through the BC_4 with field testing and phenotyping in Corvallis, Oregon from 2002 to 2005.

BC_3S_1 individuals (one per replicate per hybrid) selected to produce BC_4 populations were phenotyped for a number of branches and genotyped using a genome-wide framework of 193 simple sequence repeat (SSR) markers (Tang et al. 2002). Graphical genotypes were drawn as described by Young and Tanksley (1989) using GGT 2.0 (van Berloo 2008) with linkage groups and genetic distances previously estimated in the RHA280 × RHA801 RIL mapping population (Tang et al. 2002, 2006; Yu et al. 2003). SSR markers were genotyped as described by Tang et al. (2002).

We produced two BC_4S_1 populations by self-pollinating one CMS-HA383 × PI-AZ and one CMS-HA383 × PI-MT BC_4 individual. We field tested 94 BC_4S_1 individuals from each population in Corvallis, Oregon in 2005. The 192 progeny were phenotyped for the presence or absence of branching, number of branches, seed oil content, 100-seed weight, capitulum diameter, plant height, stem diameter and days to flowering. BC_4S_1 progeny was also genotyped using 15 SSR or insertion-deletion (INDEL) markers distributed from the upper to the lower end of linkage group 10. *B* locus genotypes were inferred from branching phenotypes; branched individuals were classified as *bb*, whereas unbranched individuals were classified as *B_* (*BB* or *Bb*). The two BC_4S_1 populations were graphically genotyped using GGT 2.0 (van Berloo 2008). Concordance of the observed segregation ratio for branching with the expected segregation ratio (3 *B_*:1*bb*) was tested using χ^2 statistics (Griffiths et al. 2002). Exact *p* values for tests of the null hypothesis were derived from the PROB-CHI function in SAS 9.1. QTL analyses were performed using ORS1088, a codominant SSR marker locus, and the *B* locus as independent variables in linear model analyses, with fixed DNA or phenotypic marker locus effects and random residual effects. Type III *F* statistics, least square means, and coefficients of determination (R^2) were estimated for an SSR marker locus (ORS1088) and *B* locus genotypes using PROC GLM. We estimated the additive (*a*) and dominant (*d*) effects of ORS1088 for each trait using linear contrasts among least square means using CONSTRAST statements in PROC GLM (Lynch and Walsh 1998). The degree of dominance (|*d*/*a*|) was also estimated for each trait. In addition, the additive effect of the *B* locus was estimated for each trait using linear contrasts between least square means for dominant (*B_*) and recessive (*bb*) genotypes.

Using graphical genotypes for DNA marker loci previously mapped on linkage group 10 (Tang et al. 2002), we selected a single CMS-HA383 × PI-MT BC_3S_1 individual (PI-MT-1-1-2) with an unknown branching phenotype carrying a wild introgression spanning the *B* locus. PI-MT1-1-2 was heterozygous for two SSR marker loci upstream (HT347 and ORS1088) and two SSR marker loci downstream (ORS591 and ORS595) of the *B* locus and homozygous for recurrent parent (CMS-HA383) alleles for SSR and INDEL markers upstream and downstream of the introgressed segment. The NIL was developed by genotyping the four SSR markers upstream of the *B* locus (HT347 and ORS1088) and downstream of the *B* locus (ORS815 and ORS595). Two cycles of MAS were conducted for the wild introgression (spanning HT347–ORS595) by genotyping greenhouse-grown plants (without phenotyping), selecting an individual carrying the wild introgression, and backcrossing the selected individual to CMS-HA383. BC_4S_1 and BC_5 progeny were produced in the last cycle of MAS. The 47 BC_4S_1 and 47 BC_5 progeny were field tested in Corvallis, Oregon in 2006, phenotyped for the presence or absence of branching, and genotyped

using SSR markers spanning the wild introgression (HT347, ORS1088, ORS591, and ORS595). *B* locus genotypes were inferred from branching phenotypes (*B_* = unbranched and *bb* = branched) of the BC_4S_1 and BC_5 populations. The expected segregation ratio of the *B* locus (3*B_*:1*bb*) was tested using χ^2 statistics (Griffiths et al. 2002). The SSR marker loci and *B* locus were genetically mapped in the BC_4S_1 population using Mapmaker/Exp 3.0 (Lander et al. 1987).

Results

Seed trait segregation and heterosis in the TC-RIL population

Testcross hybrids developed by crossing branched (*bb*) and unbranched (*BB*) RHA280 × RHA801 RILs to an unbranched (*BB*) tester (CMS-HA372) were unbranched. The parents of the TC-RIL population had significantly different seed trait phenotypes (Fig. 1). The seed oil content of the unbranched oilseed B-line (HA372) was closer to the branched oilseed R-line (RHA801) than the unbranched confectionery R-line (RHA280). We phenotyped the fertile isogenic B-line (HA372) instead of the sterile A-line (CMS-HA372) to eliminate sterility and pollination biases. The inbred parent and tester lines had significantly lower seed yields than the TC-RILs. The phenotypic differences between HA372 and RHA801 were typical of unbranched oilseed B-lines and branched oilseed R-lines in sunflower (Fick and Miller 1997; Cheres et al. 1999, 2000). HA372 produced a single large capitula and larger seeds with slightly lower seed oil concentrations than RHA801, which produced a smaller primary capitula and numerous secondary and tertiary capitula (Fig. 1). The phenotypic ranges were wide and the phenotypic distributions were approximately normal for each trait among TC-RILs. Seed oil content was negatively correlated with 100-seed weight ($r_p = -0.64$; $p < 0.0001$) and seed yield ($r_p = -0.31$; $p < 0.0001$), while 100-seed weight and seed yield were positively correlated ($r_p = 0.49$; $p < 0.0001$).

Significant transgressive segregation was observed for seed yield among the TC-RILs, partly because of heterosis and partly because of the absence of branching (Figs. 1, 2). Heterosis for seed yield ranged from 57 to 253%, and mean heterosis was 135% among testcross hybrids (Fig. 2). The tester and parents of the RIL population had lower seed yields than hybrids, as is typical of other inbred A-, B-, and R-lines in sunflower (Fick and Miller 1997; Cheres et al. 1999, 2000). Moreover, branched R-lines typically have much lower seed yields than unbranched B-lines, as typified by the seed yield differences between RHA801 and HA372 in the present study (Fig. 1); heterosis percentages were

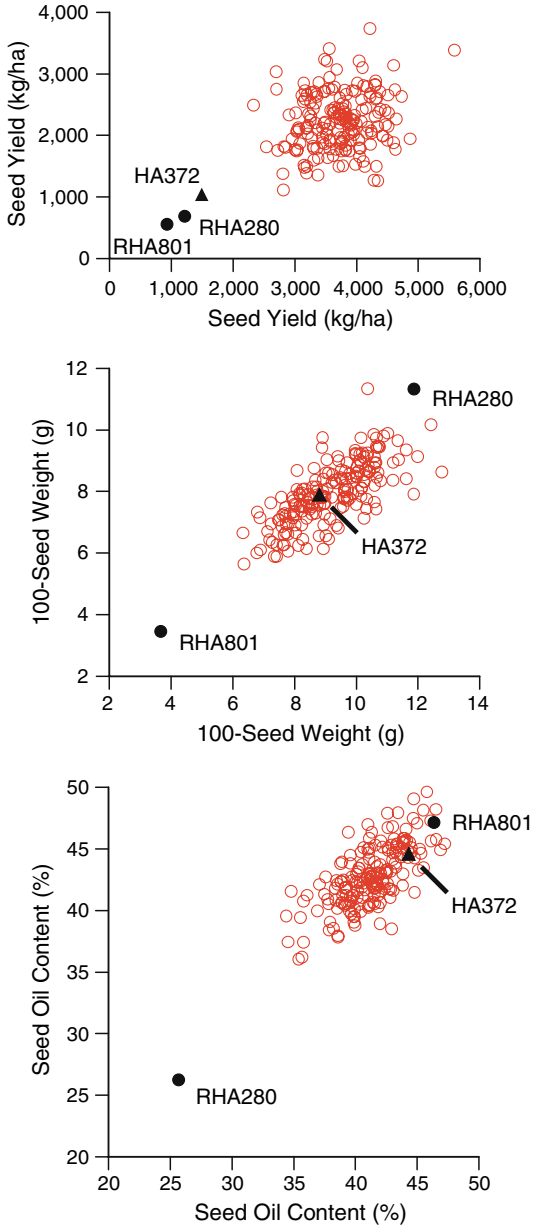


Fig. 1 Seed oil content, 100-seed weight and seed yield distributions for 173 CMS-HA372 × (RHA280 × RHA801) testcross hybrid recombinant inbred lines (TC-RILs) field tested and phenotyped in Corvallis, Oregon in 2001 (*x*-axis) and 2002 (*y*-axis). TC-RIL phenotypes (trait means) are identified by *open red circles*, B-line (HA372) phenotypes of the near-isogenic A-line tester (CMS-HA372) are identified by a *filled black triangle*, and phenotypes of the low-oil parent (RHA280) and high-oil parent (RHA801) of the RIL population are identified by *filled black circles* (color figure online)

estimated using the high parent (the unbranched oilseed tester) to eliminate the branching bias (Fig. 2). Seed oil content minima and maxima were greater in the TC-RIL than the RIL population (Tang et al. 2006). The mean of the low parent and lower tails of the TC-RIL distributions for seed oil content (where RHA280 was low) and 100-seed weight (where RHA801 was low) were widely separated

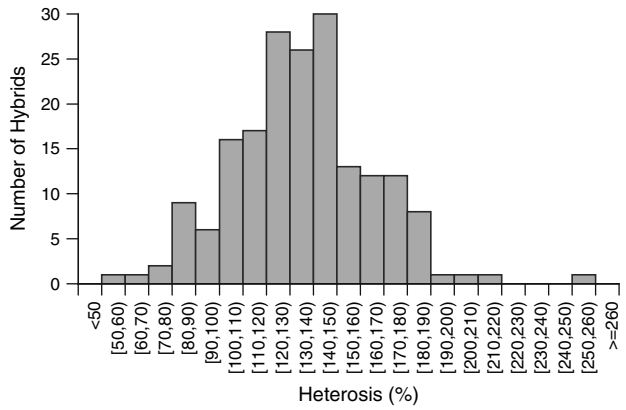


Fig. 2 Heterosis (%) distribution across years for 173 CMS-HA372 × (RHA280 × RHA801) testcross hybrid recombinant inbred lines field tested and phenotyped in Corvallis, Oregon in 2001 and 2002

because the testcross hybrids were produced using a high-oil tester, thereby shifting the phenotypic distributions for both traits upward and away from the low parent (Fig. 1).

Significant differences were observed among TC-RILs for each trait within and across years with Type III F statistic probabilities ranging from 0.0014 to <0.0001. TC-RIL-mean heritabilities were high for seed oil content ($h^2 = 0.83$) and 100-seed weight ($h^2 = 0.84$) and low for seed yield ($h^2 = 0.24$). RIL mean heritabilities were similarly high for seed oil content and 100-seed weight (Tang et al. 2006). While broad-sense TC-RIL and narrow-sense RIL mean heritabilities have different theoretical interpretations (Bernardo 2002), non-genetic variances were significantly smaller than genetic variances for seed oil content and 100-seed weight in both the RIL and TC-RIL populations. TC-RIL × year interactions were non-significant for seed oil content ($p = 0.50$) and 100-seed weight ($p = 0.50$) and highly significant ($p = 0.0015$) for seed yield, as depicted in the between-year phenotypic distributions for each trait (Fig. 1). The phenotypic distribution between years was more diffuse and circular for seed yield than seed oil content and 100-seed weight, both of which were tighter and more elliptical (the x- and y-axis scales of Fig. 1 are uniform). Seed yield ranks for TC-RILs were uncorrelated between years ($r = 0.09$; $p = 0.25$), whereas rank correlations for TC-RILs between years were significant for seed oil content ($r = 0.75$; $p < 0.0001$) and 100-seed weight ($r = 0.69$; $p < 0.0001$). Hence, the lower heritability for seed yield was caused, at least in part, by the significant TC-RIL × year differences.

QTL identified in the TC-RIL population

QTL for seed oil content were identified on linkage groups 1, 3, 9, 10, 16 and 17 in the TC-RIL population and collectively explained 30–40% of the phenotypic variation within and between years (Table 1; Fig. 3; Suppl. Tables 1, 2). Of the seed oil content QTL identified by TC-RIL testing, two (*soc16.1** and *soc17.1**) were directly centered on previously identified QTL (*soc16.1* and *soc17.1*), two (*soc1.1** and *soc9.1**) mapped in close proximity to previously identified QTL (*soc1.1* and *soc9.1*), and two (*soc3.1** and *soc10.1**) were not previously identified in the RIL population (Tang et al. 2006). High-oil alleles were transmitted by the high-oil parent for five of the six QTL. The low-oil parent (RHA280) transmitted the high-oil allele for *soc10.1**, a QTL linked to the *B* locus.

LOD maxima for QTL on linkage groups 16 and 17 (*soc16.1** and *soc17.1**) were found in the same positions as LOD maxima for QTL previously identified by RIL per se testing (*soc16.1* and *soc17.1*). Testcross effects for *soc16.1** ($a^* = -1.04$) and *soc17.1** ($a^* = -0.45$) were in the same direction and approximately half the magnitude of additive effects for *soc16.1* ($a = -1.70$) and *soc17.1* ($a = -0.90$). The QTL on linkage group 16 was centered on *Hyp*, a hypodermis pigment locus previously shown to be tightly linked to a large-effect QTL for seed oil content (Leon et al. 1995, 2003; Tang et al. 2006). The *Hyp*-linked QTL explained 27.5% of the variation for seed oil content in the TC-RIL population (Suppl. Table 2). The QTL on linkage group 17 mapped near *P*, a phytomelanin pigment locus previously shown to be tightly linked to a QTL for seed oil content (Tang et al. 2006). Both QTL explained a larger percentage of the seed oil content variation in the TC-RIL than the RIL population, partly because *soc10.1*, a large-effect QTL ($a = -2.3$) centered on the *B* locus and previously identified by RIL per se testing, had no effect in the TC-RIL population (Table 1; Fig. 3; Suppl. Tables 1, 2).

QTL on linkage groups 1 and 9 (*soc1.1** and *soc9.1**) mapped in close proximity to QTL identified in the RIL population (*soc1.1* and *soc9.1*). The LOD maxima of these QTL were in different intervals and the 1.0-LOD support intervals were non-overlapping. However, the high-oil alleles for both QTL were transmitted by RHA801 (the high-oil parent) in the TC-RIL and RIL populations (testcross and additive effects were negative for both traits) and the testcross effects, *soc1.1** ($a^* = -0.69$) and *soc9.1** ($a^* = -0.60$), were two- to threefold smaller than the additive effects, *soc1.1* ($a = -1.30$) and *soc9.1* ($a = -1.50$) (Table 1; Tang et al. 2006).

The QTL on linkage group 3 (*soc3.1**) and coincident large-effect QTL for 100-seed weight (*swt3.1**) were not identified in the RIL population per se. The *swt3.1** QTL explained 27.4% of the phenotypic variation in the TC-RIL population (Table 1; Fig. 3; Suppl. Tables 1, 2). The high-oil/high-oil (CMS-HA372/RHA801) heterozygote (*A*a*) had significantly smaller seeds and greater seed oil content than the high-oil/low-oil (CMS-HA372/RHA280)

Table 1 Testcross hybrid effect, Type III *F* statistic probability ($Pr > F$), genotype mean and coefficient of determination (R^2) estimates for seed trait QTL mapped in the CMS-HA372 × (RHA280 × RHA801) testcross hybrid recombinant inbred line population ($n = 173$)

Trait	QTL	Locus[a]	Genotype mean		Effect[b]	$Pr > F$
			*A*A*	*A*a*		
Seed oil content (%)	*soc1.1**	ORS716	41.3	42.7	−0.69	<0.0001
	*soc3.1**	ORS134	41.5	42.5	−0.53	0.0002
	*soc9.1**	CRT250	41.4	42.6	−0.60	<0.0001
	*soc10.1**	ORS1088	42.5	41.5	0.49	0.0005
	*soc16.1**	*Hyp*	41.0	43.1	−1.04	<0.0001
	*soc17.1**	ORS1245	41.6	42.5	−0.45	0.0011
		R^2			0.30	
100-seed weight (g)	*swt3.1**	ORS134	8.9	8.0	0.47	<0.0001
	*swt4.1**	ORS334	8.6	8.3	0.18	0.004
	*swt5.1**	ORS864	8.2	8.7	−0.25	<0.0001
	*swt9.1**	ORS188	8.7	8.2	0.22	0.0006
	*swt10.1**	ORS613	8.7	8.2	0.29	<0.0001
	*swt12.1**	ORS810	8.6	8.3	0.19	0.002
	*swt14.1**	ORS1079	8.8	8.1	0.36	<0.0001
	*swt17.1**	ORS561	8.8	8.1	0.35	<0.0001
		R^2			0.34	
Seed yield (kg/ha)	*yld10.1**	ORS437	3,069.5	2,812.1	128.69	<0.0001
	*yld13.1**	CRT76	3,037.0	2,844.7	96.17	0.0003
		R^2			0.03	

[a] Type III *F* statistics were estimated for the intralocus effects (least square mean differences) of DNA and phenotypic marker loci (independent variables) in mixed linear model analyses. DNA and phenotypic marker loci selected as independent variables were closest to the LOD maxima for QTL identified by composite interval mapping (see Suppl. Table 1)

[b] Testcross hybrid RIL effects (*a**) were estimated by least square mean differences between RHA280 and RHA801 testcross hybrid genotypes, where $a^* = (y_{A^*A} - y_{A^*a})/2$, the *A** allele was transmitted by the tester (CMS-HA372), the *A* allele was transmitted by RHA280, the *a* allele was transmitted by RHA801, y_{A^*A} is the least square mean of the *A*A* genotype, and y_{A^*a} is the least square mean of the *A*a* genotype; hence, the RHA280 allele in hybrid combination with the CMS-HA372 allele (*A*A*) increased the trait mean when the effect was positive, whereas the RHA801 allele in hybrid combination with the CMS-HA372 allele (*A*a*) increased the trait mean when the effect was negative

heterozygote (*A*A*). Hence, *soc3.1** and *swt3.1** were either masked by the pleiotropic effects of the *B* locus in the RIL per se analysis or caused by the heterotic effects of the underlying loci or both.

QTL for 100-seed weight were identified on linkage groups 4, 5, 9, 10, 12, 14 and 17 in the TC-RIL population, in addition to *swt3.1**, and they collectively explained 34–45% of the phenotypic variation within and between years (Table 1; Fig. 3; Suppl. Tables 1, 2). Of the 100-seed weight QTL identified by TC-RIL testing, only one was (*swt9.1**) was centered on a QTL (*swt9.1*) previously identified by RIL per se testing, whereas three (*swt5.1**, *swt14.1**, and *swt17.1**) mapped in close proximity to previously identified QTL (*swt5.1*, *swt14.1*, and *swt17.1*) and three (*swt4.1**, *swt10.1**, and *swt12.1**) were not previously identified in the RIL per se analysis. High-seed weight alleles were transmitted by the high-seed weight parent for six of the seven QTL. The low-seed weight parent (RHA801) transmitted the high-seed weight allele for *swt5.1**, which mapped in close proximity to the only other QTL identified in the RIL population where the low-seed weight parent transmitted the high-seed weight allele (*swt5.1*); hence, *swt5.1** and *swt5.1* effects may have been produced by the same locus. One additional 100-seed weight QTL was identified on linkage group 4 (*swt4.1**) and mapped close to a seed oil content QTL (*soc4.1*). The effect of *swt4.1** was significant in 2001 and across years, but not in 2002 (Table 1; Suppl. Table 1).

Seed yield QTL (*yld10.1** and *yld13.1**) were identified on linkage groups 10 and 13 and collectively explained 3–13% of the phenotypic variation within and across years (Table 1; Fig. 3; Suppl. Tables 1, 2). TC-RIL effects for both were significantly larger in 2001 than 2002 (Suppl. Table 1). The high-seed yield alleles for both were transmitted by the low-oil parent (RHA280) and increased seed yields by 167–348 kg/ha when combined with the CMS-HA372 allele in testcross hybrids. The seed yield QTL on linkage group ten mapped in close proximity to the *B* locus and a 100-seed weight QTL (*swt10.1**). Because 100-seed weight and seed yield were positively correlated, this QTL

could be either tightly linked or caused by a single locus. The QTL on linkage group 13 (*yld13.1**) was not associated with a 100-seed weight QTL and neither of the seed yield QTL were associated with seed oil content QTL, even though these traits were strongly negatively correlated. Low heritability, coupled with a small population size ($n = 173$), limited the power to identify and accurately map seed yield QTL (Lynch and Walsh 1998; Kearsey and Farquhar 1998; Bernardo 2002); however, both QTL (*yld10.1** and *yld13.1**) produced economically significant changes in seed yield.

Testing for the pleiotropic effects of the branching locus (*B*) in the unbranched testcross hybrid RIL population

The elimination of branching in the TC-RIL population eliminated large-effect *B*-linked seed oil content and 100-seed weight QTL previously identified in the RIL population per se and uncovered a small-effect QTL for seed oil content (*soc10.1**) located slightly upstream and a small-effect 100-seed weight (*swt10.1**) located slightly downstream of the *B* locus, in addition to uncovering a large-effect QTL for 100-seed weight (*swt3.1**) and small-effect QTL for seed oil content (*soc3.1**) on linkage group 3 (Table 1; Fig. 3; Suppl. Tables 1, 2) and increasing the magnitude of the effect of the 100-seed weight QTL (*soc5.1**) on linkage group 5. The testcross effect for *swt5.1** ($a^* = -2.75$) was fourfold greater than the additive effect for *swt5.1* ($a = -0.64$) and opposite in sign to the additive effect of *swt10.1* ($a = 1.81$), the *B*-centered QTL identified by RIL per se testing. Hence, the *B* locus appears to have pleiotropically affected seed trait QTL on linkage groups 3, 5 and 10.

The seed oil content and 100-seed weight QTL identified on linkage group 10 (*soc10.1** and *swt10.1**) were not previously identified and appear to have been masked by the pleiotropic effects of the *B* locus in the RIL per se analysis. The testcross effect of *soc10.1** ($a^* = 0.60$) was opposite in sign to the additive effect of *soc10.1* ($a = -2.40$), whereas the signs of the testcross effect of *swt10.1** ($a^* = 2.84$) and additive effect of *swt10.1* ($a = 1.81$) were both positive. The RHA280 *soc10.1** allele increased seed oil content, whereas the RHA801 *swt10.1** allele decreased 100-seed weight and concomitantly increased seed oil content through the negative genetic correlation between these traits (Tang et al. 2006). Hence, *soc10.1** was linked in repulsion, whereas *swt10.1** was linked in coupling to the *B* locus. If the *B*-centered seed oil content and 100-seed weight QTL previously identified by RIL per se testing were caused by the pleiotropic effects of the *B* locus, the dominant allele transmitted by RHA280 (*B*) should decrease seed oil content and increase 100-seed weight.

Fig. 3 One-LOD support intervals for seed oil content (*soc*), 100-seed weight (*swt*) and seed yield (*yld*) QTL identified by composite interval mapping in the CMS-HA372 × (RHA280 × RHA801) testcross hybrid recombinant inbred line (TC-RIL) population. *Filled red bars* identify seed oil content and 100-seed weight QTL identified by TC-RIL testing, whereas *filled yellow bars* identify seed oil content and 100-seed weight QTL identified by RHA280 × RHA801 RIL per se testing. *Filled gray bars* identify seed yield QTL identified by TC-RIL testing (color figure online)

Branching in elite × wild and elite × land race hybrids

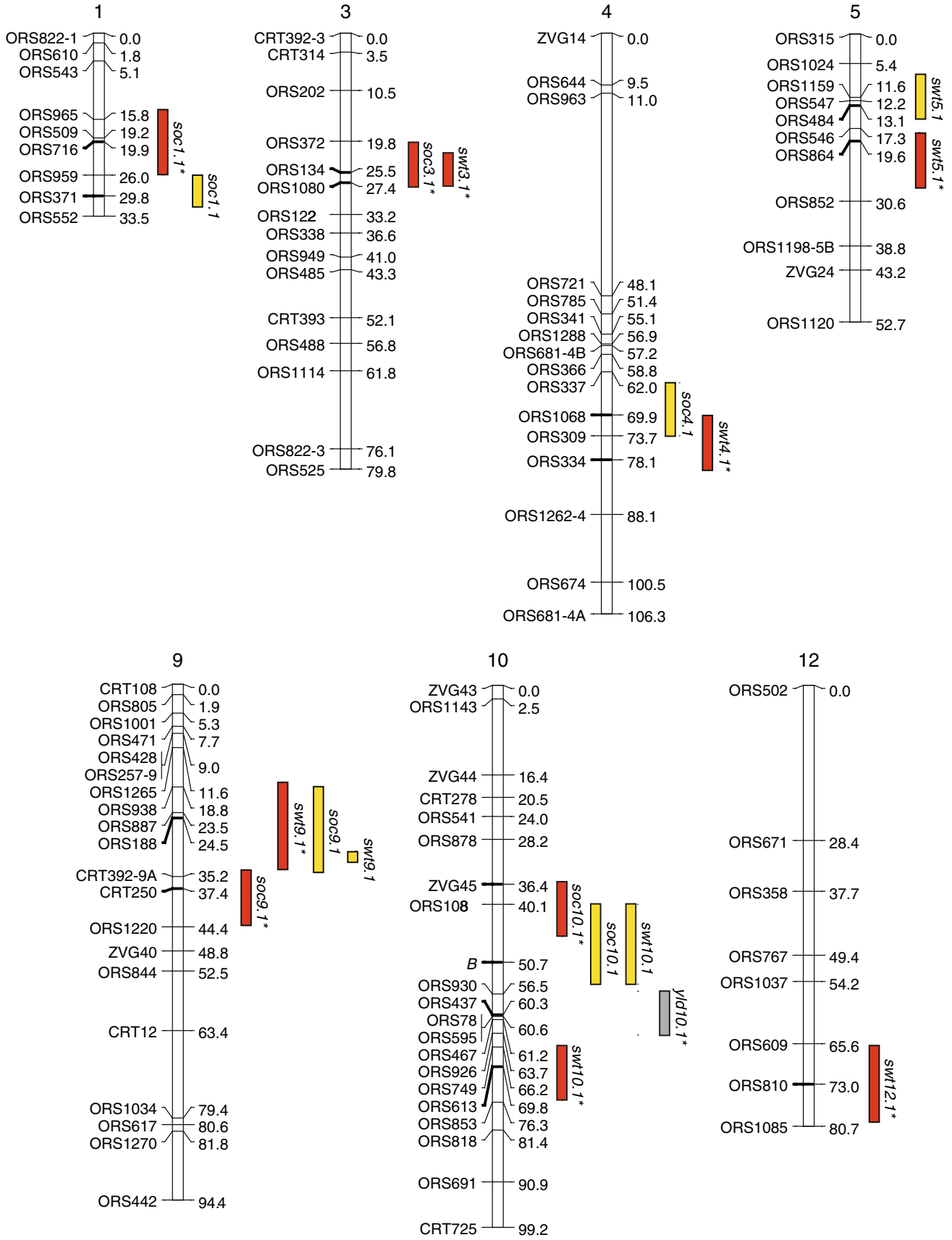
Single-cross hybrids between monocephalic elite A-lines and Native American land races and monocephalic elite A-lines and polycephalic wild ecotypes were phenotyped for branching, fertility and self-compatibility (Suppl. Table 3). Of the 22 elite × wild hybrids, 19 were completely and uniformly branched (four out of four replicate hybrids), whereas two wild ecotypes produced branched and unbranched hybrids (50–75% of the replicate hybrids of PI-CA and PI-SD were branched). Hence, the effect of the *B* allele, which is dominant in hybrids between elite unbranched (*BB*) A- or B-lines and elite branched (*bb*) R-lines (Putt 1943; Hockett and Knowles 1970; Fick et al. 1974b), was masked by alleles at *B* or other branching loci transmitted by each of the wild ecotypes tested. PI-CA and PI-SD transmitted alleles sufficient for producing unbranched hybrids, although both produced branched and unbranched hybrids (Suppl. Table 3). PI-CA and PI-SD could carry alleles introduced by migration from modern cultivars. The six CMS-HA372 and CMS-HA383 × land race hybrids were unbranched. Hence, the four land races are apparently homozygous for dominant *B* alleles and do not carry alleles at other branching loci sufficient for masking the effects of the *B* allele.

Discovery of self-compatible wild ecotypes

Two self-compatible wild ecotypes (PI-AZ and PI-MT) were discovered by manually selfing the 22 hybrids between unbranched A-lines and geographically diverse wild ecotypes (Suppl. Table 3). Only hybrids with PI-AZ and PI-MT were strongly self-compatible (hybrids with both ecotypes produced several hundred of F_2 seeds when manually selfed). Because CMS-HA383 and CMS-HA372 are self-compatible and carry self-compatibility alleles (*s*) at the *S* locus on linkage group 17 (Gandhi et al. 2005), PI-AZ and PI-MT transmitted either self-compatibility alleles or weak self-incompatibility alleles, whereas the other wild ecotypes transmitted strong self-incompatibility alleles.

Testing for branching locus (*b*) pleiotropy in advanced backcross progeny

The pleiotropic effects of the *B* locus were further analyzed by QTL analyses of advanced backcross progeny


1
ORS822-1 0.0
ORS610 1.8
ORS543 5.1
ORS965 15.8
ORS509 19.2
ORS716 19.9
ORS959 26.0
ORS371 29.8
ORS552 33.5
soc1.1*
soc1.1
3
CRT392-3 0.0
CRT314 3.5
ORS202 10.5
ORS372 19.8
ORS134 25.5
ORS1080 27.4
ORS122 33.2
ORS338 36.6
ORS949 41.0
ORS485 43.3
CRT393 52.1
ORS488 56.8
ORS1114 61.8
ORS822-3 76.1
ORS525 79.8
soc3.1*
swt3.1*
4
ZVG14 0.0
ORS644 9.5
ORS963 11.0
ORS721 48.1
ORS785 51.4
ORS341 55.1
ORS1288 56.9
ORS681-4B 57.2
ORS366 58.8
ORS337 62.0
ORS1068 69.9
ORS309 73.7
ORS334 78.1
ORS1262-4 88.1
ORS674 100.5
ORS681-4A 106.3
soc4.1
swt4.1*
5
ORS315 0.0
ORS1024 5.4
ORS1159 11.6
ORS547 12.2
ORS484 13.1
ORS546 17.3
ORS864 19.6
ORS852 30.6
ORS1198-5B 38.8
ZVG24 43.2
ORS1120 52.7
swt5.1
swt5.1*
9
CRT108 0.0
ORS805 1.9
ORS1001 5.3
ORS471 7.7
ORS428 9.0
ORS257-9
ORS1265 11.6
ORS938 18.8
ORS887 23.5
ORS188 24.5
CRT392-9A 35.2
CRT250 37.4
ORS1220 44.4
ZVG40 48.8
ORS844 52.5
CRT12 63.4
ORS1034 79.4
ORS617 80.6
ORS1270 81.8
ORS442 94.4
soc9.1*
swt9.1*
soc9.1
swt9.1
10
ZVG43 0.0
ORS1143 2.5
ZVG44 16.4
CRT278 20.5
ORS541 24.0
ORS878 28.2
ZVG45 36.4
ORS108 40.1
B 50.7
ORS930 56.5
ORS437 60.3
ORS78 60.6
ORS595
ORS467 61.2
ORS926 63.7
ORS749 66.2
ORS613 69.8
ORS853 76.3
ORS818 81.4
ORS691 90.9
CRT725 99.2
soc10.1*
soc10.1
swt10.1
yld10.1*
swt10.1*
12
ORS502 0.0
ORS671 28.4
ORS358 37.7
ORS767 49.4
ORS1037 54.2
ORS609 65.6
ORS810 73.0
ORS1085 80.7
swt12.1*

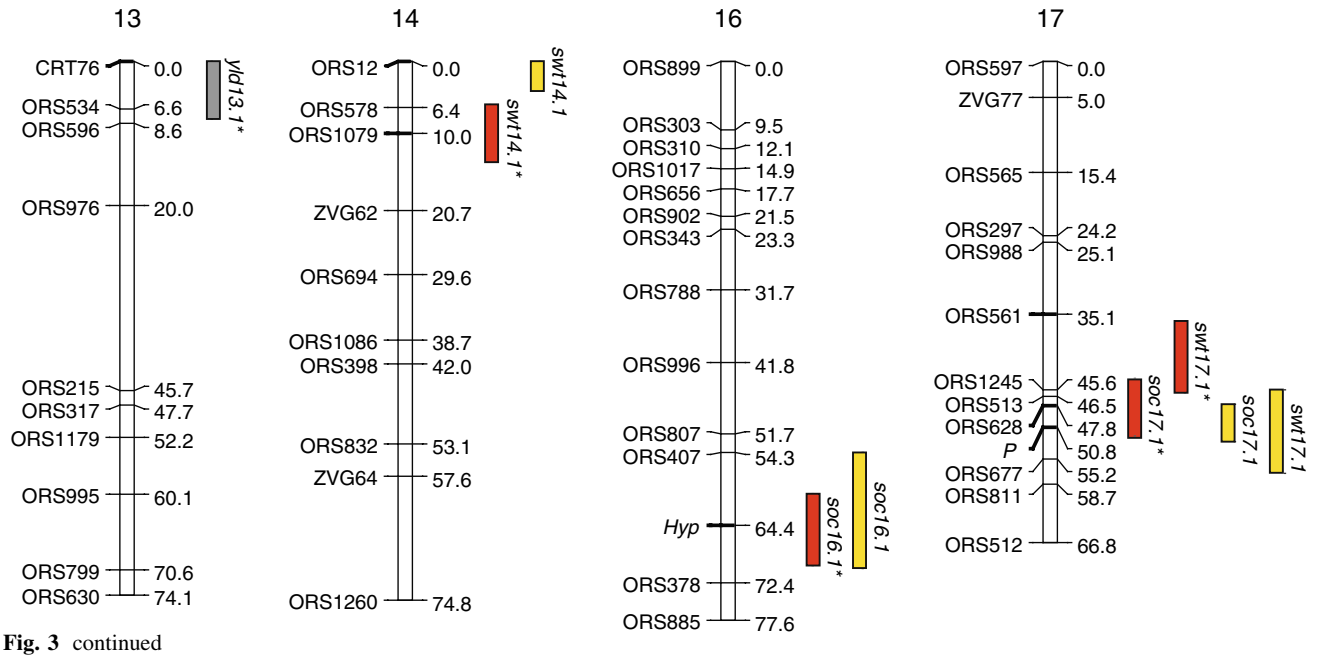


Fig. 3 continued

developed by phenotypic selection for profuse, whole plant branching in two elite × wild hybrid populations, CMS-HA383 × PI-AZ and CMS-HA383 × PI-MT (Tables 2, 3; Fig. 4). Four BC_3S_1 populations (206–226 individuals per population) were phenotyped for branching. One branched BC_3S_1 plant was selected from each population and graphically genotyped using 193 SSR markers distributed throughout the sunflower genome (Tang et al. 2002). One genomic segment (ORS541–ORS815) carried wild alleles among the four graphically genotyped BC_3S_1 individuals. The introgressed segment spanned the *B* locus and encompassed the upper arm of the chromosome in both PI-AZ introgressions and a shorter segment in both PI-MT introgressions (Fig. 4). The four selected BC_3S_1 individuals were selfed and 100% of the progeny within the four BC_3S_2 lines were branched. While wild introgressions were found on a few other linkage groups, none were common across replicates and populations. Nonetheless, introgressions on linkage groups 3, 4, 13, 14 and 16 were identified for both replicates developed from one wild donor or the other and warrant further study.

Two BC_4S_1 populations were subsequently developed, one from each wild donor (CMS-HA383 × PI-AZ and CMS-HA383 × PI-MT). We field tested, phenotyped and genotyped 94 progeny from both populations with DNA markers spanning linkage group 10 (Fig. 4). Both populations segregated with a ratio of 74 unbranched:20 branched, which was not significantly different from the expected segregation ratio (3 unbranched:1 branched) for the *B* locus ($\chi^2 = 0.69$; $p = 0.41$). Hence, wild *B* locus alleles introgressed from PI-AZ and PI-MT were recessive

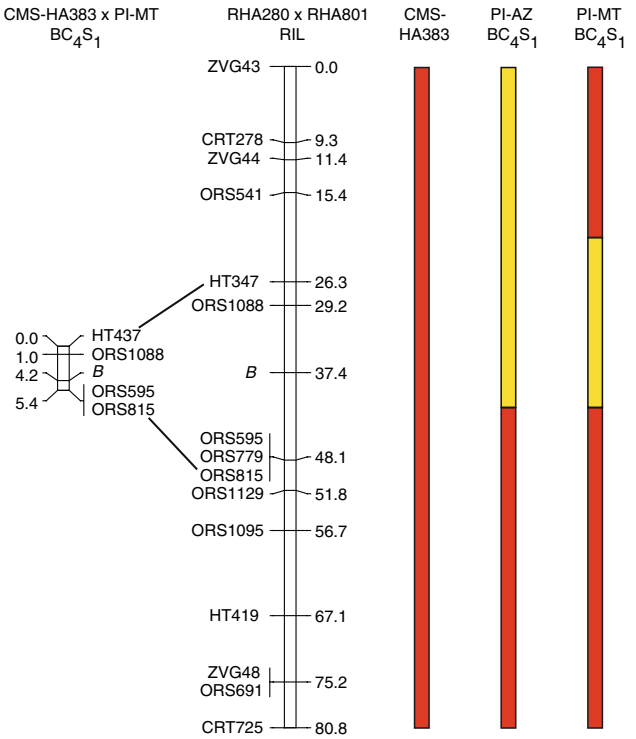


Fig. 4 Graphical genotypes for the recurrent parent CMS-HA383 (*red bars*) and introgressed wild donor (PI-AZ or PI-MT) segments (*yellow bars*) segregating in BC_4S_1 populations developed from hybrids between an unbranched A-line (CMS-HA383) and branched wild ecotypes (PI-AZ and PI-MT). Graphical genotypes for replicate BC_4S_1 individuals within each hybrid were identical; therefore, a single replicate per hybrid is displayed. Genetic distances are shown for DNA markers and the *B* locus on linkage group 10 in the CMS-HA383 × PI-MT BC_4S_1 and RHA280 × RHA801 RIL populations (color figure online)

Table 2 Genotype mean, additive (a) and dominance (d) effect, Type III F statistic probability ($Pr > F$), degree of dominance ($|d/a|$) and coefficient of determination (R^2) estimates for several traits for a DNA marker locus (ORS1088) tightly linked to the branching locus (B) in introgressed segments on linkage group 10 segregating in two elite × wild BC_4S_1 populations, CMS-HA383 × PI-AZ ($n = 94$) and CMS-HA383 × PI-MT ($n = 94$)

Trait	Donor	Genotype mean[a]			Additive (a)[b]		Dominant (d)		$\|d/a\|$	R^2
		AA	*Aa*	*aa*	Effect	$Pr > F$	Effect	$Pr > F$		
Number of branches	PI-AZ	0.0	0.5	12.1	−6.1	<0.0001	5.6	<0.0001	0.9	0.62
	PI-MT	0.8	0.7	18.8	−9.0	<0.0001	9.1	<0.0001	1.0	0.75
Seed oil content (%)	PI-AZ	38.8	38.8	44.6	−2.9	<0.0001	3.1	0.003	1.0	0.21
	PI-MT	36.9	38.2	45.3	−4.2	<0.0001	2.9	0.02	0.7	0.22
Capitula diameter (cm)	PI-AZ	22.2	21.8	14.8	3.7	<0.0001	−3.3	0.0001	0.9	0.37
	PI-MT	24.6	23.2	15.5	4.6	<0.0001	−3.2	<0.0001	0.7	0.32
100-seed weight (g)	PI-AZ	8.1	8.1	5.9	1.1	<0.0001	−1.1	0.0007	1.0	0.26
	PI-MT	8.0	7.5	5.2	1.4	<0.0001	−0.9	0.002	0.6	0.36
Stem diameter (mm)	PI-AZ	22.1	22.6	19.8	1.2	0.08	−1.7	0.09	1.4	0.05
	PI-MT	24.6	23.5	23.7	0.5	0.5	0.7	0.5	1.4	0.02
Plant height (cm)	PI-AZ	124.2	123.8	119.5	2.4	0.05	−2.0	0.3	0.8	0.05
	PI-MT	125.7	126.7	128.1	−1.2	0.5	0.2	0.9	0.2	0.01
Days to flower (days)	PI-AZ	72.7	72.7	72.3	0.2	0.7	−0.2	0.7	1.0	0.00
	PI-MT	73.4	72.7	72.5	0.5	0.3	0.3	0.7	0.6	0.02

The introgressed segments were heterozygous for elite recurrent parent (CMS-HA383) and wild donor (PI-AZ or PI-MT) alleles in the BC_4 and segregated in BC_4S_1 populations

[a] The *A* allele was transmitted by the unbranched recurrent parent (CMS-HA383) and the *a* allele was transmitted by the wild donors (PI-AZ or PI-MT)

[b] The additive effects (a) of the ORS1088 locus were estimated by least square mean differences between CMS-HA383 and wild donor (PI-AZ or PI-MT) genotypes as $a = (y_{AA} - y_{aa})/2$, where y_{AA} is the least square mean for *AA* genotypes and y_{aa} is the least square mean for *aa* genotypes. The additive effect was positive when the CMS-HA383 allele (*A*) increased and negative when the PI-AZ or PI-MT allele (*a*) increased the trait mean

(b^w) and produced phenotypic effects analogous to the recessive allele (b) found in elite R-lines (Putt 1964). Significant variability was observed for seed oil content, 100-seed weight and capitula diameter within and among branched and unbranched BC_4S_1 progeny and number of branches among branched BC_4S_1 progeny.

The additive and dominant effects of ORS1088, an SSR marker locus tightly linked to the *B* locus, were estimated among BC_4S_1 progeny (Table 2). One recombinant was observed between *B* and ORS1088 in the CMS-HA383 × PI-AZ BC_4S_1 population and three recombinants were observed between these loci in the CMS-HA383 × PI-MT BC_4S_1 population. Hence, the effect of ORS1088, a codominant DNA marker, should closely approximate the effect of *B*, a dominant phenotypic marker (Tables 2, 3). The codominant DNA marker facilitated indirect estimation of additive and dominance effects of the *B* locus. ORS1088 and *B* significantly affected the number of branches, seed oil content, 100-seed weight and capitula diameter. Stem diameter, plant height and days to flowering were not affected by the *B* locus or *B*-linked QTL in either population. QTL in the introgressed segment were strongly or completely dominant ($|d/a| = 1$) for traits affected by branching (Table 2). ORS1088 effects were in the direction predicted from *B* locus genotypes and phenotypes; the dominant allele transmitted by the CMS-HA383 (*B*) eliminated branching, decreased seed oil content and increased capitula diameter and 100-seed weight (Tables 2, 3). The *B* locus per se significantly affected the number of branches, seed oil content, 100-seed weight and capitula diameter (Table 3). Because the *B* allele is dominant, differences between unbranched (*B_*) and branched (*bb*) genotypes were approximately equal to twice the additive effects of the ORS1088 locus for traits affected by branching (Tables 2, 3).

Development of a near-isogenic line carrying a wild *B* allele

We developed a NIL carrying a wild introgression spanning the *B* locus through marker-assisted selection (MAS) of the CMS-HA383 × PI-MT population (Fig. 4). BC_4S_1 progeny segregated 32 unbranched to 15 branched, which was not significantly different from the expected segregation ratio (3 unbranched to 1 branched) for the *B* locus ($\chi^2 = 1.20$; $p = 0.55$). Four SSR markers spanning the

Table 3 Genotype mean, Type III F statistic probability ($Pr > F$) and coefficient of determination (R^2) estimates for several traits for the B locus in two elite × wild BC_4S_1 populations, CMS-HA383 × PI-AZ ($n = 94$) and CMS-HA383 × PI-MT ($n = 94$), segregating for wild introgressions spanning the B locus

Trait	Donor	Genotype mean[a]		Effect[b]	$Pr > F$	R^2
		$B\_$	b^wb^w			
Number of branches	PI-AZ	0.0	14.2	−14.2	<0.0001	0.85
	PI-MT	0.0	18.9	−18.9	<0.0001	0.92
Seed oil content (%)	PI-AZ	38.7	45.4	−6.7	<0.0001	0.27
	PI-MT	37.3	45.8	−8.5	<0.0001	0.31
Capitula diameter (cm)	PI-AZ	21.8	14.4	7.4	<0.0001	0.38
	PI-MT	24.2	15.3	8.9	<0.0001	0.41
100-Seed weight (g)	PI-AZ	8.1	5.4	2.7	<0.0001	0.32
	PI-MT	7.8	5.1	2.7	<0.0001	0.45
Stem diameter (mm)	PI-AZ	22.1	20.0	2.1	0.07	0.03
	PI-MT	24.1	23.3	0.8	0.4	0.01
Plant height (cm)	PI-AZ	123.5	120.7	2.8	0.2	0.02
	PI-MT	126.4	127.5	−1.1	0.7	0.00
Days to flower (days)	PI-AZ	72.6	72.7	−0.1	1.0	0.00
	PI-MT	73.0	72.7	0.3	0.6	0.00

The introgressed segments were heterozygous for elite recurrent parent (CMS-HA383) and wild donor (PI-AZ or PI-MT) alleles in the BC_4 and segregated in BC_4S_1 populations

[a] The B allele was transmitted by the unbranched recurrent parent (CMS-HA383), whereas the b^w allele was transmitted by the wild donors (PI-AZ or PI-MT)

[b] The effect of the B locus was estimated by least square mean differences between unbranched ($B\_$) and branched (b^wb^w) genotypes. The B locus effect was calculated as ($y_{B\_} - y_bw_bw$), where $y_{B\_}$ is the least square mean for unbranched progeny (BB and Bb^w genotypes), and y_bw_bw is the least square mean for branched progeny (b^wb^w genotype). The B locus effect was positive when the CMS-HA383 allele (B) increased and negative when the PI-AZ or PI-MT allele (b) increased the trait mean

wild introgression were genetically mapped in the BC_4S_1 population and produced locus orders identical to those previously reported in the RHA280 × RHA801 RIL population (Fig. 4; Tang et al. 2006). Genetic distances were fourfold shorter in the HT347–ORS815 DNA marker interval in the elite × wild (CMS-HA383 × PI-MT) BC_4S_1 population than the RHA280 × RHA801 RIL population, presumably because recombination was suppressed in the heterozygous (elite/wild) genomic DNA segment, as previously observed across the genome in a different elite × wild (HA89 × ANN1238) population (Burke et al. 2002). Of the 15 branched BC_4S_1 progeny, 14 were homozygous for the wild ORS1088 and ORS595 alleles, whereas one was homozygous for the wild ORS1088 allele and heterozygous for ORS595 alleles. The 32 unbranched plants were homozygous for the recurrent parent (CMS-HA383) alleles. The 47 BC_5 plants were heterozygous for the HT347–ORS595 interval and unbranched. Hence, the wild B allele introgressed from PI-MT was recessive (b^w), as were wild alleles (b^w) carried by the other four introgression lines (Tables 2, 3; Fig. 4).

Discussion

The pleiotropic effects of the branching locus (B), as opposed to loci tightly linked to B, were substantiated by comparative mapping of QTL in branched and unbranched populations and B-locus NILs. The pleiotropic effects of B were logical and in the direction predicted by many earlier phenotypic and QTL analyses in sunflower (Fick et al. 1974b; Dedio 1980; Mestries et al. 1998; Bert et al. 2003; Tang et al. 2006). While the pleiotropic effects of B were highly predictable, the genomic region surrounding the B locus apparently harbors additional loci affecting a wide range of domestication and post-domestication traits, linked in coupling, repulsion, or both across populations (Mestries et al. 1998; Burke et al. 2002; Bert et al. 2003; Tang et al. 2006). The power of identifying such loci depends on their effects and spatial separation (genetic distances), segregation of the B locus and sampling (number of recombinants) (Lynch and Walsh 1998; Yi and Xu 2002; Yi 2004; Liu et al. 2007; Banerjee et al. 2008).

The effects of the B locus were hypothesized to mask the effects of both linked and unlinked QTL pleiotropically affected by branching in the RHA280 × RHA801 RIL population (Tang et al. 2006). The search for linked QTL in the present study was partly stimulated by the findings of Bert et al. (2003), where the effects of B-linked QTL for seed oil content and 100-seed weight, in an F_2 population (XRQ × PSC8) segregating for branching, were in the opposite direction of predictions from earlier phenotypic and QTL mapping analyses (Fick et al. 1974b; Dedio 1980; Mestries et al. 1998; Tang et al. 2006). Unless the effects of the B-linked QTL identified in the XRQ × PSC8 F_2 population were caused by the B locus per se, which seems improbable, the findings of Bert et al. (2003) predict the presence of QTL linked in repulsion with positive effects large enough to compensate for the negative effects of the B locus. We tested the hypothesis of linked QTL by eliminating the effects of the B locus, which had been previously mapped in the RIL population. This approach uncovered a seed oil content QTL linked in repulsion and a 100-seed weight QTL linked in coupling to the B locus, both of which appear to have been masked by the pleiotropic effects of the B locus in the RIL population (Table 1; Fig. 3; Tang et al. 2006).

These results highlight one of the shortcomings of low-resolution QTL analyses in small populations; specifically the inability to identify and resolve the effects of multiple linked QTL and distinguish between linkage and pleiotropy

when one or both are operating on correlated traits (Lynch and Walsh 1998; Kearsey and Farquhar 1998; Bernardo 2002; Yi and Xu 2002; Yi 2004; Liu et al. 2007; Banerjee et al. 2008). Linkage and pleiotropy both appear to play a role in the genetics of correlated seed traits in sunflower. Here, our approach was to phenotype unbranched testcross hybrid progeny of an RIL population segregating for branching to disentangle the effects of multiple QTL, which were confounded by the pleiotropic effects of the *B* locus; but such solutions are obviously not possible for most traits. The effects of loci linked in repulsion can cancel or obscure one another, as was found in the present study for the *soc10.1** QTL. This QTL appears to be slightly upstream of the *B* locus with the favorable allele, which increased seed oil content, transmitted by the low-oil parent. Similarly, QTL linked in coupling can be obscured by the effects of pleiotropically acting loci, as for *swt10.1**, a QTL slightly downstream of the *B* locus.

Four linkage groups (3, 9, 10 and 17) harbored coincident QTL affecting both seed oil content and 100-seed weight (Table 1; Fig. 3). The QTL on linkage group 9 (*soc9.1** and *swt9.1**) coincided with QTL identified in the RIL population (*soc9.1* and *swt9.1*) and clusters of domestication trait QTL identified in elite × wild populations, e.g., QTL for leaf size, number of leaves, days to flower, number of capitula, and weight and width of achene (Burke et al. 2002; Baack et al. 2008). This region appeared to be a target of selection for flowering time genes during sunflower domestication (Baack et al. 2008). The QTL for seed oil content and 100-seed weight on linkage group 17 of the TC-RIL population (*soc17.1** and *swt17.1**) were coincident with QTL identified in the RIL population (*soc17.1* and *swt17.1*) and tightly linked to phytomelanin pigment (*P*) and self-incompatibility (*S*) loci (Gandhi et al. 2005; Tang et al. 2006). This region harbors numerous domestication and post-domestication trait QTL identified in elite × wild populations, e.g., achene weight, plant height, days to flower, seed shattering and autonomous self-pollination (Burke et al. 2002; Gandhi et al. 2005), and number of branches in a wild × land race population (Wills and Burke 2007).

While our analyses focused on the effects of the *B* locus on branching, the genetics of branching appears to be complex in populations developed from elite × wild and land race × wild hybrids (Burke et al. 2002; Wills and Burke 2007; Suppl. Table 3). Several generations of phenotypic selection for profuse branching in elite × wild populations yielded four introgression lines carrying a single common wild introgression flanking the *B* locus (Fig. 4). While the effect of *B* was masked by other loci in elite × wild hybrids (Suppl. Table 3), *B* appears to be the only branching locus transferred by phenotypic selection in the development of these introgression lines. Our hypothesis was that multiple loci would be introgressed through phenotypic selection and would be essential for whole plant branching. While the wild genetic background was not completely eliminated in the introgression lines, segregating segments on other chromosomes probably do not harbor branching loci because none were common across the four introgression lines and the *B* locus explained most of the phenotypic variation for number of branches (Tables 2, 3). Hence, phenotypic selection for profuse branching appears to have eliminated wild alleles at branching loci, which epistatically or pleiotropically masked the effect of the *B* allele in elite × wild hybrids (Suppl. Table 3). When wild alleles at these other branching loci are not present in the genetic background, the effect of the wild allele (b^w) is sufficient for branching and recessive to the dominant *B* allele found in unbranched elite lines in sunflower. Hence, the *B* locus appears to be necessary, but not sufficient for monocephaly in domesticated sunflower.

Acknowledgments This research was supported by grants from the National Science Foundation Plant Genome Research Program (No. 0421630) and United States Department of Agriculture Plant Genome Program (No. 2000-04292).

References

Baack EJ, Sapir Y, Chapman MA, Burke JM, Rieseberg LH (2008) Selection on domestication traits and quantitative trait loci in crop–wild sunflower hybrids. Mol Ecol 17:666–677

Banerjee S, Yandell BS, Yi NJ (2008) Bayesian quantitative trait loci mapping for multiple traits. Genetics 179:2275–2289

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury

Bert PF, Jouan I, de Labrouhe DT, Serre F, Philippon J, Nicolas P, Vear F (2003) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.). II. Characterisation of QTL involved in developmental and agronomic traits. Theor Appl Genet 107:181–189

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

Cheres MT, Miller JF, Knapp SJ (1999) Donor inbred lines for enhancing the performance of single-cross sunflower hybrids. Crop Sci 39:1325–1331

Cheres MT, Miller JF, Crane JR, Knapp SJ (2000) Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. Theor Appl Genet 100: 889–894

Dedio W (1980) Comparison of achene characteristics and combining ability of branching and non-branching near-isogenic sunflower restorer lines. Crop Sci 20:189–190

Fick GN (1975) Heritability of oil content in sunflower. Crop Sci 15:77–78

Fick GN, Miller JF (1997) Sunflower breeding. In: Schneiter AA (ed) Sunflower technology and production. Agronomy Society of America, Inc, Madison, pp 395–439

Fick GN, Zimmer DE, Kinman ML (1974a) Registration of six sunflower parental lines. Crop Sci 14:912

Fick GN, Zimmer DE, Zimmerman DC (1974b) Correlation of seed oil content in sunflowers with other plant and seed characteristics. Crop Sci 14:755–757

Gandhi SD, Heesacker A, Freeman CA, Argyris J, Bradford K, Knapp SJ (2005) The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. Theor Appl Genet 111:619–629

Griffiths A, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2002) An introduction to genetic analysis, 7th edn. W.H. Freeman, New York

Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH (2004) Origin of extant domesticated sunflowers in eastern North America. Nature 430:201–204

Heiser CB (1945) The Hopi sunflower. Miss Bot Garden Bull 3:163–166

Heiser CB (1951) The sunflower among the North American Indians. Proc Am Phil Soc 95:432–448

Heiser CB (1976) The sunflower. University Oklahoma Press, Norman

Hockett EA, Knowles PF (1970) Inheritance of branching in sunflowers, *Helianthus annuus* L. Crop Sci 10:432–436

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants: where are we now? Heredity 80:137–142

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln ES, Newburg L (1987) MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Leon AJ, Lee M, Rufener GK, Berry ST, Mowers RP (1995) Use of RFLP markers for genetic linkage analysis of oil percentage in sunflower seed. Crop Sci 35:558–564

Leon AJ, Andrade FH, Lee M (2003) Genetic analysis of seed-oil concentration across generations and environments in sunflower. Crop Sci 43:135–140

Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH (2005) Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 169:2225–2239

Liu J, Liu Y, Liu X, Deng H-W (2007) Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. Am J Hum Genet 81:304–320

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland

Mestries E, Gentzbittel L, de Labrouhe DT, Nicolas P, Vear F (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflowers (*Helianthus annuus* L.) using molecular markers. Mol Breed 4:215–226

Miller JF, Gulya TJ (1990) Registration of ten oilseed sunflower germplasm lines. Crop Sci 30:430–431

Miller JF, Gulya TJ (1995) Registration of 4 maintainer (HA-382 to HA-385) and 4 restorer (RHA-386 to RHA-389) sunflower germplasm lines. Crop Sci 35:286

Mokrani L, Gentzbittel L, Azanza F, Fitamant L, Al-Chaarani G, Sarrafi A (2002) Mapping and analysis of quantitative trait loci for grain oil content and agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). Theor Appl Genet 106:149–156

Putt ED (1943) Association of seed yield and oil content with other characters in the sunflower. Sci Agr 23:377–383

Putt ED (1964) Recessive branching in sunflowers. Crop Sci 4:444–445

Roath WW, Miller JF, Gulya TJ (1981) Registration of RHA 801 sunflower germplasm. Crop Sci 21:479

Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. National Sunflower Association, Bismarck

SAS Institute (2004) SAS/STAT software v. 9.1.3. SAS Institute Inc, Cary

Schneiter AA, Miller JF (1981) Description of sunflower growth stages. Crop Sci 21:901–903

Tang S, Knapp SJ (2003) Microsatellites uncover extraordinary diversity in Native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Tang SX, Leon A, Bridges WC, Knapp SJ (2006) Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci 46:721–734

van Berloo R (2008) GGT 2.0: Versatile software for visualization and analysis of genetic data. J Hered 99:232–236

Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics NCSU (ed) Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wills DM, Burke JM (2007) Quantitative trait locus analysis of the early domestication of sunflower. Genetics 176:2589–2599

Yi NJ (2004) A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. Genetics 167:967–975

Yi NJ, Xu SZ (2002) Mapping quantitative trait loci with epistatic effects. Genet Res 79:185–198

Young ND, Tanksley SD (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor Appl Genet 77:95–101

Yu JK, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Chu WC, Webb DM, Thompson L, Edwards KJ, Berry S, Leon AJ, Grondona M, Olungu C, Maes N, Knapp SL (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. Crop Sci 43:367–387

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci 90:10972–10976